The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-4/W9-2024
GeoAdvances 2024 – 8th International Conference on GeoInformation Advances, 11–12 January 2024, Istanbul, Türkiye

SEMANTIC LABELING OF 3D BUILDINGS BY USING GRAPH NEURAL NETWORK (GNN)

Hanis Rashidan[1*], Ivin Amri Musliman[1], Mohammed Jawaluddeen Sani[2], Alias Abdul Rahman[1].

[1] 3D GIS Research Lab, Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia, Johor, Malaysia.
*mhanis27@gmail.com, ivinamri@utm.my, alias@utm.my

[2] Department of Surveying and Geoinformatics, Federal Polytechnic, Bauchi, Bauchi State, Nigeria.
mjsani71@gmail.com

KEYWORDS: Semantic Labeling, 3D Building Models, CityGML

ABSTRACT:

Understanding building semantics is crucial for comprehending their structure, components, and functions. This study investigates the utilization of Graph Neural Networks (GNN) to semantically label 3D building models, aiming to optimize these labels' output within Geographic Information Systems (GIS). The methodology leverages diverse datasets comprising 3D building models, utilizing BuildingGNN algorithms to iteratively refine the labeling process. The resultant labeled components, after thorough validation, are translated into the CityGML format. The generated CityGML dataset holds promise for a wide array of 3D city model applications, heightening the utility of labeled components within GIS analyses. While exploring GNN's capabilities may reduce manual effort and time and provide standardized data representation, this study also addresses challenges in ensuring output reliability. While automation remains a goal, future research endeavors may focus on refining automatic labeling techniques and semantic translation processes to further improve accuracy and applicability.

1. INTRODUCTION

Understanding the architectural structures of buildings holds significant value within Geographic Information Systems (GIS). Incorporating 3D building models into GIS emphasizes the importance of accurately comprehending and annotating building components within a spatial context. Semantic modeling, crucial in this integration, involves assigning meaning or labels to components within datasets, such as 3D models. This process enhances the spatial representation of building elements, aiding in representing architectural features like walls, doors, and windows. Moreover, it enhances the analytical capabilities of GIS applications, providing insights beneficial for urban planning, infrastructure development, and environmental analysis.

Semantic modeling promotes a structured approach to data representation, ensuring consistency and interoperability across diverse datasets (Uceda-Sosa et al., 2011). Utilizing semantic modeling could involve creating models that not only represent the physical characteristics of buildings but also incorporate meaningful information about their components, functionalities, and relationships. This structured representation facilitates better data management and supports efficient data integration and collaboration in various applications. Thus, the integration of semantic modeling within GIS enriches spatial information, enabling comprehensive spatial analysis and informed decision-making (Stoter et al., 2020).

Current methods rely on manual annotation of building components within 3D models, which is labor-intensive and time-consuming (Selvaraju et al., 2021). Additionally, human subjectivity in this process can introduce inconsistencies, challenging the maintenance of accuracy and standardization across datasets. As datasets grow larger and the need for more detailed spatial information increases, the limitations of manual annotation may become more apparent.

To tackle these challenges, integrating automated methodologies into GIS for annotating building components may present better solutions. Technological advancements, especially in machine learning and deep learning, offer opportunities to develop algorithms capable of automatically identifying and labeling building elements within 3D models. Leveraging these advancements can significantly improve the efficiency of annotating building components, reducing analysis time and effort while enhancing accuracy and consistency.

This study aims to explore labeling of building semantics within 3D building models, employing a deep learning framework and translating the output into GIS data formats for wider applicability and benefit. The subsequent sections of the paper are organized as follows: Section 2 presents a literature review of previous works on semantic labeling of 3D building data. Section 3 outlines the proposed methodology, encompassing data preprocessing and technique implementation. Finally, Section 4 discusses preliminary results and findings, while Section 5 concludes the paper by highlighting potential contributions and discussing future research directions.

2. RELATED WORKS

2.1 Techniques and Algorithms

Research into the segmentation and labeling of building components has utilized various methodologies and technological approaches. Noteworthy studies (Alexander and Ben, 2010; Alexander and Ben, 2015) emphasized shallow pipelines employing hand-engineered point descriptors and predefined rules. Similarly, Demir et al., (2015a) developed a combinatorial algorithm focusing on grouping faces into nonlabeled components across meshes, exhibiting considerable repetition. Additionally, Demir et al., (2015b) proposed a user-assisted segmentation algorithm, while other works, such as Kobyshev et al. (2016) and Kundu et al. (2020), explored using symmetry cues to group architectural components. Recent efforts by Selvaraju et al. (2021) introduced a learning-based approach utilizing modern deep backbones for point descriptor extraction, often integrating deep learning techniques for automating the annotation process.

Moreover, Rook et al. (2016) established a workflow for semantically labeling 3D city models conforming to the CityGML standard. Khan et al. (2020) proposed an end-to-end architecture that effectively classifies raw 3D point clouds from indoor and outdoor scenes, displaying promising outcomes. Neural networks have emerged as robust tools for analyzing structured data like 3D models. PointNet++ by Qi et al. (2017) introduced a hierarchical feature learning framework, successfully applied to semantic segmentation tasks. Meanwhile, Wu et al. (2020) conducted a comprehensive survey on graph neural networks (GNNs), illustrating their ability to capture spatial dependencies and propagate information through graph structures. GNNs enable context-aware predictions by considering spatial relationships among building elements, leading to more accurate semantic labeling.

Despite advancements, challenges persist within these methodologies, particularly regarding accuracy in handling complex architectural details, computational resource demands, and standardization across datasets. Studies have highlighted difficulties in addressing diverse architectural styles, variations in data quality, and the need for real-time processing in dynamic urban landscapes. An ongoing challenge involves the generalization of labeling models across diverse datasets and urban contexts. Previous studies (Chen et al., 2009) and (Kalogerakis et al., 2010) mainly concentrated on 3D shape semantic segmentation datasets consisting of objects with relatively simple structures and a limited number of components. However, buildings exhibit much richer structures, varying in sizes, part geometry, and styles (Lun et al., 2015). Consequently, the semantic labeling of buildings poses a more intricate and demanding task.

3. METHODOLOGY

This section outlines the proposed methodology for semantic labeling of 3D building data, utilizing graph neural networks and exporting data into 3D GIS standards. The workflow consists of several steps: data collection and processing, training of the GNN model, prediction of semantic labels, validation and correction of the output semantic data, and exporting the results to GIS data format, as depicted in Figure 1 below.

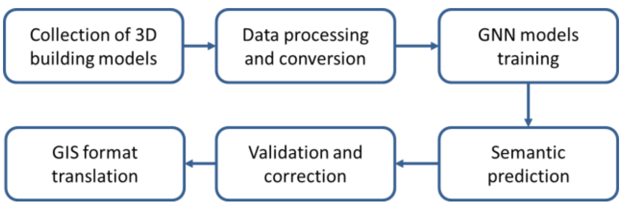


Figure 1. The workflow of the semantic labeling of 3D building data.

The following sections describe various processes related to the workflow of the semantic labeling of 3D building data to GIS data format (CityGML).

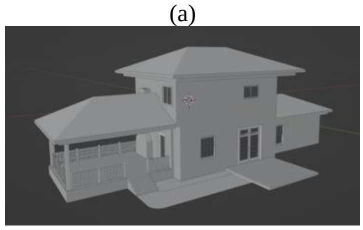
3.1 Data Collection

The primary data source for this study was BuildingNet, serving as a comprehensive repository of 3D building models available at https://buildingnet.org. These models have uniform exterior labeling, containing detailed annotations for various architectural components essential for developing the semantic labeling technique. The dataset offers a diverse collection of building structures, encompassing variability in architectural styles, sizes, and complexities crucial for training models capable of handling real-world scenarios.

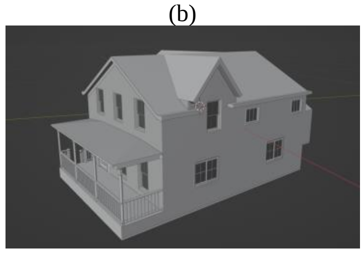
In the data preparation phase, two distinct methodologies were employed: random sampling and manual selection. Initially, a random sampling approach encompassed 100 residential building models to capture a broad spectrum of architectural diversity within the dataset. This approach aimed to ensure exposure to various building styles and complexities, contributing to the model's adaptability to different architectural structures.

Simultaneously, recognizing the variance in annotation quality within the datasets, a manual selection process was initiated. An additional set of 100 residential building models was chosen from the same repository, emphasizing quality annotations. This intentional curation aimed to enrich the training set with accurately labeled models, ensuring a higher standard of annotated data for model learning and validation. Figure 2 provides a visual representation of the sample residential building modes in their 3D form.

(a)



(b)



(c)

Figure 2. Visual representation of different styles and structures in 3D building models. (a) a single-story regular house. (b) and (c) illustrate double-story houses with increased geometric data and additional details.

3.2 Technique and Mechanism

In the selection of algorithms or techniques for labeling, several critical considerations were paramount in effectively addressing the research problem. The primary criterion involved assessing the chosen approach's ability to comprehensively analyze spatial and structural relations within 3D building meshes. Additionally, the method's capacity to handle diverse geometric primitives and accurately identify various building components played a crucial role in the selection process.

For labeling semantics in this study, the selected approach involved the utilization of a graph neural network (GNN) known as BuildingGNN (Selvaraju et al., 2021). This GNN methodology excels in labeling building meshes by intricately analyzing spatial and structural relations among geometric primitives. Within the graph, each subgroup functions as a node, leveraging relationships such as adjacency and containment between pairs of nodes. The neural message-passing mechanism within the graph facilitates the final mesh labeling, demonstrating enhanced performance compared to previous methodologies.

The GNN-based approach has showcased state-of-the-art performance in labeling 3D building data, as substantiated by prior experiments and research literature. Its ability to yield improved results triggers more research work. The BuildingGNN approach operates by segmenting distinct "subgroups" within a building, treating each subgroup as a foundational component. Initially, the GNN generates detailed representations for individual building components like walls or windows, similar to creating intricate profiles. This process enriches the GNN's understanding of each component, facilitating a more comprehensive comprehension. Subsequently, employing an edge-based approach, the GNN establishes connections between these components, where, these edges, featuring different types, serve to illustrate various relationships between components. Figure 3 presents a summary of the key steps of the BuildingGNN approach according to Selvaraju et al. (2021).

BuildingGNN

Input: 3D building mesh with subgroup
Output: Labels per subgroup
Step 1: Node Initialization
- Create nodes representing each subgroup in the mesh.
- Initialize node representations based on subgroup attributes.

Step 2: Establish edges to capture relationships
- Proximity (based on spatial distance).
- Supportive relations (upright axis correlation).
- Similarity (detecting symmetric arrangements).
- Containment (subgroup enclosure).

Step 3: GNN Processing
- Update node and edge representations.
- Employ neural message passing and MLPs.
- Decode node to assign labels through softmax.

End

Figure 3. Key steps of the BuildingGNN for labeling 3D building mesh.

3.3 Development and Training

The development and training of the BuildingGNN prediction model encompassed several sequential steps aimed at ensuring its efficacy in processing and labeling building components within 3D models. Prior to training, the data underwent preprocessing to ensure compatibility and readiness. This phase focused on formatting the 3D building data to suit the chosen approach, ensuring consistency and uniformity across the dataset. Subsequently, a subset comprising 100 residential building models was divided into training and validation sets, maintaining an 80:20 ratio. Furthermore, the model underwent a comprehensive training process using the prepared dataset. This iterative process involved presenting the dataset to the model, enabling it to recognize patterns and relationships among various building elements. Default hyperparameters were employed in this phase and validation techniques were utilized to assess the model's performance and verify its adaptability to the new data.

3.4 Translation of Semantic Label to CityGML

This study not only focuses on semantic labeling of building components within 3D models but also involves translating these labels into a GIS-compatible format. The output from the BuildingGNN, responsible for semantic labeling, undergoes further processing and exported into the CityGML format. CityGML stands as a standardized data format specifically crafted for representing and exchanging 3D city models in a GIS environment.

The choice of CityGML as the export format is due to its extensive schema, purpose-built for encapsulating semantic information in 3D urban and architectural models. Its hierarchical structure and semantic depth offer an optimal framework for encoding detailed information about building components and their interrelationships. This facilitates interoperability and seamless integration within GIS systems. Figure 4 illustrates the CityGML class modules representing 3D city objects, which are thematically decomposed into a Core module and 16 extension modules (Kutzner et al., 2020).

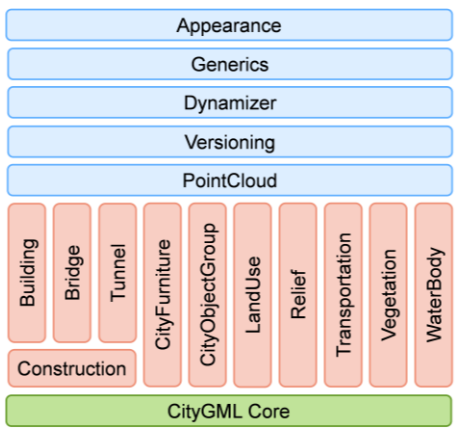


Figure 4. CityGML conceptual model.

The process of translating semantic labels into the CityGML format, as illustrated in Figure 5, involves several key steps that leverage FME software for efficient data processing. Initially, the geometry of the 3D models undergoes validation to ensure compliance with CityGML geometry standards. This validation process checks for geometric integrity, identifying missing overlaps, holes, or irregularities to ensure the creation of a seamless, watertight 3D model. Any identified issues are addressed and rectified to ensure the 3D models' suitability and reliability for integration into the CityGML framework.

https://doi.org/10.5194/isprs-archives-XLVIII-4-W9-2024-307-2024 | © Author(s) 2024. CC BY 4.0 License.

Following this, the 3D models undergo a georeferencing procedure aimed at aligning them with real-world geographic coordinates. This alignment process utilizes embedded reference or ground control points (GCPs) within the 3D models, matched to their actual geographical locations through geospatial transformation algorithms. This transformation ensures precise spatial positioning and orientation of the 3D models within the GIS environment. Subsequently, the related semantic labels are integrated and mapped to correspond with CityGML classes and attributes, ensuring coherence and alignment with CityGML's structured data model.

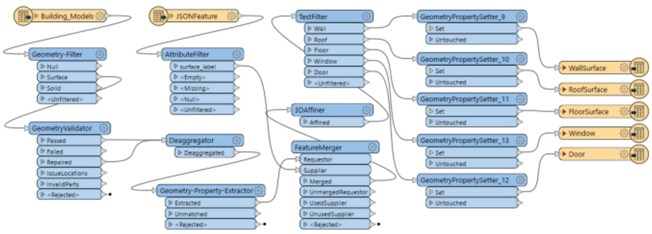


Figure 5. Workflow for translating semantic labels into CityGML format, depicting the sequential steps involving geometry validation, georeferencing, and semantic label integration.

4. RESULTS AND DISCUSSION

The training process employed for the BuildingGNN technique yielded acceptable results. Through iterative training using diverse datasets, the model showcased an improved ability to identify and label building components with a considerable level of accuracy. Table 1 illustrates an overview of the model's performance metrics across both datasets, allowing a comparative analysis of its efficacy within varied datasets. The results in Table 1 reveal interesting insights into the accuracy variation among different building components. For instance, the roof components achieved notably high accuracy, surpassing 92%. This might be attributed to the distinct and identifiable patterns associated with roof structures, facilitating their recognition by the model. Conversely, the model yielded comparatively lower accuracy, falling below 57%, in identifying door components. This lower value could come from several factors such as complexity and variability in door designs across different building types might present challenges for the model, impacting its ability to recognize and label doors accurately.

Furthermore, the utilization of diverse datasets, incorporating both randomly sampled and manually selected 3D building models, played a role in capturing the variability in annotation quality, building styles, and complexities. Consequently, the model developed an ample understanding of the building components, enhancing its capability to process and label several building elements within complex 3D models.

Class	Percentage (%)	
	Dataset 1	Dataset 2
Roof	92.7	94.3
Wall	80.3	83.2
Window	79.8	80.1
Door	54.2	56.4
Floor	74.2	81.2

Table 1. Evaluation metrics for each class.

Moreover, a collection of 3D building models is utilized in the labeling procedure, employing the trained GNN model to assign semantic classes. Subsequently, the outcomes produced through this labeling procedure undergo validation and refinement, encompassing both their semantic labeling and geometry, before their conversion into the CityGML format. This translation process aims to embed semantic information into GIS systems, enabling the integration of labeled 3D building components. Such integration could facilitate comprehensive spatial analysis and visualization capabilities. Figure 6 provides a visual representation of the final output displaying the CityGML building models, ensuring compliance with the CityGML schema classes. This schema contains information, including geometry type, level of detail (LOD), the number of parts, coordinate systems, and other essential attributes.

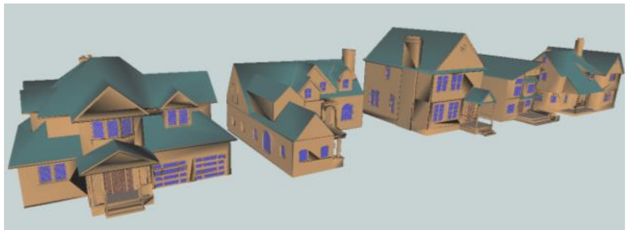


Figure 6. 3D models of CityGML with embedded semantic information.

The created CityGML 3D building models, which contain detailed information about the buildings' characteristics, can also be expanded to include terrain data which opens ways for a more comprehensive and enriched spatial representation. The presence of terrain data alongside 3D models enables a more realistic portrayal of cityscapes, facilitating a deeper understanding of the interaction between urban structures and the surrounding environment. However, challenges arise in aligning and processing these datasets, demanding precision in placing the 3D models on top of diverse terrain conditions and incorporating sub-surface components. Despite these challenges, the potential benefits, including enhanced urban analysis, and improved visualization, emphasize the benefit of exploring this integration in future research endeavors.

5. CONCLUSION

In summary, the iterative training of the BuildingGNN model exhibited acceptable outcomes, improving its accuracy in identifying building components within intricate 3D models. Notably, the utilization of diverse datasets allowed the model to comprehend various architectural styles and complexities, contributing to more precise labeling of architectural elements. Achieving better labeling may also depend on utilizing quality training datasets.

Furthermore, the output from the GNN segmentation process requires thorough validation before translation into GIS data to ensure accuracy and reliability. The direct conversion of the GNN output into CityGML format may not be straightforward due to various complexities. More focused effort is needed to process the GNN output into CityGML compliance. In several cases, the original input building models may exhibit geometry issues such as holes, missing faces, or overlapping structures, necessitating the use of additional algorithms to rectify these problems. Moreover, the semantic labels generated by the GNN might lack instance-level granularity, requiring further work to

https://doi.org/10.5194/isprs-archives-XLVIII-4-W9-2024-307-2024

identify and merge labels belonging to the same building instance and assign unique identifiers. Currently, manual intervention is necessary for these tasks. Establishing a robust workflow for seamless integration between GNN output and CityGML would greatly enhance the efficiency and accuracy of the process. Despite these challenges, the resulting CityGML dataset holds substantial potential for various 3D city model applications and spatial analysis within GIS systems.

Nevertheless, while this study showcases potential application, future research should prioritize refining the labeling technique and semantic translation processes to adhere to spatial data standards, aiming for an automated end-to-end labeling process. Addressing challenges related to handling uncommon and complex architectural details could further improve the applicability of semantic labeling. It's important to note that the trained model may have limited applicability for different types of building structures and styles if not appropriately trained and fine-tuned in the given direction.

REFERENCES

Alexander Toshev and Ben Taskar. Detecting and parsing architecture at city scale from range data. In Proc. CVPR, 2010.

Alexander Toshev and Ben Taskar. 3d all the way: Semantic segmentation of urban scenes from start to end in 3d. In Proc. CVPR, 2015.

Chen, X., Golovinskiy, A., & Funkhouser, T. (2009). A benchmark for 3D mesh segmentation. Acm transactions on graphics (tog), 28(3), 1-12.

Demir, I., Aliaga, D. G., & Benes, B. (2015a). Coupled segmentation and similarity detection for architectural models. ACM Transactions on Graphics (TOG), 34(4), 1-11.

Demir, I., Aliaga, D. G., & Benes, B. (2015b). Procedural editing of 3d building point clouds. In Proceedings of the IEEE International Conference on Computer Vision (pp. 2147-2155).

Kalogerakis, E., Hertzmann, A., & Singh, K. (2010). Learning 3D mesh segmentation and labeling. In ACM SIGGRAPH 2010 papers (pp. 1-12).

Khan, S. A., Shi, Y., Shahzad, M., & Zhu, X. X. (2020). FGCN: Deep feature-based graph convolutional network for semantic segmentation of urban 3D point clouds. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 198-199).

Kobyshev, N., Riemenschneider, H., Bodis-Szomoru, A., & Van Gool, L. (2016). Architectural decomposition for 3D landmark building understanding. In 2016 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 1-10). IEEE.

Kundu, A., Yin, X., Fathi, A., Ross, D., Brewington, B., Funkhouser, T., & Pantofaru, C. (2020). Virtual multi-view fusion for 3d semantic segmentation. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIV 16 (pp. 518-535). Springer International Publishing.

Kutzner, T., Chaturvedi, K., & Kolbe, T. H. (2020). CityGML 3.0: New functions open up new applications. PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 88(1), 43-61.

Lun, Z., Kalogerakis, E., & Sheffer, A. (2015). Elements of style: learning perceptual shape style similarity. ACM Transactions on graphics (TOG), 34(4), 1-14.

Qi, C. R., Yi, L., Su, H., & Guibas, L. J. (2017). Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Advances in neural information processing systems, 30.

Rook, M., Biljecki, F., & Diakité, A. A. (2016). Towards automatic semantic labelling of 3D city models. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, 4, 23.

Selvaraju, P., Nabail, M., Loizou, M., Maslioukova, M., Averkiou, M., Andreou, A., ... & Kalogerakis, E. (2021). BuildingNet: Learning to label 3D buildings. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 10397-10407).

Stoter, J. E., Arroyo Ohori, G. A. K., Dukai, B., Labetski, A., Kavisha, K., Vitalis, S., & Ledoux, H. (2020). State of the Art in 3D City Modelling: Six Challenges Facing 3D Data as a Platform. GIM International: the worldwide magazine for geomatics, 34.

Uceda-Sosa, R., Srivastava, B., & Schloss, R. J. (2011). Building a highly consumable semantic model for smarter cities. In Proceedings of the AI for an Intelligent Planet (pp. 1-8).

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2020). A comprehensive survey on graph neural networks. IEEE Transactions on Neural Networks and Learning Systems, 32(1), 4-24.