BWFormer: Building Wireframe Reconstruction from Airborne LiDAR Point Cloud with Transformer

Supplementary Material

In the supplementary material, we first provide some details about the model training and inference time. Then, we provide an additional ablation study on the setting of 3D corner offset prediction. Finally, we provide some more qualitative comparison results to demonstrate the effectiveness of our model design, including the effects of different model components, more 3D wireframe reconstruction results, and more LiDAR scanning simulation results.

1. Time Efficiency

During training, we utilized 8×A800 GPUs for approximately 1.5 days on the original Building3D dataset [5]. After augmenting the dataset with synthetic data, tripled relative to the real dataset's size, the training duration increased to around 6 days. For inference, the average prediction time per building is about 0.2 seconds.

2. Additional Ablation Study

During 2D-to-3D corner detection, the first two dimensions of 3D corner positions (x and y) are predicted in the 2D corner detection procedure and frozen in the 3D corner detection procedure. As a result, only the height dimension offset is predicted during 3D corner refinement. That is because the dense pixel-by-pixel 2D corner position prediction is more precise. An ablation study on the setting of 3D corner offset prediction is done to evaluate its impact on the reconstruction results. As shown in Table 1, we compare our method, which predicts only the height offset, with the other approach that predicts all three-dimensional offsets. Results show that simultaneous prediction of three-dimensional coordinate offsets would lead to a decrease in model performance.

3. More Qualitative Comparison Results.

3.1. Results on Different Model Components.

Some qualitative comparison results of the ablation studies on different model components are shown in Figure 1. The baseline method is based on the direct 3D corner detection Transformer and the vanilla deformable attention-based edge detection Transformer. With the 2D-to-3D corner detection and edge attention mechanism, both corner and edge completeness are improved, even when dealing with sparse and incomplete point clouds. Then, with the addition of data augmentation, the reconstruction performance is further improved, enhancing the model's prediction accuracy under conditions of different LiDAR scanning qualities.

	Distance		Corner (%)			Edge (%)		
Method	WED ↓	ACO ↓	CP ↑	CR ↑	CF1 ↑	EP ↑	ER ↑	EF1 ↑
Height dimension	**0.253**	**0.207**	**94.1**	**82.8**	**88.1**	**83.7**	**73.6**	**78.3**
All three dimensions	0.276	0.225	92.5	82.2	87.1	80.5	72.2	76.2

Table 1. **Ablation study on 3D corner offset prediction setting.** In our method, only the offsets of height dimension are predicted. We compare it with the method of predicting all three dimension offsets. Results show that with frozen x and y positions, our method performs better than the comparison one.

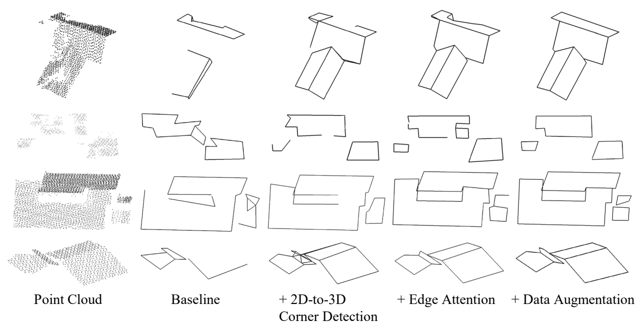

Figure 1. **Qualitative comparison on different components.** With the 2D-to-3D corner detection module and edge attention mechanism, the 3D wireframe reconstruction results are more complete even with the low-quality point clouds. Besides, the data augmentation also increases the reconstruction performance with more precise models from sparse and incomplete point clouds.

3.2. Results in 3D Wireframe Reconstruction

We present additional qualitative results on the Building3D dataset [5], comparing our approach with other baseline methods in Figure 2. While traditional methods [4, 6] produce mesh-based outputs, we include their wireframes for a more comprehensive comparison. The wireframes generated by these traditional methods tend to lack compactness, miss finer details, and are highly sensitive to noise, leading to poorer performance in complex scenarios. For deep-learning-based methods [1–3], BWFormer surpasses other SOTAs particularly in terms of completeness, capturing more intricate details and achieving superior results.

3.3. Results in LiDAR Scanning Simulation

We provide more qualitative results of LiDAR scanning simulation in Figure 3. The simulated scanning positions by the proposed method are similar to the real ones with different levels of sparsity and completeness, which could not be achieved by the uniform sampling-based method.

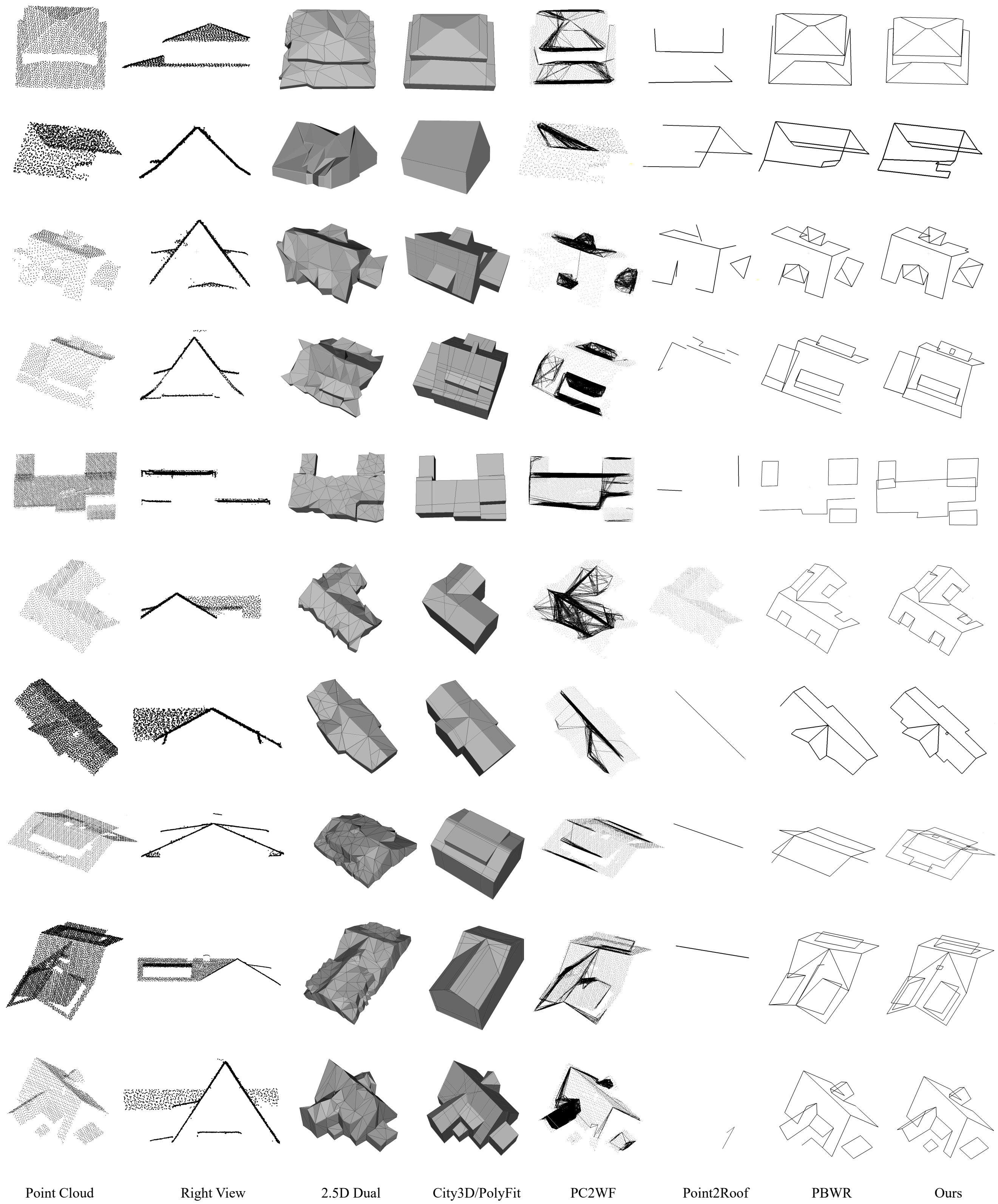

Figure 2. **More quantitative comparison results in 3D wireframe reconstruction.** The 1-st column is the point clouds segmented with the single building, however, the segmentation is not perfect and the buildings are still messed with some noise. The 2-nd column is the right views of the original point clouds in the 1-st column without removing the non-building parts. The results of the traditional methods (3-rd and 4-th columns) are mesh models, but overlaid with the wireframes. The results of the deep learning-based methods (5-th to 8-th columns) are 3D wireframes.

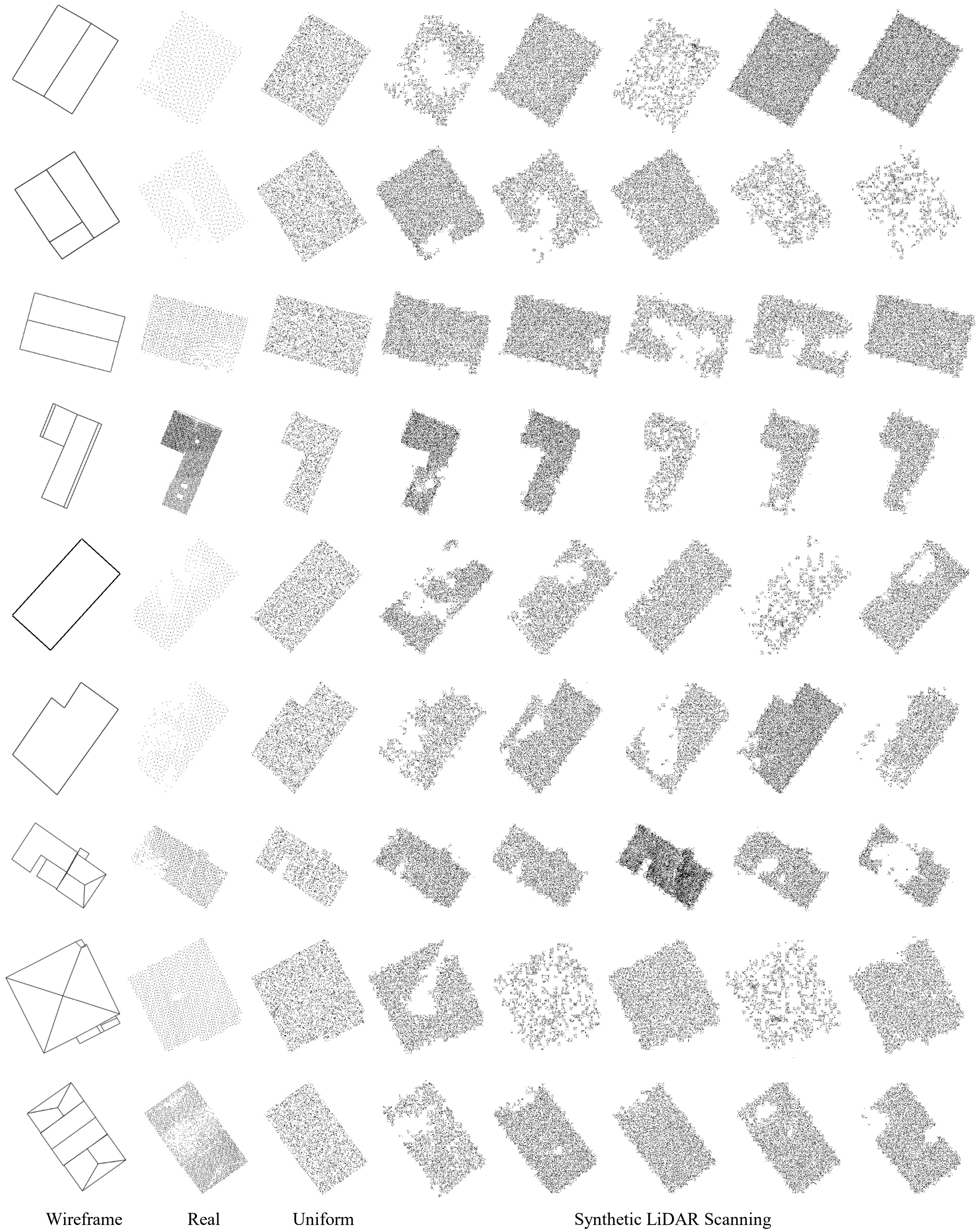

Figure 3. **More quantitative comparison results in LiDAR scanning simulation.** Compared with uniform sampling of 85% sparsity (3-th column), the simulated LiDAR scannings of our method (4-th to 8-th columns) are more diverse in both sparsity and completeness.

References

[1] Shangfeng Huang, Ruisheng Wang, Bo Guo, and Hongxin Yang. PBWR: Parametric building wireframe reconstruction from aerial LiDAR point clouds. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 27778–27787, 2024. 1

[2] Li Li, Nan Song, Fei Sun, Xinyi Liu, Ruisheng Wang, Jian Yao, and Shaosheng Cao. Point2Roof: End-to-end 3D building roof modeling from airborne LiDAR point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 193:17–28, 2022.

[3] Yujia Liu, Stefano D'Aronco, Konrad Schindler, and Jan Dirk Wegner. PC2WF: 3D wireframe reconstruction from raw point clouds. In *International Conference on Learning Representations (ICLR)*, 2021. 1

[4] Liangliang Nan and Peter Wonka. PolyFit: Polygonal surface reconstruction from point clouds. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2353–2361, 2017. 1

[5] Ruisheng Wang, Shangfeng Huang, and Hongxin Yang. Building3D: A urban-scale dataset and benchmarks for learning roof structures from point clouds. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 20076–20086, 2023. 1

[6] Qian-Yi Zhou and Ulrich Neumann. 2.5D building modeling by discovering global regularities. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 326–333, 2012. 1