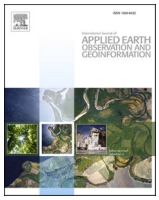
International Journal of Applied Earth Observations and Geoinformation 112 (2022) 102859

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag



3D building reconstruction from single street view images using deep learning

Hui En Pang [a], Filip Biljecki [b,c,*]

[a] Department of Geography, National University of Singapore, Singapore
[b] Department of Architecture, National University of Singapore, Singapore
[c] Department of Real Estate, National University of Singapore, Singapore

ARTICLE INFO

Keywords:
3D geoinformation
GeoAI
Urban morphology
Digital twin
Google Street View
3D GIS

ABSTRACT

3D building models are an established instance of geospatial information in the built environment, but their acquisition remains complex and topical. Approaches to reconstruct 3D building models often require existing building information (e.g. their footprints) and data such as point clouds, which are scarce and laborious to acquire, limiting their expansion. In parallel, street view imagery (SVI) has been gaining currency, driven by the rapid expansion in coverage and advances in computer vision (CV), but it has not been used much for generating 3D city models. Traditional approaches that can use SVI for reconstruction require multiple images, while in practice, often only few street-level images provide an unobstructed view of a building. We develop the reconstruction of 3D building models from a single street view image using image-to-mesh reconstruction techniques modified from the CV domain. We regard three scenarios: (1) standalone single-view reconstruction; (2) reconstruction aided by a top view delineating the footprint; and (3) refinement of existing 3D models, i.e. we examine the use of SVI to enhance the level of detail of block (LoD1) models, which are common. The results suggest that trained models supporting (2) and (3) are able to reconstruct the overall geometry of a building, while the first scenario may derive the approximate mass of the building, useful to infer the urban form of cities. We evaluate the results by demonstrating their usefulness for volume estimation, with mean errors of less than 10% for the last two scenarios. As SVI is now available in most countries worldwide, including many regions that do not have existing footprint and/or 3D building data, our method can derive rapidly and cost-effectively the 3D urban form from SVI without requiring any existing building information. Obtaining 3D building models in regions that hitherto did not have any, may enable a number of 3D geospatial analyses locally for the first time.

1. Introduction

3D building models continue to prove themselves useful for a wide range of applications, across real estate, urban planning, and disaster management (Elfouly and Labetski, 2020; Stoter et al., 2020; Beran et al., 2021; Jang et al., 2021; Turan et al., 2021). Their applications can be classified into either visual or non-visual instances. In the former, attention is placed on the visual fidelity and aesthetic appeal and not necessarily much on accuracy and quality, e.g. visualisation of apartments for real estate (Cohen et al., 2016). In the latter, the focus is on using data for quantitative operations (Florio et al., 2021; Gassar and Cha, 2021; Bizjak et al., 2021; Chen et al., 2020a; Palliwal et al., 2021). In particular, 3D building models are convenient for estimating the volume and envelope area of buildings, operations previously not possible with traditional building information such as footprints, which are without a volumetric representation (Sindram et al., 2016; Doan et al., 2021; Rosser et al., 2019; Eicker et al., 2014; Braun et al., 2021). Therefore, they earned an important role in various use cases. For example, at the urban scale, the volume and surface area of buildings are useful for estimating energy consumption (Bahu et al., 2014; Kaden and Kolbe, 2014), population estimation (Sridharan and Qiu, 2013; Szarka and Biljecki, 2022), urban planning (Ahmed and Sekar, 2015), estimation of urban heat island intensity (Li et al., 2020), and assessing solar energy resources in cities (Eicker et al., 2014).

In the recent years, great strides have been made in the 3D acquisition domain with the advancement of the reconstruction of building models using lidar, photogrammetry, and novel approaches such as regression and analysing satellite signal obstructions, and the

* Corresponding author at: Department of Architecture, National University of Singapore, Singapore.
E-mail addresses: e0008082@u.nus.edu (H.E. Pang), filip@nus.edu.sg (F. Biljecki).

https://doi.org/10.1016/j.jag.2022.102859
Received 13 December 2021; Received in revised form 25 May 2022; Accepted 2 June 2022

(a) (b) (c)

Fig. 1. Despite their dense geographical coverage and volume of imagery, SVI can be challenging for inferring information on buildings, as images often contain incomplete views on them due to occlusion by vegetation and other objects (examples (a) and (b)). Therefore, often only one image of a clear view of a particular building is available (example (c)). These images have been sourced from Mapillary, a volunteered SVI platform.

development of tools (Dehbi et al., 2020; Milojevic-Dupont et al., 2020; Nys et al., 2020; Biljecki, 2020; Cao and Huang, 2021; Lines and Basiri, 2021; Ledoux et al., 2021; Gui and Qin, 2021). Despite the advancements, there are two constraints, which we believe will largely continue to persist for some time. First, geographical coverage is still limited — 3D building models remain available in a minority of regions, primarily due to lack of essential data such as airborne point clouds to generate them. Second, their level of detail (LoD) remains basic — most of the models are in LoD1 (block models) according to the definition of CityGML (Gröger and Plümer, 2012; Kutzner et al., 2020). While they have demonstrated their value for many use cases, there is an increasing demand for data in higher level of detail (Wysocki et al., 2021; Biljecki et al., 2021; Virtanen et al., 2021; Yamani et al., 2021; Peters et al., 2022). Generating higher-LoD data remains constrained, and there is little research on enhancing the level of detail of existing, low-LoD 3D models such as block models. Third, efforts to characterise the 3D urban form at the large-scale do not regard individual buildings, i.e. they provide aggregate values at a coarse grid from satellite observations (Geis et al., 2019; Chen et al., 2020b; Li et al., 2020; Frantz et al., 2021; Esch et al., 2022; Zhu et al., 2022).

Simultaneously, the geospatial landscape has been witnessing a surge in another instance of spatial data — street view imagery (SVI) (Biljecki and Ito, 2021; He and Li, 2021). Commercial services such as Google Street View (GSV)[1] and their crowdsourced counterparts, e.g. Mapillary[2] and KartaView[3], now supply an enormous amount of georeferenced images around the world, often at a dense geographical resolution (Ma et al., 2019; Zhang et al., 2020; Ding et al., 2021). This source of data has been thoroughly exploited for a range of urban studies, some of which involve extracting information of buildings (Kruse et al., 2021; Ito and Biljecki, 2021; Rosenfelder et al., 2021; Yohannes et al., 2021; Kang et al., 2021; Helbich et al., 2021; Szcześniak et al., 2021; Cinnamon and Gaffney, 2021; Zhang et al., 2021a; Zhang et al., 2021b; Yin et al., 2021).

Much of the imagery covers urban areas that have not been subject of 3D city modelling, including cities that do not have even building footprints to begin with. As such, SVI has been employed for 3D building reconstruction using traditional techniques (Cavallo, 2015; Torii et al., 2009; Micusik and Kosecka, 2009). However, these approaches utilising SVI often require multiple images to form a dense correspondence, which is often not suitable for SVI as buildings are often partially or fully occluded by vegetation, vehicles, and other objects (Zhang et al., 2021c) (Fig. 1), and therefore are not available in more than one or two unobstructed images. Furthermore, because imagery is taken from roads, usually not all sides of a building are captured, presenting a considerably limited view of buildings unlike in counterparts derived from aerial or satellite platforms.

In recent years, 3D object reconstruction from single or few images has been gaining traction in computer vision (CV) (Han et al., 2019; Fu et al., 2021). Even with single 2D images, nascent methods have achieved remarkable results in inferring the 3D geometry of an object (Fu et al., 2021). However, these 3D object reconstruction methods were trained largely on simple symmetric household objects from indoor scenes such as chairs, and large irregular objects in outdoor scenes such as buildings have not been in their focus.

Connecting the dots described above, this research aims to bridge the notable gap of 3D building reconstruction from single SVI using recent developments in CV such as image-to-mesh reconstruction techniques. As additional data such as satellite imagery and 2D building footprints are becoming increasingly available around the world (Xie et al., 2019; Huang and Wang, 2020; Li et al., 2020; Jochem and Tatem, 2021; Fleischmann et al., 2021; Sirko et al., 2021; Leonard et al., 2022), we investigate whether they can be used to aid the reconstruction. Furthermore, as block (LoD1) 3D models are already available in some cities, we also endeavour on understanding the benefit of SVI and CV in enhancing the LoD of existing coarse 3D building models. Considering a broader context, the research aims to understand whether recent CV techniques designed for indoor scenes can be adopted in the geospatial (i.e. outdoor and urban scale) domain, enabling cross-fertilisation among the two fields, providing also input to the CV community whether the developed approaches are applicable on other objects such as buildings and in the geospatial realm.

Hence, this multi-pronged work focuses on using a single street-level image of the building and investigates whether (1) it can be used to directly reconstruct 3D building models with no other data; (2) the availability of a top view outlining the building footprint improves the performance of the 3D reconstruction; and (3) a single SVI can be used to enhance coarse 3D building models potentially resulting in their improved usability and increased visual appeal. Once the 3D models are generated, on top of evaluating the geometric accuracy, the subsequent objective is to assess their usability in spatial analyses — estimating the envelope area and volume of buildings, providing an evaluation of the performance that may be easier to interpret in the geospatial context.

2. Related work

2.1. Approaches in 3D building reconstruction

Existing approaches to generate 3D building models are primarily photogrammetry, laser scanning, and procedural modelling (Suveg and Vosselman, 2004; Vosselman and Dijkman, 2001; Demir and Baltsavias, 2012; Jovanović et al., 2020; Goetz, 2013; Martinovic, 2015; Bshouty et al., 2020). Each method is characterised by the level of detail it can achieve, which has an impact on its usability. The process of generation

[1] https://www.google.com/maps
[2] https://www.mapillary.com
[3] https://kartaview.org

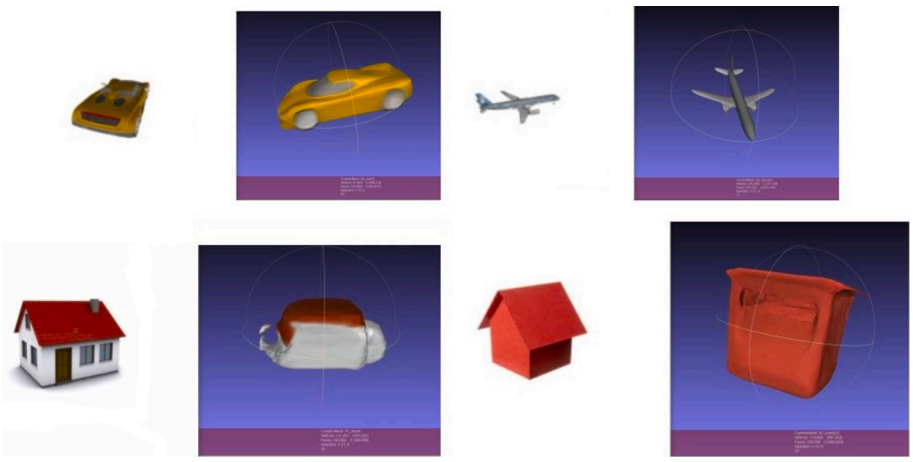


Fig. 2. The top row displays the inference results using DVR model by Niemeyer et al. (2020) trained on 13 categories (bench, cabinet, car, chair, monitor, lamp, speaker, firearm, couch, table, cellphone, plane, watercraft) of the ShapeNet dataset (Chang et al., 2015). The bottom row demonstrated the outputs of pretrained DVR model when inferred on building imagery, suggesting the need to adapt and advance the state of the art to suit 3D bui.lding modelling.

is also subject to the availability of existing data, equipment, and budget.

Photogrammetric and airborne survey methods are usually able to generate photorealistic 3D building models. However, acquisition of aerial images and lidar point clouds often utilises equipment that is expensive and not widely accessible, and it requires further processing using specialised software to obtain building models (Fan et al., 2021; Jovanović et al., 2020).

3D building models could also be generated via extrusion from 2D footprints (Dukai et al., 2020; Stoter et al., 2020). Nonetheless, such a method can only generate low-LoD models and relies on the availability and accuracy of building footprint and height information. Therefore, the barrier to achieve higher-LoD 3D models remains even in such locations. Further, while footprint data might be attainable from OpenStreetMap in many locations around the world (Fan et al., 2014; Li et al., 2022; Wang et al., 2021; Komadina and Mihajlovic, 2022) and height data can be approximated from other building attributes such as the number of storeys, data from non-authoritative sources might contain errors that propagate to geometrically inaccurate 3D models, potentially leading to unreliable results when used for spatial analyses. If height and/or footprint data are unavailable, which is often the case, procuring such information might be complex, motivating the development of a standalone method that does not require such information.

2.2. Approaches in 3D building reconstruction and extracting building information using SVI

Generating 3D building models from SVI has been of continuous interest (Zhang et al., 2021a), dating back to the work by Torii et al. (2009). Structure from Motion (SfM) techniques have been employed to reconstruct buildings by stitching a series of GSV images with known GPS location and camera internal parameters (Lee, 2009; Torii et al., 2009). For example, Bruno and Roncella (2019) investigated reconstruction using GSV photogrammetric strip but reported hit-or-miss results. SfM methods could only reconstruct the visible portions of the building, are computationally inefficient, and strongly rely on multiple images (the more the better) from different angles (Fan et al., 2021).

SVI has also been used to infer building characteristics such as number of storeys and elevation of the ground floor (Kim and Han, 2018; Taubenböck et al., 2018; Kraff et al., 2020; Rosenfelder et al., 2021; Pelizari et al., 2021; Ning et al., 2021), which may be indirectly used to reconstruct 3D building models via extrusion when their footprints are available. For example, Chu et al. (2016) used street-level imagery to approximate the floor height and location of building features such as windows and doors, and use that information to reconstruct buildings procedurally. For a recent related work, see the publication of Fan et al. (2021). However, these approaches still rely on having a building footprint, inhibiting the replicability of the method in most parts of the world.

2.3. 3D building reconstruction using deep learning methods

Convolutional Neural Networks (CNNs) have been used in the computer vision domain to tackle problems such as image classification (Chen et al., 2018), segmentation (Badrinarayanan et al., 2017; Wang et al., 2017), object detection (Song et al., 2017), and image super-resolution (Johnson et al., 2016; Kim et al., 2016).

Image segmentation has been predominantly applied to derive building footprints from aerial and satellite images (Alidoost and Arefi, 2015; Mahmud et al., 2020). Yu et al. (2021) combine 5 or more aerial images to reconstruct detailed (LoD2) 3D models by estimating roof planes. Both approaches require height information and are developed for top view (aerial or satellite) imagery. Bacharidis et al. (2020) reconstruct detailed 3D building surfaces from a single RGB image by estimating depth and incorporating façade segmentation based on generative adversarial networks. Yet, this method can only reconstruct the visible surface, resulting in an incomplete 3D building model. The increasing prominence of point cloud data has also given rise to deep learning applications such as classification of roof types and building elements using 3D point clouds (Wichmann et al., 2019).

2.4. 3D reconstruction of objects using deep learning

The success of deep learning approaches applied on 2D images, coupled with large amounts of openly available 3D data, spurred the progress in 3D reconstruction tasks (Fu et al., 2021). Trained models demonstrate strong reconstruction ability by being able to infer the 3D geometry given only a single back or side-view image. Similar to how humans can infer 3D shapes from images based on our own rich experience accumulation, training a model to reconstruct 3D geometries from 2D images requires a sufficiently large dataset of 3D models (Han et al., 2019). Most of the work is focused on indoor scenes and objects such as furniture and cars.

Nonetheless, trained models are often unable to generalise well to new unseen categories (Tatarchenko et al., 2019). This corrugates to our findings when we applied the same pre-trained model to building images in the exploratory phase of this research (Fig. 2).

To the extent of our knowledge, there are no known studies that have applied image-to-mesh 3D object reconstruction techniques in the CV domain to outdoor buildings in the function of generating 3D building models. Reconstruction using *in-the-wild* images can be rather

Table 1
Investigated scenarios using a single street view image.

Approach	Input		Output
	SVI	Aiding data	
1	Single image	–	3D building mesh
2	Single image	Top view/ footprint	3D building mesh
3	Single image	Block (LoD1) model	Enhanced 3D building mesh

challenging, and we investigate the validity of applying and modifying models developed in the CV domain for real-world outdoor images on buildings that may suffer from various imperfections. While this method presents a potential to tap on readily available real-world SVI for 3D building reconstruction, the lack of quality training data is an obstacle. We have supplemented our training using synthetic models, and this presents a potential to use low-cost parametric models to aid the model in learning geometries that are out of distribution from the ShapeNet dataset. Domain generalisation (from synthetic to real world scenes) is also explored in the work.

3. Method

3.1. Scenarios

Based on the context described in the previous two sections, we are particularly interested in the following scenarios (Table 1), which are an important consideration in devising the method:

1. Reconstruction of a 3D building mesh only from a single street-level image of the building.
2. Reconstruction of a building mesh from two images — a single street-level image of the building supplemented with its top view (e.g. available from segmented satellite imagery). Such an image is intended to represent an insight in the building footprints, and technically images of footprints derived with other techniques can be used (e.g. rendered footprint data).
3. Enhancing a coarse LoD1 mesh model to a more detailed building mesh using a single street-level image.

Each of these approaches is important in practice and has its application. First, many areas around the world do not have building footprint and/or 3D building model coverage (nor data such as point clouds or high resolution satellite imagery required to generate them), but they are dotted with high-quality SVI. Therefore, this work may be relevant in increasing the global availability of 3D geoinformation. By extension, it may contribute to providing 3D building models in areas that previously did not have any 3D coverage, at least with meshes that approximately convey the mass of a building, and — on a large scale — the 3D urban form. Second, satellite imagery and building footprints are increasing in coverage in some areas. We include them to investigate their usability as an ingredient in our method, similarly to their role in extrusion. Third, while the goal of this research is generation of 3D building models, it cannot be discounted that they already exist for many cities around the world. However, their LoD is most often simple. We aim to investigate whether we can still use SVI and CV also in such cases to have their LoD *upgraded*. In this scenario, we do not use footprint information, as LoD1 models in practice are often based on footprints.

3.2. Workflow and study area

The developed methodology consists of three steps: data collection and processing (Section 3.3), modelling (Section 3.4), and evaluation (Section 3.5) (Fig. 3).

The selected study area of this project is Helsinki, the capital and most populous city of Finland, primarily owing to the rich availability of open data required for the different steps in the method and evaluation of the results.

3.3. Data sources and data preparation

The datasets required for training a deep learning model are 3D building models (meshes) and street view and top view images. The sources and the preparation of the datasets for training is one of the pillars of this work, and it is described in the continuation together with the tools used in the process.

3.3.1. 3D building models

The local government of Helsinki openly released building mesh models[4], which we use in our work. They are highly detailed and geometrically accurate (Fig. 4). Further, the city provides also semantic 3D building models in both LoD1 and LoD2, a rare instance in combination with meshes, which we consider in the evaluation of reconstructed volumes and surface area (Section 3.5), since for that purpose they are more appropriate than mesh models.

The entire region of Helsinki is split into tiles. These raw reality mesh tiles were generated from aerial images, collected from an aircraft with five cameras that provide 80% length coverage and 60% side coverage, resulting in approximately 7.5 cm ground sampling distance (GSD). Aerial triangulation, dense image matching, and mesh surface reconstruction were all performed to reconstruct meshes from these images. 196 reality mesh tiles containing rough segmentation labels for 6 classes (Ground, Vegetation, Building, Water, Vehicle, and Boat) were provided in the work of Gao et al. (2021).

The next step includes cleaning the mesh tiles. Reality mesh tiles containing rough segmentation labels obtained from Gao et al. (2021), which may include incorrect labels and require manual cleaning and annotation using the Mesh Annotator Tool developed by the authors of the cited publication.

Afterwards, geometries classified as buildings were segmented, and standalone buildings were saved individually using MeshLab (Cignoni et al., 2008).

Because watertight meshes are required for volume estimation, the bottom of all meshes was cut planarly to obtain a flat surface, and all holes were filled. This process was conducted using Rhino (McNeel, 2010), and it is illustrated in Fig. 5, which also doubles as an example of a segmented building mesh model. The final stage of data preparation of includes rescaling the meshes and generalise them to construct LoD1 models for their role in the third scenario (described in Section 3.1). The watertight building mesh models were resized to unit cube to avoid scale ambiguity, and a new origin was assigned by aligning the bounding box centre to the origin coordinate using PyMeshLab, a Python API for MeshLab (Cignoni et al., 2008). The bounding boxes representing coarse meshes (in LoD1.0 and 1.1 according to the classification devised by Biljecki et al. (2016)), a common form of 3D building models, were also saved as the departure for mesh-refinement (third scenario).

It may be observed that this data preparation process using real-world meshes is intricate and time-consuming to ensure that proper and correct data is used and to avoid errors that are common in real-world datasets (Zhao et al., 2018; Noardo et al., 2021). To limit manual work in data preparation and yet obtain a sufficiently large and diverse dataset, the reconstruction method is supplemented with synthetic building models, using an existing library which we extend, and we use the aforementioned models for testing. Following the synthetic route, 1018 building models were generated in a parametric approach using the pipeline developed by Fedorova et al. (2021). These buildings models contain relatively simple footprints and were all flat-roof type, which are not representative of the buildings in the study area. Therefore, we enhance the training dataset with more complex buildings from

[4] https://hri.fi/data/en_GB/dataset/helsingin-3d-kaupunkimalli

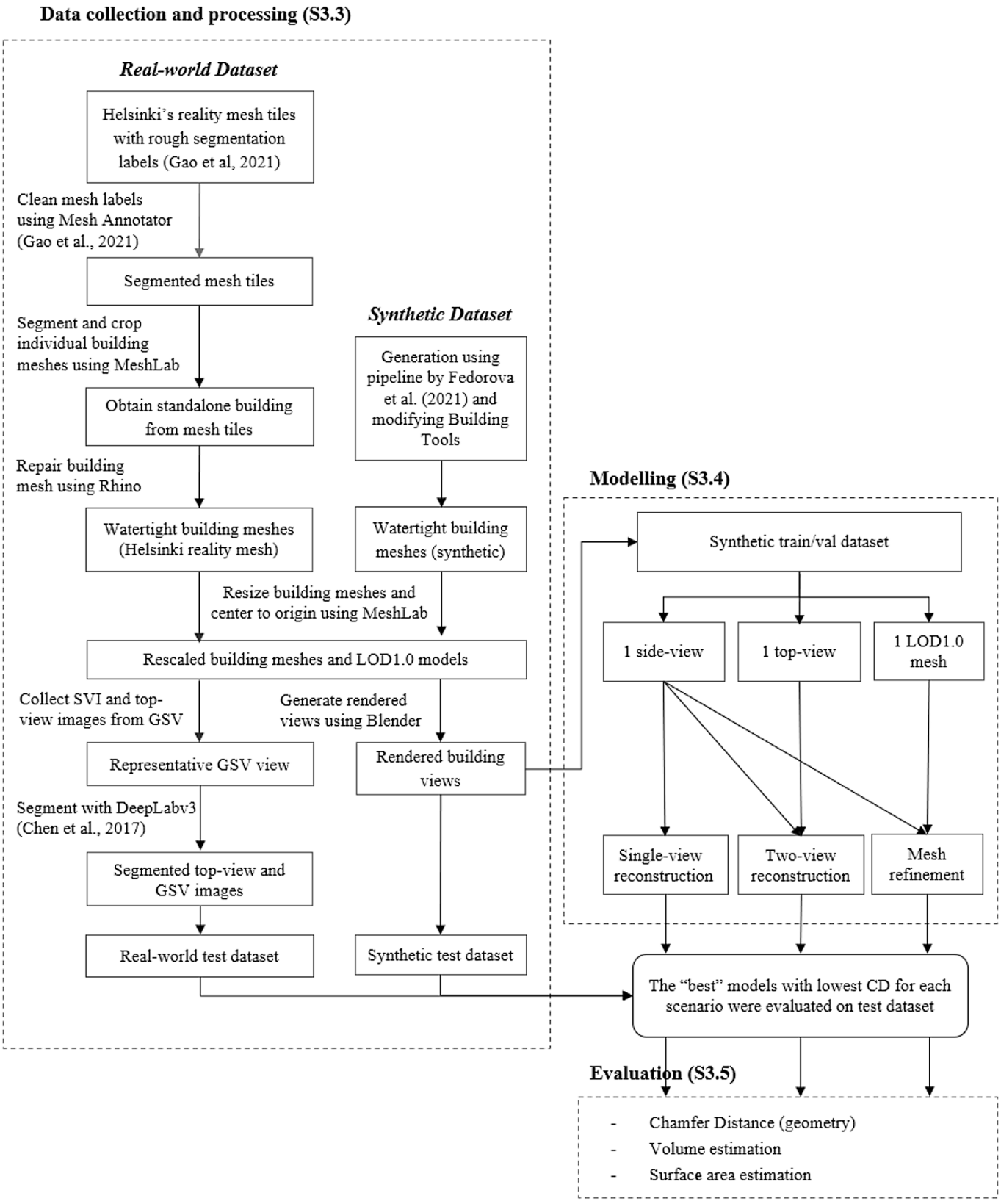


Fig. 3. Overview of the workflow of this study.

another source: using Building Tools[5], an open-source Blender package for manual building construction, which we modified to increase the diversity of the generated buildings. This allowed us to randomly generate 2770 textured building models with more elaborate footprints, height, and various roof types e.g. gable, hipped, flat, overhangs (Fig. 6)

[5] https://github.com/ranjian0/building_tools

(a) (b)

Fig. 4. 3D building models used in our work: (a) reality mesh models, used for training; and (b) semantic LoD2 models, which we use in the evaluation. Source of data: City of Helsinki.

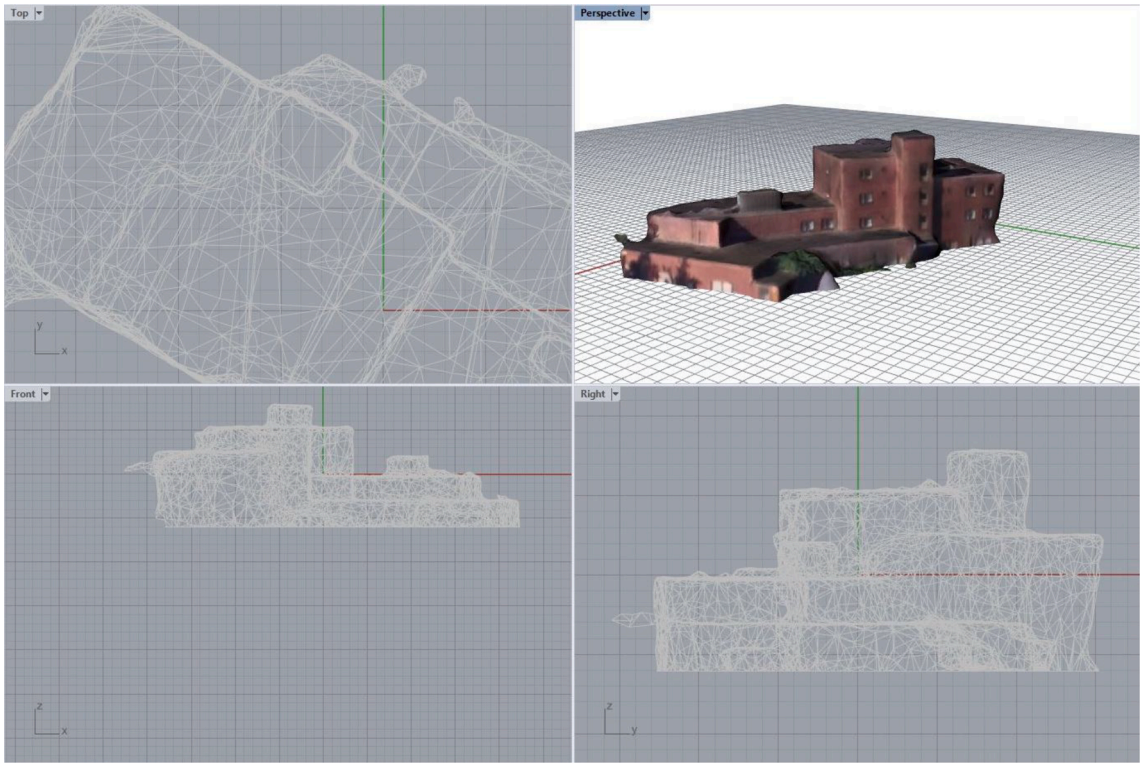


Fig. 5. Visualisation of a 3D building model during the data preparation phase after the segmentation stage to extract individual buildings and add planar ground planes to form a watertight model, enabling their use in certain applications to evaluate the usefulness of the reconstructed 3D buildings in spatial use cases.

in an automatic manner.

The subset of training data composed of synthetic instances does not require segmentation and refinement, motivating the use of the synthetic approach.

3.3.2. *Street view imagery and top views*

The geo-referenced real-world building meshes were located in Google Maps and Google Street View, and images of each building were manually sourced (street-level image from GSV and satellite image from Google Maps). After disregarding buildings that are not visible in GSV, 158 building meshes were associated with their corresponding SVI. Finding the most representative SVI was done manually as it is difficult to automate the process with the API and other steps (Fig. 7). Afterwards, collected SVI were segmented using DeepLabv3 (Chen et al., 2018) (Fig. 8), an open-source segmentation model, prior to the inference process described in the continuation.

For synthetic building meshes, 24 side-views for single-view reconstruction and their associated top-down views (for the second scenario, i.e. single SVI + top view reconstruction) were rendered using Blender. To simulate images captured from a bottom-up street-view perspective, a 25- and 35-mm field-of-view camera is preset at a −10-to-5-degree angle elevation and 0.7 to 0.95 distance, consistent with ShapeNet renderings (Choy et al., 2016). With a synthetic dataset, we are able to generate building images from a variety of viewpoints that would be difficult to collect in real world SVI, providing another reason to opt for synthetic models in the training process.

3.4. *Modelling*

Several deep learning model architectures were experimented with in this phase of the work. For single-view reconstruction, we have adopted the state-of-the-art Differentiable Volumetric Renderer (DVR) architecture by Niemeyer et al. (2020). While DVR can learn both textures and geometry, it is unable to be adapted for multiple inputs, which might help to improve the accuracy of the reconstruction mesh. Pixel2Mesh, SphereInit and VoxMesh architectures adopted from Gkioxari et al. (2019) were modified to take in multiple inputs for the latter two scenarios.

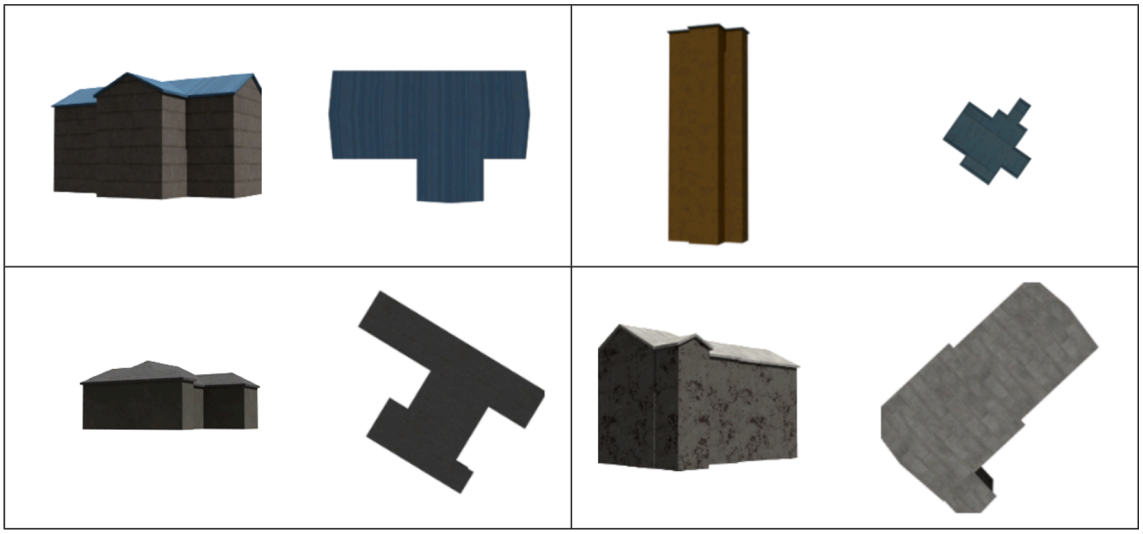


Fig. 6. Examples of building models generated from our pipeline that relies on modified open-source software, together with their side and top views used for training and testing.

Fig. 7. Retrieval of the most appropriate SVI representing a building. Many buildings are covered by multiple images from different angles, however, often only one or few images show an unobstructed and complete view of a building (affirming the motivation for this research). The imagery is obtained from Google Street View.

Fig. 8. Segmentation of a building from its street view image using *DeepLabv3*. The original image is obtained from Google Street View.

For single-view reconstruction, transfer learning was employed to fine-tune a pre-trained model (a saved network previously trained on a large dataset) to our specific task of building reconstruction. DVR was trained on 13 categories of 30 k ShapeNet meshes (Niemeyer et al., 2020). Since we have relatively fewer meshes for training, transfer learning would allow us to build upon the 3D reconstruction capabilities of the pre-trained model without having to start from scratch.

In DVR, occupancy network (Mescheder et al., 2019) and texture fields (Oechsle et al., 2019) were implemented in a single network. DVR takes as input an image x and N_p randomly sampled points. The N_p point coordinates (p_1, p_2, p_3) were encoded using a fully connected (FC) layer, five ResNet50 (He et al., 2016) blocks, followed by two FC layers. The

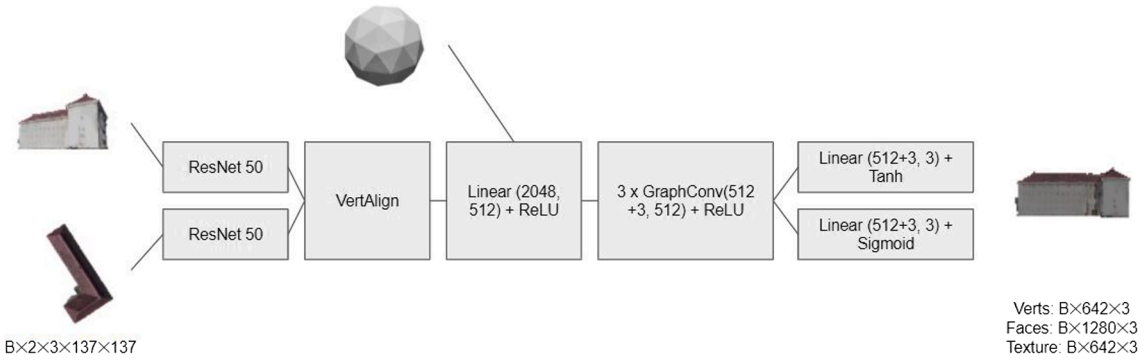


Fig. 9. Modified Pixel2Mesh architecture to take in two images.

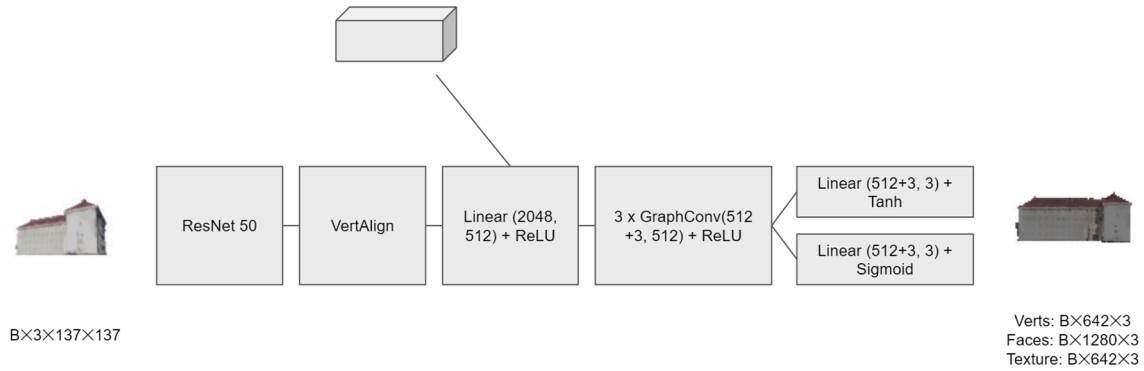


Fig. 10. Modified Pixel2Mesh model architecture to take in a single side-view and an initialised block mesh for refinement.

final output contains N_p one-dimensional occupancy probabilities and three-dimensional RGB colour values, from mesh models were extracted using marching cubes algorithm.

For transfer learning, we unfroze the last few encoder layers and all decoder layers of the pretrained model to learn higher order feature representations specifically for building reconstruction.

Moving on to the two-view reconstruction, architectures by Gkioxari et al. (2019) could only take a single image as input and requires modification for two-view reconstruction. Graph convolution networks (GCN) are commonly used for processing 3D meshes. Given feature vectors for each vertex in a mesh, graph convolutions compute new features by propagating information along mesh edges. The original Pixel2Mesh (implemented in TensorFlow by Wang et al. (2018)) learns to predict meshes by deforming and subdividing an initial sphere template using graph convolutions.

To gain more information about the building geometry, top-down images which are readily available from satellite imagery could be incorporated. To take two images as input, both the side and top-view images were encoded via a ResNet50 (He et al., 2016) backbone (initialized with ImageNet (Deng et al., 2009) weights) and image features were concatenated (Fig. 9). A sequence of 3 GCN layers each with 512 dimensions were stacked to aggregate information over local mesh regions.

Finally, we describe the approach to tackle the third scenario — mesh refinement. Instead of using an initialised sphere mesh for refinement, a coarse LoD1.0 model represented by the mesh's bounding box was taken as input (Fig. 10), with the aim of enhancing its level of detail.

3.5. Model evaluation

During training, Chamfer Distance (CD) (Fan et al., 2016) is the main loss metric to measure geometric accuracy, as it is typical in related work employed in computer vision. For each point x in pointcloud S_1 and y in pointcloud S_2, the algorithm of CD finds the nearest neighbor in the other set and sums the squared distances up. As such, the CD between pointclouds S_1 and S_2 is computed as follows:

$$d_{CD}(S_1, S_2) = \sum_{x \in S_1} \min_{y \in S_2} |x - y|_2^2 + \sum_{y \in S_2} \min_{x \in S_1} |x - y|_2^2 \quad (1)$$

For fine-tuning DVR for single-view reconstruction, a combination of CD, occupancy, and RGB loss was used. For training of two-view and mesh-refinement models, CD, normal, and edge loss were applied (more details in A).

To compute CD, 5000 points were sampled uniformly on the surface of the predicted and ground-truth meshes. For each point in the predicted point set, CD finds the nearest neighbor point in the ground-truth pointset to calculate the squared distance (Mescheder et al., 2019; Gkioxari et al., 2019) (Eq. 1). The sum of the squared distances is used to quantify the discrepancy in geometries between the predicted and ground-truth meshes.

However, a reader accustomed to the GIS domain, may notice that the metrics above might not give much insight into how do the generated models perform in spatial analyses. Therefore, on top of CD, the predicted mesh models were evaluated for volume and surface area estimation to investigate if they are suitable for particular spatial analyses. Overall, the three metrics should be read together as a whole. A reconstructed model with high accuracy in surface area and volume estimation might not necessary be geometrically accurate, which is something that could be conveyed by CD for a complete picture.

Surface area and volumetric measures were computed using MeshLab. As error would scale with mesh sizes, mean percentage error was used to quantify the discrepancy between predicted and ground-truth mesh volume and surface area. These metrics are not common in computer vision, but we introduced them to gauge the performance of the reconstructed 3D models in spatial analyses, as many geospatial use

Table 2
Distribution of building models in the train, validation, and test sets for synthetic and real-world models.

	Synthetic	Real-world
Train (70%)	2645	-
Validation (15%)	567	-
Test (15%)	566	158
Total	**3778**	**158**

Table 3
Chamfer Distance (CD) of the best model for various reconstruction methods on synthetic dataset.

Scenario	CD (5 s.f.)
1. Single-view reconstruction	0.34911
2. Two-view reconstruction	0.03571
3. Mesh-refinement	0.03449

cases require them. Also, as the models may not always be visually appealing, we investigate whether we can nevertheless use them for spatial analyses that may not be affected by the appearance of the reconstructed model (as hinted at in the first paragraph in Section 1).

4. Results

During training, all models were trained and validated on 2645 and 567 synthetic building meshes, respectively (Table 2). During test time, they were evaluated on 566 synthetic meshes and 158 real-world meshes.

Tables 3 and 4 summarise the CD, mean and standard deviation of the % errors in volume and surface area estimation for the best model of different scenarios. For all metrics, the lower the better.

4.1. Single-view reconstruction

Single-view reconstruction has the highest CD of 0.34911 and is least accurate in terms of geometry reconstruction. That is not surprising, given that this scenario is the one with least and limited input data.

The fine-tuned model achieved a lower CD (0.34911) compared to the untuned DVR model (1.5364) (Table 5), indicating that the model is more suited to the task of building reconstruction after fine-tuning. CD of models trained on ShapeNet dataset (Choy et al., 2016) range from 0.191 (Niemeyer et al., 2020) to 1.445 (Choy et al., 2016). Although CD is not directly comparable due to the use of different datasets, a CD of 0.34911 lies within the acceptable range of geometry reconstruction.

During test time, a single input image of the building's side-view was used to generate the predicted meshes. Reconstructed meshes were assessed qualitatively from both the side and top-view for synthetic (Figs. 11, 12) and real-world dataset (Fig. 13). A common observation among the reconstructed meshes is that they resemble the actual mesh from the input camera view, but not from other angles, foreseeable considering the limited view and irregular nature of buildings. In certain cases, the model can estimate the overall shape of the building, including even parts that are not visible from the input street view (Fig. 11). But in most cases, it is not able to do so, which is not surprising given the complex and unpredictable shapes building can have beyond

Table 5
Chamfer Distance (CD) for untuned and fine-tuned model.

Epoch	CD (5s.f.)
1018 (Untuned)	1.5364
1404 (Fine-tuned)	**0.34911**

the view we have from a single SVI (Fig. 12, 13), unlike symmetrical and simple objects such as furniture which are the primary subject of reconstruction in CV. This unpredictably may lead to substantial discrepancies in volume and surface area approximation.

Volume and surface area were more severely underestimated when evaluated on real-world dataset, with errors of −36.16% and −44.54% respectively. From Fig. 13, single-view reconstruction was often unable to approximate building footprint, especially when reconstructed from images where the building is front facing (Fig. 13 ID 6,7).

Due to the high errors in volume and surface area estimation in both synthetic and real-world datasets, the single-view reconstruction models would be unsuitable for such use cases. However, given that the overall mass of the building is inferred, these models could still be useful to indicate the rough 3D urban form at the urban scale.

There are two explanations for the limited effectiveness of single-view reconstruction. First, buildings are a lot more complex than most symmetric objects used in the CV community, i.e. furniture, cars, and airplanes, where the back view given a front- or side-view is predictable. Next, SVI are usually more challenging to deal with than clean and flawless imagery typically used in typical CV research. In addition, building footprint and roof shapes might be inferrable from an aerial image but would be completely inaccessible from a street-view perspective. Hence, images captured from a street-view perspective usually have greater information loss. We believe that such issues are fundamental to singular SVI, and that it may not be possible to ameliorate them in the near future.

4.2. Single SVI + top view reconstruction

The reconstruction augmented with the top view has a far better CD of 0.03571, suggesting the considerable role that the insight in the outline of the building provides. Various models with different parameters were trained (Table 6). In general, training for more epochs, slower learning rate, and larger graph convolution dimensions improve the model's reconstruction ability, as indicated by a lower CD (Table 6). Notably, addition of training samples with more complex footprints and a greater variety of roof shapes is beneficial for training, as seen from the large drop in CD from 0.0872 (E7) to 0.0357 (E8). A qualitative assessment of the best model (E8) with the lowest CD is provided in Figs. 14 and 15.

Our experiments demonstrated that the addition of the top-down view is effective in helping the model learn to predict an accurate building footprint, especially if the back-view is complicated to infer given only the front, limited view (Fig. 14).

A typical method to generate building models is from extrusion of building footprint. Such a method results in uniformly tall models (LoD1.1 or 1.2), and acquisition of external data sources such as height or number of storeys is necessary. Noticeably, having a street view

Table 4
Mean and standard deviation (SD) of errors for volume and surface area estimation for various reconstruction methods on synthetic and real-world dataset.

Attribute	Volume Estimation Error (%)				Surface Area Estimation Error (%)			
Dataset	Synthetic		Real-world		Synthetic		Real-world	
Metric	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Single-view reconstruction	−30.329	24.155	−36.167	14.241	−41.368	35.062	−44.541	22.416
Two-view reconstruction	−6.332	7.396	−10.466	8.713	−16.189	15.067	−32.579	16.881
Mesh-refinement	−6.142	6.129	−9.198	7.401	−14.512	13.774	−26.822	16.523

ID	Input image	Output model	Actual top view	Inferred top view
1				
2				

Fig. 11. Cases in Scenario 1 in which the method managed to infer the correct shape of the building.

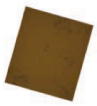
ID	Input image	Output model	Actual top view	Inferred top view
1				
2				
3				
4				
5				
6				

Fig. 12. Cases in Scenario 1 in which the method was not able to entirely infer the shape of the building.

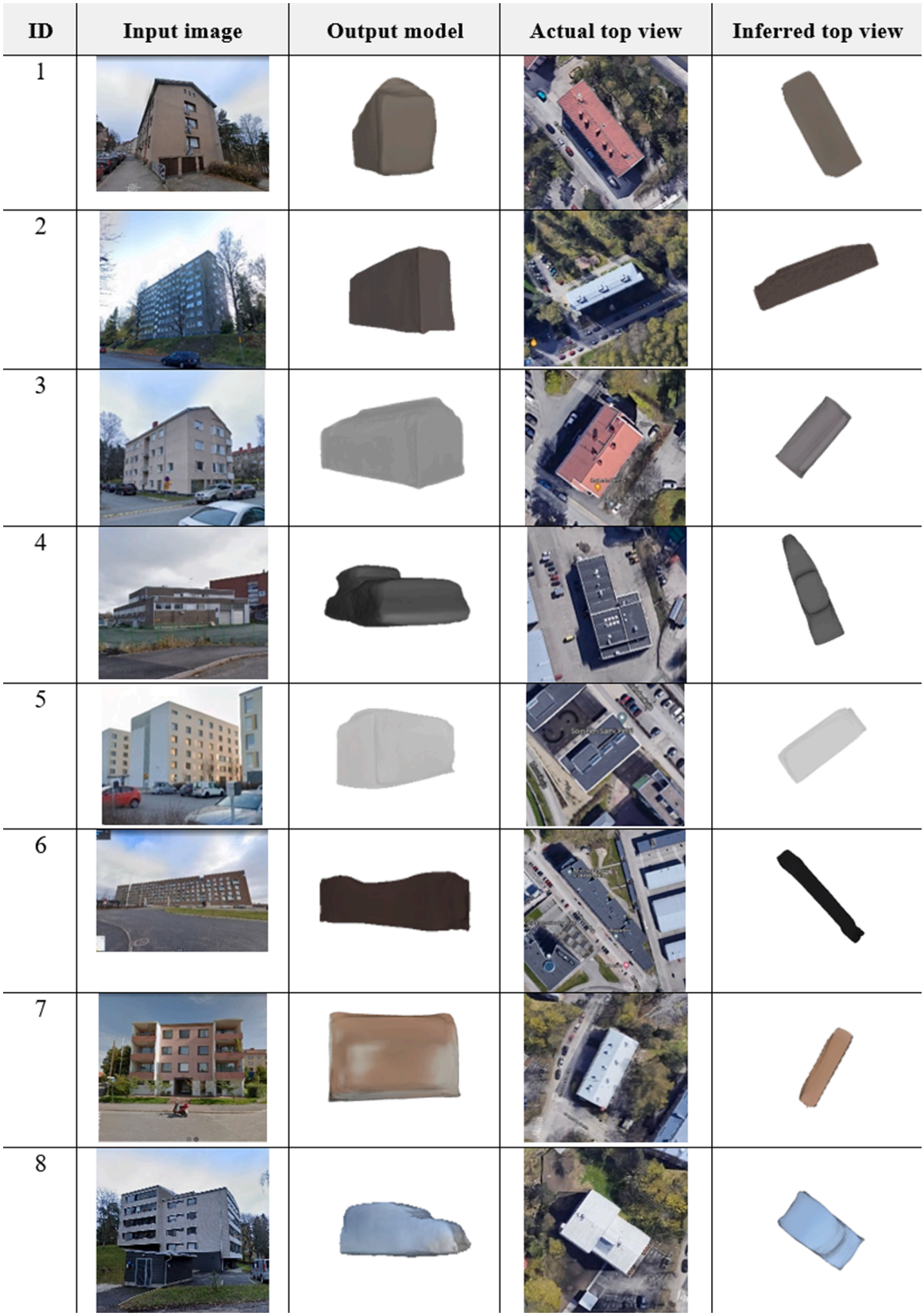


Fig. 13. Inference in Scenario 1 of fine-tuned model on a single SVI.

Table 6
Trained models and parameters for two-view reconstruction.

Experiments	Architecture	Samples	Epochs	Learning Rate	Graph Convolution dimensions	CD	Inference speed [seconds]
E1	Pix2mesh	712	20	0.0001	64	0.809234	0.36
E2	Pix2mesh	712	40	0.0001	128	0.599279	0.85
E3	Pix2mesh	712	45	0.00007	128	0.150874	0.85
E4	Pix2mesh	712	80	0.00005	128	0.246400	0.85
E5	SphereInit	712	80	0.00005	256	0.114465	0.69
E6	Voxmesh	712	40	0.00005	128	0.095553	0.96
E7	Voxmesh	712	80	0.00005	128	0.087227	0.96
E8	Voxmesh	2645	80	0.00005	128	0.035709	0.96

Fig. 14. Side and top-view of synthetic and predicted meshes (Scenario 2).

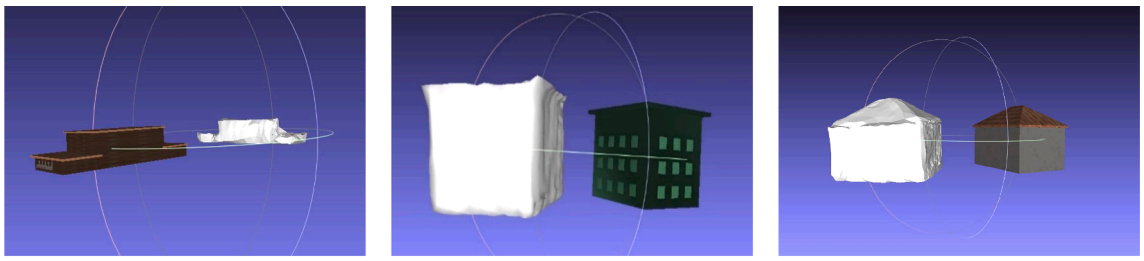


Fig. 15. Trained model was able to predict buildings (a) of non-uniform height (b) containing overhangs and (c) of non-flat roof types (Scenario 2).

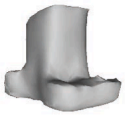
ID	Input image	Output model	Actual top view	Inferred top view
1				
2				

Fig. 16. Inferior results in reconstruction (occluded back-view and complicated footprint) (Scenario 2).

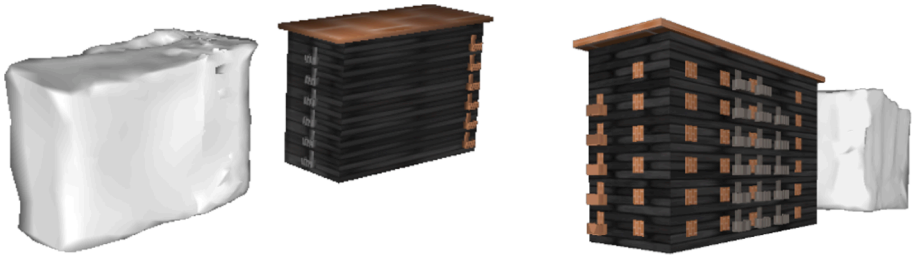


Fig. 17. The approach may not be able to derive detailed features (e.g. railings are not captured), leading to propagation of errors to spatial analyses, such as the underestimation of the envelope area.

image provides useful information to reconstruct meshes with non-uniform height, similar to that of a LoD1.3 model (Fig. 15). The trained model is also able to learn roof overhangs in some of the meshes (Fig. 15), which could avoid the problem of overestimating the volume by simply extruding based on the top-down view. Lastly, the trained model is able to predict different types of roof shapes i.e. hipped, gabled, providing models at the grade of LoD2, surpassing the LoD of most widely available models. In all cases, volume estimation from these meshes is more accurate than those extruded from building footprints with a uniform height.

When applied to volume estimation, the reconstructed mesh has a relative error of −6.332% and −10.47% on synthetic and real-world dataset, respectively. These errors may be acceptable for a number of spatial analyses, and it demonstrates that while visually the models may not be always very polished and visually appealing, this is not an issue for certain spatial analyses. In general, predicted models tend to underestimate the volume. Notably, the standard deviation is rather large, and it could be attributed to the model's inability to reconstruct complex objects such as those in Fig. 16, which may account for occasional large errors skewing the error metrics.

While the coarse predicted meshes are suitable for volume estimation, they are less ideal for envelope area estimation, having a relatively larger error of −16.19% for synthetic and −32.579% for real-world datasets (Table 4). The predicted meshes often lack finer detailed features such as the railed balconies in Fig. 17, which leads to an underestimation of surface area. Nevertheless, errors of this magnitude are not uncommon in datasets derived using traditional techniques and models generated using this approach are not necessarily inferior to many real-world datasets.

While the volume estimation is still within an acceptable range, we

Fig. 18. Side and top-view of real-world and predicted meshes (Scenario 2).

would not always recommend using the predicted models for surface area estimation, depending on the sensitivity of a particular use case to the magnitude of input errors. The higher errors when evaluated on real-world dataset could be due to the complexity of real-world building models. For instance, finer details such as balconies (Fig. 18 ID 4,6) and dormers on the roofs (Fig. 18 ID 2–3) were obviously not possible to

Table 7
Trained models and parameters for mesh-refinement.

Architecture	Epochs	# Samples	Lr	Graph Conv dimensions	CD	Inference speed [s]
Pixel2Mesh	30	712	0.00005	128	0.0949973	0.96 s
Pixel2Mesh	**80**	**2645**	**0.00005**	**128**	**0.0344915**	**0.96 s**

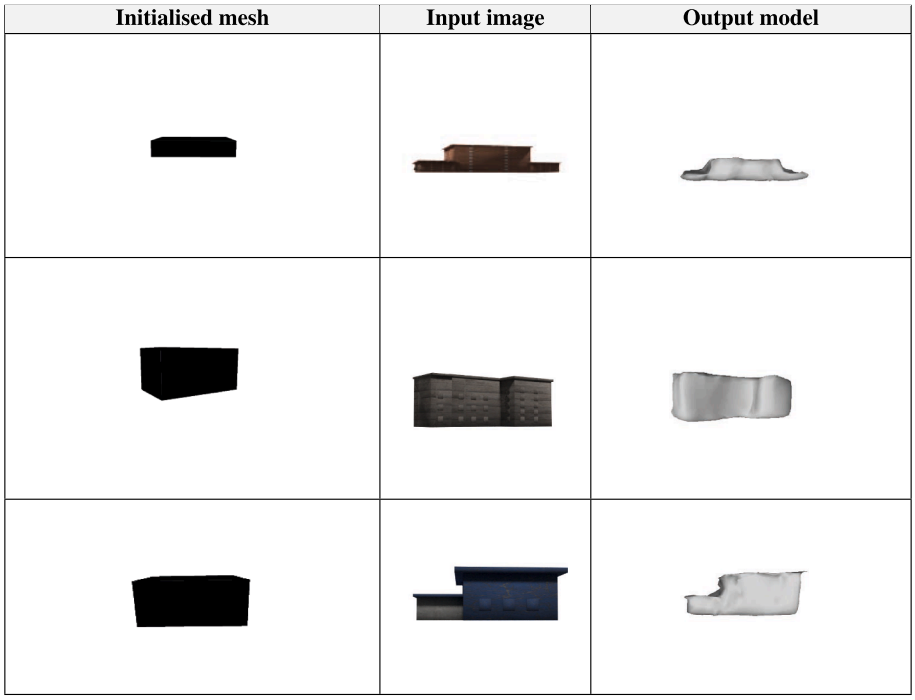


Fig. 19. Output meshes refined from LOD1.0 block model and input image.

predict, which leads to a considerable underestimation of building envelope area (-32.579%).

4.3. Mesh-refinement

The two scenarios described so far reconstruct a 3D building model from scratch. The final scenario focuses on using the approach to enhance the LoD of existing coarse 3D datasets. Similar to two-view experiments, training for longer epochs and more samples helped to improve reconstruction accuracy, as indicated by the lower CD (Table 7).

The mesh-refinement model also provides a slightly more accurate estimation of volume and surface estimation of −6.14% and −14.51%, as compared to −6.33% and −16.19% from two-view reconstruction (Table 4).

Errors for volume and surface area estimation are −9.198% and −26.822% when evaluated on real-world datasets, which are higher than synthetic dataset, but lower than that of two-view reconstruction.

The mesh-refinement method is more accurate as it takes a LoD1.0 model as input, which means that the height estimation is accurate and it does not need to be inferred from the image (Fig. 19). The implication of these results is that SVI can be used to refine the LoD of a coarse (block) 3D model, enhancing its quality and usability (Fig. 20). This approach may be scaled widely, as SVI is now available in most cities around the world, especially in those that are covered by low-LoD 3D city models.

4.4. Evaluation

Reconstruction with only a single side-view image as input proved to be a difficult task for buildings, but not impossible and the results may still be relevant for certain use cases. While the predicted models are unable to capture the finer details of the building geometry and underestimate the building surface area, meshes generated from two-view reconstruction or enhanced using mesh-refinement method could be used as a coarse approximation of the building form, akin to LoD1 models that are standard and perfectly usable in the research community and industry for a variety of use cases.

The most suitable method ultimately depends on the data availability. Although mesh-refinement fared slightly better than two-view reconstruction in terms of geometry, volume, and surface area estimation, it requires an LoD1.0 model as input. While building footprints are available widely (e.g. OpenStreetMap contains more than half billion buildings at the time of the submission of this paper), their heights are not. These buildings are concentrated in Europe and North America (Bshouty et al., 2020). Where LoD1.0 models are common, or there is data for their extrusion, we hope that our method can be used to increase their level of detail, contributing to their further application.

Considering the similar performance between both methods and data availability considerations, reconstruction aided by the delineated footprint seems to be the most promising scenario and most scalable approach for real-world applications.

In conclusion, while the results suggest that reconstructing 3D building models from single street view images will not be the most accurate method to do so, it will often be the only possible method (and providing entry-level 3D geoinformation) and the results are not unexpected given the sparsity of the input data. Nevertheless, the results may reveal the approximate urban form, which is relevant for applications such as population estimation, urban morphology, and noise propagation, in which only the coarse building mass is valuable and sufficient.

In the evaluation, for future work, we will consider involving airborne lidar data, which may provide further insights in the performance.

4.5. Limitations

The process of preparing real-world data exposes several issues, thus, together with the flexibility of procedural models and their unlimited rendering capabilities, it affirms our motivation to follow the synthetic route for much of the method. However, these benefits come at a price: since this study is mainly trained on synthetic buildings, the trained model might not generalise to some complex buildings that contain features it has not been trained on e.g. patio, dormers, windows. Additional studies could focus on generating more complex textured building models in an automatic manner or find an efficient method to collect real-world building meshes. Further, while we generate buildings as

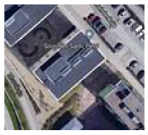
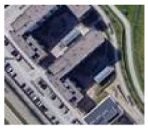
ID	Input image	Output model	Actual top view	Inferred top view
1				
2				
3				
4				
5				
6				
7				
8				
9				

Fig. 20. Side and top-view of real-world and predicted meshes (Scenario 3).

close to those in real-world using a rule-based approach, as it is the case with other work that uses synthetic data, there may be some cases of unrealistic instances that may not be found in the real-world.

In addition, the collection of representative building SVI is currently done manually. Future efforts could investigate if it would be possible to automatically determine usable SVI for reconstruction, e.g. by automatically detecting whether an image contains an obstructed view of a building.

Remaining key limitations pertain to SVI data, which often provides an incomplete view. This limitation is precisely why we embarked to conduct this research developing an approach to work with single images, but sometimes it was not possible to gather even that single usable and sufficiently clear image of a building. The scalability and coverage of the method is limited by the availability of SVI — there are still many cities without SVI, and in those cities where SVI is available, the coverage is often partial (e.g. collected from main roads, so buildings

along tertiary roads are not imaged). Thus, in practice, not all buildings in a city may be reconstructed, unlike in the case with data derived from airborne platforms.

5. Conclusion

Billions of street view images covering most of the world contain much information about buildings, which has not been fully exploited for generating 3D building models using deep learning. As of our knowledge, this work is the first application of image-to-mesh reconstruction techniques to outdoor scenes and buildings. We believe that focusing on single street view images is an important contribution because of vegetation and other obstacles in imagery in practice, as in many cases only one clear image of a particular building is available, inhibiting standard photogrammetric and other approaches that require multiple images. Further, this work is relevant as we believe that SVI as a data source for generating 3D city models will grow in importance, not only due to increasing coverage and quality but also because it offers some advantages over other sources, e.g. the ground-level perspective of SVI may provide more detailed insight into buildings and other features when they are obstructed to aerial and satellite platforms such as due to thick canopies.

The method that we have developed outputs a 3D mesh model from a single image without additional information such as building footprint or height, which are often not available, and it presents as a new method for acquiring 3D geospatial data. For cities with coarse 3D models, mesh refinement could be applied to enhance existing 3D building datasets. In addition, by coupling 3D and SVI data, we also present a contribution in the integration of these two sources.

We find that single-view building reconstruction using street-level imagery may provide models that indicate their approximate size and shape, but accuracy remains constrained primarily because of the inherent nature of SVI — gathering information on buildings is limited from a single or sparse horizontal view, and the parts of the building not visible in the image remain difficult to predict. This limitation is in contrast with the convenience of indoor scenes and symmetric objects such as furniture, which have been the main focus of such methods. However, augmenting the method with footprints, which are available in many places around the world, may provide sufficient information for 3D reconstruction, with results comparable to the mesh-refinement method.

We have demonstrated that a single SVI of the building side-view might provide vital information such as roof shape and overhangs, or if the building is of non-uniform height (improving the detail of LoD1 models, which do not regard the roof shape). This allows the trained model to reconstruct buildings that are more geometrically accurate as compared to simply extruding a basic volumetric shape from building footprints. Although the reconstructed models lack fine-grained details and are not necessarily usable for all visualisation purposes, the predicted data may be useful for spatial analyses such as volume computation, which are important to use cases in energy simulations, population estimation, and more (Mathews et al., 2019; Fibæk et al., 2021). Further, it may also support studies in urban morphology, noise propagation, and change detection (Rastiveis et al., 2013; Xu et al., 2017; Vitalis et al., 2019; Stoter et al., 2020; Meouche et al., 2021; Chen et al., 2021; Lu et al., 2021), for which such data may be of sufficient detail and quality.

As SVI is now available in most countries worldwide, the results indicate that our method can contribute towards deriving rapidly and cost-effectively the 3D urban form, paving the way to low-cost large-scale 3D reconstruction, which may serve well locations where 3D models are not available, which is — unlike SVI — the majority of the world. Obtaining 3D building models in regions that do not have any, may enable a number of 3D geospatial analyses locally for the first time and may even result in new applications catering to local challenges in regions that are seeing their first instances of 3D models. For example, the novel application of 3D building models to estimate the potential for urban farming was introduced by Palliwal et al. (2021) shortly after the first open 3D dataset of the study area became available, contributing to the topical local challenge of investigating the potential of buildings as a venue for food production (Song et al., 2021).

An idea for future work would be to generate building models with textures and finer details, as textured models are useful for visual-based applications such as augmented reality and positioning aiding. The current approach utilises supervised training which requires 3D mesh models that are time-consuming and laborious to collect. Additional research could investigate unsupervised approaches.

Further, as satellite methods to measure the height of urban blocks have been developing (Geis et al., 2019; Chen et al., 2020b; Li et al., 2020; Frantz et al., 2021; Esch et al., 2022; Zhu et al., 2022), we plan to investigate whether our work can aid such efforts, e.g. to delineate the form of individual buildings.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We gratefully acknowledge the help of Weixiao Gao (Delft University of Technology) and the valuable comments by the editor and the reviewers. We appreciate the data sources used for this research and the work of the open-source community. We thank the members of the NUS Urban Analytics Lab for the discussions. This research is part of the project Large-scale 3D Geospatial Data for Urban Analytics, which is supported by the National University of Singapore under the Start Up Grant R-295–000-171–133.

Appendix A

A.1. Additional losses for single-view reconstruction experiment – Occupancy loss and RGB Loss

For training the differentiable volumetric renderer model for single-view reconstruction, binary cross-entropy (BCE) loss will be used, and it is defined following the paper of Mescheder et al. (2019) as:

$$L_{BCE} = -\frac{1}{N}\sum_{n=1}^{N}[p_n \log q_n + (1-p_n)\log(1-q_n)]$$

Here, N represents the total number of points in the 3D occupancy grid, p_n is the ground truth probability (1 or 0) of the filled occupancy, and q_n is the prediction probability.

A.2. Additional losses for multi-view reconstruction and mesh refinement experiments – Normal and edge loss

For training the mesh deformation model for multi-view reconstruction and mesh refinement using a single image, a weighted sum of the CD, normal distance, and edge loss was used to calculate the mess loss.

The absolute normal distance is given by:

$$d_{Norm}(S_1, S_2) = -\sum_{x \in S_1} \min_{y \in S_2} |U_x - U_y| - \sum_{y \in S_2} \min_{w \in S_1} |U_x - U_y|$$

To measure the reconstruction goodness, CD and normal distances penalize mismatched positions and normals between two point clouds but minimizing these distances results in degenerate meshes (Gkioxari et al., 2019). High-quality mesh predictions require additional shape regularisers. Following Gkioxari et al. (2019), an edge loss is used,

$$L_{edge}(V, E) = -\frac{1}{|E|} \sum_{(v,v') \in E} |U_x - U_y|_2^2$$

where E contains the set of edges of the predicted mesh. During training, the objective is to minimise the mean of the mesh loss, which is a composite of CD, normal distance, and edge loss.

References

Ahmed, F.C., Sekar, S., 2015. Using three-dimensional volumetric analysis in everyday urban planning processes. Appl. Spatial Anal. Policy 8, 393–408. https://doi.org/10.1007/s12061-014-9122-2.

Alidoost, F., Arefi, H., 2015. An image-based technique for 3D building reconstruction using multi-view UAV images. Int. Arch. Photogram., Remote Sens. Spatial Inform. Sci. 40, 43. https://doi.org/10.5194/isprsarchives-XL-1-W5-43-2015.

Bacharidis, K., Sarri, F., Ragia, L., 2020. 3D building façade reconstruction using deep learning. ISPRS Int. J. Geo-Inform. 9, 322. https://doi.org/10.3390/ijgi9050322.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39, 2481–2495.

Bahu, J.M., Koch, A., Kremers, E., Murshed, S.M., 2014. Towards a 3D spatial urban energy modelling approach. Int. J. 3-D Inform. Model. (IJ3DIM) 3, 1–16. https://doi.org/10.4018/ij3dim.2014070101.

Beran, D., Jedlička, K., Kumar, K., Popelka, S., Stoter, J., 2021. The Third Dimension in Noise Visualization – a Design of New Methods for Continuous Phenomenon Visualization. The Cartographic Journal 1–17. https://doi.org/10.1080/00087041.2021.1889450.

Biljecki, F., 2020. Exploration of open data in Southeast Asia to generate 3D building models. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences VI-4/W1-2020, 37–44. doi:10.5194/isprs-annals-vi-4-w1-2020-37-2020.

Biljecki, F., Ito, K., 2021. Street view imagery in urban analytics and GIS: A review. Landscape and Urban Planning 215, 104217. https://doi.org/10.1016/j.landurbplan.2021.104217.

Biljecki, F., Ledoux, H., Stoter, J., 2016. An improved LOD specification for 3D building models. Comput. Environ. Urban Syst. 59, 25–37. https://doi.org/10.1016/j.compenvurbsys.2016.04.005.

Biljecki, F., Lim, J., Crawford, J., Moraru, D., Tauscher, H., Konde, A., Adouane, K., Lawrence, S., Janssen, P., Stouffs, R., 2021. Extending CityGML for IFC-sourced 3D city models. Automation in Construction 121, 103440. https://doi.org/10.1016/j.autcon.2020.103440.

Bizjak, M., Žalik, B., Štumberger, G., Lukač, N., 2021. Large-scale estimation of buildings' thermal load using LiDAR data. Energy and Buildings 231, 110626. https://doi.org/10.1016/j.enbuild.2020.110626.

Braun, R., Padsala, R., Malmir, T., Mohammadi, S., Eicker, U., 2021. Using 3D CityGML for the Modeling of the Food Waste and Wastewater Generation—A Case Study for the City of Montréal. Frontiers in Big Data 4, 662011. https://doi.org/10.3389/fdata.2021.662011.

Bruno, N., Roncella, R., 2019. Accuracy assessment of 3D models generated from Google Street View. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences.

Bshouty, E., Shafir, A., Dalyot, S., 2020. Towards the generation of 3D OpenStreetMap building models from single contributed photographs. Comput. Environ. Urban Syst. 79, 101421.

Cao, Y., Huang, X., 2021. 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 Sens. Environ. 264, 112590. https://doi.org/10.1016/j.rse.2021.112590.

Cavallo, M., 2015. 3D city reconstruction from Google Street View. Comput. Graph. J.

Chang, A.X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., et al., 2015. Shapenet: An information-rich 3d model repository. arXiv preprint arXiv:1512.03012.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2018. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence 40, 834–848.

Chen, S., Zhang, W., Wong, N.H., Ignatius, M., 2020a. Combining CityGML files and data-driven models for microclimate simulations in a tropical city. Build. Environ. 185, 107314. https://doi.org/10.1016/j.buildenv.2020.107314.

Chen, T.H.K., Qiu, C., Schmitt, M., Zhu, X.X., Sabel, C.E., Prishchepov, A.V., 2020b. Mapping horizontal and vertical urban densification in Denmark with Landsat time-series from 1985 to 2018: A semantic segmentation solution. Remote Sens. Environ. 251, 112096. https://doi.org/10.1016/j.rse.2020.112096.

Chen, W., Wu, A.N., Biljecki, F., 2021. Classification of urban morphology with deep learning: Application on urban vitality. Comput. Environ. Urban Syst. 90, 101706. https://doi.org/10.1016/j.compenvurbsys.2021.101706.

Choy, C.B., Xu, D., Gwak, J., Chen, K., Savarese, S., 2016. 3d-r2n2: A unified approach for single and multi-view 3d object reconstruction, in: European conference on computer vision, Springer. pp. 628–644.

Chu, H., Wang, S., Urtasun, R., Fidler, S., 2016. Housecraft: Building houses from rental ads and street views, in: European Conference on Computer Vision, Springer. pp. 500–516.

Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., Ranzuglia, G., 2008. MeshLab: An open-source mesh processing tool. 6th Eurographics Italian Chapter Conference 2008 - Proceedings, 129–136.

Cinnamon, J., Gaffney, A., 2021. Do-It-Yourself Street Views and the Urban Imaginary of Google Street View. Journal of Urban Technology 1–22. https://doi.org/10.1080/10630732.2021.1910467.

Cohen, A., Schönberger, J.L., Speciale, P., Sattler, T., Frahm, J.M., Pollefeys, M., 2016. Indoor-outdoor 3d reconstruction alignment, in: European Conference on Computer Vision, Springer. pp. 285–300.

Dehbi, Y., Henn, A., Gröger, G., Stroh, V., Plümer, L., 2020. Robust and fast reconstruction of complex roofs with active sampling from 3D point clouds. Transactions in GIS. https://doi.org/10.1111/tgis.12659.

Demir, N., Baltsavias, E., 2012. Automated modeling of 3D building roofs using image and LiDAR data, in: Proceedings of the XXII Congress of the International Society for Photogrammetry, Remote Sensing, Melbourne, Australia.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database, in: 2009 IEEE conference on computer vision and pattern recognition, Ieee. pp. 248–255.

Ding, X., Fan, H., Gong, J., 2021. Towards generating network of bikeways from Mapillary data. Comput. Environ. Urban Syst. 88, 101632. https://doi.org/10.1016/j.compenvurbsys.2021.101632.

Doan, T.Q., León-Sánchez, C., Peters, R., Agugiaro, G., Stoter, J., 2021. Volume comparison of automatically reconstructed multi-LoD building models for urban planning applications. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences V-4-2021, pp. 169–176. https://doi.org/10.5194/isprs-annals-v-4-2021-169-2021.

Dukai, B., Peters, R., Wu, T., Commandeur, T., Ledoux, H., Baving, T., Post, M., van Altena, V., van Hinsbergh, W., Stoter, J., 2020. Generating, storing, updating, and disseminating a country-wide 3D model. In: ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLIV-4/W1-2020, pp. 27–32. https://doi.org/10.5194/isprs-archives-xliv-4-w1-2020-27-2020.

Eicker, U., Nouvel, R., Duminil, E., Coors, V., 2014. Assessing passive and active solar energy resources in cities using 3D city models. Energy Procedia 57, 896–905.

Elfouly, M., Labetski, A., 2020. Flood damage cost estimation in 3D based on an indicator modelling framework. Geomatics, Natural Hazards and Risk 11, 1129–1153. https://doi.org/10.1080/19475705.2020.1777213.

Esch, T., Brzoska, E., Dech, S., Leutner, B., Palacios-Lopez, D., Metz-Marconcini, A., Marconcini, M., Roth, A., Zeidler, J., 2022. World settlement footprint 3d - a first three-dimensional survey of the global building stock. Remote Sens. Environ. 270, 112877.

Fan, H., Kong, G., Zhang, C., 2021. An Interactive platform for low-cost 3D building modeling from VGI data using convolutional neural network. Big Earth Data 5, 49–65.

Fan, H., Su, H., Guibas, L., 2016. A point set generation network for 3D object reconstruction from a single image. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017 2017-Janua, 2463–2471. doi: 10.1109/CVPR.2017.264, arXiv:1612.00603.

Fan, H., Zipf, A., Fu, Q., Neis, P., 2014. Quality assessment for building footprints data on OpenStreetMap. International Journal of Geographical Information Science 28, 700–719.

Fedorova, S., Tono, A., Nigam, M.S., Zhang, J., Ahmadnia, A., Bolognesi, C., Michels, D. L., 2021. Synthetic 3D Data Generation Pipeline for Geometric Deep Learning in Architecture. arXiv preprint arXiv:2104.12564.

Fibæk, C.S., Keßler, C., Arsanjani, J.J., 2021. A multi-sensor approach for characterising human-made structures by estimating area, volume and population based on sentinel data and deep learning. Int. J. Appl. Earth Obs. Geoinf. 105, 102628. https://doi.org/10.1016/j.jag.2021.102628.

Fleischmann, M., Feliciotti, A., Kerr, W., 2021. Evolution of Urban Patterns: Urban Morphology as an Open Reproducible Data Science. Geographical Analysis. https://doi.org/10.1111/gean.12302.

Florio, P., Peronato, G., Perera, A., Blasi, A.D., Poon, K.H., Kämpf, J.H., 2021. Designing and assessing solar energy neighborhoods from visual impact. Sustainable Cities and Society 71, 102959. https://doi.org/10.1016/j.scs.2021.102959.

Frantz, D., Schug, F., Okujeni, A., Navacchi, C., Wagner, W., van der Linden, S., Hostert, P., 2021. National-scale mapping of building height using Sentinel-1 and Sentinel-2 time series. Remote Sens. Environ. 252, 112128. https://doi.org/10.1016/j.rse.2020.112128.

Fu, K., Peng, J., He, Q., Zhang, H., 2021. Single image 3D object reconstruction based on deep learning: A review. Multimedia Tools and Applications 80, 463–498.

Gao, W., Nan, L., Boom, B., Ledoux, H., 2021. SUM: A benchmark dataset of Semantic Urban Meshes. ISPRS J. Photogram. Remote Sens. 179, 108–120. https://doi.org/10.1016/j.isprsjprs.2021.07.008.

Gassar, A.A.A., Cha, S.H., 2021. Review of geographic information systems-based rooftop solar photovoltaic potential estimation approaches at urban scales. Appl. Energy 291, 116817. https://doi.org/10.1016/j.apenergy.2021.116817.

Geis, C., Leichtle, T., Wurm, M., Pelizari, P.A., Standfus, I., Zhu, X.X., So, E., Siedentop, S., Esch, T., Taubenbock, H., 2019. Large-Area Characterization of Urban Morphology—Mapping of Built-Up Height and Density Using TanDEM-X and Sentinel-2 Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12, 2912–2927. https://doi.org/10.1109/jstars.2019.2917755.

Gkioxari, G., Malik, J., Johnson, J., 2019. Mesh r-cnn, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9785–9795.

Goetz, M., 2013. Towards generating highly detailed 3D CityGML models from OpenStreetMap. International Journal of Geographical Information Science 27, 845–865.

Gröger, G., Plümer, L., 2012. CityGML – interoperable semantic 3d city models. ISPRS J. Photogram. Remote Sens. 71, 12–33. URL: https://doi.org/10.1016. 10.1016/j.isprsjprs.2012.04.004.

Gui, S., Qin, R., 2021. Automated LoD-2 model reconstruction from very-high-resolution satellite-derived digital surface model and orthophoto. ISPRS J. Photogram. Remote Sens. 181, 1–19. https://doi.org/10.1016/j.isprsjprs.2021.08.025.

Han, X.F., Laga, H., Bennamoun, M., 2019. Image-based 3D object reconstruction: State-of-the-art and trends in the deep learning era. IEEE transactions on pattern analysis and machine intelligence 43, 1578–1604.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2016-December, 770–778. doi:10.1109/CVPR.2016.90, arXiv: 1512.03385.

He, N., Li, G., 2021. Urban neighbourhood environment assessment based on street view image processing: A review of research trends. Environmental. Challenges 4, 100090. URL: https://doi.org/10.1016. 10.1016/j.envc.2021.100090.

Helbich, M., Poppe, R., Oberski, D., van Emmichoven, M.Z., Schram, R., 2021. Can't see the wood for the trees? An assessment of street view- and satellite-derived greenness measures in relation to mental health. Landscape and Urban Planning 214, 104181. https://doi.org/10.1016/j.landurbplan.2021.104181.

Huang, X., Wang, C., 2020. Estimates of exposure to the 100-year floods in the conterminous United States using national building footprints. International Journal of Disaster Risk Reduction 50, 101731. https://doi.org/10.1016/j.ijdrr.2020.101731.

Ito, K., Biljecki, F., 2021. Assessing bikeability with street view imagery and computer vision. Transportation Research Part C: Emerging Technologies 132, 103371. https://doi.org/10.1016/j.trc.2021.103371.

Jang, Y.H., Park, S.I., Kwon, T.H., Lee, S.H., 2021. CityGML urban model generation using national public datasets for flood damage simulations: A case study in Korea. J. Environ. Manage. 297, 113236. https://doi.org/10.1016/j.jenvman.2021.113236.

Jochem, W.C., Tatem, A.J., 2021. Tools for mapping multi-scale settlement patterns of building footprints: An introduction to the R package foot. PLOS ONE 16, e0247535. https://doi.org/10.1371/journal.pone.0247535.

Johnson, J., Alahi, A., Fei-Fei, L., 2016. Perceptual losses for real-time style transfer and super-resolution, in: European conference on computer vision, Springer. pp. 694–711.

Jovanović, D., Milovanov, S., Ruskovski, I., Govedarica, M., Sladić, D., Radulović, A., Pajić, V., 2020. Building virtual 3D city model for Smart Cities applications: A case study on campus area of the University of Novi Sad. ISPRS International Journal of Geo-Information 9, 476.

Kaden, R., Kolbe, T.H., 2014. Simulation-based total energy demand estimation of buildings using semantic 3D city models. International Journal of 3-D Information Modeling IJ3DIM) 3, 35–53.

Kang, Y., Zhang, F., Gao, S., Peng, W., Ratti, C., 2021. Human settlement value assessment from a place perspective: Considering human dynamics and perceptions in house price modeling. Cities 118, 103333. https://doi.org/10.1016/j.cities.2021.103333.

Kim, H., Han, S., 2018. Interactive 3D building modeling method using panoramic image sequences and digital map. Multimedia tools and applications 77, 27387–27404.

Kim, J., Lee, J.K., Lee, K.M., 2016. Accurate image super-resolution using very deep convolutional networks, in. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1646–1654.

Komadina, A., Mihajlovic, Z., 2022. Automated 3D Urban Landscapes Visualization Using Open Data Sources on the Example of the City of Zagreb. KN - Journal of Cartography and Geographic Information 1–14. https://doi.org/10.1007/s42489-022-00102-w.

Kraff, N.J., Wurm, M., Taubenböck, H., 2020. The dynamics of poor urban areas-analyzing morphologic transformations across the globe using Earth observation data. Cities 107, 102905.

Kruse, J., Kang, Y., Liu, Y.N., Zhang, F., Gao, S., 2021. Places for play: Understanding human perception of playability in cities using street view images and deep learning. Comput. Environ. Urban Syst. 90, 101693. https://doi.org/10.1016/j.compenvurbsys.2021.101693.

Kutzner, T., Chaturvedi, K., Kolbe, T.H., 2020. CityGML 3.0: New functions open up new applications. PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation. Science 88, 43–61. URL: https://doi.org/10.1007. 10.1007/s41064-020-00095-z.

Ledoux, H., Biljecki, F., Dukai, B., Kumar, K., Peters, R., Stoter, J., Commandeur, T., 2021. 3dfier: automatic reconstruction of 3D city models. Journal of Open Source Software 6, 2866. https://doi.org/10.21105/joss.02866.

Lee, T., 2009. Robust 3D street-view reconstruction using sky motion estimation, in: 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, IEEE. pp. 1840–1847.

Leonard, A., Wheeler, S., McCulloch, M., 2022. Power to the people: Applying citizen science and computer vision to home mapping for rural energy access. Int. J. Appl. Earth Obs. Geoinf. 108, 102748. https://doi.org/10.1016/j.jag.2022.102748.

Li, H., Herfort, B., Lautenbach, S., Chen, J., Zipf, A., 2022. Improving OpenStreetMap missing building detection using few-shot transfer learning in sub-Saharan Africa. Transactions in GIS. https://doi.org/10.1111/tgis.12941.

Li, M., Koks, E., Taubenböck, H., Vliet, J.v., 2020a. Continental-scale mapping and analysis of 3D building structure. Remote Sensing of Environment 245, 111859. doi: 10.1016/j.rse.2020.111859.

Li, Y., Schubert, S., Kropp, J.P., Rybski, D., 2020. On the influence of density and morphology on the Urban Heat Island intensity. Nature communications 11, 1–9.

Lines, T., Basiri, A., 2021. 3D map creation using crowdsourced GNSS data. Comput. Environ. Urban Syst. 89, 101671. https://doi.org/10.1016/j.compenvurbsys.2021.101671.

Lu, H., Li, F., Yang, G., Sun, W., 2021. Multi-scale impacts of 2D/3D urban building pattern in intra-annual thermal environment of Hangzhou, China. Int. J. Appl. Earth Obs. Geoinf. 104, 102558. https://doi.org/10.1016/j.jag.2021.102558.

Ma, D., Fan, H., Li, W., Ding, X., 2019. The State of Mapillary: An Exploratory Analysis. ISPRS International Journal of Geo-Information 9, 10. https://doi.org/10.3390/ijgi9010010.

Mahmud, J., Price, T., Bapat, A., Frahm, J.M., 2020. Boundary-aware 3D building reconstruction from a single overhead image, in. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 441–451.

Martinovic, A., 2015. Inverse Procedural Modeling of Buildings. Ph.D. thesis. KU Leuven.

Mathews, A.J., Frazier, A.E., Nghiem, S.V., Neumann, G., Zhao, Y., 2019. Satellite scatterometer estimation of urban built-up volume: Validation with airborne lidar data. Int. J. Appl. Earth Obs. Geoinf. 77, 100–107. https://doi.org/10.1016/j.jag.2019.01.004. URL: https://www.sciencedirect.com/science/article/pii/S0303243418310420.

McNeel, R., et al., 2010. Rhinoceros 3d, version 6.0. Robert McNeel & Associates, Seattle, WA.

Meouche, R.E., Eslahi, M., Ruas, A., 2021. Investigating the Effects of Population Growth and Urban Fabric on the Simulation of a 3D City Model, pp. 1344–1358. doi: 10.1007/978-3-030-66840-2_102.

Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., Geiger, A., 2019. Occupancy networks: Learning 3d reconstruction in function space, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4460–4470.

Micusik, B., Kosecka, J., 2009. Piecewise planar city 3D modeling from street view panoramic sequences, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE. pp. 2906–2912.

Milojevic-Dupont, N., Hans, N., Kaack, L.H., Zumwald, M., Andrieux, F., de Barros Soares, D., Lohrey, S., Pichler, P.P., Creutzig, F., 2020. Learning from urban form to predict building heights. PLOS ONE 15, e0242010. https://doi.org/10.1371/journal.pone.0242010.

Niemeyer, M., Mescheder, L., Oechsle, M., Geiger, A., 2020. Differentiable volumetric rendering: Learning implicit 3D representations without 3D supervision, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3504–3515.

Ning, H., Li, Z., Ye, X., Wang, S., Wang, W., Huang, X., 2021. Exploring the vertical dimension of street view image based on deep learning: a case study on lowest floor elevation estimation. International Journal of Geographical Information Science 1–26. https://doi.org/10.1080/13658816.2021.1981334.

Noardo, F., Arroyo Ohori, K., Biljecki, F., Ellul, C., Harrie, L., Krijnen, T., Eriksson, H., van Liempt, J., Pla, M., Ruiz, A., Hintz, D., Krueger, N., Leoni, C., Leoz, L., Moraru, D., Vitalis, S., Willkomm, P., Stoter, J., 2021. Reference study of CityGML

software support: The GeoBIM benchmark 2019—Part II. Trans. GIS 25, 842–868. https://doi.org/10.1111/tgis.12710.
Nys, G.A., Poux, F., Billen, R., 2020. CityJSON Building Generation from Airborne LiDAR 3D Point Clouds. ISPRS International Journal of Geo-Information 9, 521. https://doi.org/10.3390/ijgi9090521.
Oechsle, M., Mescheder, L., Niemeyer, M., Strauss, T., Geiger, A., 2019. Texture fields: Learning texture representations in function space, in. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4531–4540.
Palliwal, A., Song, S., Tan, H.T.W., Biljecki, F., 2021. 3D city models for urban farming site identification in buildings. Comput. Environ. Urban Syst. 86, 101584. https://doi.org/10.1016/j.compenvurbsys.2020.101584.
Pelizari, P.A., Geiß, C., Aguirre, P., María, H.S., Peña, Y.M., Taubenböck, H., 2021. Automated building characterization for seismic risk assessment using street-level imagery and deep learning. ISPRS J. Photogram. Remote Sens. 180, 370–386. https://doi.org/10.1016/j.isprsjprs.2021.07.004.
Peters, R., Dukai, B., Vitalis, S., van Liempt, J., Stoter, J., 2022. Automated 3D Reconstruction of LoD2 and LoD1 Models for All 10 Million Buildings of the Netherlands. Photogrammetric Engineering & Remote Sensing 88, 165–170. https://doi.org/10.14358/pers.21-00032r2.
Rastiveis, H., Samadzadegan, F., Reinartz, P., 2013. A fuzzy decision making system for building damage map creation using high resolution satellite imagery. Natural Hazards and Earth System Sciences 13, 455–472. https://doi.org/10.5194/nhess-13-455-2013.
Rosenfelder, M., Wussow, M., Gust, G., Cremades, R., Neumann, D., 2021. Predicting residential electricity consumption using aerial and street view images. Appl. Energy 301, 117407. https://doi.org/10.1016/j.apenergy.2021.117407.
Rosser, J.F., Long, G., Zakhary, S., Boyd, D.S., Mao, Y., Robinson, D., 2019. Modelling urban housing stocks for building energy simulation using CityGML EnergyADE. ISPRS International Journal of Geo-Information 8, 163.
Sindram, M., Machl, T., Steuer, H., Pültz, M., Kolbe, T.H., 2016. Voluminator 2.0–speeding up the approximation of the volume of defective 3D building models. ISPRS annals of photogrammetry, remote sensing and spatial information sciences 3, 29–36.
Sirko, W., Kashubin, S., Ritter, M., Annkah, A., Bouchareb, Y.S.E., Dauphin, Y., Keysers, D., Neumann, M., Cisse, M., Quinn, J., 2021. Continental-scale building detection from high resolution satellite imagery. arXiv:2107.12283.
Song, S., Lim, M.S., Richards, D.R., Tan, H.T.W., 2021. Utilization of the food provisioning service of urban community gardens: Current status, contributors and their social acceptance in Singapore. Sustainable Cities and Society 103368. https://doi.org/10.1016/j.scs.2021.103368.
Song, S., Yu, F., Zeng, A., Chang, A.X., Savva, M., Funkhouser, T., 2017. Semantic scene completion from a single depth image, in. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1746–1754.
Sridharan, H., Qiu, F., 2013. A spatially disaggregated areal interpolation model using light detection and Ranging-Derived building volumes. Geographical Analysis 45, 238–258.
Stoter, J., Peters, R., Commandeur, T., Dukai, B., Kumar, K., Ledoux, H., 2020. Automated reconstruction of 3D input data for noise simulation. Comput. Environ. Urban Syst. 80, 101424. https://doi.org/10.1016/j.compenvurbsys.2019.101424.
Suveg, I., Vosselman, G., 2004. Reconstruction of 3D building models from aerial images and maps. ISPRS J. Photogram. Remote Sens. 58, 202–224.
Szarka, N., Biljecki, F., 2022. Population estimation beyond counts—Inferring demographic characteristics. PLOS ONE 17, e0266484. https://doi.org/10.1371/journal.pone.0266484.
Szcześniak, J.T., Ang, Y.Q., Letellier-Duchesne, S., Reinhart, C.F., 2021. A method for using street view imagery to auto-extract window-to-wall ratios and its relevance for urban-level daylighting and energy simulations. Build. Environ. 108108. https://doi.org/10.1016/j.buildenv.2021.108108.
Tatarchenko, M., Richter, S.R., Ranftl, R., Li, Z., Koltun, V., Brox, T., 2019. What do single-view 3d reconstruction networks learn?, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3405–3414.
Taubenböck, H., Kraff, N.J., Wurm, M., 2018. The morphology of the Arrival City-A global categorization based on literature surveys and remotely sensed data. Applied Geography 92, 150–167.
Torii, A., Havlena, M., Pajdla, T., 2009. From Google Street View to 3D city models, in: 2009 IEEE 12th international conference on computer vision workshops, ICCV workshops, IEEE. pp. 2188–2195.
Turan, I., Chegut, A., Fink, D., Reinhart, C., 2021. Development of view potential metrics and the financial impact of views on office rents. Landscape and Urban Planning 215, 104193. https://doi.org/10.1016/j.landurbplan.2021.104193.
Virtanen, J.P., Jaalama, K., Puustinen, T., Julin, A., Hyyppä, J., Hyyppä, H., 2021. Near Real-Time Semantic View Analysis of 3D City Models in Web Browser. ISPRS International Journal of Geo-Information 10, 138. https://doi.org/10.3390/ijgi10030138.
Vitalis, S., Labetski, A., Arroyo Ohori, K., Ledoux, H., Stoter, J., 2019. A data structure to incorporate versioning in 3D city models. ISPRS Ann. Photogramm. Remote Sens. Spatial. Inf. Sci. IV-4/W8, 123–130. https://doi.org/10.5194/isprs-annals-iv-4-w8-123-2019.
Vosselman, G., Dijkman, S., et al., 2001. 3D building model reconstruction from point clouds and ground plans. International archives of photogrammetry remote sensing and spatial information sciences 34, 37–44.
Wang, C., Wei, S., Du, S., Zhuang, D., Li, Y., Shi, X., Jin, X., Zhou, X., 2021. A systematic method to develop three dimensional geometry models of buildings for urban building energy modeling. Sustainable Cities and Society 71, 102998. https://doi.org/10.1016/j.scs.2021.102998.
Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., Wang, X., Tang, X., 2017. Residual attention network for image classification, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3156–3164.
Wang, N., Zhang, Y., Li, Z., Fu, Y., Liu, W., Jiang, Y.G., 2018. Pixel2mesh: Generating 3d mesh models from single rgb images, in: Proceedings of the European Conference on Computer Vision (ECCV), pp. 52–67.
Wichmann, A., Agoub, A., Schmidt, V., Kada, M., 2019. RoofN3D: A Database for 3D Building Reconstruction with Deep Learning. Photogrammetric Engineering & Remote Sensing 85, 435–443.
Wysocki, O., Schwab, B., Hoegner, L., Kolbe, T.H., Stilla, U., 2021. Plastic surgery for 3D city models: A pipeline for automatic geometry refinement and semantic enrichment. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences V-4-2021, pp. 17–24. https://doi.org/10.5194/isprs-annals-v-4-2021-17-2021.
Xie, Y., Cai, J., Bhojwani, R., Shekhar, S., Knight, J., 2019. A locally-constrained YOLO framework for detecting small and densely-distributed building footprints. International Journal of Geographical Information Science 34, 1–25. https://doi.org/10.1080/13658816.2019.1624761.
Xu, Y., Ren, C., Ma, P., Ho, J., Wang, W., Lau, K.K.L., Lin, H., Ng, E., 2017. Urban morphology detection and computation for urban climate research. Landscape and Urban Planning 167, 212–224. https://doi.org/10.1016/j.landurbplan.2017.06.018.
Yamani, S.E., Hajji, R., Nys, G.A., Ettarid, M., Billen, R., 2021. 3D Variables Requirements for Property Valuation Modeling Based on the Integration of BIM and CIM. Sustainability 13, 2814. https://doi.org/10.3390/su13052814.
Yin, J., Dong, J., Hamm, N.A.S., Li, Z., Wang, J., Xing, H., Fu, P., 2021. Integrating remote sensing and geospatial big data for urban land use mapping: A review. Int. J. Appl. Earth Obs. Geoinf. 103, 102514. https://doi.org/10.1016/j.jag.2021.102514.
Yohannes, E., Lin, C.Y., Shih, T.K., Hong, C.Y., Enkhbat, A., Utaminingrum, F., 2021. Domain Adaptation Deep Attention Network for Automatic Logo Detection and Recognition in Google Street View. IEEE Access PP, 1–1. doi:10.1109/access.2021.3098713.
Yu, D., Ji, S., Liu, J., Wei, S., 2021. Automatic 3D building reconstruction from multi-view aerial images with deep learning. ISPRS J. Photogram. Remote Sens. 171, 155–170.
Zhang, C., Fan, H., Kong, G., 2021a. VGI3D: an Interactive and Low-Cost Solution for 3D Building Modelling from Street-Level VGI Images. Journal of Geovisualization and Spatial Analysis 5, 18. https://doi.org/10.1007/s41651-021-00086-7.
Zhang, F., Fan, Z., Kang, Y., Hu, Y., Ratti, C., 2021b. "Perception bias": Deciphering a mismatch between urban crime and perception of safety. Landscape and Urban Planning 207, 104003. https://doi.org/10.1016/j.landurbplan.2020.104003.
Zhang, F., Zu, J., Hu, M., Zhu, D., Kang, Y., Gao, S., Zhang, Y., Huang, Z., 2020. Uncovering inconspicuous places using social media check-ins and street view images. Comput. Environ. Urban Syst. 81, 101478. https://doi.org/10.1016/j.compenvurbsys.2020.101478.
Zhang, J., Fukuda, T., Yabuki, N., 2021c. Automatic object removal with obstructed façades completion using semantic segmentation and generative adversarial inpainting. IEEE Access 9, 117486–117495. https://doi.org/10.1109/access.2021.3106124.
Zhao, J., Ledoux, H., Stoter, J., Feng, T., 2018. HSW: Heuristic Shrink-wrapping for automatically repairing solid-based CityGML LOD2 building models. ISPRS J. Photogram. Remote Sens. 146, 289–304. https://doi.org/10.1016/j.isprsjprs.2018.09.019.
Zhu, X.X., Qiu, C., Hu, J., Shi, Y., Wang, Y., Schmitt, M., Taubenböck, H., 2022. The urban morphology on our planet – global perspectives from space. Remote Sens. Environ. 269, 112794. https://doi.org/10.1016/j.rse.2021.112794.