Automated LoD-3 Reconstruction Using Oblique UAV Images

Han Sae Kim[1], Joshua Carpenter[2], Jinha Jung[1]

[1] Lyles School of Civil and Construction Engineering, Purdue University,
550 West Stadium Mall, West Lafayette, IN 47906, USA - (kim4012, jinha)@purdue.edu

[2] Department of Civil Engineering, University of Akron,
Akron, OH 44325, USA - jcarpent4@uakron.edu

Keywords: LoD, Building model, 3D reconstruction, UAV, Photogrammetry, CityJSON

Abstract

Urban modeling has increasingly gathered attention for urban resilience simulation studies. However, existing methodologies often fall short in reconstructing detailed building facades necessary for Level of Detail (LoD) 3 modeling. This study presents an automated LoD-3 reconstruction framework using oblique UAV images, which remains underexplored for efficient and scalable LoD-3 reconstruction. The proposed method integrates photogrammetry-based processing with deep learning-based window detection. Our approach consists of extracting roof structures, generating simulated facade images, detecting windows , and mesh intersections. Based on the photogrammetry, cameras are simulated to generate facade images of buildings. The YOLOv5-based window detection is followed using these simulated images. A ray-mesh intersection algorithm is implemented by projecting detected bounding boxes of windows onto the reconstructed LoD-2 model. The final LoD-3 model is exported in CityJSON format for seamless integration into urban simulation applications. Experimental results demonstrate that this approach shows the feasibility of an efficient and scalable solution for large-scale LoD-3 reconstructions.

1. Introduction

The study of urban resilience has gained significant attention in recent years. While urbanization provides numerous societal benefits, it also presents challenges that necessitate advanced modeling and analysis. As a result, city-scale simulation studies have become a focal point of research. The significance of 3D city models stems from their wide range of applications, including seismic damage simulation (Xiong et al., 2019), building energy modeling (Gao et al., 2019), climate modeling (Murshed et al., 2018), and flood risk analysis (Kilsedar et al., 2019). These models provide insights for diverse urban studies and decision-making processes. Level of Detail (LoD) building models serve as simplified 3D representations of structures and are often described using the City Geography Markup Language (CityGML) standard (Gröger and Plümer, 2012). Unlike the detailed 3D building modeling in computer vision, LoD models are simplified but include semantic information such as geolocation, geometric primitives, and their orientation, making them suitable for large-scale urban simulations. LoD models range from level 0 to 4, with increasing detail at each level, encompassing features like roof types, facade elements, and interior objects. Each LoD supports different applications (Biljecki et al., 2016). In addition to CityGML, CityJSON has emerged as a newer OGC standard, specifically designed to facilitate the management and manipulation of 3D city models, offering a more efficient and simplified format (Ledoux et al., 2019).

Current LoD modeling primarily focuses on LoD-2 reconstruction using different sources, including satellite images and airborne LiDAR data. While LoD-2 models are effective in various applications, they lack information about building facades. Building facades include building openings, which are critical information for urban energy simulation studies (Eicker et al., 2018, Xu et al., 2023). Therefore, the need for a robust LoD-3 reconstruction method is increasing. Meanwhile, LoD-2 modeling methods face limitations in generating LoD-3 models due to lacking details within the building facade. This can be attributed to the low resolution of images or the low point density of LiDAR data to capture fine details of the building facade. However, when utilizing high-resolution images or LiDAR data, the large data size and complexity can hinder the direct implementation of existing LoD-2 modeling algorithms. Therefore, it is essential to develop methodologies that bridge this gap by developing approaches using high-resolution inputs for LoD-3 reconstruction.

In this context, UAVs offer a viable solution. Unlike satellite-based or aerial platforms, UAVs capture high-resolution data over smaller regions while preserving detailed facade information. Additionally, they are more cost-effective and labor-efficient than terrestrial data collection methods, making them well-suited for LoD-3 reconstruction. Therefore, an approach that integrates high-resolution UAV imagery is essential for achieving automated LoD-3 modeling.

2. Related Work

Various approaches have been explored for LoD reconstruction, leveraging data from satellite imagery, airborne and terrestrial-based photography, and LiDAR point clouds. These methods differ in terms of data acquisition techniques, resolution, and computational complexity, influencing their suitability for different levels of detail in 3D building modeling.

2.1 Satellite Image-based Methods

Satellite images are one of the most used data sources to reconstruct the LoD building model (Gui and Qin, 2021, Schuegraf et al., 2024). These approaches typically share similar concepts and follow a series of steps. The initial step is to segment buildings using orthoimages or the Digital Surface Model

https://doi.org/10.5194/isprs-archives-XLVIII-G-2025-797-2025

(DSM). Building segmentations can be achieved by classifying building pixels using supervised classifiers. In the recent development of artificial intelligence, deep learning approaches such as Mask R-CNN (He et al., 2017), or U-Net (Ronneberger et al., 2015) have emerged as a popular method to detect and semantically segment building areas. The second step is to polygonize the extracted building boundaries. Simplifying and regularizing polylines of building boundaries can be performed using the Douglas-Peucker algorithm (Douglas and Peucker, 1973) or Line Segment Detector (LSD) (Von Gioi et al., 2008). The final step is to create a building model using these polylines by fitting the model using a model library or directly modeling using polylines and DSM. Similarly, aerial images can be used to reconstruct the LoD-2 model (Qian-Yi Zhou and Neumann, 2012).

2.2 LiDAR-based Methods

On the other hand, LiDAR data-based building reconstruction methods were investigated in various studies (Zhang et al., 2022, Li et al., 2022, Huang et al., 2022, Jung et al., 2017, Wang et al., 2016). These methods take LiDAR point clouds as input, extracting geometric primitives and regularizing them according to various roof types. LiDAR-based reconstruction is regarded as effective and efficient, as it directly captures height values across extensive areas. However, its application typically remains limited to LoD-2 reconstruction due to the sparse point density of building facades. Additionally, the large-scale data collection required for terrestrial LiDAR mapping to complete the building facade elements is often impractical. Similarly, (Li et al., 2016) used UAV images to generate the dense point cloud and utilized it for their input.

2.3 LoD-3 Reconstruction Methods

Advancing from LoD-2 to LoD-3 reconstruction, (Pantoja-Rosero et al., 2022) utilized building facade images to detect building openings using their deep learning model. The detected corner points are then converted to 3D coordinates and projected to the LoD-2 model by calculating the homography matrix. Multi-source-based methods, including terrestrial and UAV images, are also investigated to reconstruct the LoD-3 model (Huang et al., 2020). Their pipeline uses dense point cloud from photogrammetry for building detection and used building facade images to detect and project to the LoD-2 model. Although these LoD-3 reconstruction frameworks employ data-driven methods that eliminate the need for a model library, a limitation remains in requiring high-quality building facade images captured from ground level. LoD-3 reconstruction using terrestrial LiDAR data has also been researched (Wysocki et al., 2023). 3D semantic segmentation using point cloud, 2D semantic segmentation, and probability map based on laser physics and 3D building model are fused in Bayesian network to estimate building facade elements. Estimated shapes are then fitted to a facade object library. Limitations remain in the scalability of their implementation due to as they require complete terrestrial LiDAR data, and the heavy training data for Bayesian network.

2.4 Contributions

The primary challenges hindering the development of city-scale LoD-3 models, as identified in the literature, stem from the difficulty in acquiring the necessary data and the substantial computational costs associated with processing very high-resolution images and dense point clouds. Despite significant advancements in LoD-3 reconstruction, existing methods face challenges related to data availability, computational complexity, and scalability. This study addresses these limitations by proposing an automated LoD-3 reconstruction framework that utilizes oblique UAV imagery to enhance facade detail extraction. The proposed methodology integrates deep learning-based window detection with photogrammetric processing to efficiently generate geometrically accurate facade elements. Additionally, the study introduces a novel facade image generation approach, which enables the precise positioning of building openings to the corresponding LoD-2 model. The key contributions of this research are as follows:

- Development of an automated LoD-3 reconstruction framework leveraging oblique UAV imagery.
- Introduction of a photogrammetry-based facade image generation method, facilitating the accurate extraction of window coordinates and improving the geometric accuracy of LoD-3 models.

3. Methodology

Figure 1 presents the workflow of the proposed method, which follows a similar process to LoD-3 reconstruction methods in the literature. The approach consists of six key steps: 1) preprocessing, 2) building footprint extraction, 3) primitive extraction and optimization, 4) facade image generation, 5) window detection, and 6) mesh intersection. The process begins with pre-processing UAV images to generate orthophotos and DSM. Building footprint are then extracted from orthophoto and DSM. Simple planar primitives are extracted to model the roof types using the DSM and they are optimized as LoD-2 model using an existing building modeling method. Based on the reconstructed LoD-2 model, facade images are generated by simulating a camera facing the building facades. These images are input into a deep-learning model for window detection. The corner points of the detected window bounding boxes are then projected onto the LoD-2 mesh to complete the LoD-3 reconstruction.

3.1 LoD-2 Reconstruction

LoD-2 reconstruction is a prerequisite for LoD-3 modeling. The proposed LoD-2 reconstruction method utilizes two primary inputs: building footprints and the DSM). In our implementation, the building footprint was manually extracted using DSM and orthophoto. First, the LoD-1 model can be reconstructed by utilizing these building footprints and DSM, where an average height value is calculated within the building masks. Roof modeling is performed to advance the LoD-1 model to the LoD-2 model. This initial step is to clip the DSM for each building footprint. Then, the DSM is cut by thresholding. A threshold can be determined by examining the histogram, which exhibits a bi-modal distribution between ground and non-ground points. The smoothing operation is then performed by binning height values and applying an average height to selected bins. First, height values above zero are grouped into bins with a width of 10 meters. A histogram is generated to count the number of values within each bin, and only bins containing more than 100 pixels are selected for further processing to exclude small objects. For each of these selected bins, a mask is created to identify the corresponding DSM values, and the mean height within the bin is computed. This

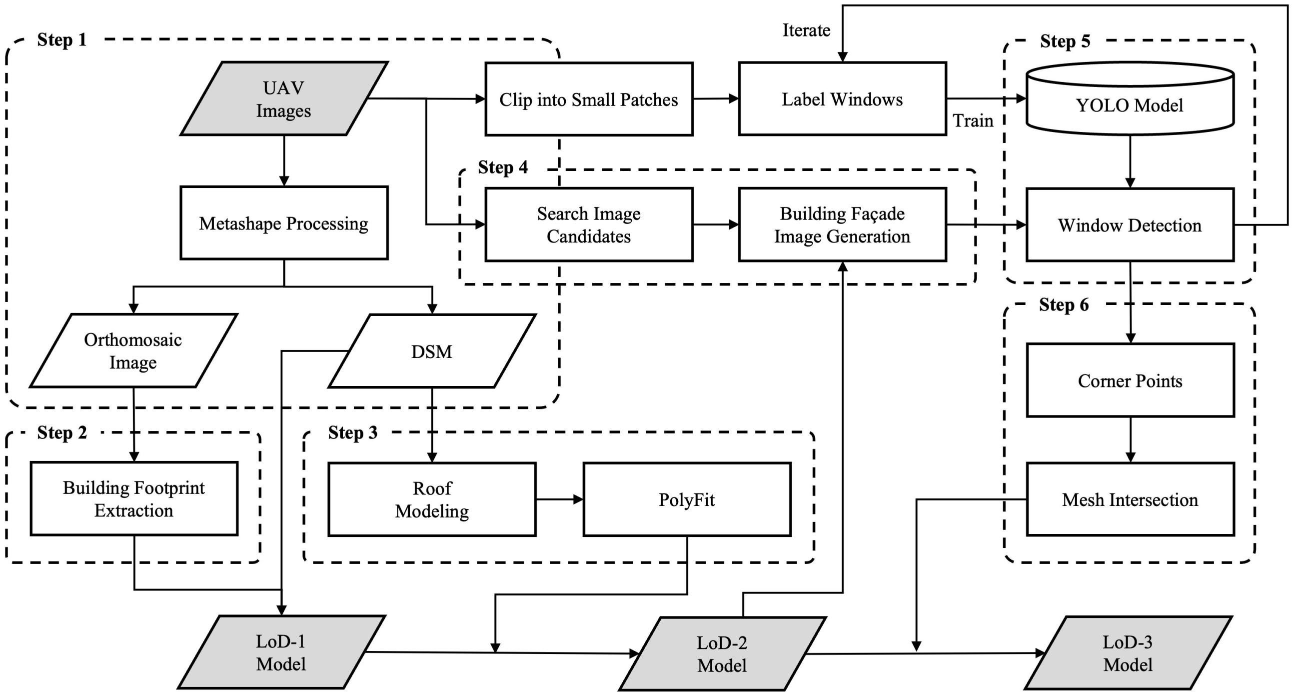
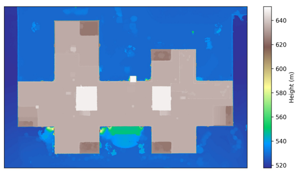
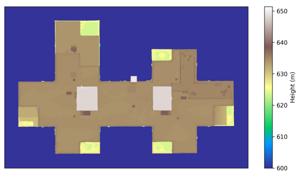


Figure 1. Flow chart of the proposed LoD-3 reconstruction.

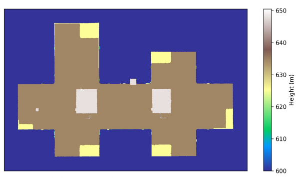
mean value is then applied to the relevant regions of a new, smoothed DSM array. The approach effectively reduces noise in the DSM while preserving important height features within the building masks. Figure 2 showcases the example of the roof modeling process. Polylines and corner points are extracted at each height level using the simplified building DSM. A mask layer is applied for each level, followed by an opening operation to eliminate thin lines along the edges. The resulting data is then vectorized. The extracted corner points are subsequently utilized to construct the mesh for the LoD-2 model.



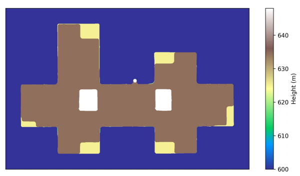
(a) DSM



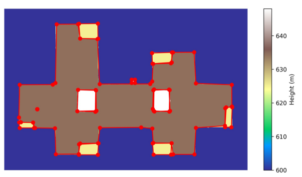
(b) Height thresholding



(c) Height smoothing



(d) Median & minimum filtering



(e) Extracted roof vertices

Figure 2. An example of the roof modeling process.

At this stage, the model may still contain imperfections, as the polylines have not yet been regularized. Following the simplification and vectorization of the roof details, the initial mesh is further modeled and optimized using PolyFit (Nan and Wonka, 2017). Unlike the original method, which applies the algorithm directly from dense point clouds, our approach utilizes simplified planar primitives for roof types. This adjustment allows faster and more accurate optimization by bypassing the RANSAC primitive fitting process, which is typically sensitive to point cloud quality, density, and noise levels.

3.2 LoD-3 Reconstruction

The proposed LoD-3 reconstruction framework consists of three major tasks: facade image generation, window detection, and mesh intersection.

3.2.1 Facade Image Generation Once the LoD-2 model is reconstructed, the next step is to generate facade images to detect and locate windows on the building facades. Facade images are critical for efficiently detecting complete windows in building facades. In our case, over 1,000 raw UAV images were captured for the entire study area, so detecting windows for all these images is redundant and computationally inefficient for scalable LoD-3 reconstruction. Even if the window detection for all images can be done, preprocessing or optimization should be followed to determine the best window coordinates of the same points that are extracted from multiple images. Therefore, generating a complete facade image for each building facade facilitates efficient and quick window detection. Unlike ortho-rectification methods that rely on point clouds and optical imagery (Zureiki and Roux, 2009), our approach is entirely photogrammetry-based. Cameras are simulated in a perpendicular direction for each building facade to generate the facade images, knowing the camera’s exterior and interior orientation parameters (EOPs and IOPs). With the aid of these parameters, we can calculate the 3D geo coordinates of window corner points using image coordinates. Figure 3 illustrates the concept of simulating camera positions and orientations for each facade and retrieving pixel data from different source images. As illustrated in Figure 3, parts of the building facade may be occluded in the UAV images due to the field of view. By combining multiple images taken from different locations and orientations, a complete facade image can be achieved for each building facade.

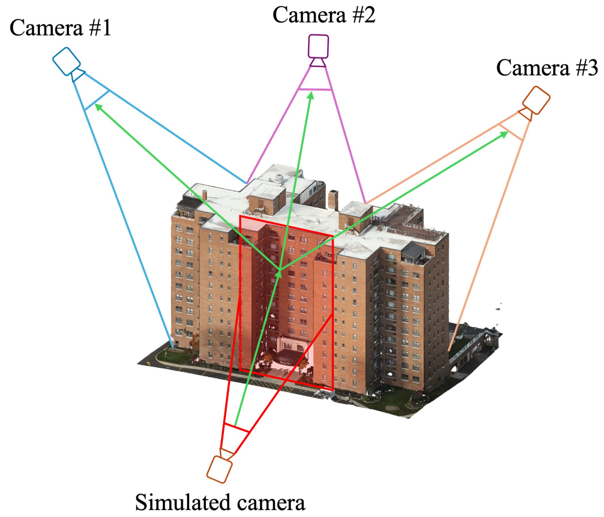


Figure 3. Illustration of ortho facade image generation concept.

3.2.2 Window Detection Once the facade images are created, window detection is performed using these facade images. For the object detection model, YOLOv5 was implemented for our case. As an improved version of YOLO (Redmon, 2016), the model size of YOLO v5 is relatively small and computationally efficient. It is extremely fast and suitable for resource-constrained environments. While YOLOv5 is known to score low accuracy compared to larger, more complex models, the model could achieve high accuracy given the task is relatively simple. Although YOLOv5 was implemented on our cases, other object detection model can be applied to train and detect the windows.

A total of 106 training samples (640 $\times$ 640 pixels) were manually labeled, including an average of 10 window objects for each training sample (Figure 4). The model was then trained using these training samples, with data augmentation techniques applied, including shear, rotation, flip, and scale transformations. From the raw UAV images, windows are not always in a horizontal or vertical direction to the image. The data augmentation ensures that the windows vary in size and orientation. This helps to infer the windows in facade images, which are the windows are located in perpendicular directions to the image. Also, since the training samples are derived from raw UAV images while the actual detection occurs in facade images, ensuring consistency in image resolution and quality is critical for optimal performance. To achieve this, the camera IOPs and the distance between the camera and the object were assumed to be identical during the generation of the facade images to match the resolution of the image and also the distortions.

3.2.3 Mesh Intersection Mesh intersection is a critical step for accurately determining the 3D coordinates of window points detected in facade images. This process integrates the photogrammetry-based LoD-2 model with the window detection results, ensuring precise spatial alignment. To begin, the LoD-2 mesh model, comprising vertices and faces representing

Figure 4. Training samples for the window detection model.

the building structure, is imported into the system. For each detected window in the facade image, rays are cast from the simulated camera's position through the window corners, following the camera's intrinsic and extrinsic parameters. These rays are projected into the 3D space where they intersect with the LoD-2 mesh, specifically targeting the building facade planes. The intersection calculation employs ray-mesh intersection algorithms using spatial data structures like KD-trees to enhance computational efficiency. Each ray's origin is defined by the simulated camera's position, and its direction vector is determined by the orientation towards the detected window corners in the facade image. The algorithm identifies intersection points where the rays meet the mesh surfaces, yielding precise 3D coordinates corresponding to each window corner. Furthermore, to mitigate errors from occlusions or misalignments, multiple intersections from different camera perspectives are aggregated. A weighted averaging method, considering the angle of incidence and distance from the camera, refines the final coordinates. This approach ensures robustness, particularly in urban environments with complex geometries. The extracted 3D coordinates are then integrated into the LoD-2 model, upgrading it to LoD-3. An illustration of the process is shown in Figure 5.

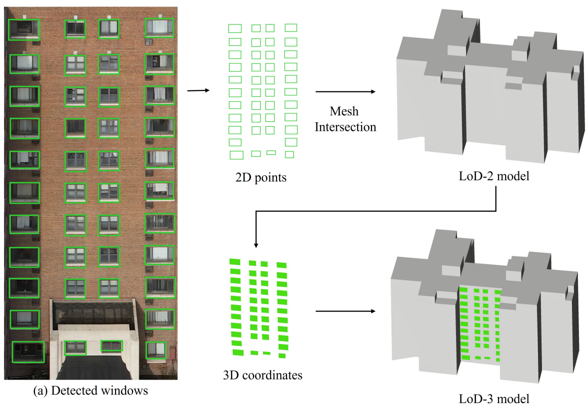


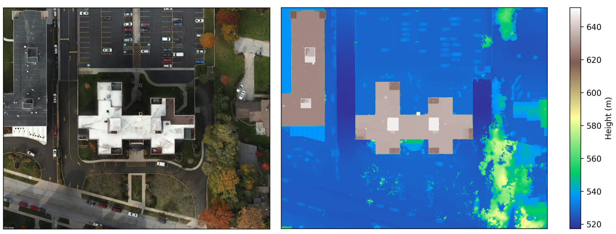
Figure 5. Illustration of mesh intersection process. The extracted 2D window points are projected onto the LoD-2 model, and the 3D coordinates are determined based on the intersection points.

3.2.4 CityJSON Standards The final LoD-3 model is exported in the CityJSON format to ensure compatibility with geospatial applications and 3D city modeling standards. CityJSON is a JSON-based encoding of the CityGML data model, designed to be lightweight and easily integrable with web technologies. The conversion process begins by structuring the 3D model’s geometric and semantic data. The vertices and faces obtained from the mesh intersection process are first organized into the CityJSON ”vertices” and ”boundaries” arrays. Each vertex represents a 3D coordinate, while the boundaries define the relationships between vertices, forming surfaces like walls, roofs, and windows. The model is encapsulated under a ”CityObjects” dictionary, where each building is represented as an object of type ”Building.” The geometry is described using the ”Solid” type with nested ”boundaries” that specify different building components. Semantic information, such as surface types (e.g., RoofSurface, WallSurface, GroundSurface, and Window), is embedded within the ”semantics” section, providing contextual details about each surface. For accurate spatial referencing, a transformation matrix is applied to the model’s coordinates. This involves scaling and translating the coordinates based on the original data’s reference system, ensuring geographic accuracy when visualized in GIS platforms.

4. Experiments

4.1 Study Area and Material

The data for this study was collected on November 10th, 2021, in the Gold Coast area of Cleveland, Ohio. A DJI M300 equipped with a P1 sensor was flown over the area in an oblique configuration, with 85% side and forward overlap. The altitude of flying height was set to 120 meters above the tallest building, using a camera sensor with a 35mm focal length, resulting in a ground sampling distance (GSD) of 0.06 meters. Each raw UAV images have 8192×5460 pixel resolution. A total of 4704 images were captured and processed using Metashape v.2.0.3 to generate both the orthophoto and DSM. Figure 6 shows a small region of generated orthophoto and DSM.



(a) Orthomosaic image (b) Digital surface model

Figure 6. A subset of orthomosaic image and DSM generated from UAV images.

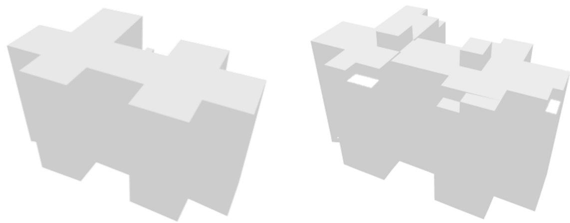
4.2 Computational Environment

All experiments were performed on a machine with the following hardware and software configurations: a 13th Gen Intel(R) Core(TM) i9-13900K CPU @ 5.80 GHz with 24 cores (Intel Corporation, Santa Clara, CA, USA), an NVIDIA GeForce RTX 4090 GPU with 24 GB VRAM running CUDA Version 12.2 (NVIDIA Corporation, Santa Clara, CA, USA), and 128 GB of DDR4 RAM. The system operated on Ubuntu 22.04.3 LTS with Kernel Version 6.8.0-40-generic. The experiments were conducted in a Python 3.9.18 environment with PyTorch 2.4.0, leveraging CUDA 12.4 for GPU-accelerated computations.

5. Results and Discussion

5.1 LoD-1 & LoD-2 model

Our proposed framework can reconstruct from LoD-1 to LoD-3, as the framework sequentially utilizes the model as input for the update. Figure 7 shows the example of reconstructed LoD-1 and LoD-2 models. The LoD-1 model was successfully reconstructed using the building footprints extracted from the DSM and orthomosaic image. This process involved calculating an average height value within the building masks. For LoD-2 reconstruction, roof-type modeling was applied using the DSM, followed by vectorization of the building boundaries and corners. By utilizing PolyFit using planar primitives from roof type modeling, the initial LoD-2 model was refined to represent building roof structures accurately. The simplification of roof details, coupled with the elimination of RANSAC primitive fitting, allowed for faster and more reliable results. The generated LoD-2 model serves as a foundation for further LoD-3 reconstruction.



(a) LoD-1 model (b) LoD-2 model

Figure 7. LoD-1 and LoD-2 model reconstruction result.

5.2 Building Facade Image

The example of the generated facade image can be found in Figure 8. As demonstrated in Figure 8, the facade images exhibit sufficient quality for subsequent window detection tasks, with minimal distortion and occlusion. In cases where occlusion occurred, mesh indices from the simulated and source cameras were matched, allowing for the recovery of hidden facade areas. The camera is positioned at the center of the wall and set at a distance corresponding to the drone’s flight altitude. This setup produces a facade image that can be considered a vertical photograph captured from a distance with a nadir orientation toward the building wall. However, since the image is generated using a LoD-2 mesh model as a depth map, it can be termed an facade image. This image will serve as the basis for window detection, with intersections performed using the same LoD-2 mesh model utilized as the depth reference. The resulting LoD-3 model is expected to yield consistent and reliable outcomes, maintaining the integrity of the LoD framework.

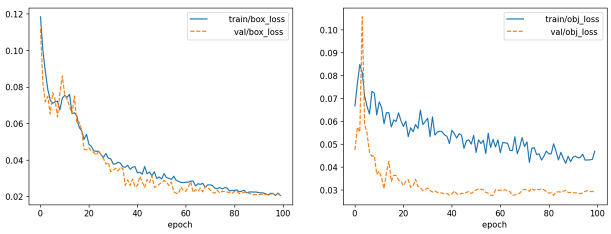
5.3 Performance of the Window Detection Model

Based on the YOLO v5 architecture, the window detection model was trained with manually labeled samples. Figure 10 shows the training and validation loss, indicating a well-converged model. Box loss was converged with both training and validation loss, while object loss shows a discrepancy between training and validation loss. This may be due to the data augmentation, which makes the training samples more difficult to learn than the validation samples. The model demonstrated a precision of 0.96 and a recall of 0.47, aligning with the

(a) Source Image (b) Selected mesh face (c) Generated face image (d) Resulted façade image

Figure 8. Example of facade image generation.

model loss interpretation. While augmented data could cause lower recall, generalization could have caused the lower recall as well. The shape and the size of the window look quite different throughout the buildings. This can be addressed by adding more training samples of various windows. Some failure cases were also observed where the model struggled to detect windows due to occlusions, such as branches or objects obstructing the facade. In this context, the model showed strong performance in window detection (high precision) but faced difficulties identifying windows across different types (low recall). To improve the model's performance and enhance its generalization capability, incorporating more diverse window labels from various building types is essential. As outlined in our pipeline in Figure 1, the model can be iteratively refined and trained to achieve these improvements. The extracted windows can be further optimized using geometrical constraints as described in (Zhuo et al., 2023).

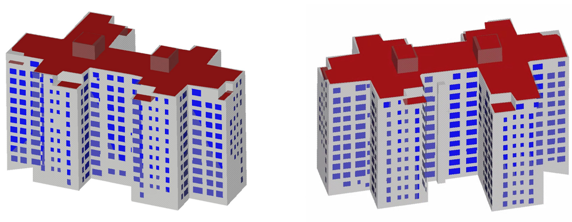


(a) Training and validation loss of box (b) Training and validation loss of object

Figure 9. Training and validation loss of the window detection model.

5.4 LoD-3 Reconstruction Results

The final LoD-3 model was reconstructed by integrating the detected windows into the optimized LoD-2 mesh (Figure 10). The detected windows were projected onto the LoD-2 mesh. Projecting the window bounding boxes onto the LoD-2 mesh allowed for the accurate placement of facade details, contributing to the semantic information of the 3D model. Out of approximately 782 windows in total, 668 windows (True Positives) were accurately detected and projected onto the LoD-2 model. However, 4 windows (False Positives) were incorrectly detected, and 114 windows (False Negatives) were missed. The primary reason for the False Negatives was the presence of occlusions caused by adjacent buildings, trees, and other urban elements, which led to obstructed building facade images that the window detection model could not effectively process. Our evaluation reported a precision of 0.99, a recall of 0.85, and a F_1 score of 0.91. The results demonstrate that the proposed framework successfully captures detailed building features, including roof types and windows, which are critical for advanced urban simulations. Using UAV oblique images for LoD-3 reconstruction proves to be a feasible and efficient method, especially for applications requiring high levels of detail.



(a) Front view (b) Rear view

Figure 10. LoD-3 model reconstruction result.

6. Limitations and Future Work

Although this study proposes an automated LoD-3 reconstruction framework, some cases have still not been explored for large-scale implementation. For example, our roof modeling process has only been tested on the flat-type roof, which would fail in oblique roof planes. This hinders the implementation of the proposed framework for a city-wide application. Future work will primarily focus on dealing with various roof types and improving the generalization capacity of the window detection model.

7. Conclusion

This study introduces an automated framework for LoD-3 building reconstruction using high-resolution UAV oblique images. By combining roof type simplification with the photogrammetry-based facade image generation, the method successfully enhances building facade details on LoD-2 model. The use of deep learning for window detection further refines the LoD-3 model, overcoming limitations seen in traditional LoD-2 models. The proposed approach streamlines the LoD-3 reconstruction process and provides a practical solution for creating detailed 3D models suitable for large-scale urban applications. Future work will focus on further improving the quality of the facade image and exploring the integration of additional building elements to enhance the semantics of the buildings.

References

Biljecki, F., Ledoux, H., Stoter, J., Vosselman, G., 2016. The variants of an LOD of a 3D building model and their influence on spatial analyses. *ISPRS Journal of Photogrammetry and Remote Sensing*, 116, 42–54.

Douglas, D. H., Peucker, T. K., 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: the international journal for geographic information and geovisualization*, 10(2), 112–122.

Eicker, U., Zirak, M., Bartke, N., Romero Rodríguez, L., Coors, V., 2018. New 3D model based urban energy simulation for climate protection concepts. *Energy and Buildings*, 163, 79–91.

Gao, H., Koch, C., Wu, Y., 2019. Building information modelling based building energy modelling: A review. *Applied Energy*, 238, 320–343.

Gröger, G., Plümer, L., 2012. CityGML – Interoperable semantic 3D city models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 71, 12–33.

Gui, S., Qin, R., 2021. Automated LoD-2 model reconstruction from very-high-resolution satellite-derived digital surface model and orthophoto. *ISPRS Journal of Photogrammetry and Remote Sensing*, 181, 1–19.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. *Proceedings of the IEEE international conference on computer vision*, 2961–2969.

Huang, H., Michelini, M., Schmitz, M., Roth, L., Mayer, H., 2020. LOD3 building reconstruction from multi-source images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 427–434.

Huang, J., Stoter, J., Peters, R., Nan, L., 2022. City3D: Large-Scale Building Reconstruction from Airborne LiDAR Point Clouds. *Remote Sensing*, 14(9), 2254.

Jung, J., Jwa, Y., Sohn, G., 2017. Implicit Regularization for Reconstructing 3D Building Rooftop Models Using Airborne LiDAR Data. *Sensors*, 17(3), 621. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

Kilsedar, C. E., Fissore, F., Pirotti, F., Brovelli, M. A., 2019. Extraction and visualization of 3D building models in urban areas for flood simulation. *The international archives of the photogrammetry, remote sensing and spatial information sciences*, 42, 669–673.

Ledoux, H., Arroyo Ohori, K., Kumar, K., Dukai, B., Labetski, A., Vitalis, S., 2019. CityJSON: a compact and easy-to-use encoding of the CityGML data model. *Open Geospatial Data, Software and Standards*, 4(1), 4.

Li, L., Song, N., Sun, F., Liu, X., Wang, R., Yao, J., Cao, S., 2022. Point2Roof: End-to-end 3D building roof modeling from airborne LiDAR point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 193, 17–28.

Li, M., Nan, L., Smith, N., Wonka, P., 2016. Reconstructing building mass models from UAV images. *Computers & Graphics*, 54, 84–93.

Murshed, S. M., Picard, S., Koch, A., 2018. Modelling, Validation and Quantification of Climate and Other Sensitivities of Building Energy Model on 3D City Models. *ISPRS International Journal of Geo-Information*, 7(11), 447. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

Nan, L., Wonka, P., 2017. PolyFit: Polygonal Surface Reconstruction from Point Clouds. *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, Venice, 2372–2380.

Pantoja-Rosero, B., Achanta, R., Kozinski, M., Fua, P., Perez-Cruz, F., Beyer, K., 2022. Generating LOD3 building models from structure-from-motion and semantic segmentation. *Automation in Construction*, 141, 104430.

Qian-Yi Zhou, Neumann, U., 2012. 2.5D building modeling by discovering global regularities. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Providence, RI, 326–333.

Redmon, J., 2016. You only look once: Unified, real-time object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, Springer, 234–241.

Schuegraf, P., Shan, J., Bittner, K., 2024. PLANES4LOD2: Reconstruction of LoD-2 building models using a depth attention-based fully convolutional neural network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 211, 425–437.

Von Gioi, R. G., Jakubowicz, J., Morel, J.-M., Randall, G., 2008. LSD: A fast line segment detector with a false detection control. *IEEE transactions on pattern analysis and machine intelligence*, 32(4), 722–732.

Wang, J., Xu, Y., Remil, O., Xie, X., Ye, N., Yi, C., Wei, M., 2016. Automatic Modeling of Urban Facades from Raw LiDAR Point Data. *Computer Graphics Forum*, 35(7), 269–278.

Wysocki, O., Xia, Y., Wysocki, M., Grilli, E., Hoegner, L., Cremers, D., Stilla, U., 2023. Scan2lod3: Reconstructing semantic 3d building models at lod3 using ray casting and bayesian networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6548–6558.

Xiong, C., Lu, X., Huang, J., Guan, H., 2019. Multi-LOD seismic-damage simulation of urban buildings and case study in Beijing CBD. *Bulletin of Earthquake Engineering*, 17(4), 2037–2057.

Xu, H., Kim, J. I., Zhang, L., Chen, J., 2023. LOD2 for energy simulation (LOD2ES) for CityGML: A novel level of details model for IFC-based building features extraction and energy simulation. *Journal of Building Engineering*, 78, 107715.

Zhang, P., He, H., Wang, Y., Liu, Y., Lin, H., Guo, L., Yang, W., 2022. 3D Urban Buildings Extraction Based on Airborne LiDAR and Photogrammetric Point Cloud Fusion According to U-Net Deep Learning Model Segmentation. *IEEE Access*, 10, 20889–20897.

Zhuo, X., Tian, J., Fraundorfer, F., 2023. Cross Field-Based Segmentation and Learning-Based Vectorization for Rectangular Windows. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16, 431–448.

Zureiki, A., Roux, M., 2009. Ortho-rectified facade image by fusion of 3D laser data and optical images. *Laserscanning09*, 38, 305–310.