Edges Compete for Trust: Group Relative Edge Optimization for Building Reconstruction from Point Clouds

Supplementary Material

A. Overview

The supplementary material is organized into the following sections:

- **Sec. B.** Training loss formulation and ablation studies on loss weights.
- **Sec. C.** User study evaluating perceived reconstruction quality and usability.
- **Sec. D.** Qualitative comparisons with baselines.
- **Sec. E.** Comparison with mesh reconstruction methods.
- **Sec. F.** City-scale reconstruction on the entire Tallinn dataset.

B. Training Loss

We propose Group Relative Edge Optimization (GREO), a plug-and-play training strategy that seamlessly integrates into edge-based wireframe reconstruction methods. During training, the overall optimization objective combines the baseline losses with our proposed GREO loss, ensuring consistent training dynamics and fair comparison.

For PBWR [3], the complete training objective is formulated as:

$$\begin{aligned}\mathcal{L} = \lambda_{\text{mid}}\mathcal{L}_{\text{mid}} + \lambda_{\text{comp}}\mathcal{L}_{\text{comp}} + \lambda_{\text{con}}\mathcal{L}_{\text{con}} \\ + \lambda_{\text{quad}}\mathcal{L}_{\text{quad}} + \lambda_{\text{sim}}\mathcal{L}_{\text{sim}} + \lambda_{\text{greo}}\mathcal{L}_{\text{GREO}},\end{aligned} \tag{1}$$

where $\mathcal{L}_{\text{mid}}$ employs L1 distance to supervise edge midpoint positions, $\mathcal{L}_{\text{comp}}$ regresses the component lengths along the x-, y-, and z-axes, $\mathcal{L}_{\text{con}}$ optimizes confidence scores through cross-entropy with edge-similarity-based soft labels to alleviate class imbalance, $\mathcal{L}_{\text{quad}}$ applies cross-entropy loss for quadrant classification to determine edge orientation, and $\mathcal{L}_{\text{sim}}$ directly computes the average edge-similarity scores across positive predictions to enforce geometric consistency. Following PBWR [3], the weights are set to $\lambda_{\text{mid}} = 5.0$, $\lambda_{\text{comp}} = 1.0$, $\lambda_{\text{con}} = 1.0$, $\lambda_{\text{quad}} = 2.0$, $\lambda_{\text{sim}} = 2.0$, and $\lambda_{\text{greo}} = 0.1$.

For EdgeDiff [4], the training objective is extended with an additional attention alignment term:

$$\begin{aligned}\mathcal{L} = \lambda_{\text{mid}}\mathcal{L}_{\text{mid}} + \lambda_{\text{comp}}\mathcal{L}_{\text{comp}} + \lambda_{\text{con}}\mathcal{L}_{\text{con}} \\ + \lambda_{\text{quad}}\mathcal{L}_{\text{quad}} + \lambda_{\text{sim}}\mathcal{L}_{\text{sim}} + \lambda_{\text{attn}}\mathcal{L}_{\text{attn}} \\ + \lambda_{\text{greo}}\mathcal{L}_{\text{GREO}},\end{aligned} \tag{2}$$

where $\mathcal{L}_{\text{attn}}$ supervises the edge attention during iterative refinement, with $\lambda_{\text{attn}} = 5.0$. All other loss components and their weights remain identical to PBWR [3].

The GREO loss is decomposed as:

$$\mathcal{L}_{\text{GREO}} = \mathcal{L}_{\text{CE}} + \mathcal{L}_{\text{entropy}}, \tag{3}$$

where the cross-entropy loss $\mathcal{L}_{\text{CE}} = -\sum_i q_i \log \pi_i$ aligns predicted confidence π_i with target distribution q_i derived from group-normalized rewards, and the entropy regularization $\mathcal{L}_{\text{entropy}} = \eta \sum_i \pi_i \log \pi_i$ with $\eta = 0.01$ prevents over-concentration and stabilizes optimization.

Effect of the loss weight λ_{greo}. We analyze different loss weights to understand how GREO influences the training process (see Tab. 1). A very small weight such as 0.01 produces almost no meaningful reward signal, so the performance stays close to the baseline. A moderate value like 0.05 brings clear gains, as the reward begins to adjust confidence predictions while keeping optimization stable. The best results occur at 0.10, where the reward strength is well balanced with the regression loss. A large weight such as 1.00 reduces accuracy because the reward dominates the gradients and introduces unstable confidence updates. These results show that GREO improves optimization in a wide range, and only extreme weights disrupt the balance between reward shaping and regression loss.

λ_{greo}	WED↓	CR↑	CF1↑	ER↑	EF1↑
0.00	0.255	0.74	0.84	0.70	0.79
0.01	0.252	0.74	0.84	0.71	0.79
0.05	0.235	0.75	0.85	0.73	0.81
0.10	**0.225**	**0.76**	**0.86**	**0.75**	**0.82**
1.00	0.245	0.74	0.84	0.72	0.80

Table 1. **Ablation study on the GREO loss weight λ_{greo}.** Results are based on the EdgeDiff [4]. Default settings of GREO are marked in gray, and best results are shown in **bold**.

Effect of the entropy weight η. We evaluate different values of η to study how entropy regularization influences training stability (see Tab. 2). Without entropy regularization at $\eta = 0.00$, the model relies only on reward shaping and already improves over the baseline. A small weight such as 0.01 delivers the best results because it smooths confidence updates while preserving clear distinctions among edges. Increasing the weight to 0.05 or 0.10 keeps the model stable, but the confidence distribution becomes slightly over-smoothed, which leads to minor drops in recall and F1. A large value such as 1.00 further reduces performance because the regularization weakens useful re-

η	WED↓	CR↑	CF1↑	ER↑	EF1↑
0.00	0.242	0.75	0.85	0.73	0.81
0.01	**0.225**	**0.76**	**0.86**	**0.75**	**0.82**
0.05	0.232	0.75	0.85	0.74	0.81
0.10	0.236	0.75	0.85	0.73	0.80
1.00	0.245	0.74	0.84	0.72	0.79

Table 2. **Ablation study on the entropy regularization weight η.** Results are based on the EdgeDiff [4]. Default settings of GREO are marked in gray, and best results are shown in **bold**.

ward signals. Overall, GREO remains effective across reasonable settings, and entropy regularization improves reliability without requiring precise hyperparameter tuning.

C. User Study

Standard metrics evaluate geometric accuracy but struggle to capture subjective qualities such as visual coherence and practical usability. Since building wireframes serve as foundational assets for digital-twin applications, we conduct a user study to assess perceived reconstruction quality. We recruit 10 participants, including experts in computer graphics, photogrammetry, architecture, and Geographic Information System (GIS), alongside non-experts for diversity.

Each participant views 20 randomly selected buildings from the Tallinn test set. For each building, we present the input point cloud and four wireframes: PBWR [3], PBWR+GREO, EdgeDiff [4], and EdgeDiff+GREO. All wireframes are displayed without labels, using identical colors and random presentation order to eliminate bias. Participants adopt the perspective of digital-twin designers and answer: *Which reconstruction would you directly reuse for visualization, or which requires minimal editing?* They also evaluate whether the shape aligns with realistic building structures. This yields pairwise comparisons within each baseline method. To reduce sampling variance, we repeat the process on a second independent set of 20 buildings, collecting 200 pairwise choices per baseline.

As shown in Tab. 3, GREO is strongly favored across both baselines. For PBWR, participants prefer PBWR+GREO in 82.0% and 81.0% of cases. For EdgeDiff, EdgeDiff+GREO is chosen in 86.5% and 84.0% of cases. These results confirm that GREO not only improves quantitative metrics but also enhances structural completeness and geometric fidelity. Users consistently perceive GREO-enhanced reconstructions as more directly deployable in digital-twin scenarios. Notably, GREO achieves these improvements as a plug-and-play training strategy, introducing no inference overhead while delivering substantial gains in both objective accuracy and subjective usability.

Method	Module	User Study↑ Sample 1	Sample 2
PBWR [3]	-	18.0%	19.0%
	+ GREO	**82.0%**	**81.0%**
EdgeDiff [4]	-	13.5%	16.0%
	+ GREO	**86.5%**	**84.0%**

Table 3. **User study results.** Percentages indicate the proportion of participants who prefer each method for direct reuse or minimal editing in digital-twin applications. Sample 1 and Sample 2 represent two independent sets of 20 buildings each, with 200 pairwise comparisons per baseline. GREO (marked in gray) is consistently favored across both baselines and both sample sets.

D. Qualitative Comparison with Baselines

Fig. 1 presents visual comparisons on the Building3D dataset [5]. PBWR [3] and EdgeDiff [4] frequently produce fragmented or misplaced edges, especially around complex roof ridges and junctions. These errors stem from sparse Hungarian matching, which provides insufficient gradients for unmatched proposals, causing unstable confidence calibration. GREO substantially improves structural consistency across both baselines. PBWR+GREO reduces missing ridges and distorted polygons, yielding cleaner roof outlines. EdgeDiff+GREO enhances fine-grained structures, suppressing spurious edges while preserving consistent topology. GREO consistently produces geometrically plausible configurations without altering network architecture or inference cost, demonstrating the efficacy of dense optimization.

E. Qualitative Comparison with Mesh Reconstruction Methods

Fig. 2 compares EdgeDiff+GREO with mesh reconstruction methods, including City3D [2], 2.5D Dual Contouring [6], and KSR [1]. Our method focuses on high-precision wireframe reconstruction. The reconstructed wireframes can be converted into mesh structures through coplanarity analysis. Traditional mesh methods face a trade-off between redundancy and fidelity. Dense meshes preserve details but introduce excessive faces and storage overhead. Sparse meshes reduce complexity but sacrifice fine-grained structures. In contrast, wireframe representations encode geometry with minimal primitives, offering a compact yet expressive abstraction. EdgeDiff+GREO maintains structural fidelity while achieving superior storage efficiency, making it particularly suitable for large-scale digital-twin applications where storage costs are critical.

F. Large-Scale Urban Reconstruction

Fig. 3 demonstrates the scalability of our approach on the entire Tallinn dataset, encompassing 36,084 buildings. The top row presents the complete city-scale reconstruction, showcasing the capability of EdgeDiff+GREO to handle large-scale urban scenarios. The bottom row zooms into a 2 km $\times$ 2 km region, revealing the density and diversity of reconstructed meshes across diverse architectural structures. The bottom row further magnifies individual buildings, displaying the input point clouds alongside their corresponding lightweight mesh models. These results demonstrate that our wireframe-based approach maintains high geometric fidelity while producing compact representations. The reconstructed models are directly suitable for downstream digital-twin applications, where storage efficiency and structural interpretability are critical requirements.

References

[1] Jean-Philippe Bauchet and Florent Lafarge. Kinetic shape reconstruction. *ACM Transactions on Graphics (TOG)*, 39(5): 1–14, 2020. 2

[2] Jin Huang, Jantien Stoter, Ravi Peters, and Liangliang Nan. City3d: Large-scale building reconstruction from airborne lidar point clouds. *Remote Sensing*, 14(9):2254, 2022. 2

[3] Shangfeng Huang, Ruisheng Wang, Bo Guo, and Hongxin Yang. Pbwr: Parametric-building-wireframe reconstruction from aerial lidar point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27778–27787, 2024. 1, 2

[4] Yujun Liu, Ruisheng Wang, Shangfeng Huang, and Guorong Cai. Edgediff: Edge-aware diffusion network for building reconstruction from point clouds. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 17008–17018, 2025. 1, 2

[5] Ruisheng Wang, Shangfeng Huang, and Hongxin Yang. Building3d: A urban-scale dataset and benchmarks for learning roof structures from point clouds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 20076–20086, 2023. 2, 4

[6] Qian-Yi Zhou and Ulrich Neumann. 5 d dual contouring: A robust approach to creating building models from aerial lidar point clouds. In *European conference on computer vision*, pages 115–128. Springer, 2010. 2

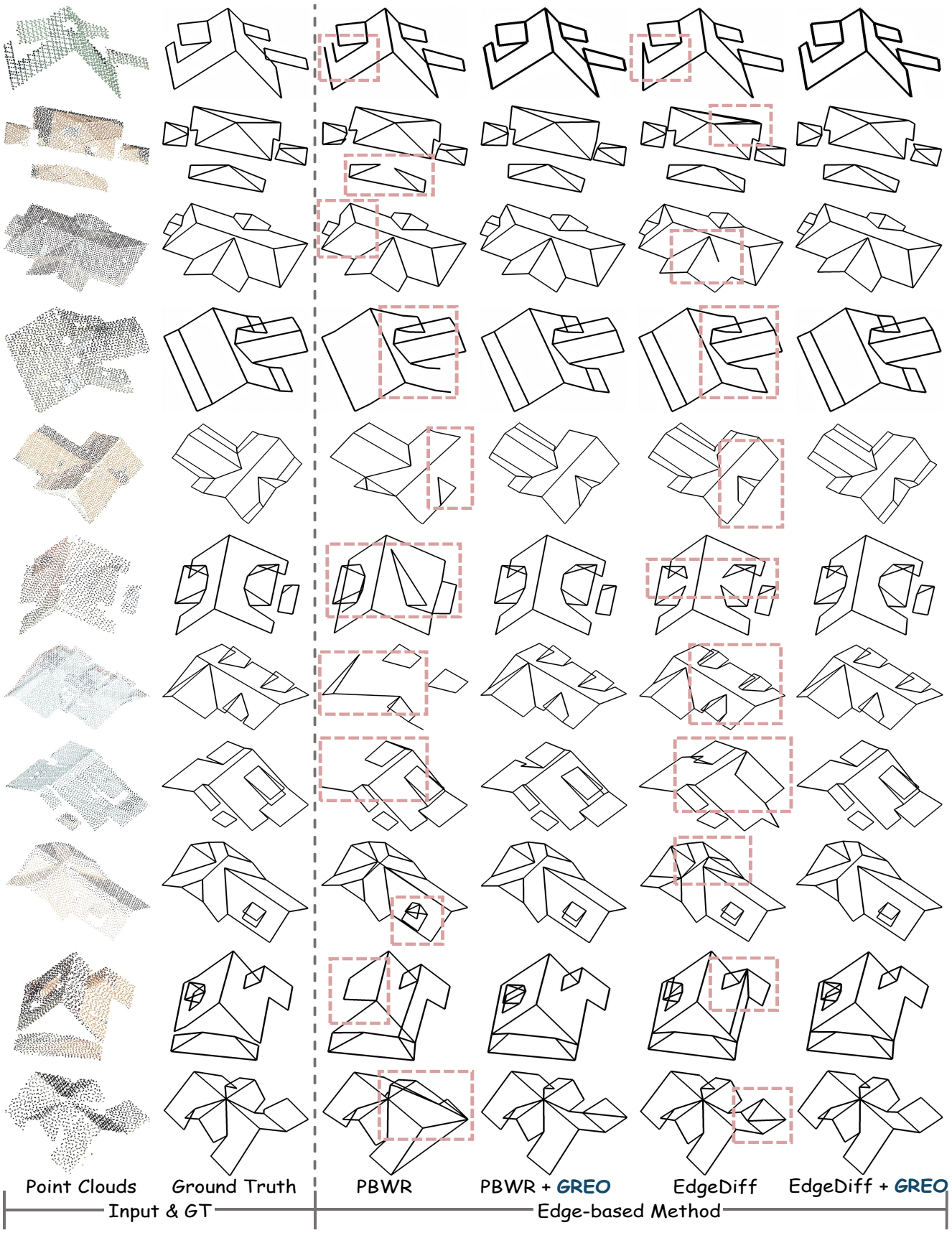

Figure 1. Qualitative comparison on the Building3D dataset [5].

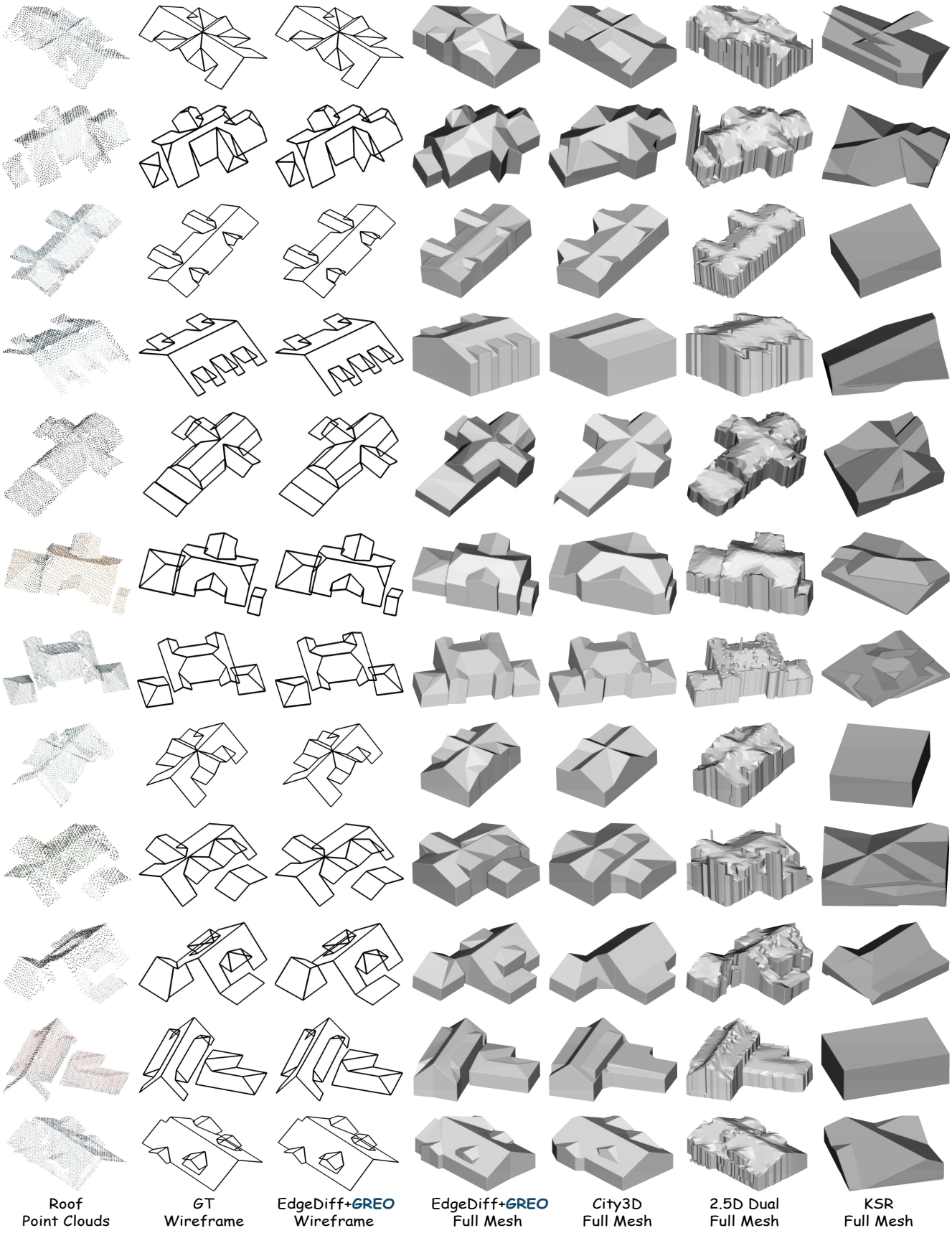

Figure 2. Comparison with mesh reconstruction methods.

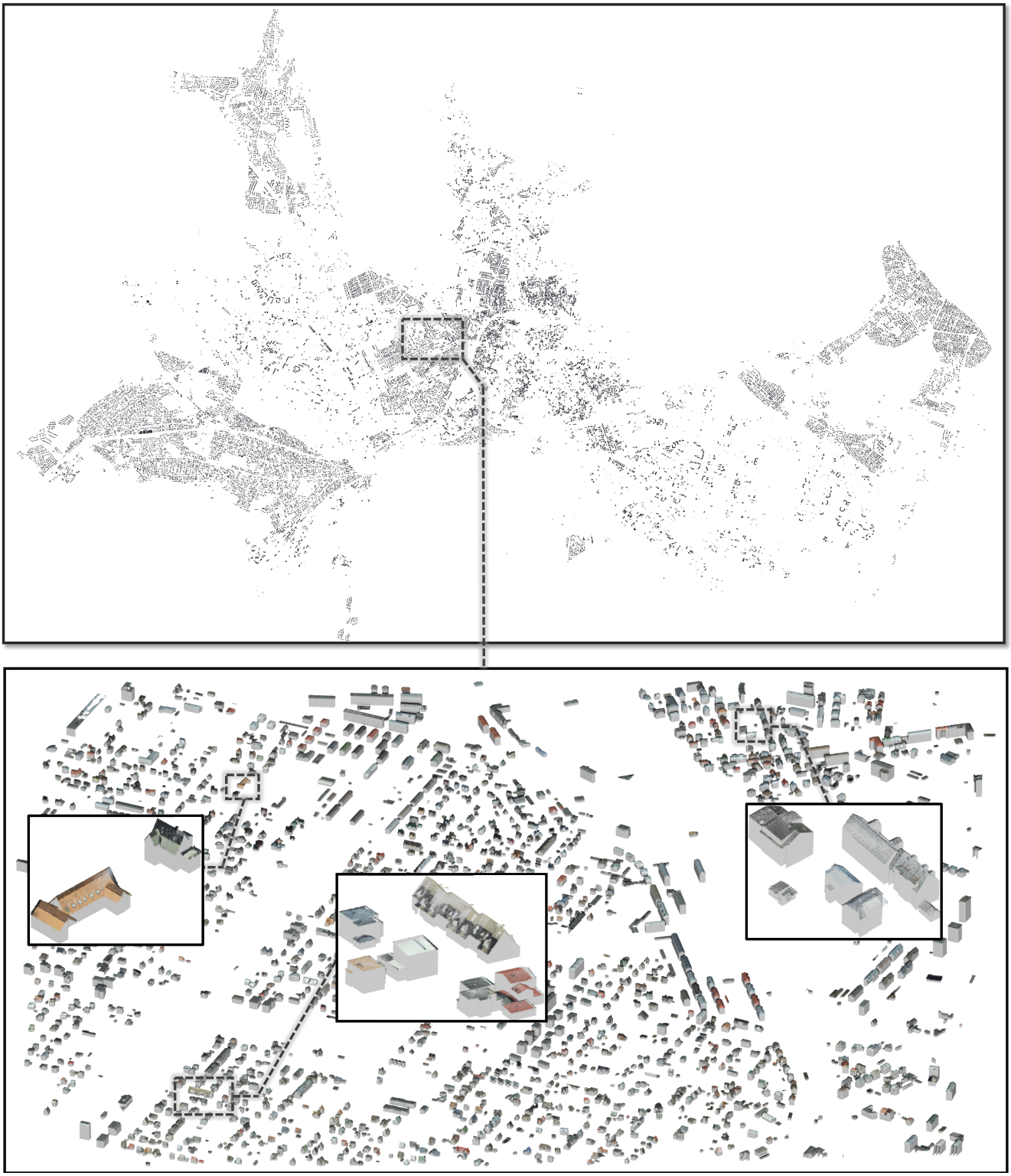

Figure 3. Large-scale reconstruction of Tallinn city.