Graph-Based Deep Learning for Mesh-Based 3D Building Reconstruction

Hossein Zavar*[1], Mohammad Saadatseresht[2], Hossein Arefi[3]

1. PhD Student, Department of Photogrammetry and Remote Sensing, School of Surveying Engineering and Spatial Information, Faculty of Engineering, University of Tehran

2. Department of Photogrammetry and Remote Sensing, School of Surveying Engineering and Spatial Information, Faculty of Engineering, University of Tehran

3. i3mainz, Institute for Spatial Information and Surveying Technology, School of Technology, Mainz University of Applied Sciences, D-55118 Mainz, Germany

ABSTRACT

Historically, three-dimensional (3D) geospatial data were mainly used for visualization, while two-dimensional (2D) data underpinned spatial analyses. With recent advances in sensing and computation, as well as the demands of smart-city applications, 3D urban data, especially for building objects, has become a valuable source for processing and interpretation. This study proposes a deep-learning approach that exploits neighborhood relations on urban triangle meshes to reconstruct simplified 3D building models. Neural networks help overcome mesh-specific challenges such as irregular connectivity and heterogeneous sampling, enabling robust geometric analysis. Experimental results indicate that the proposed network can effectively leverage mesh data for building reconstruction. Furthermore, the results suggest that additional spatial analyses can be performed directly on mesh data, enabling the production of high-quality products. Overall, the approach delivers accurate building abstractions from real-world meshes, reduces manual post-editing, and supports downstream urban analytics. The findings highlight the growing potential of mesh-native learning for scalable 3D city modeling.

KEY WORDS: 3D city model, deep learning, building reconstruction, curvature, corner detection, LOD simplification

1. INTRODUCTION

The importance and use of 3D city models are rapidly increasing. With advances in hardware and software, as well as the emergence of smart cities, the need for richer data, particularly in 3D, is increasingly felt. Representative applications include noise-propagation simulation, urban planning, crisis management, rapid training simulations, and indoor navigation.

In parallel, the number of 3D city models produced by municipalities, national mapping agencies, government bodies, commercial companies, and other organizations has steadily increased, mainly due to significant progress in 3D reconstruction over the past decade. In many applications, purely geometric/graphic aspects are insufficient; semantic segmentation (object classification) is also crucial, e.g., the building usage type plays a key role in acoustic simulation.

2. BACKGROUND AND RELATED WORK

Two primary data sources are commonly used for large-scale urban 3D reconstruction (Poullis 2013; Zhou and Neumann 2010; Wilk et al. 2022a; Yi et al. 2017 and Han et al. 2021). Airborne LiDAR was among the first technologies deployed at the city scale. The resulting point clouds are dense and accurate, and mostly contain geometric information. However, LiDAR acquisition requires costly equipment, and due to the flight altitude, points on façades are difficult to obtain.

More recently, with advancements in unmanned aerial vehicles (UAVs) and image-based 3D reconstruction—specifically, multi-view stereo (MVS) and structure-from-motion (SfM)—oblique or multi-view aerial imagery has been widely adopted for urban modeling. Compared to LiDAR point clouds, aerial imagery offers higher spatial resolution, lower cost, and more efficient data collection; moreover, textured triangle meshes reconstructed from imagery provide appearance cues not present in pure point clouds.

Nonetheless, textured meshes are voluminous, redundant and usually lack high-level semantics and structure (Bouzas, Ledoux, and Nan 2020; Liu et al. 2022 and Yan et al. 2023). As a result, they cannot be directly integrated into modern GIS, where entities are predominantly vector—based, motivating research on converting meshes into compact, semantically enriched vector representations.

Broadly, 3D building modeling approaches can be grouped into data-driven (bottom-up) methods, model-driven (top-down) methods, and hybrid data–model approaches (Xiao & Furukawa, 2012, 2014). Data-driven methods extract roof vertices, edges, or planar patches from images or DSMs and treat them as primitives to be merged and

analyzed topologically to form building shapes. In theory, they can reconstruct arbitrary buildings; however, performance degrades when extracted data are incomplete or noisy due to poor input quality, complex topology, or weak geometric detectors in cluttered scenes (Bouzas et al. 2020; Yu et al. 2021).

Converting building data into polygonal meshes is a central challenge in urban modeling. Classical methods generate polygonal meshes from inputs such as point clouds and typically require careful parameter tuning and heavy pre-processing; performance can degrade in complex scenes. By contrast, deep learning using convolutional neural networks (CNNs) and graph-based neural networks can integrate geometric and appearance cues and has shown strong potential for mesh understanding.

2. 1 Classical methods

Before the widespread use of neural networks, researchers proposed heuristic yet effective pipelines. In Guo et al. (2022) and Nan and Wonka (2017) A lightweight polygonal surface model is reconstructed by detecting line segments in overlapping SfM images; 2D lines from all views are reprojected and lifted to 3D using camera parameters and coplanarity constraints. Intersecting planes form candidate cells; a Bayesian formulation selects plausible planes and, following, yields a building model. While line-based cues capture façade details, the evaluation covers only a limited set of buildings.

Point clouds are a key input for building reconstruction. Nan and Wonka (2017) used an energy formulation with three terms—fidelity to input data, plane coverage over the point cloud, and model complexity—selects a near-optimal 3D model from RANSAC-initialized planes and their intersections, forming volumetric cells. Due to its flexibility across various building styles, this method is widely used as a baseline, although it may be less accurate than some specialized approaches.

In Li and Shan (2022) After detecting roof parts from a building point cloud, a small set of primitive roof shapes is scored, and the best-fitting shape is chosen to reconstruct the 3D model. Because the set of reference shapes is limited, the method applies only to specific building types; nevertheless, the results are strong with minor gaps in some cases.

Bouzas et al. (2020) Using photogrammetric meshes, also reconstructs building models via energy minimization. An initial classification is produced by region-growing and refined through graph-based neighborhood analysis.

2. 2 Deep learning methods

With the advent of deep learning, 3D reconstruction methods increasingly adopt neural architectures. Chen et al. (2023) Reconstructs a graph-based neural network by building models directly from point clouds. Initial planes define cells, whose neighborhood tree is fed to a network that performs binary classification over cells to extract the final model.

Mesh data can likewise serve as input to neural networks; textures in meshes provide additional information that can improve performance. In Yang et al. (2023) and Zhang (2023)A U-shaped network processes per-face graph features, geometric descriptors, and texture. For each face, neighboring faces are identified, and 1D convolutions with pooling and mesh resampling are applied to classify the mesh into seven semantic classes. A similar idea appears in, which combines geometric and texture features with a tri-level design to label building, ground, vegetation, and other classes.

With a different setup, classify oblique aerial images using deep CNNs and project the predictions back onto textured meshes reconstructed from the imagery. One of the most recent approaches (Liu et al. 2024; Wilk et al. 2022b and Chen et al. 2022) proposes an implicit deep model over meshes using signed distance functions (SDF) and occupancy. Alongside the network, a Markov formulation with SDF and occupancy supports the extraction of dominant planes for building reconstruction; qualitative results (Fig. 8) highlight the capability of deep models for mesh simplification and reconstruction.

Texture on meshes is an important source for downstream analysis. In Frueh, Sammon, and Zakhor (2004) Textures are added to a photogrammetric mesh by first detecting main planes and then selecting the best image for each plane based on the viewing angle. Edge features are extracted, and a UV map with texture is generated, thereby improving detail and interpretability.

3. METHODOLOGY

The main pipeline comprises three stages: (i) mesh data preparation; (ii) network training and corner detection; and (iii) construction of a simplified 3D model (LOD).

3. 1 Mesh data preparation

Roof surfaces are considered the principal structural element of a building's geometry. Roofs are separated using surface normals and geometric thresholds: horizontal and sloped roofs are distinguished by normal direction; connected roof parts are grouped via mesh neighborhood relations (Figure 3). To remove minor artifacts, components were filtered by total triangle area, leaving only the dominant roofs.

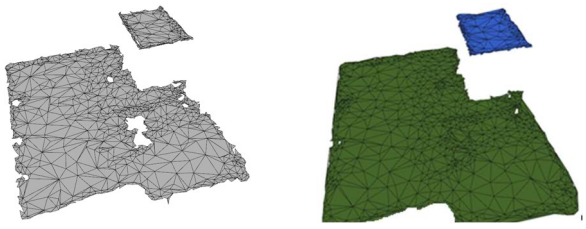

Figure 1. Left: building roof detection; right: separation of different components.

For each data, we define a neighborhood matrix by connecting the faces through their edges. This matrix is not limited to the first-order neighbors, but is also defined for faces that are neighbors to the first-order faces. In this way, the neighborhood of the data faces is categorized in a graph based on the edges that exist between the faces.

3. 2 Network and corner detection

A U-shaped (U-Net–style) deep network was used for training (Figure 3). The input data is represented as a matrix based on the neighborhood graph, where the rows correspond to the neighboring faces, and the columns represent the calculated features for each face. During downsampling, the most discriminative features are preserved, and during resampling (upsampling), essential features that may have been lost are recovered via the skip connections. The activation function employed is ReLU. Two inputs corresponding to first- and second-ring neighborhoods are fed to a multi-input network. The neighborhood rings refer to the adjacent faces at each stage. In the first stage, the nearest faces that share an edge with the current face are taken as the first ring; in the second stage, the faces that share an edge with any face in the first ring are selected as the second ring (Figure 2).

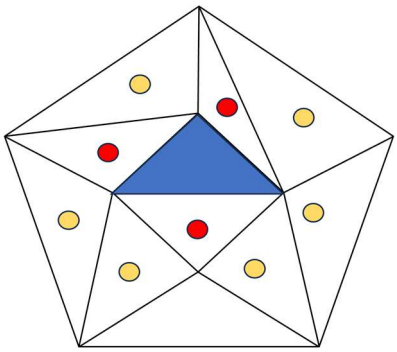

Figure 2. Blue triangle: central triangle; red circles: first-ring neighboring triangles; yellow circles: second-ring neighboring triangles.

The first input is processed through three Conv1D blocks and then fused with the second input. Batch Normalization is applied after each convolution to stabilize optimization and mitigate overfitting. Two MaxPooling stages down-sample the temporal dimension, while retaining the most discriminative features; two subsequent UpSampling stages restore resolution and merge information from earlier layers via skip connections. Finally, two Dropout layers and a Softmax classifier yield a binary decision (corner vs. non-corner).

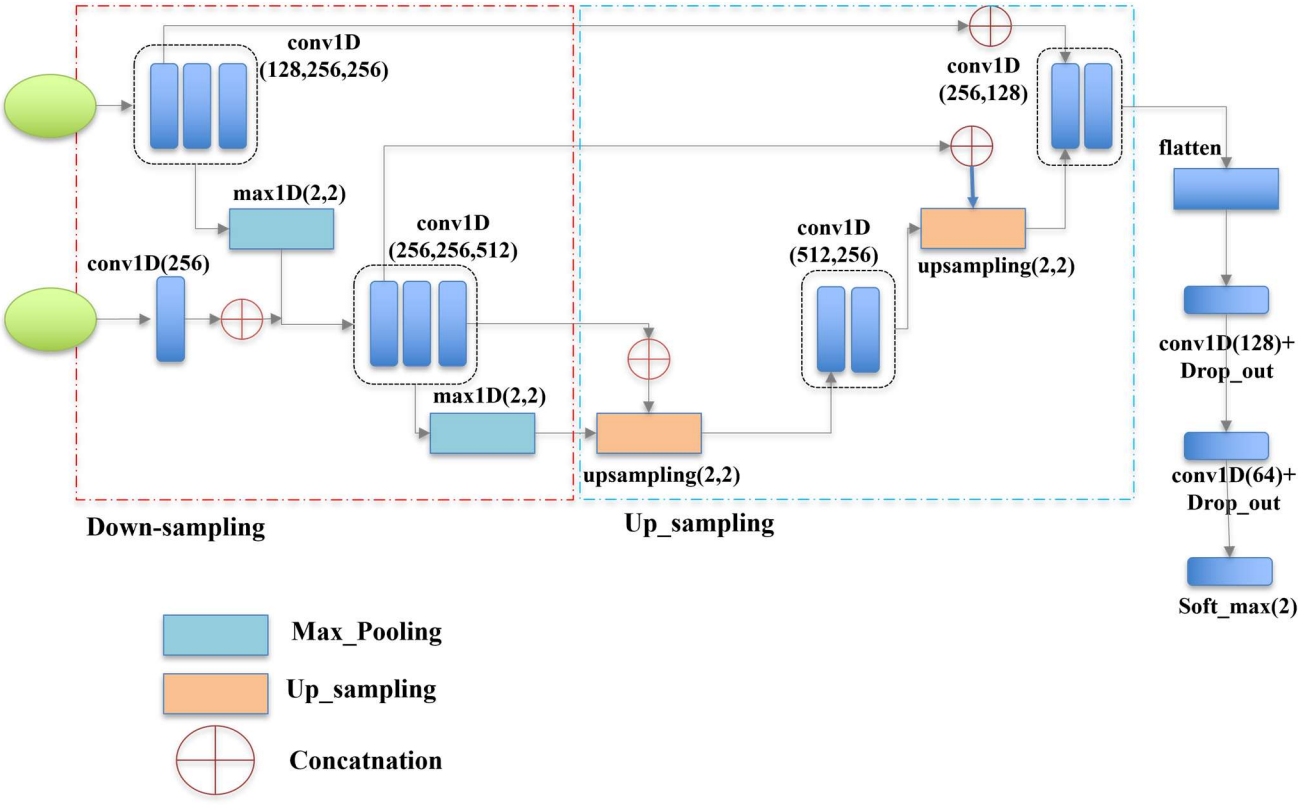

Figure 3. Architecture of the deep-learning-based graph neural network (GNN)

Because public datasets for geometric (non-semantic) mesh analysis are scarce, labels for roof corners were constructed. For each roof mesh, an expert selects a point at each corner. All faces incident to that point are labeled as corner faces (Figure 2). The dataset consists of photogrammetric mesh data reconstructed from aerial imagery of Bafq in Yazd Province, Iran. To address the imbalance between corner and non-corner faces, the non-corner class was down-sampled and the corner class was assigned a weight five times that of the non-corner class during training. In total, 1,255 faces were labeled from 45 buildings for training, of which 345 belong to the corner (class 1) category.

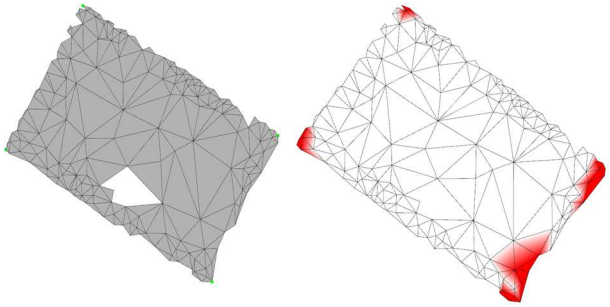

Figure 4. Detecting corner faces to generate the training dataset

3. 3 Geometric features

For each face, seven features were computed: discrete Gaussian curvature at three scales (small, medium, large), normalized triangle-centroid coordinates (x, y, z), and a cornerness (sharpness) index based on angular coverage around the centroid. The cornerness index samples points on a circle of radius $R = C \cdot \bar{d}$ ($\bar{d}$: mean nearest-neighbor distance; C: neighborhood factor) and counts how many angular samples fall outside the mesh; fewer in-mesh samples imply a sharper face. All features are min–max normalized to the range [0, 1].

3. 4 Lod model

Faces predicted as corners are grouped using mesh adjacency. For each group, the Douglas–Peucker algorithm selects salient boundary points; a convex hull delineates the final polygonal outline for each roof. After determining the outlines, small gaps between adjacent roofs in the XY plane are detected and filled to avoid protrusions or intrusions. The final building model is obtained by merging roof outlines with consistent boundaries.

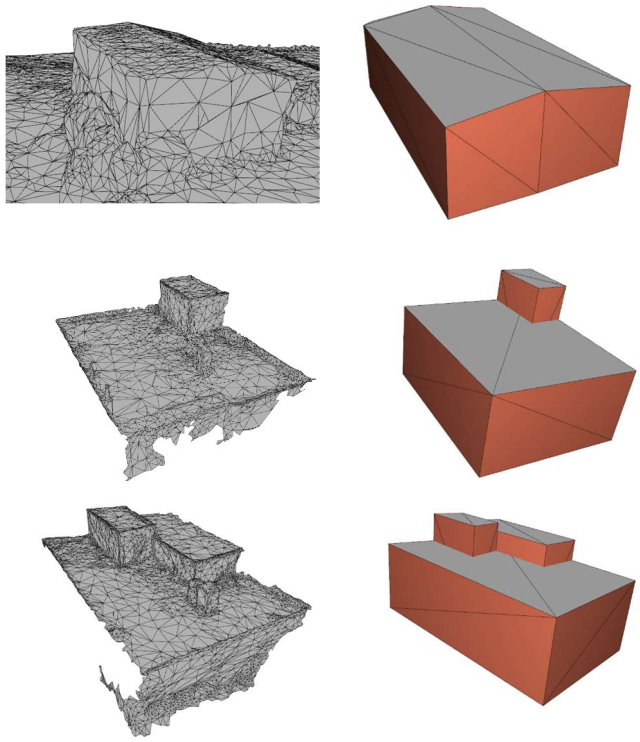

Figure 5. Left column: mesh data; right column: simplified 3D model derived from the mesh

4. VALIDATION

The absence of benchmark datasets for evaluating accuracy and comparing with prior work remains a major gap in this field. To assess the accuracy of our results, we computed the distance between the primary building roofs and the roofs produced by our method (Figure 6).

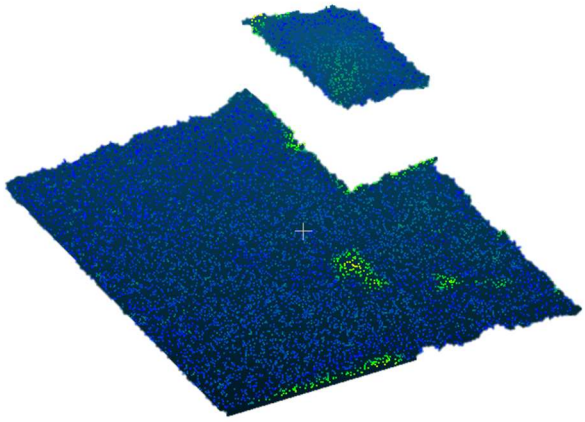

Figure 6. Distance between the point cloud of the simplified 3D model and the point cloud sampled from the original mesh (color transition from blue to green indicates increasing distance)

The discrepancy between the simplified 3D model and the original mesh was quantified by computing signed-distance function (SDF) distances from a point cloud sampled on the original mesh to the simplified surface. Across ten buildings, 95% of distances were under 15 cm, and the maximum deviation was 40 cm (Figure 7). Owing to substantial boundary noise in the meshes, the largest discrepancies were concentrated near the edges.

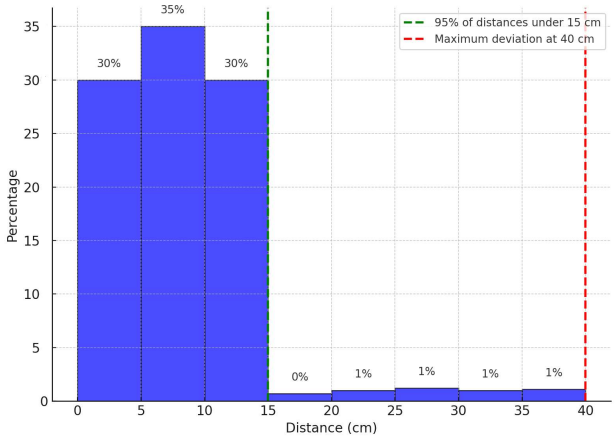

Figure 7. Distance percentage chart

5. CONCLUSION

This study presents a mesh-based pipeline for 3D building reconstruction that couples multi-scale geometric descriptors with a compact deep network. Starting from textured urban triangle meshes, seven per-face features were computed—discrete Gaussian curvature at three scales, normalized triangle-centroid coordinates (x, y, z), and a cornerness index derived from angular coverage around each face. A U-shaped (U-Net–style) multi-input 1D CNN ingests two neighborhood "rings" (first- and second-order face adjacencies), extracting and fusing features through down-/up-sampling with skip connections; Batch Normalization and Dropout improve stability and generalization. Because public datasets for geometric (non-semantic) mesh analysis are scarce, roof-corner labels are created by expert annotation and used to train a binary classifier (corner vs. non-corner). The training data prepared is based on the convex corners of buildings, which is why concave internal corners have not been trained in the neural network. Preparing training data in a way that also supports concave shapes and allows the neural network to be trained on them is one suggestion for future research in this area. Corner faces are then grouped via mesh adjacency, and salient boundary points are selected using the Douglas–Peucker algorithm. Convex-hull operations are then used to delineate roof outlines, while small gaps are filled to ensure consistent boundaries. The resulting models provide simplified, semantically meaningful building geometry suitable for downstream GIS integration, while the same mesh representation supports additional spatial analyses and the production of high-quality deliverables. The use of a graph as the input data for the neural network, based on neighborhood relationships, has led to the network exhibiting very low sensitivity to the influence of the size and orientation of neighboring triangles in corner class identification.

A current limitation is that the approach handles only convex buildings and does not support non-convex footprints. Future work will augment the training data with interior corner points and train the network to recognize them.

References

Bouzas, V., Ledoux, H., Nan, L., 2020: Structure-aware Building Mesh Polygonization. *ISPRS Journal of Photogrammetry and Remote Sensing*, *167*(March), 432–442. https://doi.org/10.1016/j.isprsjprs.2020.07.010

Chen, Z., Ledoux, H., Khademi, S., Nan, L., 2022: Reconstructing compact building models from point clouds using deep implicit fields. *ISPRS Journal of Photogrammetry and Remote Sensing*, *194*(September), 58–73. https://doi.org/10.1016/j.isprsjprs.2022.09.017

Chen, Z., Shi, Y., Nan, L., Xiong, Z., Zhu, X. X., 2023: PolyGNN: Polyhedron-based Graph Neural Network for 3D Building Reconstruction from Point Clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, *218*(PA), 693–706. https://doi.org/10.1016/j.isprsjprs.2024.09.031

Dehbi, Y., Plümer, L., 2011: Learning grammar rules of building parts from precise models and noisy observations. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(2), 166–176. https://doi.org/10.1016/j.isprsjprs.2010.10.001

Frueh, C., Sammon, R., Zakhor, A., 2004: Automated texture mapping of 3D city models with oblique aerial imagery. *Proceedings - 2nd International Symposium on 3D Data Processing, Visualization, and Transmission. 3DPVT 2004*, 396–403. https://doi.org/10.1109/TDPVT.2004.1335266

Gröger, G., Plümer, L., 2012: CityGML - Interoperable semantic 3D city models. *ISPRS Journal of Photogrammetry and Remote Sensing*, *71*, 12–33. https://doi.org/10.1016/j.isprsjprs.2012.04.004

Guo, J., Liu, Y., Song, X., Liu, H., Zhang, X., Cheng, Z., 2022: Line-Based 3D Building Abstraction and Polygonal Surface Reconstruction From Images. *IEEE Transactions on Visualization and Computer Graphics*, *PP*, 1–15. https://doi.org/10.1109/TVCG.2022.3230369

Haala, N., Kada, M., 2010: An update on automatic 3D building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*(6), 570–580. https://doi.org/10.1016/j.isprsjprs.2010.09.006

Han, J., Zhu, L., Gao, X., Hu, Z., Zhou, L., Liu, H., Shen, S., 2021: Urban Scene LOD Vectorized Modeling from Photogrammetry Meshes. *IEEE Transactions on Image Processing*, *30*, 7458–7471. https://doi.org/10.1109/TIP.2021.3106811

He, Y., Wu, X., Pan, W., Chen, H., Zhou, S., Lei, S., Gong, X., Xu, H., Sheng, Y., 2024: LOD2-Level+ Low-Rise Building Model Extraction Method for Oblique Photography Data Using U-NET and a Multi-Decision RANSAC Segmentation Algorithm. *Remote Sensing*, *16*(13), 1–18. https://doi.org/10.3390/rs16132404

Kattan, R. A., Abdulrahman, F. H., Gilyana, S. M., Zaya, Y. Y., 2022: 3D modelling and visualization of large building using photogrammetric approach. *Journal of Engineering Research (Kuwait)*, *10*(4), 1–15. https://doi.org/10.36909/jer.12167

Li, Z., Shan, J., 2022: RANSAC-based multi primitive building reconstruction from 3D point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, *185*(December 2021), 247–260. https://doi.org/10.1016/j.isprsjprs.2021.12.012

Liu, X., Liu, Z., Zhang, Y., Gao, Z., Tan, Y., 2024a: UMeshSegNet: Semantic Segmentation of 3D Mesh Generated from UAV Photogrammetry. *IEEE International Conference on Control and Automation, ICCA*, 388–393. https://doi.org/10.1109/ICCA62789.2024.10591843

Liu, X., Liu, Z., Zhang, Y., Gao, Z., Tan, Y., 2024b: UMeshSegNet: Semantic Segmentation of 3D Mesh Generated from UAV Photogrammetry. *IEEE International Conference on Control and Automation, ICCA*, 388–393. https://doi.org/10.1109/ICCA62789.2024.10591843

Liu, Y., Guo, B., Wang, S., Liu, S., Peng, Z., Li, D., 2022: Urban Building Mesh Polygonization Based on Plane-Guided Segmentation, Topology Correction, and Corner Point Clump Optimization. *Remote Sensing*, *14*(17). https://doi.org/10.3390/rs14174300

Nan, L., Wonka, P., 2017: PolyFit: Polygonal Surface Reconstruction from Point Clouds. *Proceedings of the IEEE International Conference on Computer Vision*, *2017-Octob*, 2372–2380. https://doi.org/10.1109/ICCV.2017.258

Poullis, C., 2013: A framework for automatic modeling from point cloud data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(11), 2563–2575. https://doi.org/10.1109/TPAMI.2013.64

Wilk, L., Mielczarek, D., Ostrowski, W., Dominik, W., Krawczyk, J., 2022a: Semantic Urban Mesh Segmentation Based on Aerial Oblique Images and Point Clouds Using Deep Learning. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *43*(B2-2022), 485–491. https://doi.org/10.5194/isprs-archives-XLIII-B2-2022-485-2022

Wilk, L., Mielczarek, D., Ostrowski, W., Dominik, W., Krawczyk, J., 2022b: Semantic Urban Mesh Segmentation Based on Aerial Oblique Images and Point Clouds Using Deep Learning. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *43*(B2-2022), 485–491. https://doi.org/10.5194/isprs-archives-XLIII-B2-2022-485-2022

Xiao, J., Furukawa, Y., 2012: *Reconstructing the World's Museums*. 668–681.

Xiao, J., Furukawa, Y., 2014: *Reconstructing the World's Museums*. *February*. https://doi.org/10.1007/s11263-014-0711-y

Yan, L., Li, Y., Dai, J., Xie, H., 2023: UBMDP: Urban Building Mesh Decoupling and Polygonization. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1–16. https://doi.org/10.1109/TGRS.2023.3288590

Yang, Y., Tang, R., Xia, M., Zhang, C., 2023: A surface graph-based deep learning framework for large-scale urban mesh semantic segmentation. *International Journal of Applied Earth Observation and Geoinformation*, *119*(April), 103322. https://doi.org/10.1016/j.jag.2023.103322

Yi, C., Zhang, Y., Wu, Q., Xu, Y., Remil, O., Wei, M., Wang, J., 2017: Urban building reconstruction from raw LiDAR point data. *CAD Computer Aided Design*, *93*, 1–14. https://doi.org/10.1016/j.cad.2017.07.005

Yu, D., Ji, S., Liu, J., Wei, S., 2021: Automatic 3D building reconstruction from multi-view aerial images with deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, *171*(July 2020), 155–170. https://doi.org/10.1016/j.isprsjprs.2020.11.011

Zhang, G., Zhang, R., 2023: MeshNet-SP: A Semantic Urban 3D Mesh Segmentation Network with Sparse Prior. *Remote Sensing*, *15*(22). https://doi.org/10.3390/rs15225324

Zhou, Q., Neumann, U., 2010: *2. 5D Dual Contouring : A Robust Approach to Creating Building Models from Aerial LiDAR*. 115–128.