This is the Accepted Manuscript version of an article published by Elsevier in the journal Automation in Construction in 2026, which is available at: https://doi.org/10.1016/j.autcon.2026.106842

Cite as: Ma R, Yang C, Chen J, Biljecki F, Li X (2026). Street view imagery-based method for reconstructing 3D building façade openings. Automation in Construction 184: 106842.

Street View Imagery-Based Method for Reconstructing 3D Building Façade Openings

Rui Ma[a,b], Chendi Yang[c], Jiayu Chen[d], Filip Biljecki[e,f,*], Xin Li[b]

[a] School of Management, Zhengzhou University, Zhengzhou, China
[b] Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong SAR, China
[c] School of Architecture, Zhengzhou University, Zhengzhou, China
[d] School of Civil Engineering, Tsinghua University, Beijing, China
[e] Department of Architecture, National University of Singapore, Singapore
[f] Department of Real Estate, National University of Singapore, Singapore

[*]Corresponding author: filip@nus.edu.sg (Filip Biljecki)

Abstract

The availability of 3D building models has been increasing, but they often lack detail at the architectural scale. This paper presents a method for reconstructing façade openings in 3D building models by integrating Street View imagery (SVI). Methodologically, the paper advances opening reconstruction in two key ways: first, by introducing a mathematically derived method for estimating unknown intrinsic camera parameters, enabling metric 2D-to-3D projection without relying on multi-view imagery or pre-existing depth information. Second, the method extends single-image photogrammetry to accurately measure detailed façade openings, converting pixel coordinates into spatial coordinates. The proposed method is validated through case studies in Amsterdam. Quantitative evaluation using the Façade Re-projection Dice Score (FRDS) shows high spatial consistency between reconstructed openings and reference opening geometries, with most scores ranging from 0.84 to 0.98. Given the broad coverage of SVI, there is a significant potential for enhancing 3D city models in diverse urban contexts where current representations remain geometrically basic.

Keywords: Street view imagery; façade opening reconstruction; image formation; 3D model augmentation

1. Introduction

Highly detailed 3D building models provide precise representations of architectural elements. They offer substantial potential for a wide range of applications, spanning from individual buildings to entire urban landscapes. These applications include building energy modeling, indoor thermal comfort assessment, photovoltaic power analysis, illumination evaluation, and 3D visualization [1]. To assess the quality and detail level of building models derived from various production workflows, the concept of "level of detail (LOD)" has been established in CityGML, a standardized data format for storing digital 3D city models [2,3]. For building exteriors, the specification starts with LOD1, the basic grade in which buildings are represented as prismatic blocks and are usually derived by extruding footprints to a singular height. They have been popular and widely available thanks to the favorable relationship between ease of acquisition and wide usability. LOD2 models are also relatively basic, but they add generalized roof shapes that may contribute to their appearance and some use cases. LOD3 describes the highest level of detail in façade openings, playing a critical role in establishing physical connections between buildings and the outdoor environment, i.e., detailed façade elements such as windows are included [4]. These connections are indispensable for processes such as heat transfer, solar heat gain, air leakage, and ventilation [5]. However, building-related studies in real urban environments typically lack detailed façade opening information. For example, numerous urban building energy studies rely on roughly assumed window-to-wall ratios [6–9] or pre-defined ratios for specific prototype building models [10,11]. These assumptions can propagate non-negligible simulation errors, as window size, distribution, and performance significantly affect heating and cooling demand. Recent findings show that façade openings are among the most sensitive envelope parameters influencing building energy use, with window upgrades often yielding the largest single-measure savings in urban-scale simulations [10]. Therefore, there is a need to enhance widely used building models by incorporating realistic façade openings. The enhanced buildings align with LOD3.1 defined by Biljecki et al [12], enabling unattainable or insufficiently reliable applications with lower LODs.

Recent research highlights the growing interest in upgrading LOD2 to LOD3 representations to support urban-scale digital twins and simulation. For example, Xia et al. [13] introduced a pipeline that enriches LOD2 models with façade openings using oblique imagery and deep learning, while Hanke et al. [14] proposed the CM2LoD3 approach, leveraging semantic conflict maps for

large-scale automated reconstruction. UAV (unmanned aerial vehicle) -based frameworks have also been tested for automated LOD3 reconstruction [15], and Tang et al. [16] developed the Texture2LoD3 method that integrates panoramic imagery with low-level models to refine façade details. These advances indicate that LOD3 generation is increasingly feasible through diverse data sources, yet challenges remain in ensuring geometric consistency, semantic consistency, and computational scalability.

Current methods for reconstructing façade openings encounter two main challenges. First, while numerous studies have achieved satisfactory reconstruction outcomes, the availability and quality of relevant data in actual reconstruction scenarios are difficult to guarantee. The fidelity to reconstructed images based on airborne or terrestrial imagery can be sensitive to the number of images captured, the degree of overlap, and where the images were captured [17], all of which vary across projects and reduce reproducibility. In addition, high-quality UAV imagery and LiDAR point clouds are frequently unavailable in many urban regions due to regulatory restrictions, high acquisition costs, limited technical capacity, or incomplete spatial coverage, particularly in developing regions and rapidly urbanizing cities. By contrast, street view imagery (SVI) is collected through a standardized and globally consistent acquisition pipeline. This systematic process ensures stable coverage and greatly reduces variability in data quality. Second, most existing approaches require either advanced equipment or dedicated surveying and mapping work, making them resource-intensive. Consequently, there is a critical need to develop strategies that exploit publicly available data to enable cost-effective and efficient reconstruction processes.

The method proposed in this paper builds on these needs by applying easily accessible and widely available SVI to extract façade information and integrate detailed façade openings into existing building models. Even though SVI is broadly used for several tasks, including estimates of building height, materials, or land-use typology, existing façade reconstruction methodologies using ground-level imagery are not directly applicable to freely available SVI. This is because most current approaches require either known camera intrinsic parameters, dense multi-view coverage, or dense point cloud generation, conditions that are rarely satisfied in typical street view data. In practice, a building is often captured by only one or a few panoramas. In addition, camera intrinsic parameters are typically not disclosed. These limitations make metrically meaningful reconstruction of façade openings from a single panorama infeasible using conventional pipelines. As a result, existing street-view-based methods are largely limited to non-metric or approximate

representations, which constrain their applicability in downstream tasks, such as building energy simulation and urban digital twin applications. This limitation leads to several specific research questions: (1) How can the unknown intrinsic parameters of SVI cameras be estimated in a reproducible way without access to proprietary metadata? (2) How can façade-opening geometries be reconstructed from a single SVI panorama in a metrically consistent manner? (3) How can the recovered 3D openings be integrated consistently with existing city-scale building models to overcome the known limitations of LOD2 representations in façade-sensitive applications?

Methodologically, this paper advances façade reconstruction in two distinct ways. First, it includes a mathematically derived method for estimating the unknown intrinsic parameters of panoramic SVI, allowing metric 2D-to-3D projection, without relying on multi-view imagery or pre-existing depth information. This paper explicitly establishes the geometric relationship between image space, camera space, and world coordinates, providing an analytically grounded solution to a long-standing limitation of publicly available street view data. Second, it allows for extending single-image photogrammetry reconstruction to measure detailed façade openings, converting pixel coordinates to spatial coordinates. This paper differs from conventional techniques, since these methods require a massive collection of multi-view images or a dense cloud of points. To our knowledge, no prior work has proposed a mathematically grounded estimation method that achieves metric projection of façade openings using just a single publicly available panorama. The contribution of this paper lies in a reproducible reconstruction workflow tailored for freely accessible SVI, without relying on depth maps, multi-view inputs, or the camera's internal parameters.

The rest of this paper is organized as follows. We begin with a literature review (Section 2), followed by detailed explanations of the proposed SVI-based method (Section 3). Section 4 implements the method: in Section 4.1, mathematical derivation is employed to validate the assumptions made during the estimation of camera parameters, while Section 4.2 reports three case studies in Amsterdam that assess the reconstruction quality. Additionally, this paper demonstrates the benefits that enhanced building models bring to applications (Section 4.3). The final section examines the contributions and limitations of the proposed method and concludes with a concise summary of the main findings.

2. Literature review

Research on façade opening reconstruction covers a broad spectrum of themes and data sources, reflecting the diversity of objectives across architectural, photogrammetric, and urban modeling studies. Before examining specific techniques for reconstructing façade openings, it is necessary to outline the main methodological directions that shape current façade reconstruction research. The following subsections summarize the key approaches relevant to understanding how façade openings are detected and modeled.

2.1 Façade opening reconstruction

Research on façade opening reconstruction generally follows two complementary paradigms: top-down model-driven approaches and bottom-up data-driven approaches. Top-down methods rely on predefined template libraries, which contain reusable façade primitives such as windows and doors, or on parametric façade grammars to guide the reconstruction process. Representative examples include template-assembly frameworks [18], semantically enriched LOD3 façade templates [19], and parametric multi-source façade modeling [20]. These approaches benefit from structural regularity and interactive adjustment but lack flexibility when applied to heterogeneous building types because their predefined primitives restrict generalization across diverse architectural forms [18].

In contrast, bottom-up approaches extract façade components directly from observations. These techniques utilize BIM data [21,22], airborne/terrestrial laser scanning [23,24], or image-based point clouds [25,26] to derive LOD3 façade details. While accurate, they typically require specialized hardware, controlled acquisition settings, dense multi-view coverage, or high-quality BIM files, which limit their scalability for city-scale reconstruction. Despite extensive advances, existing LOD3 façade opening reconstruction workflows remain constrained by data cost, equipment requirements, geometric completeness, and reproducibility. These limitations motivate the need for methods that operate robustly under sparse and heterogeneous data conditions. This gap directly supports the motivation, which aims to reconstruct façade openings using widely accessible street-view imagery without relying on dense point clouds or known camera parameters.

2.2 Application of SVI in building reconstruction

SVI is a collection of sequential ground-level images taken along a street, which usually includes buildings. Companies such as Google and Baidu have collected billions of such images covering thousands of cities around the world, while crowdsourced platforms such as Mapillary and KartaView gather them from volunteers and also contain a large collection. Due to their high coverage and image quality [27], and offering detailed insights into buildings, they may be a viable data source to collect data on buildings. Existing studies have shown that SVI can be used to extract various geometric or semantic attributes, including building height [28], window/door elevations [29], window-to-wall ratios [30], etc. These works demonstrate the value of SVI for extracting individual façade attributes, but they do not address the full 3D reconstruction of façade openings.

Early attempts to employ SVI for reconstruction primarily focused on point cloud generation from street-level imagery [31,32], which is computationally intensive and often unsuitable when only limited views of a building façade are available. Kim and Han [33] generated textured prismatic 3D models from panoramic sequences, while Pang and Biljecki [34] produced plausible single-image-based massing models but without reconstructing façade openings. More recently, Tang et al. [16] introduced the Texture2LoD3 framework to transform panoramic textures into semantically enriched LOD3 façades. However, the method still relies on high-quality textured LOD2 models and was mainly validated at the building scale.

Across these studies, the core challenge remains: freely available SVI typically lacks known camera parameters and multi-perspective imagery, as a building may only appear in one or two panoramas. Consequently, most existing SVI-based methods are either (1) limited to coarse geometry or texture transfer, or (2) require auxiliary data such as multi-view images, point clouds, or textured 3D shells. This motivates the need for approaches capable of converting single SVI images into metrically meaningful 3D façade opening geometries, which is precisely the contribution of this paper.

2.3 Single image-based methods for building reconstruction

Single-image 3D reconstruction is a longstanding problem in computer vision [35], and building-related applications often combine footprint extraction and height estimation from remote sensing images [36]. Other efforts generate point clouds or mesh representations from single-view images [37,38]. While these methods demonstrate the feasibility of inferring spatial information

from single images, most of them are designed for remote sensing or controlled aerial imagery, where camera models and imaging geometry are well understood.

With the growing availability of SVI, several recent works have explored single-image reconstruction from street-level views. Pang and Biljecki [34] showed that SVI can support rapid massing reconstruction. Wang et al. [39] used façade textures from LOD2 models to generate more detailed LOD3 façades, though the method depends on the availability of textured 3D models. Neural implicit representations such as NeRBuilder [40] and scalable façade parsers like SI3FP [41] further improve robustness through semantic priors or orthographic projections, but they still rely heavily on large annotated training sets and do not explicitly tackle 3D metric recovery of façade openings from single panoramic images.

Overall, existing single-image approaches either require strong priors or textured 3D models or focus on massing rather than façade openings. The gap is no existing single-SVI workflow capable of transforming 2D opening detections into accurate 3D opening geometries without known camera parameters.

3. Methodology

Building upon the challenges and opportunities discussed in the previous sections, this paper proposes a systematic workflow for reconstructing façade openings using widely accessible SVI in combination with existing 3D building models. The methodology is designed to be both modular and scalable, enabling its application to individual buildings, groups of adjacent structures, or larger urban blocks, depending on the availability and quality of SVI.

The process operates on a building-by-building basis to ensure precise geometric alignment between the extracted façade features and their corresponding 3D models. Taking a single target façade as an illustrative example, Fig. 1 outlines the overall workflow, which consists of six major stages: (1) data preprocessing to prepare SVI and building model inputs; (2) pixel-level localization to detect façade openings and their corners; (3) spatial coordinate calculation to project 2D image features into accurate 3D positions; (4) storage and integration of reconstructed façade openings into the original building model; (5) quantitative evaluation of reconstruction quality; and (6) application of the enhanced models to potential use cases, such as building energy simulation.

In scenarios where a building has multiple façades facing different streets, each façade is treated as an independent target, and the reconstruction workflow can be repeated for each one. This modular design allows the method to accommodate varied urban forms and data conditions, making it adaptable for both small-scale and city-scale façade reconstruction tasks.

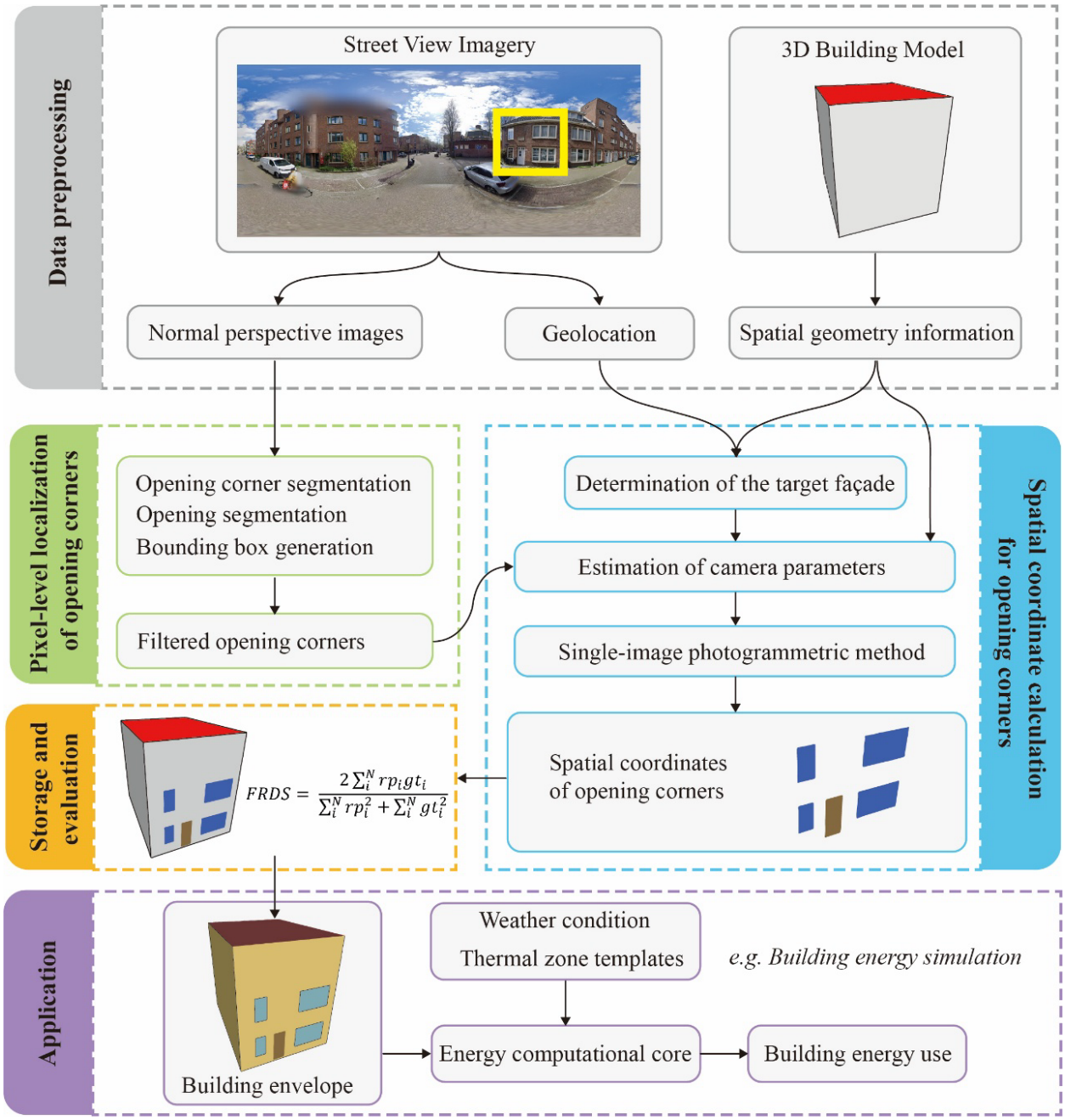

Fig. 1. Workflow of the proposed method.

To ensure the reproducibility and interpretability of the proposed workflow, three key assumptions adopted during camera parameter estimation and façade opening reconstruction are summarized below:

First, it is assumed that the target façade is largely planar and clearly visible in at least one street view panorama, with limited occlusion from vegetation, vehicles, or pedestrians. This assumption ensures that façade opening boundaries can be reliably detected and that the façade plane can be consistently identified in the 3D building model. Second, the workflow assumes that the selected and cropped perspective image represents a single dominant façade. The cropping step is designed to minimize panoramic distortion, privacy masking, and irrelevant scene content, and to ensure that the façade occupies the majority of the image area. While this step is currently performed manually to guarantee geometric reliability, it does not alter the mathematical formulation of the reconstruction and can be automated in future implementations. Third, during camera parameter estimation, the center pixel of the cropped perspective image is assumed to approximately correspond to the geometric center of the target façade. This façade center alignment simplifies the estimation of the translation vector between the camera and façade plane. Deviations from this assumption may introduce lateral shifts in reconstructed façade opening positions, particularly for façades observed at oblique angles.

3.1 Data preprocessing

The proposed façade opening reconstruction workflow primarily relies on two main data sources. The first is an existing 3D building model, which can be at LOD1 or LOD2, depending on data availability. Regardless of the level of detail, it is essential that the spatial location of the buildings and their associated coordinate system are clearly defined, as these provide the geometric reference framework for the reconstruction process. In this paper, the building models used already include an explicitly defined coordinate reference system provided by the dataset creator. Each CityJSON file specifies its reference system in the metadata field, which ensures that the geometry is correctly positioned in geographic space and can be aligned with the spatial information extracted from SVI.

The second source is panoramic SVI, which offers ground-level visual information critical for detecting façade openings. Each SVI contains essential metadata, including a unique panorama ID, geographic coordinates, capture date, image resolution, and the north rotation parameter, all of which facilitate accurate spatial alignment between the imagery and the building model. To provide a reconstruction with confidence, the SVI used in this paper must fulfill a small set of simple requirements. The façade of interest should be clearly visible in at least one panorama with minimal obstruction from vegetation, vehicles, or pedestrians so that the segmentation network can accurately capture façade opening boundaries. The image resolution must be sufficient for window and door corner localization, and the SVI metadata must provide the camera location and the coordinate reference information.

We used the *Equirec2Perspec* package [42] to split the equirectangular panorama into normal perspectives according to the given FOV, theta, phi, as shown in Fig. 2. FOV refers to the field-of-view angle in degrees. Theta and phi are polar coordinates in degrees, where theta corresponds to rotation around the vertical axis and phi represents rotation around the horizontal axis. In this paper, the *Equirec2Perspec* package is used only to generate perspective views from the panoramic SVI. The tool applies a standard spherical projection based on the specified field of view and viewing angles to create a rectified perspective image. Since this procedure is part of the package's built-in functionality and not the methodological focus of our workflow, only a concise description is provided here.

A notable aspect of our workflow involves manually selecting and cropping the SVI perspective to ensure that the target façade is captured clearly and with minimal occlusion. This step is necessary because panoramic street-level imagery often contains distortions, privacy masks, blurriness, or physical obstructions such as pedestrians, vehicles, and vegetation, all of which may propagate errors into subsequent automated stages. By manually refining the crop, the façade boundaries can be accurately aligned before entering the photogrammetric estimation process, ensuring that the extracted image provides reliable reference points for the 2D-to-3D spatial reconstruction. While this manual operation ensures high quality in the presence of inconsistent SVI quality and typical urban occlusions, it is not an intrinsic requirement of the workflow. Automating that manual cropping step could be implemented with two primary developments. First, automated façade-detection models to identify which façade segments are visible in each SVI panorama, and quantitative quality-screening indicators to evaluate occlusion, blur, privacy

masking, or extreme distortion. Second, incorporating such detection results would enable automated, robust view selection and cropping, allowing the workflow to scale efficiently to large urban areas. Future versions of the pipeline will integrate such automated modules to support city-scale deployment. The cropped image provides the reference points for calculating the spatial coordinates of façade openings described in Section 3.3.

Fig. 2. Preprocessing of a panoramic street view image to facilitate the acquisition of the façade.

3.2 Pixel-level localization of façade-opening corners

In this stage, the SVI cropped from the previous step is processed to localize the corners of façade openings. The localization follows a four-stage deep learning–based workflow.

Stage 1 and 2: Semantic segmentation of façade openings and corner candidates. The first two stages apply semantic segmentation to detect façade openings (windows and doors) and the potential positions of their corners, as shown in Fig. 3(a) and Fig. 3(b). We adopted TernausNet [43], a U-Net variant with strong performance in pixel-level tasks, trained on the Amsterdam

façade dataset from the City of Amsterdam's open data portal (https://data.amsterdam.nl/). The training dataset contains pixel-level annotations for windows and doors, which are treated as the representative façade openings in this paper. Other façade elements, such as balconies, ventilation outlets, or shopfront glazing, were not labeled and are therefore excluded from both model training and evaluation, as illustrated in Appendix A. Such detailed segmentation provides a robust basis for the subsequent corner extraction and 3D re-projection steps. To ensure adequate generalisation, the dataset was randomly divided into a training set comprising 90% of the samples and a validation set comprising the remaining 10%. TernausNet was trained using the Adam optimizer with an initial learning rate of 1×10^{-4}, a batch size of 4, and cross-entropy loss. Following common practice in façade segmentation, the network was fine-tuned for 50 epochs. Training and evaluation followed the COCO evaluation protocol, reporting the mean Average Precision (AP) across IoU thresholds from 0.50 to 0.95 in 0.05 increments, along with AP50 and AP75 to represent more lenient and stricter localization criteria. The best-performing TernausNet model achieved AP = 65.94%, AP50 = 90.47%, and AP75 = 76.04%. In façade segmentation tasks, an AP50 above 85% is generally regarded as excellent, indicating high reliability in capturing façade-opening boundaries [44].

Stage 3: Bounding box generation. One challenge in direct corner prediction is the absence of grouping information, and each detected corner is treated as an independent point without explicit association to its corresponding façade opening. To address this, we introduced a bounding box computation step: for each segmented opening, the extremal coordinates of its region are used to form a bounding box, as illustrated in Fig. 3(c). This enables grouping and spatial association between candidate corners within the same façade opening.

Stage 4: Corner assignment and refinement. In the final stage, bounding box geometry is integrated with the detected corner set. For each façade opening, the detected corner closest to each bounding box vertex is assigned as that opening's corner. In cases where a corner is missing, the bounding box vertex itself is used as a substitute, ensuring completeness of the structural representation. The filtered and assigned corners for each opening are shown in Fig. 3(d).

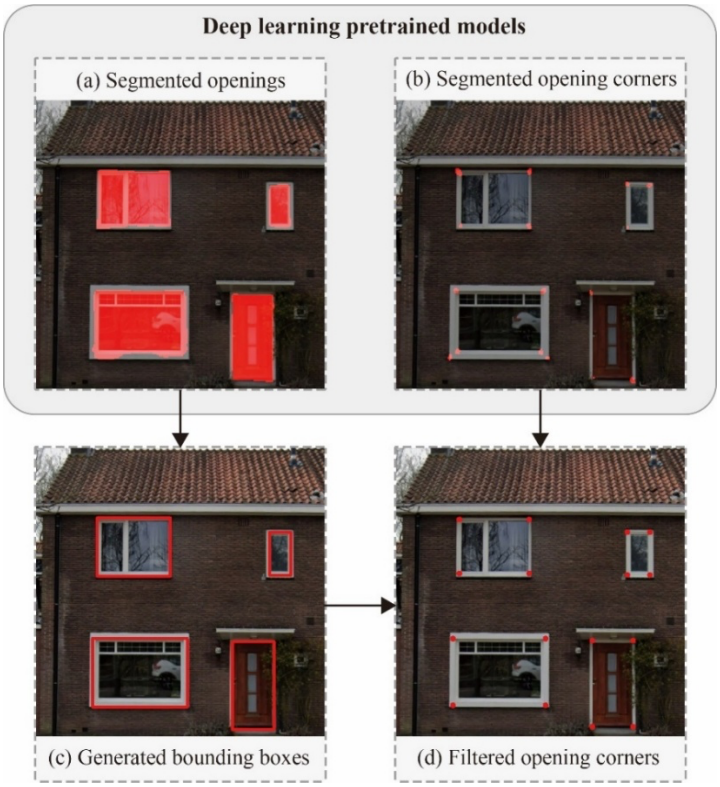

Fig. 3. Localization process of opening corners.

3.3 Spatial coordinate calculation for façade-opening corners

The objective of this section is to determine onto which façade the identified corners should be projected and to propose a projection method for calculating the spatial coordinates of these corners.

3.3.1 Determination of the target façade in building models

The determination process involves three steps. Initially, the normalized normal vector is calculated for each surface of the target building, pointing outward from the building. The surface with the normal vector (0, 0, -1) is considered as the building's footprint. Then, the view vector needs to be computed, which represents the direction from the building to the SVI camera location. Finally, the angle α, which is defined as the angle between the normal vector and the view vector, is calculated for each surface. The normal vector with the smallest α is recognized as the normal vector for the target façade.

3.3.2 Estimation of camera parameters

This section details the process of estimating SVI camera parameters and transforming recognized 2D opening positions into corresponding 3D spatial coordinates on the determined target façade using the single-image photogrammetric method. This technique utilizes four distinct coordinate systems: the world coordinate system (WCS), camera coordinate system (CCS), image coordinate system (ICS), and pixel coordinate system (PCS) [45,46], as shown in Fig. 4.

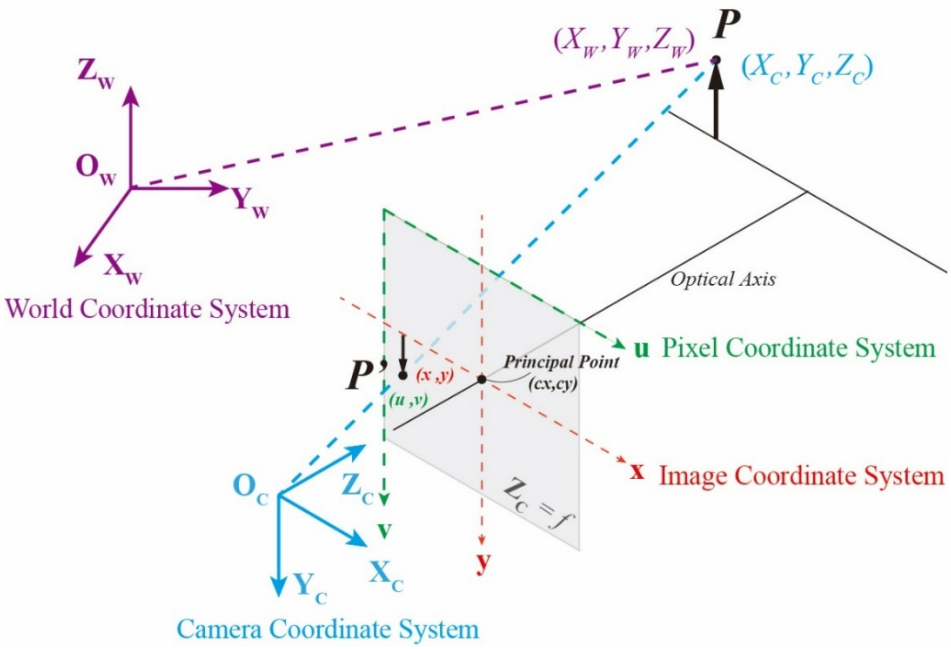

Fig. 4. Projection of point P onto the image plane.

The camera imaging process can be interpreted as a transformation from WCS to PCS. In this paper, the WCS aligns with the spatial coordinate system used by the building model. A point P in the 3D space is represented by coordinates $(X_w,\ Y_w,\ Z_w)$ in WCS, corresponding to the pixel location (u, v) in PCS. The projection process in Eq. (1) can be understood intuitively as a sequence of coordinate transformations. A point in the real world is first represented in the camera coordinate system, which describes its relative position to the camera. K denotes the intrinsic matrix, containing the internal parameters of the camera. The extrinsic matrix $[R|t]$, comprising a rotation matrix R and a translation vector t, describes the camera's position and orientation relative

to the WCS [47]. Definitions of all parameters and detailed derivations are provided in Appendix B.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K[R|t] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \left(R \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + t \right) \quad (1)$$

Now, moving on to the single-image photogrammetric process, which derives façade-opening corner locations in WCS from the respective positions in PCS. Theoretically, if (u, v) of a point in PCS is known, $(X_w,\ Y_w,\ Z_w)$ of this point in WCS can be derived from Eq. (2), transformed from Eq. (1), where R^{-1} and K^{-1} are inverse matrixes of R and K, respectively. However, R, t, and K are unknown, estimating these parameters becomes a necessary part of the workflow.

$$\begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = R^{-1}K^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} - R^{-1}t \quad (2)$$

R can be calculated based on the view vector obtained in Section 3.3.1, as the direction of the Z_C axis is opposite to that of the view vector. The calculation formula is $R = R_x \cdot R_y \cdot R_z$, where R_x, R_y and R_z are the rotation vectors of WCS around the x, y, and z axes, respectively. After rotating CCS 90° counter-clockwise around the X_C axis, the spatial distribution of CCS and WCS can be obtained, as illustrated in Fig. 5. (X_{wn}, Y_{wn}, Z_{wn}) denotes the normal vector of the target façade. This figure also illustrates four possible orientations of the normal vector of the target façade on the $X_wO_wY_w$ plane. To align the coordinate axes of CCS and WCS, rotating around the Z_C axis is sufficient at present. In each subfigure in Fig. 5, the angle required for counter-clockwise rotation around the Z_C axis is highlighted in red. This angle value is associated with β, which represents the angle between the orientation of the normal vector on the $X_wO_wY_w$ plane and the Y_w axis.

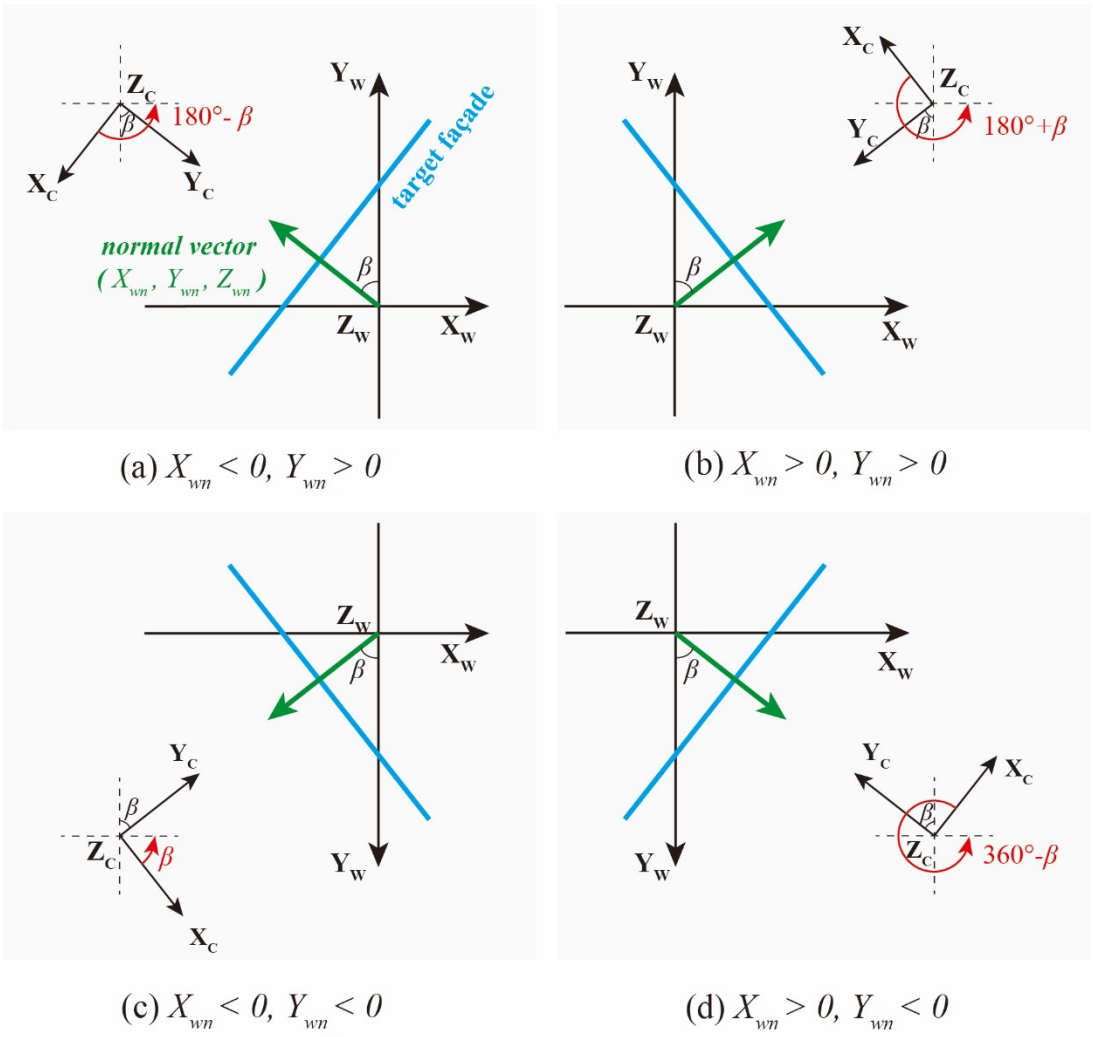

Fig. 5. Four possible orientations of the normal vector of the target façade.

The translation vector t represents the spatial position of the camera relative to the façade. Since intrinsic parameters are not available in SVI metadata, we assume that the center pixel location (c_x, c_y) corresponds to the center of the façade (X_{w0}, Y_{w0}, Z_{w0}). This assumption simplifies the reconstruction process but may lead to inaccuracies when the camera is not positioned directly in front of the façade. In such cases, deviations from this assumption could cause shifts in the reconstructed façade openings, particularly when viewed at oblique angles. These distortions become more pronounced for façades viewed at sharper angles, potentially affecting the alignment of the reconstructed façade openings. Substituting both into Eq. (1), Eq.

(3) can be derived. Initial values of f_x and f_y in intrinsic matrix K in Eq. (3) is 100, and Section 4.1 verifies that these initial values do not affect the result of t.

$$t = K^{-1}\begin{bmatrix} c_x \\ c_y \\ 1 \end{bmatrix} - R\begin{bmatrix} X_{w0} \\ Y_{w0} \\ Z_{w0} \end{bmatrix} \tag{3}$$

The intrinsic matrix K can be further estimated by assuming the bottom left corner (u_1, v_1) of the cropped street image in PCS is the bottom left corner of the corresponding target façade in WCS. The coordinate (X_{C1}, Y_{C1}, Z_{C1}) in CCS can be represented by substituting (X_{w1}, Y_{w1}, Z_{w1}) into Eq. (1), as shown in Eq. (4), and then the estimated values of f_x and f_y in the intrinsic matrix K are calculated by $f_x = \frac{u_1 - c_x}{X_{C1}}$ and $f_y = \frac{v_1 - c_y}{Y_{C1}}$.

$$\begin{bmatrix} X_{C1} \\ Y_{C1} \\ Z_{C1} \end{bmatrix} = [R|t]\begin{bmatrix} X_{w1} \\ Y_{w1} \\ Z_{w1} \\ 1 \end{bmatrix} \tag{4}$$

Once R, t, and K are computed, Eq. (2) is used to convert the 2D pixel positions of all detected façade-opening corners into their corresponding 3D coordinates in WCS. This enables the reconstruction of façade openings from a single SVI panorama.

3.4 Storage of corner information

Once the spatial coordinates of the façade-opening corners have been calculated, the next step is to integrate this geometric information into the corresponding 3D building model. This stage serves as a bridge between the computational derivation of corner positions (Section 3.3) and the generation of an enhanced, application-ready building model.

The integration process begins with extracting the geometric representation of the target building from the original dataset, ensuring consistency with the coordinate system used during the reconstruction process. Based on the calculated corner positions, façade openings are then created on the target façade by sequentially "punching" holes that correspond to the spatial layout of each reconstructed façade opening. Following this, the façade openings are refined by inserting the corresponding corner geometries, which ensure precise alignment and preserve the intended façade structure. Through this process, the original LOD1 or LOD2 building is transformed into a

seamless model that incorporates realistic façade openings, resulting in a higher level of detail and improved geometric fidelity.

3.5 Evaluation metrics

This paper uses the Façade Re-projection Dice Score (FRDS) [25] as the primary quantitative metric for evaluating the quality of reconstructed façade openings. FRDS measures the spatial overlap between two binary images by comparing the shape and distribution of objects, making it particularly suitable for assessing the consistency between reconstructed façade openings and their corresponding reference representations in image space. When the reprojection result perfectly matches the reference image, the FRDS value equals 1. In Eq. (5), rp_i and gt_i are the pixel values of reprojection and image reference, respectively.

$$FRDS = \frac{2\sum_i^N rp_i gt_i}{\sum_i^N rp_i^2 + \sum_i^N gt_i^2} \tag{5}$$

The choice of FRDS is motivated by both the objectives of this paper and the characteristics of the available reference data. The proposed workflow focuses on reconstructing façade openings from single street view images and integrating them into existing 3D building models at scale. In such settings, metrically surveyed ground truth for façade openings is rarely available, and manual annotation of accurate 3D geometries is infeasible for large urban areas. Instead, reference façade opening geometries are obtained through image-based semantic segmentation of street-level imagery, which provides a scalable, consistent, and visually grounded benchmark for evaluation. FRDS directly quantifies how well the reconstructed façade openings, when re-projected to the camera viewpoint, reproduce the spatial extent and distribution of façade openings observed in the original imagery. Nevertheless, FRDS has inherent limitations that should be explicitly acknowledged. Because the reference data are derived from semantic segmentation rather than direct physical measurement, FRDS does not represent absolute geometric accuracy in real-world units. Its achievable upper bound is constrained by segmentation quality, including boundary smoothness, occlusion effects, and potential misclassification of façade elements.

The building model generated in Section 3.4 is utilized as the subject for evaluation. Based on the calculated R, t, and K, as well as Eq. (1), a virtual photo of the target façade taken at the

street view camera location can be obtained, where the pixel value is 1 for façade opening (white), and the rest is 0 (black), as shown in Fig. 6(a). This process can be called re-projection. Then, the façade opening binary images identified in Section 3.2 are treated as image reference in this paper, as shown in Fig. 6(b). By inputting both images into Eq. (5), the distribution of the calculated façade openings on the target façade is evaluated through FRDS.

(a) Re-projection result of target façade in PCS (b) Image reference of target façade in PCS

Fig. 6. FRDS score representation.

4. Validation and results

This section presents the validation and results of the proposed façade opening reconstruction workflow. It first examines the mathematical properties of the camera parameter estimation to verify the robustness of the underlying assumptions (Section 4.1). It then evaluates the reconstruction performance through a series of case studies in Amsterdam, Netherlands, covering individual buildings, multiple adjacent buildings, and larger urban blocks, with quantitative assessment using the FRDS (Sections 4.2.1–4.2.3). Finally, the section demonstrates the practical implications of the enhanced building models through a comparative building energy simulation before and after model augmentation (Section 4.3).

4.1 Influence of initial values of f_x and f_y in intrinsic matrix K

From a mathematical point of view, this section aims to prove that the initial values of f_x and f_y do not have any effect on the estimation of t in Section 3.3.2. In Eq. (3), this mathematical problem can be further transformed into the need to prove that $K^{-1}\begin{bmatrix} c_x \\ c_y \\ 1 \end{bmatrix}$ is a unique matrix. At present, K can be represented by $\begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$. According to the inverse matrix theorem, $K^{-1}K = I$, where I denotes the identity matrix. Therefore, when K^{-1} is assumed to be $\begin{bmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{bmatrix}$, $K^{-1}K$ can be rewritten as Eq. (6):

$$\begin{bmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{bmatrix}\begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} k_{11}f_x & k_{12}f_y & k_{11}c_x + k_{12}c_y + k_{13} \\ k_{21}f_x & k_{22}f_y & k_{21}c_x + k_{22}c_y + k_{23} \\ k_{31}f_x & k_{32}f_y & k_{31}c_x + k_{32}c_y + k_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (6)$$

Since the focal length of the camera does not equal to 0, if the result of multiplying a variable by f_x or f_y is zero, the value of this variable must be 0. Based on this, K^{-1} can be represented by only four variables, k_{11}, k_{22}, c_x and c_y, as shown in Eq. (7).

$$K^{-1} = \begin{bmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{bmatrix} = \begin{bmatrix} k_{11} & 0 & -k_{11}c_x \\ 0 & k_{22} & -k_{22}c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (7)$$

Therefore, it can be seen from Eq. 8 that $K^{-1}\begin{bmatrix} c_x \\ c_y \\ 1 \end{bmatrix}$ is a 3×1 unique matrix $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, which has no effect on the estimation of the translation vector t in the estimation of camera parameters.

$$K^{-1}\begin{bmatrix} c_x \\ c_y \\ 1 \end{bmatrix} = \begin{bmatrix} k_{11} & 0 & -k_{11}c_x \\ 0 & k_{22} & -k_{22}c_y \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} c_x \\ c_y \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad (8)$$

4.2 Case studies on model augmentation

To comprehensively evaluate the effectiveness and robustness of the proposed façade opening reconstruction method, we designed and implemented three representative case studies in Amsterdam, Netherlands. These case studies were structured to test the method's applicability across varying spatial contexts: (1) individual buildings, (2) multiple adjacent buildings, and (3) larger, more complex urban blocks. The selection of Amsterdam as the study area was motivated by the ready availability of high-quality 3D building models (Peters et al., 2022), which can be directly obtained from the open-access 3DBAG platform (https://3dbag.nl/). In case studies, the criteria for selecting the target building are that the building's street façade must be almost completely captured by cameras without being obstructed by trees, cars, or other urban objects, and the building should have a corresponding building model available. To better maintain the model, this paper adopted the CityJSON format for storing building models. CityJSON is a JSON-based encoding for a subset of the CityGML data model, which promotes the creation, reading, and modification of city models in a more accessible and manageable manner [48]. Then, we used *cjio*, a Python package, to extract the spatial geometry information of each target building [48]. For street-level imagery, Google Street View (GSV) served as the primary source, providing panoramic views enriched with metadata essential for spatial alignment. In the following case studies, GSV is used only as an illustrative example of a widely available panoramic imagery platform rather than as a strict requirement of the workflow. The proposed method is fully compatible with any street-view service that supplies panoramic imagery together with basic metadata such as geographic coordinates and viewing direction, including platforms such as Mapillary, KartaView, Baidu Street View, or other municipal street-level imagery archives.

The three case studies are organized as follows. Section 4.2.1 focuses on reconstructing façade openings for individual buildings, labeled with the prefix "B." Section 4.2.2 extends the approach to multiple adjacent buildings within the same block, labeled with the prefix "MB." The selected cases, including each building's panorama ID, camera location, capture date of the SVI, and the corresponding CityJSON building ID (CJID), are summarized in Appendix C. The uniqueness of the panorama ID and CJID ensures that both the imagery and 3D model can be retrieved by any user for verification. Notably, multiple adjacent buildings may map to several distinct entries in the original 3D city model, for example, case MB0 consists of MB0-0, MB0-1, and MB0-2. To provide a clearer overview of the case-study locations, Fig. 7 presents the spatial distribution of all selected buildings, offering an intuitive geographic reference that complements

the information listed in Appendix C. Finally, Section 4.2.3 addresses more complex urban contexts involving multiple urban blocks. These blocks were sourced as a single tile, randomly downloaded from the 3DBAG platform. The selected file, identified as tile 10-432-712, contains numerous LOD2 building models and is defined within the EPSG:7415 coordinate reference system, with a bounding box of (121570.340625, 484879.6545, 122106.952625, 485402.7195). This larger-scale case study tests the method's scalability and adaptability to heterogeneous urban morphologies.

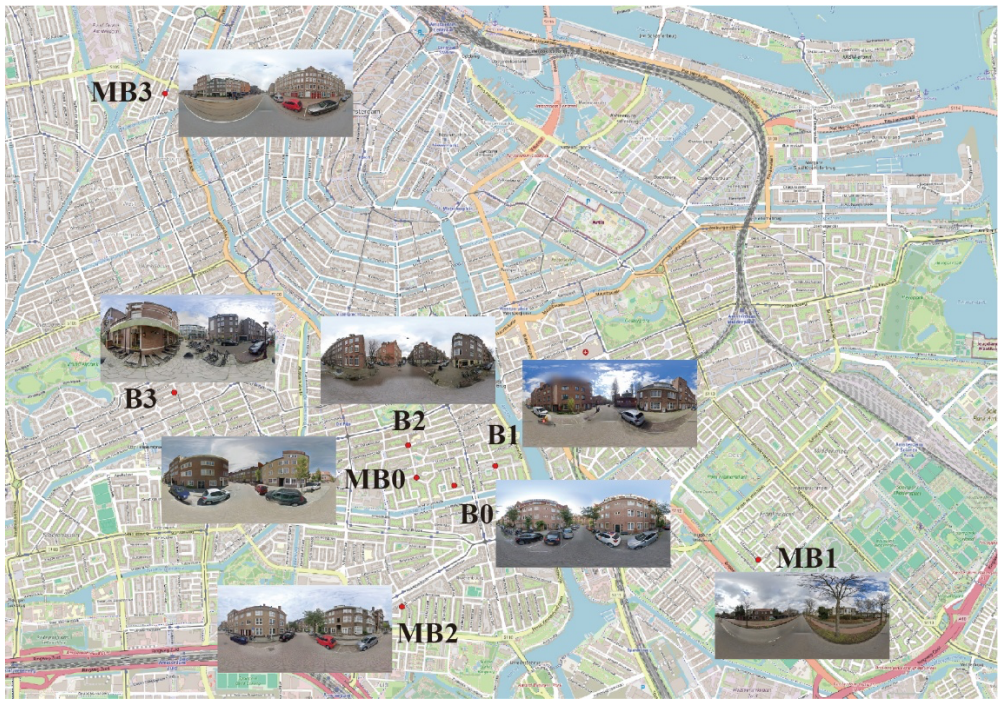

Fig. 7. Spatial distribution of the selected case-study buildings.

4.2.1 Façade opening reconstruction of individual buildings

The target buildings in this section are B0-B3. Fig. 8 shows the preprocessing results of SVI and CityJSON information related to these target buildings, including cropping the perspective from SVI and extracting the geometric information from the original building model through CJID.

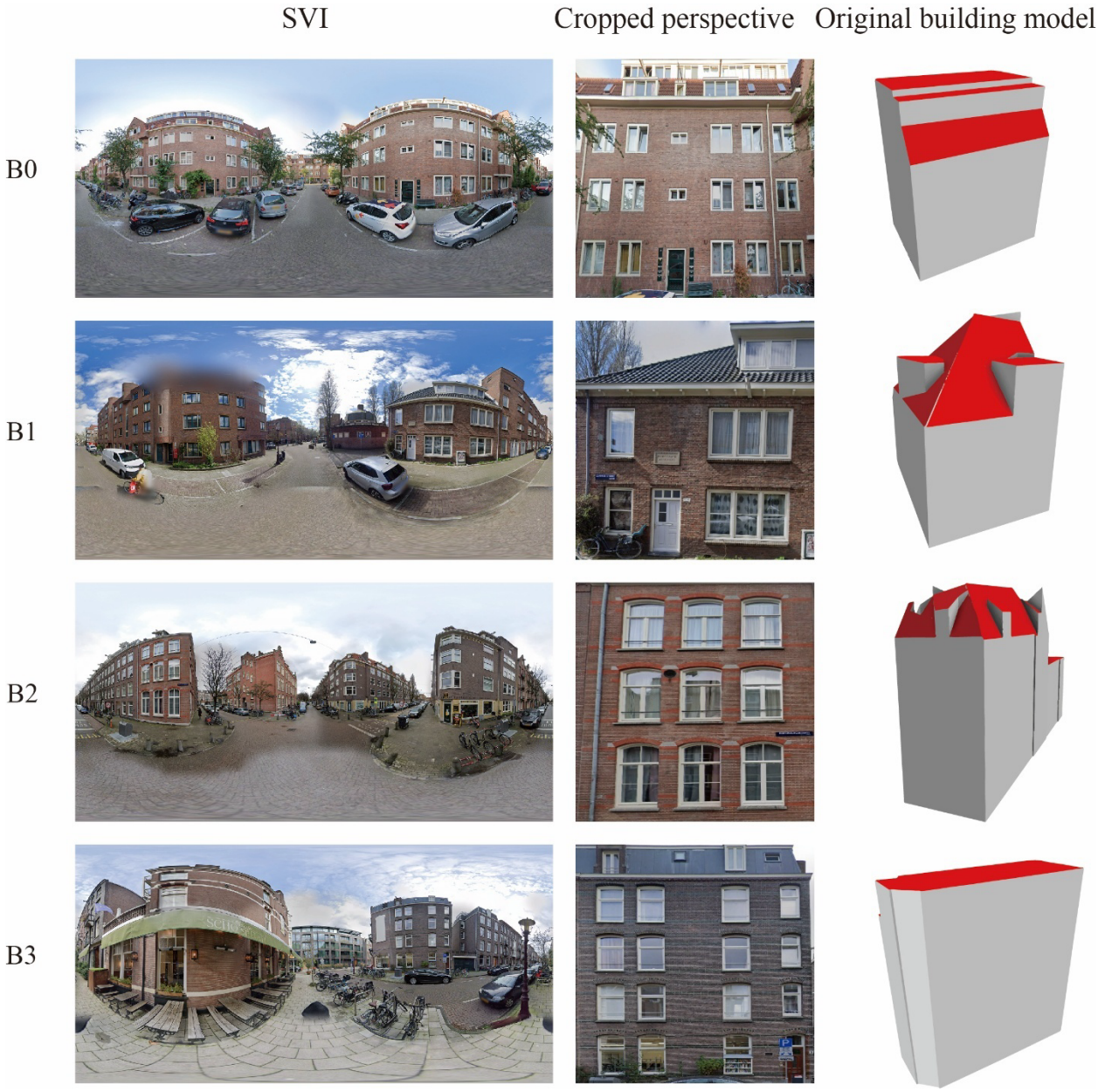

Fig. 8. Data preprocessing results of B0-B3.

The TernausNet models were then utilized to sequentially identify the façade openings and façade-opening corners of the cropped perspectives. The results of semantic segmentation for the four target buildings are shown in Fig. 9, with two columns representing segmented openings and corners. Based on segmented openings, the bounding boxes were further generated and served as

the prototype for identified façade openings. Combining these boxes with segmented corners, the final filtered corners in Fig. 9 were obtained.

Fig. 9. Pixel-level localization of façade-opening corners for B0-B3.

By utilizing the camera location of SVI along with the original building model as inputs, the target façade of each building can be determined. Subsequently, the filtered corners were projected onto the target façade. Through the single-image photogrammetric method, the rotation matrix R, translation vector t, and intrinsic matrix K were back-calculated. By applying Eq. (2), the corresponding coordinates of the façade-opening corners in WCS were then determined. Finally, these façade openings were integrated into the original building model, resulting in the visualization presented in Fig. 10. All updated building models can be validated by *val3dity* [49].

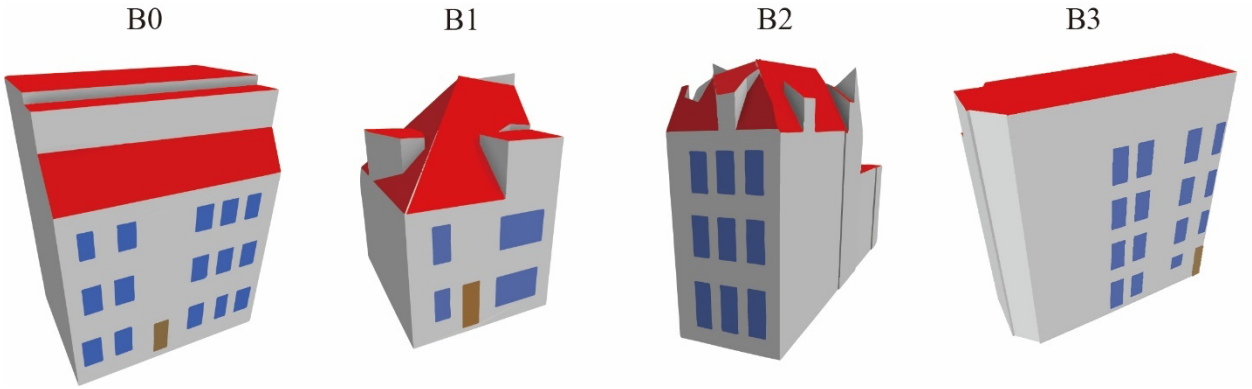

Fig. 10. Visualization of B0-B3 with reconstructed façade openings.

The binary images representing the image reference were created using the segmented opening in Fig. 9. The reprojection binary images were generated by simulating the perspective of standing on a camera location and capturing the 3D building depicted in Fig. 10. This camera image formation process was achieved by inputting the previously calculated R, t and K into Eq. (1). Fig. 11 displays the two binary images generated for each building. By substituting both images (matrices) into Eq. (5) to obtain the FRDS, the results are shown in Table 1. The FRDS values range mostly between 0.86 and 0.92, indicating that the calculated spatial locations of the façade openings are relatively accurate. As for errors, the segmented openings may not perfectly represent the actual geometric shape of the façade openings, as they are usually irregular and only provide a rough outline. This discrepancy between segmented openings and boundary boxes used for the final corner generation could be a potential source of error.

It is important to clarify that the quantitative evaluation presented in this paper is designed to assess relative improvement over low-LOD baseline building models rather than to establish absolute geometric correctness of reconstructed façade openings. Consequently, FRDS values should be interpreted as indicators of how effectively the reconstructed façade openings enhance the spatial realism and façade-level detail of existing low-LOD models. Within this context, the consistently high FRDS scores demonstrate that the proposed method yields substantial relative improvement in façade representation while acknowledging that absolute geometric accuracy remains constrained by the quality of the reference data and imaging conditions.

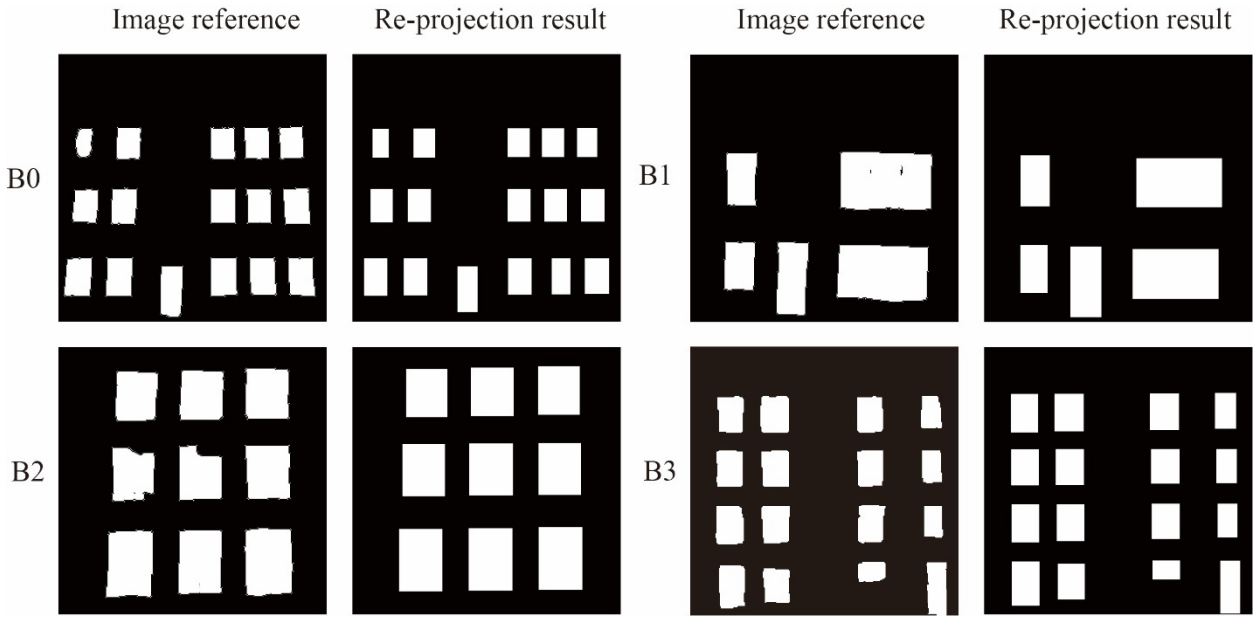

Fig. 11. Binary images representing image reference and reprojection of B0-B3.

Table 1.

Façade Re-projection Dice Score of B0-B3.

Building ID	FRDS
B0	0.872
B1	0.921
B2	0.865
B3	0.920

4.2.2 Façade opening reconstruction of multiple adjacent buildings

In the bustling cityscape of Amsterdam, it is common for buildings within the same block to be closely situated next to each other. Interestingly, from a human visual perception standpoint, a single building with a uniform color and architectural style may correspond to multiple CityJSON buildings. In this case, one question arises: how can the façade openings of these buildings be reconstructed solely using a panoramic street view image? The preprocessing results of Fig. 12 lay the foundation for this work. The reconstruction process proceeds building by building, requiring a meticulous match between building instances in the normal perspectives and those in CityJSON data. Note that the quantity of target buildings coincides with the number of buildings specified in the CityJSON model.

Fig. 12. Data preprocessing results of MB0-MB3.

According to the workflow of Section 3.2, Fig. 13(a) showcases the outcomes of the semantic segmentation process applied to the façade openings and opening corners on each building façade.

This process leads to the generation of bounding boxes and filtered corners. By employing Eq. (2), the corresponding coordinates in WCS of the filtered corners were obtained. These coordinates were then integrated into the original building model, as depicted on the left side of Fig.13(b). Subsequently, this paper utilized *cjio* to merge each building model containing façade openings. As a result, the reconstruction of the façade openings for multiple adjacent buildings has been completed. The final multiple adjacent building models can also be checked by *val3dity* to ensure they are free of any geometric errors. Additional reconstruction examples for MB1–MB3 are provided in Appendix C.

The evaluation procedure involves re-projecting the target façade of each building onto a virtual camera and utilizing the segmented openings presented as the image reference, as shown in Fig.13(c). After substituting both into Eq. (5), FRDS scores were calculated, listed in Table 2. The findings reveal that the FRDS scores surpass 0.84 for all façades, with some façades even exceeding 0.92, indicating a strong similarity between the plane distribution of façade openings captured by the camera in the real world and the virtual photos obtained from the generated 3D building models. The value of FRDS is largely affected by the angle α between the view vector and the normal vector of the target façade. The larger α is, the higher probability that the photo of the corresponding building would be stretched and deformed. An example can be observed in Fig. 13(a), where the generation of the boundary box for MB0-2 exhibits a noticeable deviation from the original segmented opening, resulting in a reduction in FRDS. Compared with Section 4.2.1, Section 4.2.2 shows that the reconstruction process of individual buildings is not fundamentally different from that of multiple adjacent buildings, except that the latter requires integrating the reconstruction results of multiple buildings.

In addition, the derived camera parameters directly influence the reprojection-based evaluation because they determine the geometric mapping between 3D façade openings and their 2D projections. Although the parameter estimation procedure is designed to be stable, small deviations in the estimated intrinsic parameters may slightly shift the projected opening boundaries, which can lower the FRDS value. These deviations are most evident when the façade is viewed from an oblique angle, where projection distortions are more sensitive to parameter inaccuracies. The consistently high FRDS values across all cases indicate that such parameter errors remain limited and do not materially affect the overall reconstruction quality.

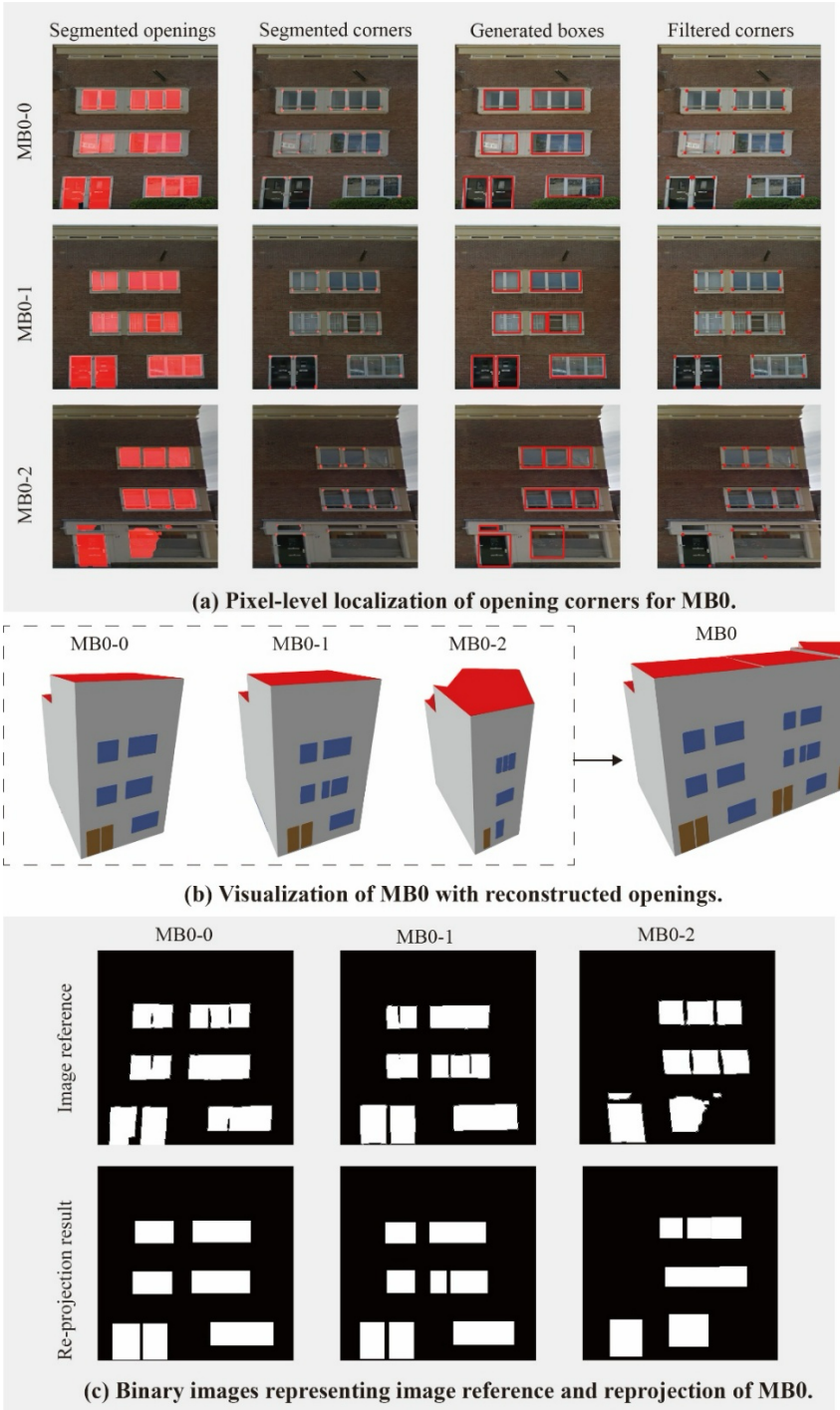

Fig. 13. Segmentation, corner detection, 3D integration and re-projection evaluation for MB0.

Table 2.

Façade Re-projection Dice Score of MB0 - MB3.

Building ID	FRDS	Multiple building ID	FRDS
MB0-0	0.910	MB0	0.894
MB0-1	0.928		
MB0-2	0.845		
MB1-0	0.922	MB1	0.887
MB1-1	0.852		
MB2-0	0.888	MB2	0.889
MB2-1	0.859		
MB2-2	0.920		
MB3-0	0.893	MB3	0.888
MB3-1	0.881		
MB3-2	0.889		

4.2.3 Façade opening reconstruction of urban blocks

To further test the robustness and generalization capability of the proposed reconstruction method, this section applies the workflow to a larger and more heterogeneous urban context comprising multiple urban blocks in Amsterdam. Unlike the previous case studies, which focused on individual or adjacent buildings, this large-scale experiment evaluates the method's performance when handling dense and morphologically diverse city environments. The dataset for this case was sourced from the 3DBAG platform (https://3dbag.nl/) as a single CityJSON tile, randomly selected to avoid bias in building morphology or SVI coverage. The chosen file contains 628 LOD2 building models, as shown in Fig. 14(a). Within the surrounding street network of these buildings, a total of 2,200 SVI panoramas were available, providing a rich but varied set of inputs for façade detail enhancement.

The sampling mechanism followed a rule-based and reproducible selection procedure. For each building, all associated SVI panoramas were examined and evaluated against a predefined set of quality criteria. A building was considered suitable for reconstruction if at least one panorama satisfied the following conditions: the target façade was largely visible with limited occlusion from trees, vehicles, or street furniture; the angle between the camera view vector and the façade normal vector was within a moderate range to avoid excessive perspective distortion; and the camera was not located at extreme proximity to the façade, which would otherwise result in severe geometric deformation after perspective projection. When multiple panoramas satisfied these criteria for a given building, the panorama with the smallest angle between the façade normal vector and the

camera view vector was selected to minimize projection distortion. If only a single panorama was available and it failed to meet the quality requirements, or if none of the panoramas provided an adequately observable façade, the building was excluded from the reconstruction process. These exclusion rules were applied uniformly across the entire CityJSON tile. Following this systematic filtering process, 460 out of the 628 buildings were retained for façade opening reconstruction. The resulting sample, therefore, represents all buildings within the selected urban block that are observable under realistic street view imaging conditions and meet the methodological assumptions of the proposed workflow.

Validation for this experiment was conducted on a substantial dataset of building models across multiple urban blocks, with reference façade opening geometries obtained through a semi-automated image-based façade analysis of street-level imagery. These reference datasets do not represent flawless ground truth. Nevertheless, they provide a scalable, consistent, and practically applicable benchmark for evaluating reconstruction performance at the city scale. The final enhanced building models are illustrated in Fig. 14(b), and the corresponding distribution of FRDS scores for all reconstructed openings is shown in Fig. 14(c). Most FRDS values fall within the range of 0.84 to 0.98, with a strong concentration above 0.90. This indicates that the majority of reconstructed façade openings exhibit a high degree of spatial correspondence with their reference geometries under diverse and realistic urban conditions. The mean FRDS score of 0.9148 and the median score of 0.9160 represent the central tendency of the reconstruction quality, demonstrating that most reconstructed façade openings exhibit high spatial consistency with their reference geometries. The variance of 0.0003 reflects the low degree of variability across the dataset, indicating that the reconstruction method performs consistently well across different urban blocks. Given the inherent uncertainties in the reference data, these values should be interpreted as indicators of relative reconstruction accuracy. They should not be regarded as absolute error metrics. Nevertheless, the consistently high FRDS scores across a dense and morphologically heterogeneous urban setting demonstrate that the proposed method maintains robust performance and scalability from individual buildings to large, complex city environments without significant degradation in reconstruction quality.

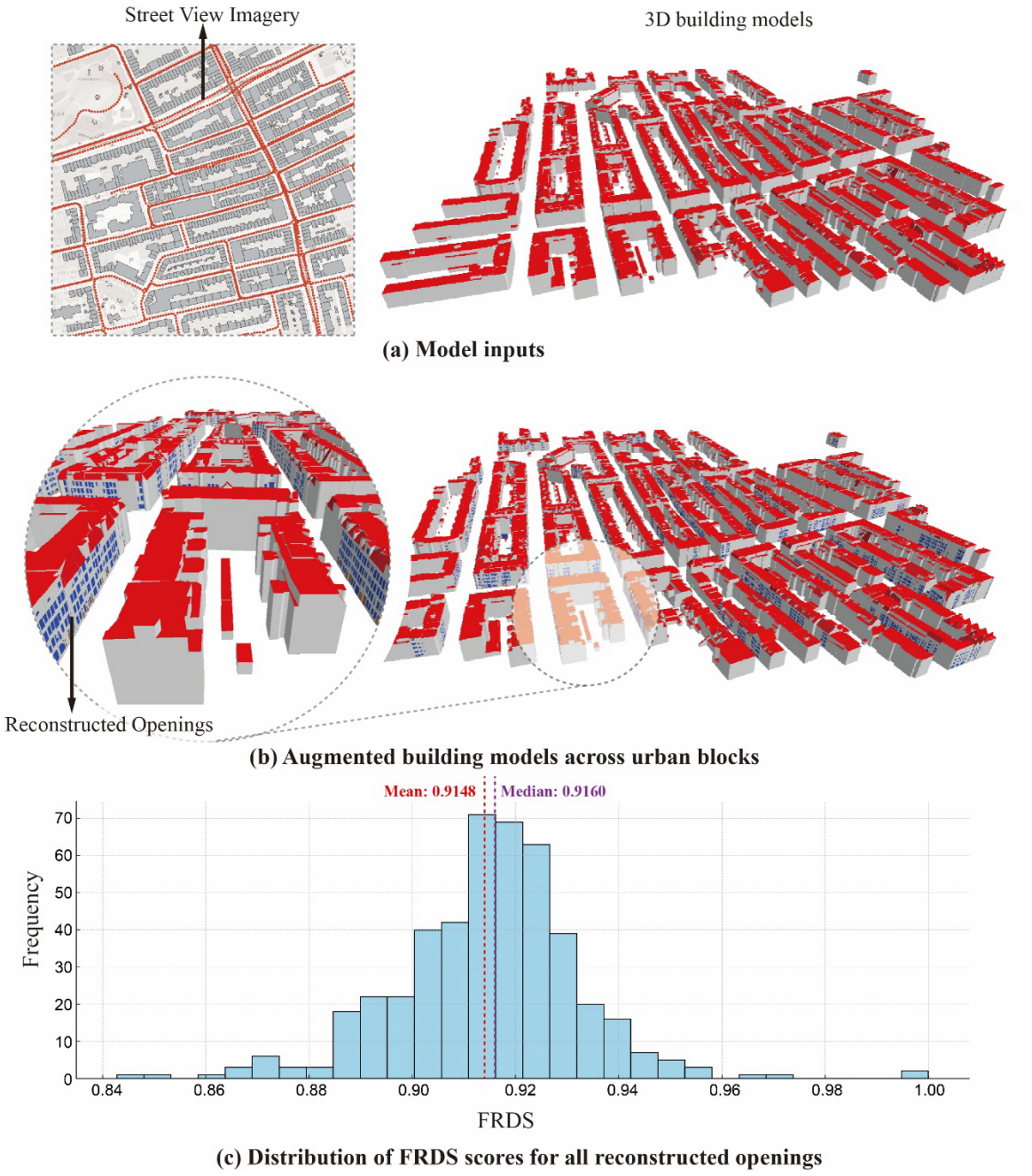

Fig. 14. Façade opening reconstruction of urban blocks in Amsterdam.

4.3 Comparative analysis of building energy simulation before and after model augmentation

This section evaluates the performance of augmented building models in practical applications, using building energy simulation as an example. Incorporating façade openings alters

the material composition of the building envelope and facilitates thermal exchange between indoor and outdoor environments [50], thereby providing more realistic physical conditions for simulation.

Using the B0–B3 examples from Section 4.2, each CityJSON model was imported into SketchUp and converted to an EnergyPlus Input Data File (IDF) via the OpenStudio plug-in. The Python package eppy was then used to batch-assign thermal zone settings based on the large-office prototype model from the U.S. Department of Energy (DOE). The simulations were run in EnergyPlus with the Amsterdam typical meteorological year (TMY) weather file, ensuring all building models were water-tight. While enriching publicly available 3D city models with reconstructed façade openings holds substantial potential for energy modeling, daylight analysis, and broader urban simulation applications, this paper positions the simulation primarily as a methodological demonstration. Integrating such enriched models into operational urban datasets requires coordination with data providers and rigorous quality control, which is beyond the scope of the present work.

Fig. 15 compares monthly heating and cooling loads before and after augmentation. Results show that heating demand dominates in Amsterdam, while augmented models produce higher heating and cooling usage. Annual heating energy increased by 0.93–6.10%, and cooling energy by 49.04–73.53%. These findings align with Troup et al. [5], confirming that higher window-to-wall ratios elevate total energy use, especially cooling loads. In cooling-dominated climates such as Hong Kong or Singapore, this effect would be even greater. Importantly, these results should not be interpreted as evidence of improved accuracy with respect to real-world energy consumption. Building energy performance depends on many interacting factors beyond façade geometry, including occupancy behavior, internal loads, system efficiencies, and operational schedules. Instead, the purpose of this experiment is to demonstrate improved physical realism and increased sensitivity of the simulation outcomes to façade opening characteristics. By replacing generic envelope assumptions with geometry derived from observable façade evidence, the augmented models respond more realistically to variations in window-to-wall ratio, a parameter widely recognized as one of the most influential factors in building energy performance. Overall, the augmented models produce thermal behavior more consistent with physical expectations for buildings with higher window-to-wall ratios, suggesting improved physical realism, though they also reveal the environmental implications of increased openings.

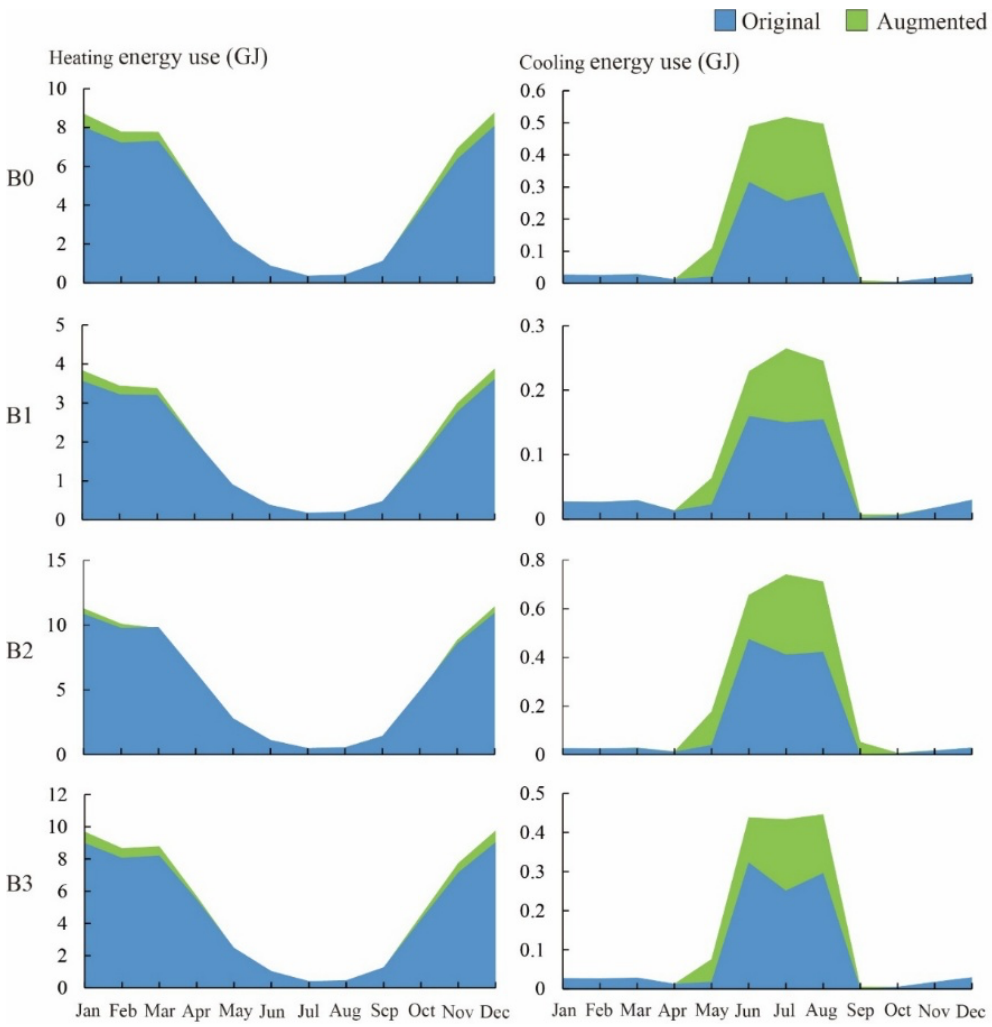

Fig. 15. Monthly heating and cooling energy use before and after the B0-B3 model augmentation, suggesting the benefit of our work.

5. Discussion

The absence of façade opening information remains a critical limitation in most existing 3D building datasets. As highlighted by Wysocki et al. [51], only five of fifty-eight global datasets include such details, largely due to the significant acquisition time, cost, and data preparation effort required [12,52]. This constraint severely limits the capacity of urban models to support fine-grained simulations and applications. From a practical perspective, the reliance on globally available street view imagery represents a key advantage of the proposed workflow. Many state-

of-the-art façade reconstruction approaches depend on UAV imagery, terrestrial laser scanning, or airborne LiDAR data, which require specialized equipment, regulatory approval, and substantial financial investment. As a result, their applicability is often limited to well-funded projects or cities with advanced geospatial infrastructures.

This paper addresses this gap by leveraging widely accessible panoramic SVI in combination with 3D building models. This approach enables scalable façade opening reconstruction without the prohibitive overhead of survey-based methods. Reconstruction quality, evaluated using the FRDS, consistently exceeded 0.84 across three case studies in Section 4.2, spanning individual buildings, multiple adjacent buildings, and urban blocks. These results demonstrate that the reconstructed façade openings align closely with their reference geometries derived from image-based façade analysis, despite the inherent limitations of street-level imagery, which is generally captured from drivable roads and may not provide perfectly frontal views. To clarify how accuracy should be interpreted in light of these limitations, we distinguish spatial consistency and geometric agreement. The FRDS metric quantifies how well reconstructed façade openings match the reference shapes within the façade regions actually visible in the SVI. Although occluded or unobserved façade sections cannot be reconstructed, this incompleteness does not affect the correctness of the façade openings recovered from visible areas. By examining FRDS values across all buildings, we also capture the variability of reprojection performance under different geometries and viewing conditions. The consistently high scores reported in Table 1, Table 2, and Fig. 14 indicate low statistical variance and stable reprojection behaviour. The remaining uncertainties primarily stem from segmentation quality, the degree of façade visibility, and the SVI viewing angle, which together define the confidence bounds of the reconstruction. While there is no established single-image benchmark for direct comparison, the upper bound of achievable FRDS is constrained by the segmentation accuracy of the underlying model, reflected by AP, AP50, and AP75, which offers an indirect reference for assessing reproducibility.

In addition to spatial consistency and geometric agreement, the computational efficiency of the workflow provides further evidence of its scalability. Using a standard laptop (12th Gen Intel i7-12700H, 20 cores, NVIDIA RTX 3060), the largest case in Section 4.3 was processed, involving 460 buildings. The semantic segmentation required 406.83 seconds, the single-image photogrammetric conversion required 146.62 seconds, and the integration of façade openings into the CityJSON model took 4.75 seconds, resulting in a total of approximately 9.3 minutes for the

entire urban block. Because each module in the workflow is independently parallelizable, the pipeline can be further accelerated and extended to city-scale processing without the need for high-performance computing clusters. As illustrated in Fig. 16, the proposed workflow achieves a favorable balance between accuracy, efficiency, and scalability when compared with three state-of-the-art approaches [16,25,39], highlighting its capability to deliver precise and computationally efficient façade reconstruction across large urban areas.

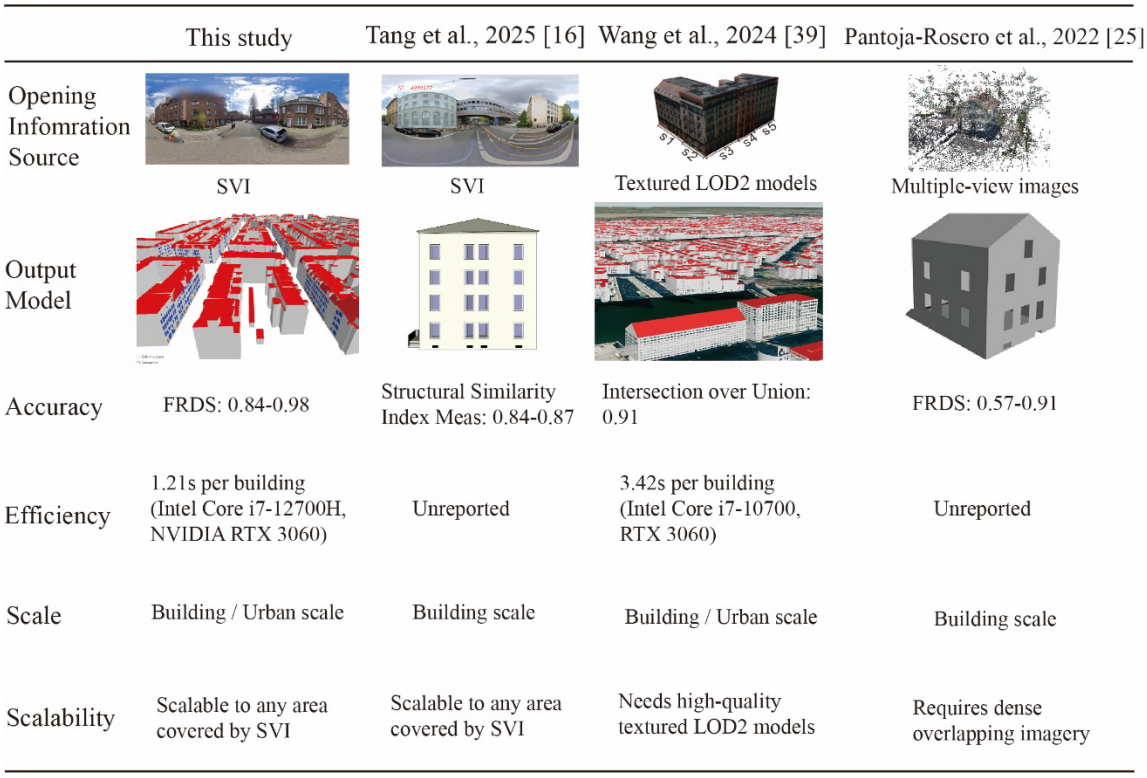

Fig. 16. Comparison with state-of-the-art façade reconstruction approaches.

Furthermore, comparative building energy simulations confirmed that the augmented models enhance the representation of façade characteristics, improving heating and cooling load estimations and underscoring the method's practical utility. From a broader perspective, the single-image photogrammetric approach developed here is not limited to SVI. It can be applied to any image source, provided that clear façade photographs and accurate capture locations are available. This flexibility allows for supplementing or replacing SVI with alternative imagery, such as handheld camera photos, to address coverage gaps or occlusions caused by urban elements. Moreover, with the integration of LOD2 building models in this paper, the method now supports

structures with complex roof geometries. In practice, when a LOD2 model contains multiple sub-façades, additional preprocessing is required to ensure correct correspondence between each sub-façade and its associated imagery.

6. Conclusion

This paper presented an effective and scalable workflow for enriching existing 3D building models with façade-level opening information by integrating panoramic SVI with a single-image photogrammetric approach. Although widely available low LOD models provide essential geometric foundations for urban analysis, their lack of façade openings has long limited their usefulness in applications requiring realistic building-envelope representation, such as energy simulation, thermal assessment, and microclimate modeling. The proposed workflow reconstructs metrically meaningful window and door geometries from a single SVI perspective and integrates them into CityJSON building models, enabling façade-aware applications at the city scale. The absence of intrinsic camera parameters in publicly available street view data represents a fundamental barrier to metric reconstruction. This challenge is addressed in this paper through the rigorous mathematical framework that establishes explicit geometric relationships between image space, camera space, and world coordinates. By grounding the reconstruction process in reproducible mathematical formulations, the proposed approach enables robust metric projection and façade opening reconstruction from a single panorama.

Validation across diverse buildings within Amsterdam demonstrates a high degree of spatial consistency and geometric agreement between reconstructed façade openings and reference geometries derived from image-based façade analysis. The reported FRDS values range from 0.84 to 0.98, with the majority exceeding 0.90, indicating that the reconstructed façade openings reliably reproduce the spatial distribution and relative geometry observed in street view imagery across different urban contexts. These results should be interpreted as measures of relative geometric agreement rather than absolute physical accuracy, as the evaluation relies on segmentation-derived reference data rather than independently surveyed ground truth.

The performance of this method is highly dependent on the quality of the input SVI and the conditions under which the façades are captured. The method works particularly well when the façade is clearly visible and captured from moderate viewing angles. However, the reconstruction

quality may decrease if the façade is obstructed, highly distorted, or viewed from extreme angles, such as sharp oblique perspectives. Users should expect FRDS values in the range of 0.84 to 0.98, with most results exceeding 0.90 under ideal conditions. For façades with limited visibility or those captured at high angles, lower FRDS values may occur due to projection distortions. These findings underline the importance of ensuring sufficient image quality and camera alignment for optimal reconstruction results.

The innovations brought forward in this paper can be summarized as follows: (1) a reproducible method for estimating intrinsic parameters of panoramic SVI to enable metric 3D reconstruction from widely accessible imagery; (2) a single-image photogrammetric workflow that converts 2D façade opening detections into 3D façade geometries without requiring multi-view observations or depth data, and integrates them directly into CityJSON building models to enhance their usefulness for downstream analytical applications. These contributions highlight the practical value of the presented method and position SVI as a viable data source for façade-level model enhancement.

Despite the demonstrated effectiveness of the proposed workflow, several limitations should be acknowledged to clarify its applicability boundaries. First, reconstruction completeness depends on the quality and viewing conditions of street view imagery. The method performs reliably when façades are clearly visible and captured under moderate viewing angles, even with partial occlusion. However, it becomes unreliable under severe obstruction, extreme camera proximity causing strong distortion, heavy blurring, extensive privacy masking, or highly oblique viewpoints, where insufficient geometric cues prevent stable corner extraction. Accordingly, acceptable reconstruction requires a largely unobstructed façade view, adequate image resolution, and accurate camera metadata. Second, the current workflow relies on manual selection and cropping of façade views from panoramic imagery. This step was adopted to ensure geometric reliability under variable image quality, but it introduces sensitivity to user judgment and limits full automation and scalability. While this does not affect the methodological validity, automated façade visibility assessment and view selection remain necessary for large-scale deployment. Third, the deep learning model is optimized for detecting predominantly rectangular façade openings and shows reduced effectiveness for non-rectangular or architecturally complex façade elements, such as arched or irregular openings. These geometries are underrepresented in the training data and are therefore not consistently captured by the current segmentation strategy. Fourth, the segmentation

model was trained exclusively on façade imagery from Amsterdam. Although the geometric reconstruction framework itself is city-agnostic, segmentation performance may degrade in urban contexts with substantially different architectural styles, materials, or opening configurations. Such domain shifts can propagate uncertainty into subsequent reconstruction stages, partially constraining generalization. Fifth, reconstruction quality is evaluated using segmentation-derived reference data rather than surveyed ground truth. While this enables scalable city-level evaluation, it inherently limits the upper bound of measurable accuracy. Reported metrics, therefore, reflect relative geometric consistency rather than absolute physical precision. Finally, although panoramic images are converted into perspective views to mitigate distortion, the simplified pinhole camera model does not explicitly account for non-linear distortions from equirectangular projection or multi-lens panoramic systems. Residual geometric deviations may persist, particularly for façades captured under high obliquity.

Overall, this work provides a reproducible, scalable, and data-efficient solution for enhancing low-LOD building models with detailed façade openings. Given the increasing global availability of SVI and the applicability of the proposed workflow across different urban contexts, the method offers a promising direction for future research and practical deployment in large-scale urban modeling.

CRediT authorship contribution statement

Rui Ma: Conceptualization, Methodology, Formal analysis, Validation, Writing- Original draft, Writing - Review & Editing.

Chendi Yang: Methodology, Visualization, Writing- Original draft.

Jiayu Chen: Investigation, Resources, Writing - Review & Editing.

Filip Biljecki: Supervision, Funding acquisition, Writing - Review & Editing.

Xin Li: Supervision, Funding acquisition, Writing - Review & Editing.

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.

Acknowledgments

This work is part of the project Multi-scale Digital Twins for the Urban Environment: From Heartbeats to Cities, which is supported by the Singapore Ministry of Education Academic Research Fund Tier 1. The support of the General Research Fund of the University Grants Committee of Hong Kong (#11216422) and City University of Hong Kong (#9239027, #9239042) is gratefully acknowledged. The conclusions herein are those of the authors and do not necessarily reflect the views of the sponsoring agencies. We express gratitude to the members of the Urban Analytics Lab at National University of Singapore and the Department of Architecture and Civil Engineering at the City University of Hong Kong for their valuable discussions.

Data availability

Data will be made available on request.

Declaration of generative AI use

During the preparation of this work, the authors used ChatGPT in order to proofread. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

[1] F. Biljecki, J. Stoter, H. Ledoux, S. Zlatanova, A. Çöltekin, Applications of 3D City Models: State of the Art Review, ISPRS International Journal of Geo-Information 4 (2015) 2842–2889. https://doi.org/10.3390/ijgi4042842.

[2] T.H. Kolbe, G. Gröger, L. Plümer, CityGML: Interoperable Access to 3D City Models, in: P. Van Oosterom, S. Zlatanova, E.M. Fendel (Eds.), Geo-Information for Disaster Management, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005: pp. 883–899. https://doi.org/10.1007/3-540-27468-5_63.

[3] G. Gröger, L. Plümer, CityGML – Interoperable semantic 3D city models, ISPRS Journal of Photogrammetry and Remote Sensing 71 (2012) 12–33. https://doi.org/10.1016/j.isprsjprs.2012.04.004.
[4] M. Barzegar, A. Rajabifard, M. Kalantari, B. Atazadeh, A framework for spatial analysis in 3D urban land administration – A case study for Victoria, Australia, Land Use Policy 111 (2021) 105766. https://doi.org/10.1016/j.landusepol.2021.105766.
[5] L. Troup, R. Phillips, M.J. Eckelman, D. Fannon, Effect of window-to-wall ratio on measured energy consumption in US office buildings, Energy and Buildings 203 (2019) 109434. https://doi.org/10.1016/j.enbuild.2019.109434.
[6] R. Ma, C. Geng, Z. Yu, J. Chen, X. Luo, Modeling city-scale building energy dynamics through inter-connected distributed adjacency blocks, Energy and Buildings 202 (2019) 109391. https://doi.org/10.1016/j.enbuild.2019.109391.
[7] T. Dogan, C. Reinhart, Shoeboxer: An algorithm for abstracted rapid multi-zone urban building energy model generation and simulation, Energy and Buildings 140 (2017) 140–153. https://doi.org/10.1016/j.enbuild.2017.01.030.
[8] R. Ma, B. Ren, D. Zhao, J. Chen, Y. Lu, Modeling urban energy dynamics under clustered urban heat island effect with local-weather extended distributed adjacency blocks, Sustainable Cities and Society 56 (2020) 102099. https://doi.org/10.1016/j.scs.2020.102099.
[9] R. Ma, T. Wang, Y. Wang, J. Chen, Tuning urban microclimate: A morpho-patch approach for multi-scale building group energy simulation, Sustainable Cities and Society 76 (2022) 103516. https://doi.org/10.1016/j.scs.2021.103516.
[10] Z. Deng, Y. Chen, J. Yang, F. Causone, AutoBPS: A tool for urban building energy modeling to support energy efficiency improvement at city-scale, Energy and Buildings 282 (2023) 112794. https://doi.org/10.1016/j.enbuild.2023.112794.
[11] L. Carnieletto, M. Ferrando, L. Teso, K. Sun, W. Zhang, F. Causone, P. Romagnoni, A. Zarrella, T. Hong, Italian prototype building models for urban scale building performance simulation, Building and Environment 192 (2021) 107590. https://doi.org/10.1016/j.buildenv.2021.107590.
[12] F. Biljecki, H. Ledoux, J. Stoter, An improved LOD specification for 3D building models, Computers, Environment and Urban Systems 59 (2016) 25–37. https://doi.org/10.1016/j.compenvurbsys.2016.04.005.
[13] Y. Xia, W. Gao, J. Stoter, Enriching LoD2 Building Models with Façade Openings Using Oblique Imagery, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. X-4/W6-2025 (2025) 225–232. https://doi.org/10.5194/isprs-annals-X-4-W6-2025-225-2025.
[14] F. Hanke, A. Bieringer, O. Wysocki, B. Jutzi, CM2LoD3: Reconstructing LoD3 Building Models Using Semantic Conflict Maps, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. X-4/W6-2025 (2025) 81–88. https://doi.org/10.5194/isprs-annals-X-4-W6-2025-81-2025.
[15] H.S. Kim, J. Carpenter, J. Jung, Automated LoD-3 Reconstruction Using Oblique UAV Images, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLVIII-G-2025 (2025) 797–804. https://doi.org/10.5194/isprs-archives-XLVIII-G-2025-797-2025.
[16] W. Tang, W. Li, X. Liang, O. Wysocki, F. Biljecki, C. Holst, B. Jutzi, Texture2LoD3: Enabling LoD3 Building Reconstruction with Panoramic Images, in: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE,

Nashville, TN, USA, 2025: pp. 2007–2017. https://doi.org/10.1109/CVPRW67362.2025.00189.
[17] S.I. Deliry, U. Avdan, Accuracy of Unmanned Aerial Systems Photogrammetry and Structure from Motion in Surveying and Mapping: A Review, J Indian Soc Remote Sens 49 (2021) 1997–2017. https://doi.org/10.1007/s12524-021-01366-x.
[18] L. Nan, C. Jiang, B. Ghanem, P. Wonka, Template Assembly for Detailed Urban Reconstruction, Computer Graphics Forum 34 (2015) 217–228. https://doi.org/10.1111/cgf.12554.
[19] A. Gruen, S. Schubiger, R. Qin, G. Schrotter, B. Xiong, J. Li, X. Ling, C. Xiao, S. Yao, F. Nuesch, Semantically Enriched High-Resolution LOD3 Building Model Generation, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLII-4/W15 (2019) 11–18. https://doi.org/10.5194/isprs-archives-XLII-4-W15-11-2019.
[20] H. Huang, M. Michelini, M. Schmitz, L. Roth, H. Mayer, LOD3 Building Reconstruction from Multi-Source Images, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLIII-B2-2020 (2020) 427–434. https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-427-2020.
[21] F. Biljecki, J. Lim, J. Crawford, D. Moraru, H. Tauscher, A. Konde, K. Adouane, S. Lawrence, P. Janssen, R. Stouffs, Extending CityGML for IFC-sourced 3D city models, Automation in Construction 121 (2021) 103440. https://doi.org/10.1016/j.autcon.2020.103440.
[22] J. Zhu, P. Wu, C. Anumba, A Semantics-Based Approach for Simplifying IFC Building Models to Facilitate the Use of BIM Models in GIS, Remote Sensing 13 (2021) 4727. https://doi.org/10.3390/rs13224727.
[23] E. Lewandowicz, F. Tarsha Kurdi, Z. Gharineiat, 3D LoD2 and LoD3 Modeling of Buildings with Ornamental Towers and Turrets Based on LiDAR Data, Remote Sensing 14 (2022) 4687. https://doi.org/10.3390/rs14194687.
[24] F. Tarsha Kurdi, E. Lewandowicz, Z. Gharineiat, J. Shan, Modeling Multi-Rotunda Buildings at LoD3 Level from LiDAR Data, Remote Sensing 15 (2023) 3324. https://doi.org/10.3390/rs15133324.
[25] B.G. Pantoja-Rosero, R. Achanta, M. Kozinski, P. Fua, F. Perez-Cruz, K. Beyer, Generating LOD3 building models from structure-from-motion and semantic segmentation, Automation in Construction 141 (2022) 104430. https://doi.org/10.1016/j.autcon.2022.104430.
[26] F. Frank, V. Shah, S. Worbis, L. Hoegner, 3D Façade Element Extraction from Image-based Instance Segmentation and Scale-Invariant Object Contour Points, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLVIII-G-2025 (2025) 473–480. https://doi.org/10.5194/isprs-archives-XLVIII-G-2025-473-2025.
[27] Y. Hou, F. Biljecki, A comprehensive framework for evaluating the quality of street view imagery, International Journal of Applied Earth Observation and Geoinformation 115 (2022) 103094. https://doi.org/10.1016/j.jag.2022.103094.
[28] Z. Xu, F. Zhang, Y. Wu, Y. Yang, Y. Wu, Building height calculation for an urban area based on street view images and deep learning, Computer Aided Civil Eng 38 (2023) 892–906. https://doi.org/10.1111/mice.12930.
[29] H. Chu, S. Wang, R. Urtasun, S. Fidler, HouseCraft: Building Houses from Rental Ads and Street Views, in: B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), Computer Vision – ECCV 2016, Springer International Publishing, Cham, 2016: pp. 500–516. https://doi.org/10.1007/978-3-319-46466-4_30.

[30] J.T. Szcześniak, Y.Q. Ang, S. Letellier-Duchesne, C.F. Reinhart, A method for using street view imagery to auto-extract window-to-wall ratios and its relevance for urban-level daylighting and energy simulations, Building and Environment 207 (2022) 108108. https://doi.org/10.1016/j.buildenv.2021.108108.
[31] A. Torii, M. Havlena, T. Pajdla, From Google Street View to 3D city models, in: 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, IEEE, Kyoto, 2009: pp. 2188–2195. https://doi.org/10.1109/ICCVW.2009.5457551.
[32] N. Bruno, R. Roncella, Accuracy Assessment of 3D Models Generated from Google Street View Imagery, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLII-2/W9 (2019) 181–188. https://doi.org/10.5194/isprs-archives-XLII-2-W9-181-2019.
[33] H. Kim, S. Han, Interactive 3D building modeling method using panoramic image sequences and digital map, Multimed Tools Appl 77 (2018) 27387–27404. https://doi.org/10.1007/s11042-018-5926-4.
[34] H.E. Pang, F. Biljecki, 3D building reconstruction from single street view images using deep learning, International Journal of Applied Earth Observation and Geoinformation 112 (2022) 102859. https://doi.org/10.1016/j.jag.2022.102859.
[35] C.B. Choy, D. Xu, J. Gwak, K. Chen, S. Savarese, 3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction, in: B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), Computer Vision – ECCV 2016, Springer International Publishing, Cham, 2016: pp. 628–644. https://doi.org/10.1007/978-3-319-46484-8_38.
[36] Z. Li, S. Ji, D. Fan, Z. Yan, F. Wang, R. Wang, Reconstruction of 3D Information of Buildings from Single-View Images Based on Shadow Information, IJGI 13 (2024) 62. https://doi.org/10.3390/ijgi13030062.
[37] Y. Rezaei, S. Lee, sat2pc: Generating Building Roof's Point Cloud from a Single 2D Satellite Images, in: Proceedings of the ACM/IEEE 14th International Conference on Cyber-Physical Systems (with CPS-IoT Week 2023), ACM, San Antonio TX USA, 2023: pp. 221–230. https://doi.org/10.1145/3576841.3585924.
[38] Y. Wei, G. Vosselman, M.Y. Yang, BuilDiff: 3D Building Shape Generation using Single-Image Conditional Point Cloud Diffusion Models, in: 2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), IEEE, Paris, France, 2023: pp. 2902–2911. https://doi.org/10.1109/ICCVW60793.2023.00313.
[39] Y. Wang, W. Jiao, H. Fan, G. Zhou, A framework for fully automated reconstruction of semantic building model at urban-scale using textured LoD2 data, ISPRS Journal of Photogrammetry and Remote Sensing 216 (2024) 90–108. https://doi.org/10.1016/j.isprsjprs.2024.07.019.
[40] J. Gu, C. Liu, A. Akbar, Y. Li, H. Ren, Z. Guo, Bridging implicit neural representations and semantic LoD3 for automated 3D building reconstruction under challenging visual conditions, Automation in Construction 178 (2025) 106382. https://doi.org/10.1016/j.autcon.2025.106382.
[41] Y. Yu, A. Gonzalez-Caceres, S. Scheidegger, S. Somanath, A. Hollberg, Deep learning-based Scalable Image-to-3D Facade Parser for generating thermal 3D building models, Automation in Construction 179 (2025) 106449. https://doi.org/10.1016/j.autcon.2025.106449.
[42] F.-E. Wang, Equirec2Perspec, (2021). https://github.com/fuenwang/Equirec2Perspec.
[43] V. Iglovikov, A. Shvets, TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation, (2018). https://doi.org/10.48550/ARXIV.1801.05746.

[44] J. Tian, Q. Jin, Y. Wang, J. Yang, S. Zhang, D. Sun, Performance analysis of deep learning-based object detection algorithms on COCO benchmark: a comparative study, J. Eng. Appl. Sci. 71 (2024) 76. https://doi.org/10.1186/s44147-024-00411-z.
[45] Y.M. Wang, Y. Li, J.B. Zheng, A camera calibration technique based on OpenCV, in: The 3rd International Conference on Information Sciences and Interaction Sciences, IEEE, Chengdu, China, 2010: pp. 403–406. https://doi.org/10.1109/ICICIS.2010.5534797.
[46] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, 2nd ed., Cambridge University Press, 2004. https://doi.org/10.1017/CBO9780511811685.
[47] S.K. Ong, M.L. Yuan, A.Y.C. Nee, Registration using projective reconstruction for augmented reality systems, in: Proceedings of the 2004 ACM SIGGRAPH International Conference on Virtual Reality Continuum and Its Applications in Industry - VRCAI '04, ACM Press, Singapore, 2004: p. 286. https://doi.org/10.1145/1044588.1044650.
[48] H. Ledoux, K. Arroyo Ohori, K. Kumar, B. Dukai, A. Labetski, S. Vitalis, CityJSON: a compact and easy-to-use encoding of the CityGML data model, Open Geospatial Data, Softw. Stand. 4 (2019) 4. https://doi.org/10.1186/s40965-019-0064-0.
[49] H. Ledoux, val3dity: validation of 3D GIS primitives according to the international standards, Open Geospatial Data, Softw. Stand. 3 (2018) 1. https://doi.org/10.1186/s40965-018-0043-x.
[50] R. Ma, Q. Li, B. Zhang, H. Huang, C. Yang, An ontology-driven method for urban building energy modeling, Sustainable Cities and Society 106 (2024) 105394. https://doi.org/10.1016/j.scs.2024.105394.
[51] O. Wysocki, B. Schwab, B. Willenborg, OloOcki/awesome-citygml: Release, (2022). https://doi.org/10.5281/ZENODO.5899096.
[52] R. Ma, J. Chen, C. Yang, X. Li, OSMsc: a framework for semantic 3D city modeling using OpenStreetMap, International Journal of Geographical Information Science 38 (2024) 1–26. https://doi.org/10.1080/13658816.2023.2266824.