PolyRoof: Precision Roof Polygonization in Urban Residential Building with Graph Neural Networks

1st Chaikal Amrullah 2nd Daniel Panangian 3rd Ksenia Bittner

Remote Sensing Technology Institute (IMF)
German Aerospace Center (DLR)
Oberpfaffenhofen, Germany
chaikal.amrullah@dlr.de daniel.panangian@dlr.de ksenia.bittner@dlr.de

***Abstract*—The growing demand for detailed building roof data has driven the development of automated extraction methods to overcome the inefficiencies of traditional approaches, particularly in handling complex variations in building geometries. Re:PolyWorld, which integrates point detection with graph neural networks, presents a promising solution for reconstructing high-detail building roof vector data. This study enhances Re:PolyWorld's performance on complex urban residential structures by incorporating attention-based backbones and additional area segmentation loss. Despite dataset limitations, our experiments demonstrated improvements in point position accuracy (1.33 pixels) and line distance accuracy (14.39 pixels), along with a notable increase in the reconstruction score to 91.99%. These findings highlight the potential of advanced neural network architectures in addressing the challenges of complex urban residential geometries.**

***Index Terms*—Urban Residential, Building Geometry Complexity, Polygonal Roof Extraction, Graph Neural Networks, Aerial Imagery.**

I. Introduction

The demand for high-Level of Detail (LoD) building data has grown significantly, driven by the need to accurately capture complex geometric features [6]. This data is essential for a wide range of applications, including urban development planning, architectural design, construction, and infrastructure monitoring [3], [12]. However, producing detailed and high-quality building data remains a significant challenge. Conventional methods, such as stereo-plotting from aerial imagery, are inefficient, labor-intensive, and time-consuming [1].

In recent years, automatic building data extraction methods, particularly for roof structures, have gained increasing attention across various fields [3], [12], [1], [8], [13], [14]. Extracting roof structures involves capturing fine-grained variations in size, shape, and spatial distribution [8], alongside complex features such as edges, corners, inscriptions, and roof components. A key challenge in this process lies not only in managing the diversity of object textures but also in addressing the inherent complexity of building geometries. Other challenges also arise when discussing the complexity variations that also change based on the type of construction activity, be it commercial, industrial, health or residential activities (see Figure 1). These geometric challenges include irregular shapes, varying roof angles, and overlapping components [7], all of which require precise modeling to achieve accurate and reliable reconstruction.

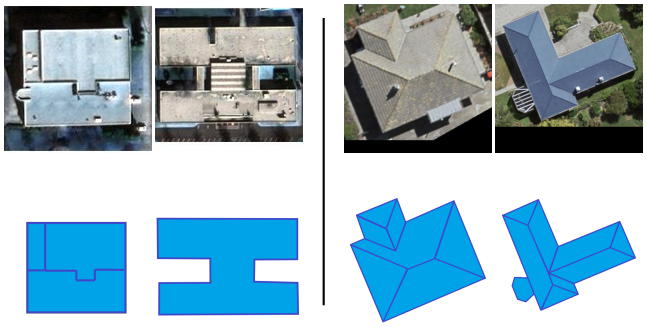

Fig. 1. Difference of industrial (column one and two) and urban residential (column three and four) building complexity

High-LoD building data has been pursued through various methods, with texture-based approaches such as semantic segmentation using Fully Convolutional Networks (FCNs) and U-shape Neural Networks (U-Nets) architectures [11] improving boundary delineation and adjacent structure separation. Instance-level segmentation has advanced extraction capabilities but struggles in scenarios with densely populated areas, complex junctions, or internal courtyards [5], [12]. As high-LoD building data is often represented as vector polygons, methods targeting the geometric complexity of these representations have gained prominence. Approaches like frame field-guided polygonization and Graph Neural Networks (GNNs) for modeling roof-lines [13] have proven effective. Notably, GNN-based frameworks [14] excel in capturing spatial relationships and enabling detailed building reconstructions tailored to Geographic Information System (GIS) applications.

In this study, we extend Re:PolyWorld [14], which combines point detection with graph neural networks to address both the texture- and geometry-based complexities of building roof data extraction. By automatically detecting object vertices and solving an optimal transport problem, this model generates accurate polygonal representations of roof structures.

We propose PolyRoof, by implementing model architec-

ture improvement, data augmentation and tailored additional area segmentation loss to improve building reconstruction using datasets of urban residential complexities [10]. A major challenge is addressing the diverse geometric characteristics of different scenes, from simpler, repetitive industrial structures [4] to more complex and relatively irregular urban residential with densely packed roofs as shown in Figure 1. PolyRoof aims to tackle this challenge, advancing automated building data extraction for detail-oriented applications across varied urban landscapes.

II. Dataset

For this study, we used the RoofOpt dataset [10] for its Very High Resolution (VHR) RGB aerial imagery and detailed annotations of residential roofs with complex geometries, including non-convex shapes and irregular angles. These challenges make it ideal for evaluating polygonization models. In contrast, the original Re:PolyWorld utilized the HEAT dataset [4], which includes industrial roof segment data called outdoor and floor plan data, both annotated with vector representations of edges while outdoor has RGB aerial image and floor plan has density image. Re:PolyWorld demonstrated strong performance on both HEAT components.

Given the varying geometric complexities across different scenes, we compared our urban residential RoofOpt dataset with the HEAT dataset to better understand these differences. Key geometric attributes—number of vertices, point degree, convexity, compactness, and the second component of Principal Component Analysis (PCA) (referred to here as the PCA score)—were analyzed to evaluate their impact on model performance. The number of vertices indicates polygon intricacy, while point degree reflects structural connectivity. Convexity and compactness measure shape regularity and circularity, offering insights into variability. The PCA score consolidates these attributes into a single orthogonal metric, highlighting geometric diversity across datasets.

This analysis aims to statistically quantify the differences in geometric complexity between the datasets used in this study and those used previously summarized in Table I.

TABLE I
Average Building Geometry Complexity.

Geometry Property	Num. Vertices ↑	Point Degree ↑	Convexity ↑	Compactness ↓	PCA Score ↑
RoofOpt [10]	**36.32**	6.09	**90.26**	**11.56**	**42.83**
Outdoor [4]	12.55	5.42	87.06	61.07	29.21
Floorplan [4]	21.95	**6.97**	86.17	65.29	38.80

The dataset analysis reveals significant geometric differences: RoofOpt, with the highest mean vertex count (36.32) and mean convexity (90.26) but lowest compactness (11.83), reflects complex roof geometries in relatively sparse area. In contrast, Outdoor data, with fewer vertices and higher compactness, represents simpler structures with low spread. PCA scores highlight RoofOpt's complexity (42.83), followed by Floorplan (38.80) and Outdoor (29.21), emphasizing the challenge of handling high-detail datasets like RoofOpt.

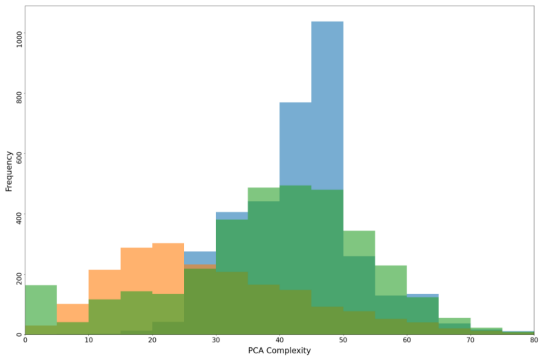

Fig. 2. Dataset histogram based on PCA complexity score of geometry complexity. RoofOpt [10] dataset with blue while HEAT [4] dataset for outdoor is in orange and floorplan with green.

Figure 2 illustrates PCA score distributions. RoofOpt's histogram shows a medium-high concentration with high kurtosis, indicating uneven data distribution. In contrast, HEAT exhibits a near-normal distribution with presence of skewness, suggesting better uniformity.

Using PCA complexity scores, a balanced split into training, validation, and test sets ensured representation of all complexity levels. This strategy enhances model generalization and performance across varying geometric complexities.

III. Methodology

From the dataset's geometric complexity analysis, we hypothesize that improving model enhancement strategies and evaluation metrics is crucial for urban residential building reconstruction. Enhancements include attention mechanisms and data augmentation for point detection, along with an additional loss function for matching-optimization, evaluated using metrics tailored to capture high-LoD intricacies.

A. Model Enhancement Strategies

1) How data augmentation techniques and backbone architecture variations influence the generalization ability and precision of the PolyRoof Model?: To enhance generalization in roof extraction models and address dataset size, variation, and distribution limitations, we employed extensive augmentation strategies, including rotations, flips, and adjustments to brightness, contrast, and sharpness, simulating real-world conditions. Additionally, we compared backbone architectures—Residual Recurrent U-Net (R2U-Net) [2] and Residual Recurrent Attention U-Net (R2AU-Net) [9]—to assess the impact of attention mechanisms in refining spatial precision. This analysis aimed to identify the optimal architecture balancing efficiency and segmentation accuracy.

2) How additional loss functions enhance the model's performance in detail point detection and segmentation tasks?: Optimizing the learning process with advanced loss functions is crucial for guiding and enhancing model performance. Traditional loss functions—such as vertex detection and matching,

also angle and segmentation by option, losses—effectively evaluate corner point accuracy and relationships but struggle to address precise segmentation and the irregularities of complex roof geometries.

To address these gaps, we propose an additional loss functions: *Area Segmentation Loss* ($\mathcal{L}_{segm}$). This value complements this by evaluating the F1-Score across three dimensions: building instances, roof segment instances, and a reconstruction index, which reflects the harmonic mean ratio of the building instances and the roof segment instances. This combined approach enhances both point accuracy and segmentation consistency, fostering improvements in the detailed reconstruction of building and roof structures.

$$\mathcal{L}_{segm} = \text{F1}_{\text{building}} + \text{F1}_{\text{roof segment}} + \text{F1}_{\text{reconstruction}} \quad (1)$$

B. Evaluation Metrics

Recognizing the limitations of Average Precision (AP) and Average Recall (AR) scores in capturing the complexities of high-resolution data, we introduced alternative metrics to evaluate geometric primitives at multiple levels: Point Position Accuracy, Line Distance Accuracy, Building Instance F1-Score, Roof Segment F1-Score, and Reconstruction Score. These metrics, all within 50% Intersection over Union (IoU) threshold framework, provide a holistic assessment of prediction quality, aligning with the model's additional loss functions to evaluate accuracy, structural fidelity, and segmentation quality.

a) Point Position Accuracy: This metric computes the RMSE between predicted and ground truth roof corner points, providing a measure of positional precision in pixel units.

b) Line Distance Accuracy: By averaging the Hausdorff and Fréchet distances, this metric captures both maximum deviation and sequential alignment, offering a comprehensive view of structural similarity.

c) Building Instance and Roof Segment F1-Scores: These metrics evaluate segmentation quality based on IoU, focusing on both the overall building shape (Building Instance F1) and individual roof segments (Roof Segment F1).

d) Reconstruction Score: The harmonic mean of the Building Instance and Roof Segment F1-Scores, this score balances segmentation performance between full building shapes and their roof segments.

IV. Result and Discussion

Table II presents the quantitative evaluation of each experimental configuration using the defined metrics, with Re:PolyWorld serving as the Baseline for improvement variations. Figure 3 illustrates qualitative prediction samples, anticipate a challenge linked to dataset geometry complexity in achieving high accuracy. Roof segment color indicates segmentation ID.

The Baseline model showed balanced performance across our evaluation metrics, but its results were lower compared to those reported in the original paper [14], which used a dataset with lower geometric complexity [4] (Figure 2). The original paper's The $F1\text{-}Score_{50}$, was 91.89%, while the Building F1-Score in this paper was 89.38%. Although Roof Segment performance remained similar, the model produced incomplete building reconstructions, leading to a lower Reconstruction Score of 88.57%. This highlights the difficulty of reconstructing entire buildings with high accuracy on more complex datasets.

TABLE II
Results of different experiment configurations

Configuration	Point Pos. Acc. Pixel(s) ↓	Line Dist. Acc. Pixel(s) ↓	Building F1-Score (%) ↑	Roof F1-Score (%) ↑	Recon Score (%) ↑
Baseline (B)	1.34	18.31	89.38	**89.68**	88.57
B +Att	1.26	17.01	88.36	89.06	87.75
B +Aug	1.34	20.72	62.28	88.39	72.11
B +$\mathcal{L}_{\text{segm}}$	1.35	16.79	90.05	89.32	88.72
B +Att +$\mathcal{L}_{\text{segm}}$	**1.33**	**14.39**	**96.61**	89.55	**91.99**

Adding attention mechanisms (+Att) improved point position accuracy, initially producing the best RMSE of 1.26 pixels. In the same time, incorporating $\mathcal{L}_{segm}$ improved Line Distance Accuracy to 16.79 pixels. When both techniques were combined, the result improved to 14.39 pixels. However, dataset augmentation led to a decrease in overall performance, suggesting caution when applying augmentation in point detection pipelines. On the other hand, $\mathcal{L}_{segm}$ showed the best Building Segmentation Score (90.05%) and competitive Roof F1-Score (89.32%), which directly impacted the Reconstruction Score (88.72%).

When combining +Att with $\mathcal{L}_{segm}$, we observed clear improvements across metrics, with Point Position Accuracy improving to 1.33 pixels, Line Distance Accuracy to 14.39 pixels, Building Instance F1-Score to 96.61%, and the Reconstruction Score rising to 91.99%. These were the best results overall, except for Roof F1-Score (89.55%), which ranked second compared to the Baseline with fraction of gap. The combined configuration clearly boosted position accuracy and segmentation precision, resulting in a better overall score.

Qualitatively, while the combined approach outperformed the Baseline in general segmentation precision score, some roof segment edges were missing, leading to merged segments and duplicated roof segment IDs indicating completeness issues. Additionally, smaller or less common roof components were occasionally missed, indicating some consistency issues. These challenges highlight areas that require further refinement.

V. Conclusion

The complexity of the dataset, with intricate roof structures and diverse shapes, significantly impacts model performance. While the model demonstrates potential for high Level of Detail (LoD) building predictions, challenges remain in handling complex geometries.

Quantitatively, the combination of attention mechanisms and area segmentation loss achieved a notable reconstruction score

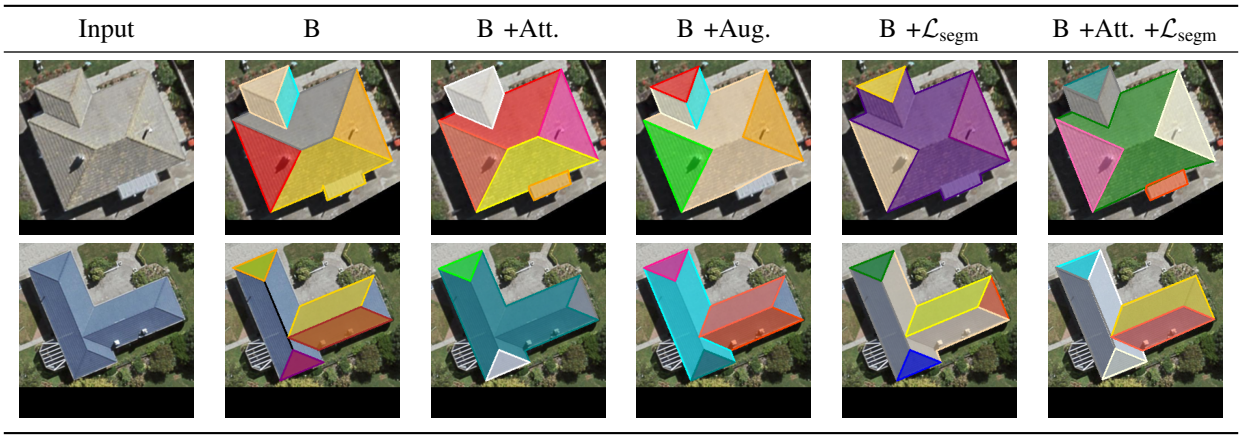

Fig. 3. Visual comparison of RGB with various experiment scenario. All examples are taken from the test set.

of 91.99%. Qualitatively, the model reconstructs buildings with detail and accuracy, though inconsistencies persist. These results underscore the need for further refinement, considering geometric complexity to enhance generalization and reliability in high LoD building model reconstruction.

ACKNOWLEDGMENT

Chaikal Amrullah is currently funded by a DLR-DAAD Research Fellowship (No. 57681552) to pursue his PhD studies.

REFERENCES

[1] Antonio Almagro. Simple methods of photogrammetry: Easy and fast, 2002.

[2] Md Zahangir Alom, Mahmudul Hasan, Chris Yakopcic, Tarek M. Taha, and Vijayan K. Asari. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation, 2018.

[3] Ksenia Bittner, Fathalrahman Adam, Shiyong Cui, Marco Körner, and Peter Reinartz. Building footprint extraction from vhr remote sensing images combined with normalized dsms using fused fully convolutional networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(8):2615–2629, 2018.

[4] Jiacheng Chen, Yiming Qian, and Yasutaka Furukawa. Heat: Holistic edge attention transformer for structured reconstruction, 2022.

[5] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn, 2018.

[6] Tatjana Kutzner, Kanishk Chaturvedi, and Thomas H Kolbe. Citygml 3.0: New functions open up new applications. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 88(1):43–61, 2020.

[7] Evangelos Pantazis and David Jason Gerber. Beyond geometric complexity: a critical review of complexity theory and how it relates to architecture engineering and construction. *Architectural science review*, 62(5):371–388, 2019.

[8] Zhen Qian, Min Chen, Teng Zhong, Fan Zhang, Rui Zhu, Zhixin Zhang, Kai Zhang, Zhuo Sun, and Guonian Lü. Deep roof refiner: A detail-oriented deep learning network for refined delineation of roof structure lines using satellite imagery. *International Journal of Applied Earth Observation and Geoinformation*, 107:102680, 2022.

[9] Zuo Qiang, Chen Songyu, and Wang Zhifang. R2au-net: Attention recurrent residual convolutional neural network for multimodal medical image segmentation. *Security and Communication Networks*, 2021:1–10, 06 2021.

[10] Jing Ren, Biao Zhang, Bojian Wu, Jianqiang Huang, Lubin Fan, Maks Ovsjanikov, and Peter Wonka. Intuitive and efficient roof modeling for reconstruction and synthesis, 2021.

[11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.

[12] Philipp Schuegraf and Ksenia Bittner. Automatic building footprint extraction from multi-resolution remote sensing images using a hybrid fcn. *ISPRS International Journal of Geo-Information*, 8(4):191, 2019.

[13] Wufan Zhao, Claudio Persello, and Alfred Stein. Extracting planar roof structures from very high resolution images using graph neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 187:34–45, 2022.

[14] Stefano Zorzi and Friedrich Fraundorfer. Re: Polyworld-a graph neural network for polygonal scene parsing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16762–16771, 2023.