Roof Geometrical Component Extraction Using Bimodal Data and Graph Neural Network

Faezeh Soleimani Vostikolaei *[1], Shabnam Jabari [1]

[1] Department of Geodesy & Geomatics Engineering, University of New Brunswick, Fredericton, NB, Canada- (fsoleima, sh.jabari)@unb.ca

Keywords: Roof Line Detection, 3D City Modeling, Building Wireframe Modeling, Graph Neural Network, Building Segmentation

Abstract

Accurate extraction of roof geometrical elements is essential for creating 3D building models, which play a critical role in urban planning, city management, infrastructure development, and disaster management. Roof geometrical elements consist of lines, which represent the intersections of roof planes, and vertices, which define the intersections of roof lines. Due to the presence of shadowed areas or poor contrast in optical images, roof geometrical elements cannot be extracted efficiently in all areas. This study proposes a novel framework using optical imagery and Digital Surface Models (DSM) to extract these elements and construct 3D building models. The proposed approach uses convolutional neural networks (CNNs) to extract roof features from both RGB and DSM data. Next, a graph-based methodology is employed to create roof models, where roof lines and vertices are represented as nodes, and their spatial relationships are captured through an adjacency matrix. Finally, a Graph Neural Network (GNN) is used to analyze these relationships and refine roof component connectivity. In the first stage, the framework was evaluated on a dataset comprising 1,300 buildings in Fredericton, New Brunswick, achieving an Intersection over Union (IoU) of 0.73, an F1-score of 0.7645, and an F2-score of 0.7641. The mAP results of the second stage, 28.3, demonstrate the effectiveness of a graph-based approach in extracting and reconstructing roof components, contributing to more accurate and automated 3D city modeling.

1. Introduction

3D building models are the key element in the digital representation of urban landscapes. These models provide detailed spatial information about buildings and are crucial for various applications, including urban planning and management, construction, green energy efficiency, navigation, and disaster management (Soleimani Vostikolaei & Jabari, 2023). Among the building components, roofs play a key role in defining the building's structure, and accurately modeling the roofs is the key step toward 3D building reconstruction. The complexity of roof structures—from simple gable and hip roofs to complex designs—makes automated extraction challenging. While significant progress has been made in reconstructing building models and extracting roof structures (Xiong et al., 2014; Xu et al., 2018; Zhao et al., 2022), developing automated and robust methods for roof component detection remains challenging.

Various types of remote sensing data, including orthoimages, stereo optical images, point clouds, and DSM data, have been widely used for extracting building footprints and roof components (Awrangjeb et al., 2013; Zhao et al., 2022). In recent years, however, RGB imagery has emerged as the primary data source due to its high spectral information, making it suitable for detecting roof geometrical elements. These elements include roof planes, lines, and vertices which are the intersection of roof lines. For example, Campoverde et al. (2024) utilized airborne optical images with cadastral information to extract roof geometrical planes. To generate 3D building models, they converted the 2D planes into vectorized format and integrated DSM to create 3D roof planes. Similarly, (Schuegraf & Bittner, 2019) used optical images to extract roof sections and then vectorized them to generate 3D building models. While RGB imagery provides valuable optical and spectral information on buildings and roof elements, its performance is limited in areas with occlusions and shadows. To overcome these limitations, this study combines DSM data with RGB imagery, leveraging both optical and elevation information to enhance roof element extraction and improve the accuracy of 3D roof reconstruction.

Roof element recognition is a higher-level task than building footprint detection because it requires distinguishing and classifying multiple structural elements, such as ridges, valleys, eaves, and junction elements, rather than just detecting a boundary of buildings (Zhao et al., 2022). Moreover, in order to generate 3D building and wireframe models, the roof elements need to be converted to vector format. Hence, vectorization and finding the topological connections between roof elements is another high-level reconstruction task.

Traditional methods for extracting roof elements typically involve multiple steps, such as feature extraction, morphological processing, and data format conversion. These approaches are often time-consuming and need labor work, which makes them difficult for large-scale 3D reconstructions. The advancement of deep neural networks (DNNs) has drawn significant attention to segmentation purposes. These models enable the fusion of multimodal data by allowing for the extraction of 2D roof components from optical and height data sources and combining these two features in the backbone of the model. Additionally, leveraging Graph Neural Networks (GNNs) to model relationships between roof lines and vertices has introduced new possibilities for automated 3D reconstruction, making end-to-end trainable methods. However, to the best of our knowledge, extracting roof lines and vertices from bimodal data and converting the pixel-level masks to the vector format has not been thoroughly explored in existing literature.

In this study, we propose an automated approach for extracting building roof structures using high-resolution airborne images and DSMs derived from LiDAR point clouds. The main contribution of this work is the development of an end-to-end framework that extracts roof lines and vertices RGB and height

* Corresponding (presenting) author

features through a convolutional neural network (CNN) and combines them before passing them to the segmentation head. The extracted pixel-based roof components are then converted into a structured, vectorized format using GNN, enabling more accurate roof modeling.

2. Related Works

This work is classified into two stages; in the first stage, the roof elements are extracted using a CNN approach, while in the second stage, the extracted roof components are converted to the vector format. The literature review of each stage is presented in the following sections.

2.1 Convolutional Neural Networks (CNNs) for Roof Element Extraction

Using convolutional neural networks (CNNs) has significantly advanced the extraction of building footprints and geometrical elements directly from high-resolution aerial photos, satellite images, and point clouds.
Several studies have explored CNN-based approaches for recognizing and segmenting building geometries. While some works focused on building footprint extraction(Crawshaw, 2020; Liu et al., 2022; Sang & Minh, 2018; Soleimani Vostikolaei & Jabari, 2023, 2024a), recent studies have shifted towards identifying detailed roof structures using different data sources. For example, a group of works utilized just aerial or satellite data to detect buildings and their footprints (Benjamin Bischke et al., 2019; Das & Chand, 2021; Robinson et al., 2022) while other studies utilized LiDAR point clouds or DSM to detect building outlines(Li et al., 2019; Pawłowski et al., 2023; Soleimani Vostikolaei & Jabari, 2024a; Zhuang et al., 2019). To detect the roof geometrical elements, Campoverde et al. (2024) employed CNNs to segment roof planes from aerial imagery, integrating cadastral data, they also used digital elevation models to generate structured 3D representations. Similarly, Schuegraf, Shan, and Bittner (2024) utilized CNN-based feature extraction from RGB images to identify and vectorize roof sections, producing LoD2-level building models. (Soleimani Vostikolaei & Jabari, 2024b) used Unet to extract roof lines and plane masks that were further used for reconstructing building wireframes. They utilized LiDAR point clouds to assign elevation to each roof line segment and create 3D wireframes of buildings. Zhao et al., (2022) used a multi-task learning method to extract planar structures from optical images and the attraction field map (AFM) to convert the masks to vector format.
While CNNs perform exceptionally at extracting visual features and segmenting roof components, they result in pixel-based segments, making it challenging to generate vectorized rooflines directly. Moreover, RGB imagery alone struggles in occluded or low-contrast areas, limiting its effectiveness in highly dense urban environments. To address these challenges, hybrid approaches integrating CNNs with additional data sources—DSMs—have been proposed to extract both optical and height features of building elements and concatenate them before feeding them to the segmentation head.

2.2 Graph Neural Networks (GNNs) for Roofline and Vertices Connectivity and Vectorization

Beyond pixel-based segmentation, accurately modeling roof structures requires detecting spatial relationships between roof elements. Graph-based approaches have gained increasing attention for their ability to capture complex geometric dependencies in structured data. Graph Neural Networks (GNNs), in particular, provide a robust framework for reasoning about connections between building line segments and vertices, which help to convert the segmented roof features into a vector format.
GNNs have been successfully applied in various geometric reasoning tasks, including scene parsing and architectural structure reconstruction (Wu et al., 2020; Xiong et al., 2014). For the purpose of building modeling, GNNs provide a robust solution for learning spatial relationships between rooflines and connecting key junctions. Some works, such as Nauata & Furukawa, (2019), used graphical relationship concept to recognize roof topology by constructing adjacency matrices that encode connectivity between detected line segments. These methods showed the potential of GNNs for structured roof reconstruction but often relied on multi-stage pipelines and handcrafted constraints, limiting their efficiency and scalability. Zhao et al., (2022) also used GNN to find the relation between roof junctions and connect them to vectorize roof lines.
In contrast, our approach integrates GNNs directly into an end-to-end learning framework, where rooflines and vertices are represented as graph nodes, and their relationships are learned through the network. By using GNNs for roofline and vertices connectivity reasoning, we aim to overcome the pixel-based limitations of conventional segmentation-based methods, ensuring that extracted roof components form topologically accurate structures.

3. Methodology

The proposed approach for roof element extraction consists of two main stages: (1) Deep learning-based bimodal feature extraction for roof element segments and (2) Graph-based neural network to find the relation between roof elements. This section describes the methodology in detail.

3.1 Stage 1: Bimodal Feature Extraction and Segmentation

The overall pipeline of the stage1 method is shown in **Figure 1**.

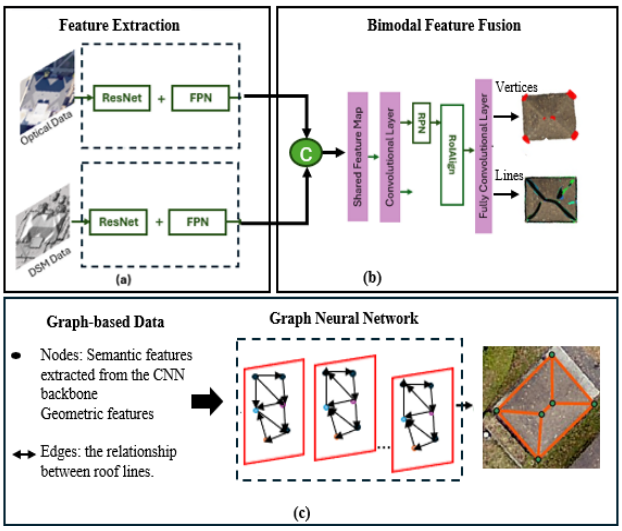

Figure 1. The overall pipeline of the proposed method. (a) represents the feature extraction section from two different branches. (b) shows the combination of the extracted features in each modality.

3.1.1 Feature Extraction with ResNet Backbone

To extract low- and mid-level spatial features, we employ a ResNet-50 backbone pre-trained on ImageNet. This allows for hierarchical semantic feature extraction using multiple convolutional layers. Each dataset is fed to the ResNet backbone, resulting in different depths of optical and height feature maps. The RsNet operation can be formulated as follows:

$$F_{res} = \text{ResNet}(I) \in \mathbb{R}^{C \times H/4 \times W/4}$$

Where H, and W are the height and width of image, tensor $I = [I_1, I_2]$, and C is the number of feature channels after convolutional processing.

3.1.2 Feature Pyramid Network (FPN) for Multi-Scale Representation

To upsample and fuse features from different levels of ResNet, a Feature Pyramid Network (FPN) is applied to each branch. FPN constructs multi-scale pyramidal feature maps by fusing top-down and bottom-up optical and height features separately.
This structure enhances the detection of both fine roof junctions and large-scale structural lines, crucial for subsequent graph-based reasoning.

$$F_{FPN} = \text{FPN}(F_{res}),\ \{P_i = [P_2, P_3, P_4, P_5]\}$$

Where P_k represents feature maps at different scales.

3.1.3 Feature Fusion

To have a unified feature map containing both modality attributes, we combined the optical and height features resulting from FPN layers.
The concatenation operation can be formulated as:

$$F_{concat} = \text{Concat}(F_{res\_RGB}, F_{res\_DSM})$$

Where Concat(·) represents the concatenation operation F_{RGB} and F_{DSM} represents the optical and height features extracted from ResNet followed by FPN backbones, respectively.

3.1.4 Region Proposal Network (RPN) and RoIAlign

To segment individual roof elements, we employ a Region Proposal Network (RPN), followed by RoIAlign, for precise feature extraction. The RPN generates candidate bounding boxes for objects, and RoIAlign ensures high-resolution feature extraction for each region:
Each bounding box is classified into:

- Roof Junctions
- Roof Lines

The segmentation mask is generated through the segmentation head, producing pixel-wise delineations.

3.2 Stage 2: Graph-Based Structural Reasoning

The flowchart of the second stage of the proposed method is presented in Figure **1**, section c.

3.2.1 Constructing the Graph from Detected Roof Elements

Once roof junctions and lines are segmented, we construct a graph representation where:

- Nodes (V): Represent roof line segments.
- Edges (E): Encode geometric and structural relationships between roof line elements.

The features of each node are the semantic features extracted from final feature maps of the previous stage and geometric features of roof lines, which are:

- Start and end coordinates
- Midpoint coordinates
- Line length (euclidean distance between endpoints)
- Optical RGB values of the midpoint of each line segment
- Feature map aggregation along the segment:

3.2.2 Edge Construction and Feature Extraction

Edges are formed based on geometric constraints. The two line segments are connected if they share an endpoint and form a valid structural connection.
Each edge is assigned features:

- Euclidean Distance: Between nodes.
- Angle Difference (θ): Between connected lines:

$$\theta_{i,j} = \left| \arctan\left(\frac{y_2 - y_1}{x_2 - x_1}\right) - \arctan\left(\frac{y_4 - y_3}{x_4 - x_3}\right) \right|$$

Where i and j are two connected line segments.

3.2.3 Graph Neural Network (GNN) for Structural Connectivity

To infer relationships between roof elements, we use a Graph Convolutional Network (GCN), where node embeddings are iteratively updated using neighboring connections:

$$H^{(l+1)} = \sigma\left(D^{-\frac{1}{2}} A D^{-\frac{1}{2}} H^{(l)} W^{(l)}\right)$$

Where $H^{(l)}$ represents node embeddings at layer l, A is the adjacency matrix, D is the degree matrix, diagonal matrix with node degrees, $W^{(l)}$ is a learnable weight matrix, and $D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ is the normalized adjacency matrix which ensures feature aggregation is scale-invariant and prevents node influence from growing arbitrarily.
We train the GNN with a binary classification objective in which positive edges represent the valid connections between roof elements and negative edges meaning that there are non-connected elements. Using a contrastive loss function, we optimize the model to learn meaningful structural relationships.

$$L = -\sum_{(I,j)\in E^{+}} \log\left(\sigma\left(H_i . H_j\right)\right) - \sum_{(I,j)\in E^{-}} \log(1 - \sigma\left(H_i . H_j\right))$$

Where E^{+} and E^{-} represent positive and negative edges, respectively.

4. Experiment

4.1 Data Preparation

We created a rooftop dataset to extract all roof elements.
Each input consists of:

- **RGB Image ($I_1 \in \mathbb{R}^3 \times H \times W$):** Three-channel data representing color intensity in red, green, and blue.
- **DSM Map ($I_2 \in \mathbb{R}^1 \times H \times W$):** Single-channel DSM representing elevation values.
- **Mask ($I_3 \in \mathbb{R}^1 \times H \times W$):** single-channel masks of roof lines and vertices.

The dataset was gathered from buildings located in New Brunswick, Canada. To generate a high-resolution digital surface model, we interpolated the first returns of LiDAR point clouds with a density of six points per square meter, resulting in a 41 cm spatial resolution. In addition, we incorporated high-resolution airborne orthophotos with a 7.2 cm spatial resolution, which were annotated to mask out roof lines and junctions. These annotated orthoimages were used to construct an optical dataset. The dataset consists of 1,300 buildings, of which 70% were assigned for training and validation, while the remaining 30% was held out for testing the model's performance.

To mitigate overfitting in deep learning, we used data augmentation for the training samples. Specifically, augmentation included vertical and horizontal flipping of the images, ensuring the model learned robust and invariant representations of roof structures.

We also used Roof3D[1] to further assess the performance of our proposed method. This dataset includes optical and DSM photos of roofs but doesn't contain the label required for segmentation purposes. Hence, we manually labeled 50 data samples for testing the performance.

4.2 Evaluation Metrics

To assess the performance of the proposed multi-output segmentation network, in the first stage, we utilized three key metrics: Intersection over Union (IoU), Dice Coefficient (F1-score), and F2-score. These metrics provide an evaluation by considering the accuracy of predicted roof vertices and line elements and their alignment with ground truth data.

Intersection over Union (IoU) measures the degree of overlap between the predicted roof elements and the ground truth. It is computed by comparing the shared area of correctly identified roof elements to the total area covered by both predictions and ground truth. A higher IoU value indicates a greater level of agreement between the predicted and actual roof structures.

F1-score serves as an alternative to IoU, placing more emphasis on the overlapping regions. Unlike IoU, it assigns double weight to the correctly predicted areas, making it a balanced measure of both precision and recall. This metric ensures that both false positives and false negatives are considered when evaluating the quality of roof segmentation.

The F2-score is a variation of the F1-score that prioritizes recall over precision. This adjustment is particularly useful in applications where missing parts of roof elements (false negatives) are more detrimental than predicting extra, incorrect components (false positives). By increasing the importance of recall, the F2-score helps ensure that essential roof details are accurately captured in the segmentation process.

To evaluate the performance of the graph neural network in the second stage, we used the mean Average Precision (mAP) metric. This metric evaluates how well the GNN predicts valid edges compared to the invalid edges.

$$\mathrm{mAP} = \frac{1}{|N|}\sum_{i=1}^{N} \mathrm{AP_i}$$

Where AP is the average precision for each IoU threshold and N is the number of IoU thresholds.

4.3 Network Details

We developed a bimodal multi-output segmentation framework to separately extract and fuse optical and elevation features for detecting essential rooftop elements required for roof modeling in two stages.

In the first stage, RGB and DSM data were processed through two parallel ResNet branches followed by a feature pyramid network to extract multi-scale features, as illustrated in Figure 1, section a. The extracted bimodal feature maps were then concatenated to form a unified representation, which was subsequently fed into the region proposal network. The RPN generated region proposals, which were refined using RoIAlign, ensuring precise spatial alignment and accurate segmentation of roof components. The backbone architecture consisted of ResNet-50 with FPN, and the network was trained for 200 epochs using the Adam optimizer on an NVIDIA GeForce RTX 3090 GPU with 64 GB memory. During both training and inference, input images were resized to 256 × 256 pixels.

To find the relation between line segments, we employed a graph convolutional network to model relationships between roof lines and junctions extracted from bimodal aerial data. The graph was constructed by defining roof lines and junctions as nodes, while edges captured geometric and topological dependencies between them. GCN was trained for 200 epochs using the Adam optimizer with a learning rate of 0.01.

5. Results and Discussion

In our previous work (Soleimani Vostikolaei & Jabari, 2024b), we extracted roof planes and lines and then segmented point clouds using the 2D detected planes and used the point segments to create 3D lines. However, one of the challenges was converting pixel-based segmentation into vector representations and accurately finding relationships between line segments. In this study, we address this limitation by using a graph convolutional network to model the connections between 2D line segments.

The raster-based segmentation of roof junctions and lines resulting from stage 1 of the proposed method is shown in Figure 2.

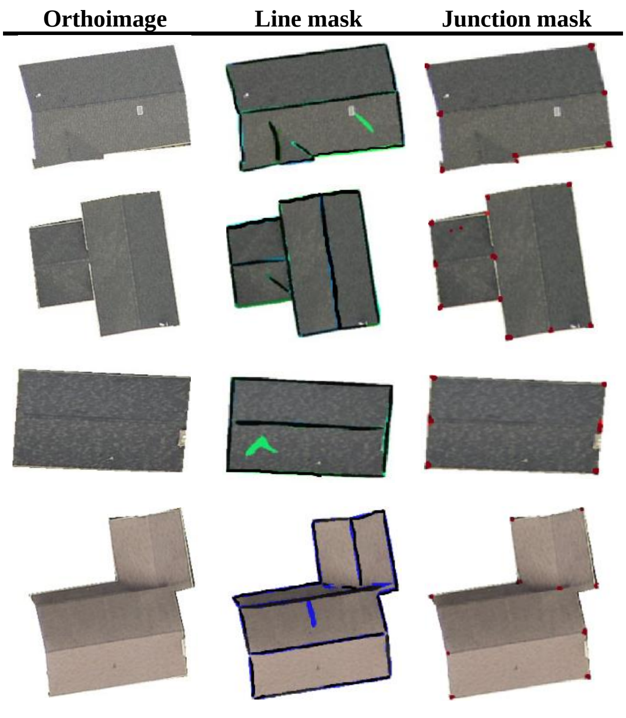

Figure 2 . Line and junction masks resulting from the proposed network. Where the green and blue lines represent different instances of line segments and red points show the vertices detected from network.

To assess the performance of our approach, we trained our segmentation model on a custom rooftop dataset and further evaluated it on the Roof3D benchmark dataset. Table **1** presents a comparative analysis of the segmentation performance for the

[1] https://github.com/dlrPHS/GPUB?tab=readme-ov-file

two datasets. The performance difference between these two datasets is attributed to differences in pixel size and spatial resolution between them. As we can see from Figure 2, our method demonstrates robust line and junction detection capabilities across diverse roof types.

We used multi-output segmentation in phase 1. As demonstrated in prior research (Soleimani Vostikolaei & Jabari, 2024b), multi-output segmentation consistently outperforms compared to single-output segmentation. A key advantage of the multi-output segmentation approach is the shared feature extraction backbone, which enhances learning across different output types. This shared learning network helps the model better recognize common features like edges and corners, leading to improved segmentation quality.

	IoU	F1-Score	F2-Score
Custom dataset	0.73	0.7645	0.7641
Roof3D	0.69	0.7123	0.7056

Table 1. Performance analysis of the proposed method

In the second stage, we employ a graph convolutional network to find the relationships between roof line segments. The results, illustrated in Figure 3, show that our method successfully reconstructs connectivity between the roof structures.

However, some challenges remain—a few line segments were not correctly detected in the segmentation stage, while others were misclassified as roof lines, introducing ambiguity in the graph construction process. the misclassified line segments usually have a image coordinates far from the existing lines, so the network can not find their relationship. Despite these challenges, our approach effectively captures topological relationships and provides a reliable basis for structured 3D roof modeling.

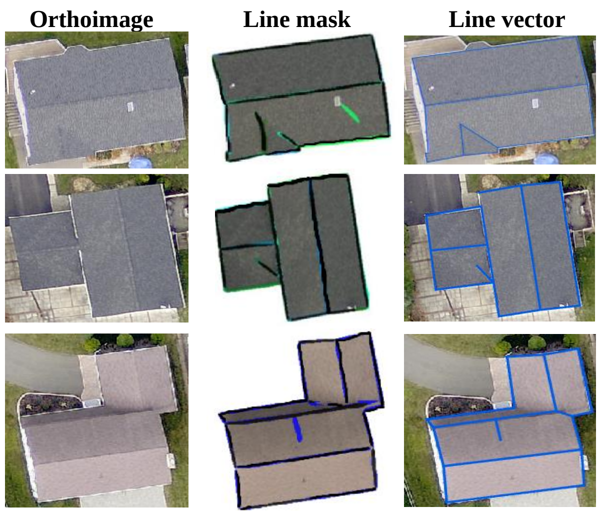

Figure 3. The vector output resulted from the GCN network compared to the orthoimage and masks.

To assess the accuracy of the second phase, we used the mAP metric. The mAP result shows an accuracy of 28.3 which proves the trustable connection among different line segments.

6. Conclusion

In this work, we developed an end-to-end learning bimodal segmentation framework to extract the optical and elevation features of roof elements and fuse them before feeding them to the segmentation head. The objective was to evaluate whether integrating both optical and DSM data enhances segmentation performance compared to processing each modality independently. The performance of the method was tested on two different datasets, one custom dataset created from rooftops of the eastern part of Canada and also Roof3D dataset.

In the second stage of the proposed method, we converted the line segments to a graph structure and then used a graph-based convolutional network to find the relationship between line segments and connect them. We explored converting pixel-based segmentations into vectorized representations, which are crucial for 3D reconstruction.

The original DSM data have an offset relative to the optical images, which causes an error in the segmentation stage. In future work, we aim to enhance the framework by integrating an end-to-end learning approach to automatically correct this misalignment and improve the feature fusion accuracy.

7. References

Awrangjeb, M., Zhang, C., & Fraser, C. S. (2013). Automatic extraction of building roofs using LIDAR data and multispectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *83*, 1–18. https://doi.org/10.1016/j.isprsjprs.2013.05.006

Benjamin Bischke, Patrick Helber, Joachim Folz, Damian Borth, & Andreas Dengel. (2019). Multi-task Learning for Segmentation of Building Footprints with Deep Neural Networks. *2019 IEEE International Conference on Image Processing (ICIP)*.

Crawshaw, M. (2020). Multi-Task Learning with Deep Neural Networks: A Survey. *ArXiv Preprint ArXiv:2009.09796*. http://arxiv.org/abs/2009.09796

Das, P., & Chand, S. (2021). AttentionBuildNet for building extraction from aerial imagery. *Proceedings - IEEE 2021 International Conference on Computing, Communication, and Intelligent Systems, ICCCIS 2021*, 576–580. https://doi.org/10.1109/ICCCIS51004.2021.9397178

Li, W., He, C., Fang, J., Zheng, J., Fu, H., & Yu, L. (2019). Semantic segmentation-based building footprint extraction using very high-resolution satellite images and multi-source GIS data. *Remote Sensing*, *11*(4). https://doi.org/10.3390/rs11040403

Liu, T., Yao, L., Qin, J., Lu, N., Jiang, H., Zhang, F., & Zhou, C. (2022). Multi-scale attention integrated hierarchical networks for high-resolution building footprint extraction. *International Journal of Applied Earth Observation and Geoinformation*, *109*. https://doi.org/10.1016/j.jag.2022.102768

Nauata, N., & Furukawa, Y. (2019). Vectorizing World Buildings: Planar Graph Reconstruction by Primitive Detection and Relationship Inference. *16th European Conference, Glasgow, UK*, 23–28. http://arxiv.org/abs/1912.05135

Pawłowski, M., Wróblewska, A., & Sysko-Romańczuk, S. (2023). Effective Techniques for Multimodal Data Fusion: A Comparative Analysis. *Sensors*, *23*(5). https://doi.org/10.3390/s23052381

Robinson, C., Ortiz, A., Park, H., Bank, W., Lozano, N., World Bank, G., Kaw, J. K., Sederholm, T., Dodhia, R., & Ferres, J. M. L. (2022). Fast building segmentation from satellite imagery and few local labels. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1463–1471.

Sang, D. V., & Minh, N. D. (2018). Fully residual convolutional neural networks for aerial image segmentation. *ACM*

International Conference Proceeding Series, 289–296. https://doi.org/10.1145/3287921.3287970

Schuegraf, P., & Bittner, K. (2019). Automatic building footprint extraction from multi-resolution remote sensing images using a hybrid FCN. *ISPRS International Journal of Geo-Information*, *8*(4). https://doi.org/10.3390/ijgi8040191

Soleimani Vostikolaei, F., & Jabari, S. (2023). Large-Scale LoD2 Building Modeling using Deep Multimodal Feature Fusion. *Canadian Journal of Remote Sensing*, *49*(1). https://doi.org/10.1080/07038992.2023.2236243

Soleimani Vostikolaei, F., & Jabari, S. (2024a). Multimodal Building Footprint Extraction from Orthophotos and Lidar Point Clouds Using Deep Learning Frameworks. *2024 IEEE International Geoscience and Remote Sensing Symposium*, 8214–8217. https://doi.org/10.1109/igarss53475.2024.10641225

Soleimani Vostikolaei, F., & Jabari, S. (2024b). Reconstruction of Building LoD2 Wireframe Models Using Semantic Segmentation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLVIII-M-4–2024*, 55–61. https://doi.org/10.5194/isprs-archives-XLVIII-M-4-2024-55-2024

Wu, T., Lu, Y., Zhu, Y., Zhang, C., Wu, M., Ma, Z., & Guo, G. (2020). GINet: Graph Interaction Network for Scene Parsing. *16th European Conference, Glasgow, UK*, 34–51. http://arxiv.org/abs/2009.06160

Xiong, B., Oude Elberink, S., & Vosselman, G. (2014). A graph edit dictionary for correcting errors in roof topology graphs reconstructed from point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, *93*, 227–242. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2014.01.007

Xu, S., Pan, X., Li, E., Wu, B., Bu, S., Dong, W., Xiang, S., & Zhang, X. (2018). Automatic building rooftop extraction from aerial images via hierarchical RGB-D priors. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(12), 7369–7387. https://doi.org/10.1109/TGRS.2018.2850972

Zhao, W., Persello, C., & Stein, A. (2022). Extracting planar roof structures from very high resolution images using graph neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, *187*, 34–45. https://doi.org/10.1016/j.isprsjprs.2022.02.022

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2019). A Comprehensive Survey on Transfer Learning. *Proceedings of the IEEE*, 43–76. http://arxiv.org/abs/1911.02685