BuildingWorld: A Structured 3D Building Dataset for Urban Foundation Models

Shangfeng Huang[1], **Ruisheng Wang**[2*], **Xin Wang**[1*],

[1]University of Calgary
[2]Shenzhen University
{shangfeng.huang, xcwang}@ucalgary.ca, ruiswang@szu.edu.cn

Abstract

As digital twins become central to the transformation of modern cities, accurate and structured 3D building models emerge as a key enabler of high-fidelity, updatable urban representations. These models underpin diverse applications including energy modeling, urban planning, autonomous navigation, and real-time reasoning. Despite recent advances in 3D urban modeling, most learning-based models are trained on building datasets with limited architectural diversity, which significantly undermines their generalizability across heterogeneous urban environments. To address this limitation, we present **BuildingWorld**, a comprehensive and structured 3D building dataset designed to bridge the gap in stylistic diversity. It encompasses buildings from geographically and architecturally diverse regions—including North America, Europe, Asia, Africa, and Oceania—offering a globally representative dataset for urban-scale foundation modeling and analysis. Specifically, BuildingWorld provides about **five million LOD2 building models** collected from diverse sources, accompanied by **real and simulated airborne LiDAR point clouds**. This enables comprehensive research on 3D building reconstruction, detection and segmentation. Cyber City, a virtual city model, is introduced to enable the generation of unlimited training data with customized and structurally diverse point cloud distributions. Furthermore, we provide standardized evaluation metrics tailored for building reconstruction, aiming to facilitate the training, evaluation, and comparison of large-scale vision models and foundation models in structured 3D urban environments.

Datasets — https://szusic.github.io/BuildingWorld/

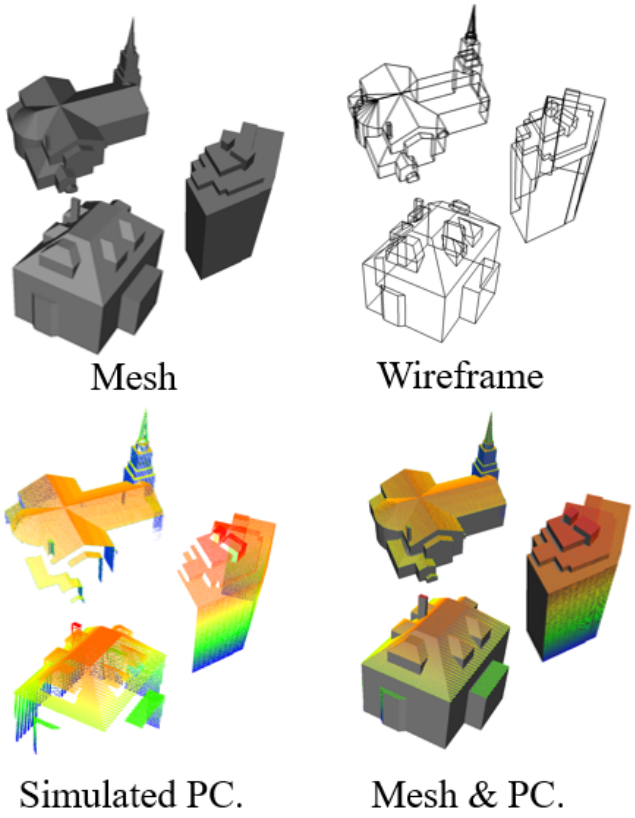

Figure 1: A glimpse of the BuildingWorld dataset.

Introduction

As urban digitization accelerates, the development of high-fidelity and continuously updatable 3D building models has become a cornerstone in enabling digital twin cities (Deng, Zhang, and Shen 2021). In digital twins, building models serve not only as geometric representations but also as integrative platforms that fuse heterogeneous urban data—ranging from energy use and structural integrity to mobility patterns and environmental conditions. These models have already demonstrated significant utility in a range of critical applications, including energy simulation (Pan et al. 2023), urban planning (Ernst et al. 2021), emergency response (Demir Ozbek et al. 2016), and virtual reality environments (Zhang, Zeng, and Liu 2021). In the past decade, the community has made remarkable progress in building reconstruction from various modalities, such as aerial imagery(Tack, Buyuksalih, and Goossens 2012), LiDAR point clouds (Bauchet and Lafarge 2020; Nan and Wonka 2017; Huang et al. 2022; Li and Shan 2022; Wang et al. 2020), multi-view stereo (MVS) (Yu et al. 2021; Luo et al. 2024). Recently, building reconstruction datasets with annotations and benchmarks (Peralta et al. 2020; Selvaraju et al. 2021;

*Corresponding author.

Wang, Huang, and Yang 2023; Government 2021) have significantly accelerated progress in this field, providing standardized resources for model training and evaluation. While these datasets have contributed significantly to progress in 3D building reconstruction, they often remain limited in architectural diversity and geographic scope, which constrains the generalizability of learned models (Liu et al. 2024b; Huang et al. 2024; Hao et al. 2025).

To this end, we propose **BuildingWorld**, a structured 3D building dataset for urban foundation models, designed to enhance the generalization capabilities of deep learning-based models. The emergence of large language models, such as ChatGPT (Achiam et al. 2023), DeepSeek (Liu et al. 2024a; Guo et al. 2025), and LLaMA (Touvron et al. 2023a,b), has demonstrated the critical role of large-scale, high-quality datasets in enabling strong reasoning and generalization capabilities. Similarly, large vision models, including SAM (Kirillov et al. 2023) and Dinov2 (Oquab et al. 2023), have achieved impressive performance by mining knowledge from massive image datasets. This paradigm highlights the need for similarly large-scale and diverse datasets in other domains, such as 3D urban modeling, to unlock comparable levels of generalization and reasoning. Therefore, the BuildingWorld dataset is constructed from about five million LOD2 building models, along with both real and simulated airborne LiDAR point clouds. These building models are collected from diverse sources and are distributed across geographically and architecturally varied regions, including North America, Europe, Asia, Africa, and Oceania. This wide coverage theoretically encompasses most major building styles found around the world. The BuildingWorld dataset aims to overcome the limitations imposed by biased data distributions and limited architectural diversity, which often hinder the generalization capabilities of current models.

Compared with text and image data, which are abundant and easily accessible on the web, the collection of 3D data, especially LiDAR point clouds, is labor-intensive, costly, and rarely available to the public. Recently, several advanced large point cloud–language models(**?**Xu et al. 2024) try to generate large amounts of synthetic text–point cloud pairs to support the training and evaluation of large models. In parallel, some methods (Huang et al. 2025; Otsuka et al. 2025) simulate point clouds to enhance model performance. Motivated by them, BuildingWorld generates simulated aerial LiDAR point clouds from LOD2 building models to approximate real-world point cloud acquisition. This allows model trained on BuildingWorld to better generalize to real-world LiDAR data. Specifically, the LiDAR simulator tool in Helios++ (Winiwarter et al. 2022) is used to simulate the acquisition of aerial point clouds, accounting for real-world factors such as occlusion, laser incidence angle, flight trajectory, and other sources of point cloud incompleteness. In addition, complete building point clouds are made available to support specific tasks, such as point cloud completion. Furthermore, to generate more realistic simulated aerial point clouds, infrastructure elements, terrain models, and vegetation such as trees are incorporated into the virtual environments in some cities.

Beyond standard benchmarks, we introduce supplementary evaluation metrics designed to more accurately assess reconstruction accuracy, completeness, and robustness under complex building conditions. A benchmark is conducted using five representative deep learning methods, validating the potential and utility of the dataset in the building reconstruction task. Furthermore, we analyze the limitations and challenge present in the dataset, and outline a range of potential downstream applications in digital twins, urban simulation, and large-scale 3D reconstruction.

Our contributions are summarized as following:

- We propose BuildingWorld, the first and largest structured 3D building dataset for urban foundation models, which contains about five million LOD2 building models from geographically and architecturally diverse regions spanning five continents, including North America, Europe, Asia, Africa and Oceania.
- We establish a comprehensive benchmark with standard and supplementary evaluation metrics, designed to assess reconstruction accuracy, completeness, and robustness.
- We create an 3D urban scene generator called Cyber City to generate 3D urban models (buildings, trees etc.) and 3D point clouds to enrich 3D dataset for large 3D foundation models.
- We simulate aerial point clouds using the Helios++ LiDAR simulator, incorporating realistic factors such as occlusion, laser incidence angle, and flight trajectory. Real aerial point clouds are also provided as part of the dataset to support evaluation and validate the effectiveness of models trained on simulated data.

Related Work

Related Datasets

In recent years, the increasing availability of annotated building datasets has significantly promoted the development of deep learning-based 3D building reconstruction methods, as summarized in 1. RoofN3D (Wichmann, Agoub, and Kada 2018), 3DBAG (León-Sánchez et al. 2021), STPLS3D (Chen et al. 2022) and City3D (Huang et al. 2022) provide large-scale aerial LiDAR point clouds and building mesh models generated through automatic reconstruction methods. However, these building models often exhibit geometric errors in local details, which makes them unsuitable as ground truth for supervised learning methods. DublinCity (Zolanvari et al. 2019) and SensatUrban (Hu et al. 2022) are two recently popular urban point cloud datasets for semantic segmentation, but they lack high-quality building mesh models. Houses3K (Peralta et al. 2020), UrbanScenes3D (Lin et al. 2022) and BuildingNet (Selvaraju et al. 2021) provide high-quality, manually crafted building models with limited architectural diversity, whereas SUM (Gao et al. 2021) offers low-quality urban meshes primarily used for mesh semantic segmentation. Due to low mesh quality or the absence of corresponding point clouds, the aforementioned datasets are rarely used for training supervised methods for building reconstruction from point clouds. Building3D (Wang, Huang,

Dataset	Scene	Sensors	#Models	Diversity	PC	Mesh	WF
RoofN3D (2018)	City	ALS	–	New York (1,009 km^2)	✓	✓	✗
DublinCity (Zolanvari et al. 2019)	Urban	ALS	–	Dublin (5.6 km^2)	✓	✗	✗
Houses3K (Peralta et al. 2020)	Object	Handcraft	3 K	600 unique buildings	✗	✓	✗
3DBAG (León-Sánchez et al. 2021)	Country	ALS	10 M	Netherlands	✓	✓	✗
SUM (Gao et al. 2021)	Urban	Aerial Photogram.	–	Helsinki (4 km^2)	✗	✓	✗
BuildingNet (Selvaraju et al. 2021)	Object	Handcraft	2 K	–	✗	✓	✗
City3D (Huang et al. 2022)	Urban	ALS	20 K	3 cities	✓	✓	✗
UrbanScene3D (Lin et al. 2022)	Urban	MVS	1.4 K	10 synthetic / 6 real scenes (55 km^2)	✗	✓	✗
STPLS3D (Chen et al. 2022)	–	UAV Photogram.	–	63 synthetic / 4 real scenes (16 km^2)	✓	✓	✗
SensatUrban (Hu et al. 2022)	Urban	UAV Photogram.	–	UK cities (7.6 km^2)	✓	✗	✗
Building3D (2023)	Country	ALS	760 K	Estonia	✓	✓	✓
BuildingWorld	Urban (Global)	ALS Simulated ALS	5 M	Five continents	✓	✓	✓

Table 1: Comparison with representative 3D building datasets. *K*: thousand, *M*: million, *PC*: point clouds, *WF*: wireframe, *ALS*: airborne laser scanning, MVS: multi-view stereo, *Photogram.*: photogrammetry.

and Yang 2023) is the first dataset specifically designed for LOD2 building reconstruction from aerial LiDAR point clouds. Notably, the provided wireframe models are manually created and well suited for training deep learning-based methods. However, these datasets are geographically limited, which constrains the generalization and robustness of models trained on them. To address this limitation, we propose BuildingWorld, the first building reconstruction dataset with architectural diversity spanning five continents, including North America, Europe, Asia, Africa and Oceania.

Related Methods

Traditional 3D building reconstruction methods can be divided into model-driven and data-driven approaches based on their reconstruction strategies. Model-driven methods (Zhang, Li, and Shan 2021; Song et al. 2020; Li and Shan 2022; Zang et al. 2024) typically rely on the assumption of structured building rules, where building models can be approximated by fitting simple parametric shapes, such as hip and flat roofs. Data-driven methods (Zhou and Neumann 2010, 2011; Chen, Wang, and Peethambaran 2017; Nan and Wonka 2017; Yang et al. 2022) focus on extracting building primitives from point clouds and reconstructing the topological relationships between these primitives to obtain watertight models. However, these traditional methods tend to be highly sensitive to parameter selection and often struggle to generalize to diverse building shapes and structural variations. Recently, several deep learning-based methods (Li et al. 2022; Yang, Huang, and Wang 2024; Hao et al. 2025; Huang et al. 2024) have been proposed to reconstruct building models from point clouds. Specifically, heuristic methods (Li et al. 2022; Yang et al. 2024; Jiang et al. 2023; Hao et al. 2025, 2024; Wang, Huang, and Yang 2023) treat building model reconstruction as a wireframe reconstruction task consisting of corners and edges. However, heuristic reconstruction strategies inherently suffer from error accumulation across different stages. Therefore, some methods (Luo et al. 2022; Yang, Huang, and Wang 2024; Huang et al. 2024) propose to regress building edges directly from point clouds, without relying on corner detection. However, these methods still rely on post-processing to generate the final wireframe models. Recently, some new building reconstruction paradigms are also proposed. Point2Building (Liu et al. 2024b) adopts a generative strategy to reconstruct building models. EdgeDiff (Liu et al. 2025a) leverages diffusion models to recover building structures from a noise field, while BWFormer (Liu et al. 2025b) employs a powerful 2D corner detector to reduce the impact of corner detection errors on the subsequent edge regression module.

BuildingWorld Dataset

Overview

The BuildingWorld dataset is constructed by collecting all publicly available LoD2 building models from around the world, with the goal of capturing diverse global architectural styles and mitigating limitations in model generalization caused by insufficient or imbalanced data distribution. Aerial point clouds are subsequently simulated using the Helios++ simulator (Ernst et al. 2021), as shown in Figure 2. The Helios++ simulator enables the generation of realistic aerial point clouds by modeling phenomena such as laser incidence angle effects, occlusions from buildings and trees, and other environmental interferences present in real-world data. To further diversify the data distribution, we construct a synthetic urban scene named *Cyber City*, which integrates buildings of diverse styles and cultural origins from

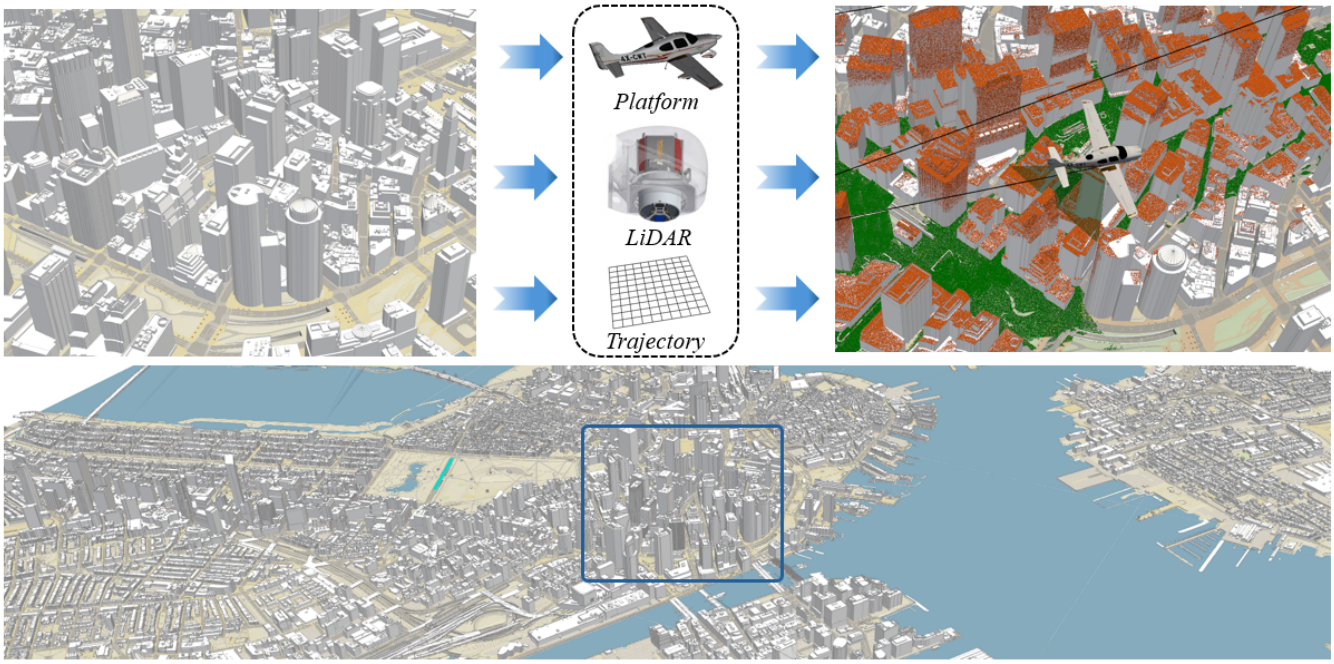

Figure 2: Illustration of the construction process of BuildingWorld dataset. A glimpse of the LOD2 digital city model of Boston is shown. The zoomed-in downtown area illustrates simulated aerial LiDAR point clouds, generated using a predefined airborne platform, LiDAR sensor, and flight trajectory.

around the world, thereby breaking geographical and temporal constraints. Theoretically, it is possible to generate an infinite variety of Cyber City instances by randomly assembling building models with diverse styles and cultural backgrounds in different spatial configurations.

Building Models

To the best of our efforts, BuildingWorld dataset are constructed by collecting LoD2 building models from 44 cities, covering an area of **21,718** km^2, resulting in a total of approximately 5 million buildings. These models span a wide variety of types and sources: some are derived from architectural blueprints provided by local governmental land management agencies, others are manually created, some are generated by integrating multiple data sources with predefined roof geometries, as shown in Figure 3. To the best of our knowledge, this is the first publicly available dataset that provides such a comprehensive and globally diverse collection of LoD2 building models. As illustrated in Figure 1, the LoD2 building models exhibit greater geometric detail, including features such as chimneys and sharply pitched roofs. The roof geometry of LoD2 models serves as an important metric for evaluating the fidelity of individual building models and the overall quality of the dataset. In the North American subset of the dataset, urban areas are predominantly characterized by small-scale residential zones covering 50–150 km² and medium-scale regions, including both residential and public buildings, ranging from 150–500 km². The building height distribution is largely concentrated in the 0–6 meter range (single-story structures) and the 6–12 meter range (two- to three-story buildings), with high-rise structures appearing relatively infrequently. Such a distribution is consistent with the urban planning and demographic realities of these countries, where ample livable land and moderate population pressures have led to a prevalence of low-density, low-rise constructions. In contrast, the Hong Kong subset exhibits a significantly higher proportion of high-rise buildings, predominantly in the 12–50 meter height range. Figure 3 demonstrates the diversity of building models within the dataset through the distribution of building footprint area and the percentage breakdown of building sizes and heights.

Simulated ALS

Some cities agencies provide not only structurally detailed LoD2 building models but also corresponding real aerial point clouds. However, our analysis reveals that the geometric consistency between the models and the real point clouds is insufficient to support their direct use in supervised learning. This misalignment may arise from various factors, such as positional inaccuracies introduced during the modeling process—especially when models are created based on architectural blueprints—or from subtle yet impactful discrepancies in roof geometry, where the government-provided models exhibit deviations from the actual roof structures observed in point clouds. Nonetheless, real aerial point clouds remain useful in reinforcement learning stage, especially when an effective reward-based ranking strategy is employed to guide the training of large 3D vision modules. Therefore, the processed real point clouds are also collected

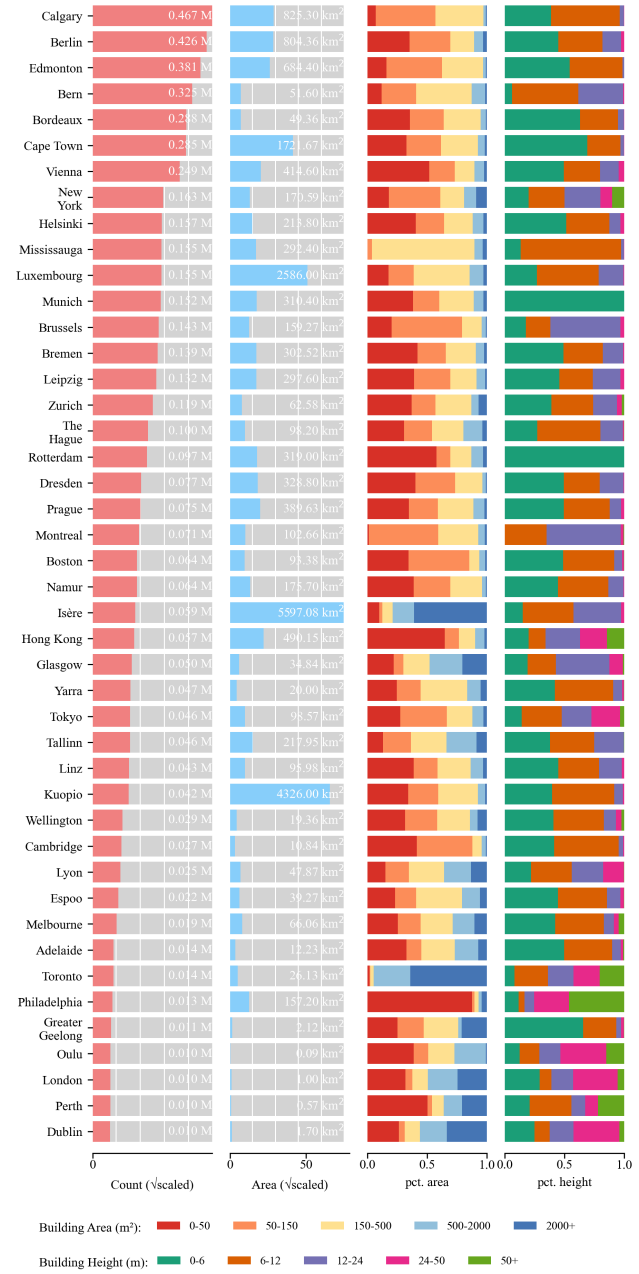

Figure 3: Statistical overview of the BuildingWorld dataset. Area bars indicate scene sizes, while percentage area and height metrics highlight the diversity of building structures.

and made available in BuildingWorld.

To construct a high-quality dataset for 3D building point cloud reconstruction, the Helios++ simulator (Ernst et al. 2021) is employed to generate realistic aerial point clouds, as shown in Figure 2. The simulator is composed of three main components—vehicle platform, LiDAR sensor, and flight trajectory—which jointly determine key properties of the generated point clouds, such as density, coverage, and structural completeness. Specifically, the vehicle platform primarily controls the flight speed, which directly influences point cloud density and the completeness of fine structural details. Lower flight speeds typically yield denser point clouds but increase the computational burden for model training. Conversely, higher speeds often result in a substantial loss of rooftop details—such as chimneys—due to sparser sampling. In practice, flight speeds are chosen based on the application scenario and task requirements. In BuildingWorld, point clouds for each scene are simulated using a randomly selected speed within the range of 185–463 km/h (100–250 knots), rather than a fixed value, to enhance the diversity of point cloud distributions. The RIEGL VQ-1560 II-S, one of the most advanced and widely used airborne LiDAR sensors, is selected as the simulated sensor in our framework. During simulation, the laser pulse repetition rates are set to 1M and 2M Hz, with a scan angle of ±30°. The scan frequencies are defined as 200, 400, and 600 Hz. These parameters are selected in combination with varying flight speeds to generate diverse point cloud distributions. Pulse repetition rate and scan frequency directly determine the density of the generated point clouds. As these values increase, the resulting point clouds capture finer rooftop details and contain a greater number of points. The flight trajectory plays a critical role not only in determining the overall density of the point cloud, but also in shaping the spatial distribution of occlusions—areas where parts of buildings are missing due to line-of-sight obstructions. In our experiments, the primary flight trajectory is aligned in the north-south direction, while the secondary trajectory follows an east-west orientation. During the simulation, the main flight path is oriented north–south, with an auxiliary trajectory in the east–west direction. To reduce occlusion-related point cloud loss, a lateral overlap rate between 40% and 60% is maintained. In addition, the altitude of the flight trajectories is set within a range of 600 to 1200 meters above sea level. Together, the flight trajectory and altitude control the distribution of point cloud coverage on building facades. At higher altitudes, the laser incidence angle relative to vertical surfaces becomes smaller, resulting in reduced interaction with facades and, consequently, lower completeness of facade point clouds compared to lower-altitude flights. Nevertheless, facade points reconstructed from aerial LiDAR data using LoD2 building models can significantly improve the estimation of building height.

BuildingWorld provides simulated aerial point clouds generated via Helios++, as shown in Figure 1. In addition, complete point clouds also are sampled directly from the 3D building models, offering comprehensive geometric data for specific tasks. For example, when training a large-scale foundation model for 3D building point clouds, aerial point clouds generated by Helios++ can be used as input, while the completed point clouds predicted by the model are supervised with ground-truth geometry sampled directly from the original building models. Specifically, uniform sampling is employed to generate points from the building models, ensuring a consistent point density of 30 $points/m^2$. Furthermore, to simulate realistic sensing noise, random perturbations are introduced along both the planar surface and normal directions during sampling, enhancing the natural vari-

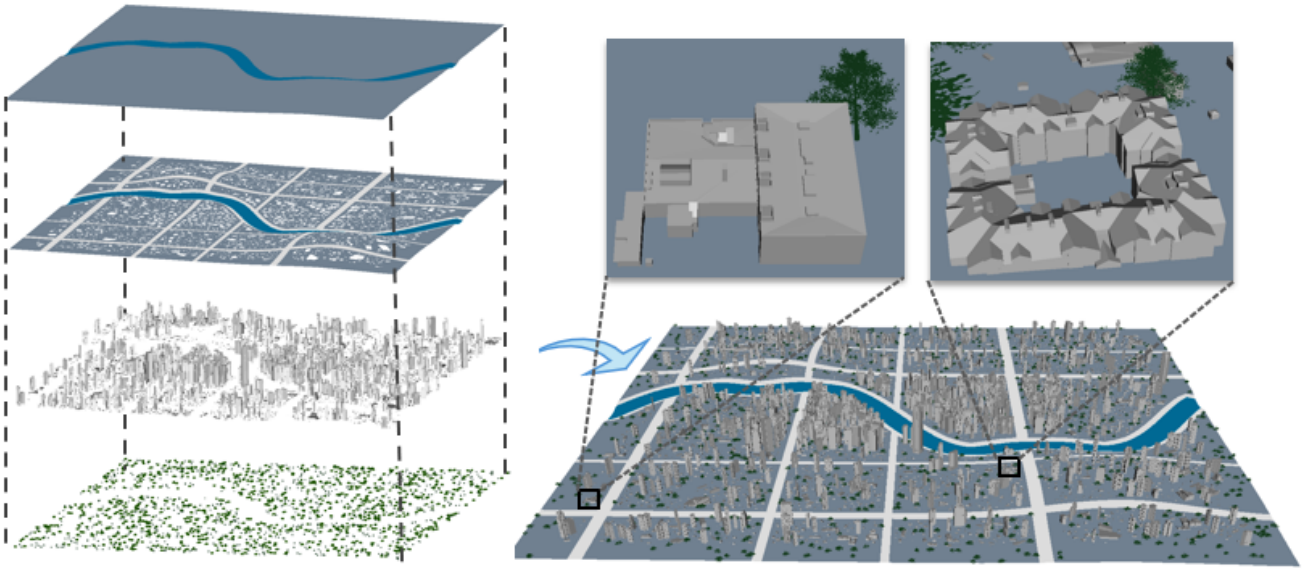

Figure 4: Illustration of Cyber City, which consists of four main components: terrain, road and building footprints, buildings, and vegetation.

ability of the resulting point clouds.

Cyber City

Across continents, architectural styles manifest distinct regional characteristics. European cities emphasize contextual modernism that integrates with historical environments. African urban landscapes reflect a coexistence of colonial-era structures and modernist glass forms, balancing cultural heritage with pragmatic functionality. Asian cities are defined by high-density, technologically driven verticality. North American architecture favors minimalist functionalism and spatial openness, while Oceania prioritizes climate-adaptive, sustainable design. Unbound by geographical location or cultural history, Cyber City serves as a virtual city where building models representing diverse architectural styles from all continents are brought together to form a globally inclusive urban composition, as shown in Figure 4. Being procedurally generated, it can produce unlimited synthetic urban configurations with diverse architectural compositions, facilitating richer data distributions for model training.

Cyber City consists of four main layers: terrain, road and building footprints, vegetation, and 3D building. In the terrain layer, we begin by randomly defining the spatial boundaries of the city, which serves as the basis for subsequent urban layout generation, as shown 4 . It defines an area of 4 × 4 km. Given that rivers frequently play a central role in the formation of real-world cities, a primary river is randomly generated to run through the synthetic urban area. Furthermore, to enhance terrain diversity, random local protrusions are applied to simulate small hills across the ground surface. The footprint layer defines the spatial organization of roads and buildings within the city. Streets are primarily arranged in an orthogonal grid pattern, while building footprints are allocated to the remaining parcels, specifying the exact locations for building placement. In Building layer, building area and height are used as heuristic indicators to differentiate functional types of buildings. As shown in Figure 4, the central region is designated as the Central Business District, where taller structures are placed to represent high-density commercial buildings, typically surrounded by lower-rise residential areas. In recreational and public activity zones, such as shopping malls and libraries, building models with relatively large footprint areas are placed to reflect their functional requirements. It is important to note that this configuration reflects conventional urban logic, but a freely designed virtual city imposes no such constraints—its structure can be arbitrarily defined to suit any desired purpose or experimental condition. In the tree layer, vegetation is distributed in a space-filling manner, with trees placed in unoccupied or residual areas of the scene to complement the urban layout. Additionally, the types and sizes of trees are randomly selected and generated to increase variability and naturalness in the scene. The tree layer is incorporated to simulate occlusions commonly encountered during point cloud acquisition, where trees partially obstruct building structures. Such occlusions are frequently observed in real-world urban environments.

Benchmark

Baseline Methods

BuildingWorld is the first dataset designed to support the training of 3D foundation models for building reconstruction. However, due to the lack of publicly available large-scale 3D building reconstruction models, a direct evaluation of BuildingWorld's effectiveness on such foundation models is currently not possible. Given this limitation, we

Method	Distance (m)	Accuracy					
	ACO	CP	CR	CF_1	EP	ER	EF_1
PBWR* (Huang et al. 2024)	0.27	0.95	0.67	0.78	0.84	0.61	0.71
EdgeDiff* (Liu et al. 2025a)	0.23	0.95	0.70	0.81	0.91	0.67	0.77
BWFormer* (Liu et al. 2025b)	0.22	0.94	0.80	0.86	0.81	0.70	0.81
PBWR (Huang et al. 2024)	0.22	0.96	0.68	0.80	0.91	0.65	0.76
PBWR* (Huang et al. 2024)	0.36	0.85	0.65	0.74	0.71	0.60	0.65
EdgeDiff (Liu et al. 2025a)	0.22	0.98	0.73	0.84	0.92	0.70	0.79
EdgeDiff* (Liu et al. 2025a)	0.23	0.94	0.63	0.76	0.85	0.57	0.68
BWFormer (Liu et al. 2025b)	0.20	0.95	0.82	0.88	0.85	0.74	0.79
BWFormer* (Liu et al. 2025b)	0.22	0.92	0.75	0.82	0.77	0.62	0.68

Table 2: Performance comparison of models on simulated and real-world point clouds. The first row block reports performance on simulated data, while the second block shows results on real aerial point clouds. * indicates models trained on simulated point clouds; otherwise, models are trained on real point clouds.

instead design a experiment to investigate how well models trained on BuildingWorld generalize to real-world aerial point clouds.

Datasets and Experimental Setup

Building3D (Wang, Huang, and Yang 2023) dataset provides real aerial point clouds paired with manually labeled LoD2 building models of Tallinn City, serving as ground truth for evaluation. In our experiments, we first train the baseline method using the simulated building data derived from Building3D, and subsequently evaluate its performance on the test set, which contains real aerial point clouds along with corresponding annotated building models. This setup allows us to assess how well models trained solely on synthetic data transfer to real-world inputs. To establish a reference point and assess the generalization ability, we compare performance with the model trained directly on the Building3D dataset. We follow the same evaluation metrics as defined in the Building3D dataset, as shown in Table 2. Specifically, ACO denotes the Average Corner Offset, measuring the mean positional deviation between predicted and ground-truth corner points, expressed in meters. CP and EP refer to the precision of predicted corners and edges, respectively, while CR and ER represent their corresponding recall values. CF_1 and EF_1 indicate the F_1-scores for corner and edge detection, respectively.

Results and Discussion

Table 2 compares the performance of recent methods (Huang et al. 2024; Liu et al. 2025a,b) trained on real data and simulated data, evaluated on both real and simulated point clouds. Experimental results indicate that the model trained on real data achieves performance comparable to that of the model (*) trained on simulated data when tested within their respective domains. It is noteworthy that the model (*) trained on simulated data also performs competitively on real-world point clouds, showing no substantial degradation in reconstruction accuracy. Table 2 indicates that models trained on BuildingWorld, using globally collected building models and simulated point clouds, are fully capable of supporting real-world applications. For example, the real point cloud data used for training the original PBWR model includes both RGB and intensity information, which contribute to improved reconstruction performance. In contrast, the model (*) trained on simulated data relies solely on the spatial structure of point clouds for building reconstruction, yet still demonstrates strong applicability to real-world data. They further suggest that large-scale 3D building reconstruction models trained on BuildingWorld have strong potential to generalize effectively and accurately reconstruct buildings across diverse geographic regions worldwide.

Challenges and Tasks

LoD3 Building Reconstruction

Beyond the comprehensive LoD2 coverage, BuildingWorld also incorporates detailed LoD3 building models for selected cities, made available through contributions from government bodies or city planning authorities, such as Glasgow and Hong Kong. Inspired by the design concept of Cyber City, scattered LoD3 building models can be aggregated into a unified LoD3 Cyber City. Furthermore, by leveraging the Helios++ simulator, we can separately simulate façade point clouds using vehicle-mounted LiDAR sensors and rooftop point clouds using airborne LiDAR. The integration of these complementary point cloud sources enables new opportunities for exploring LoD3 building reconstruction.

Semantic and Instance Segmentation

In BuildingWorld, all point clouds are annotated with unique semantic and instance-level labels, covering façades, roofs, buildings, vegetation, bridges, and terrain. Following the Cyber City design philosophy, additional elements such as vehicles and infrastructure models can be incorporated to further enrich the urban scene. Although replicating material-dependent artifacts in simulated point clouds remains challenging, BuildingWorld serves as a practical and effective resource for advancing city-scale semantic and instance segmentation tasks.

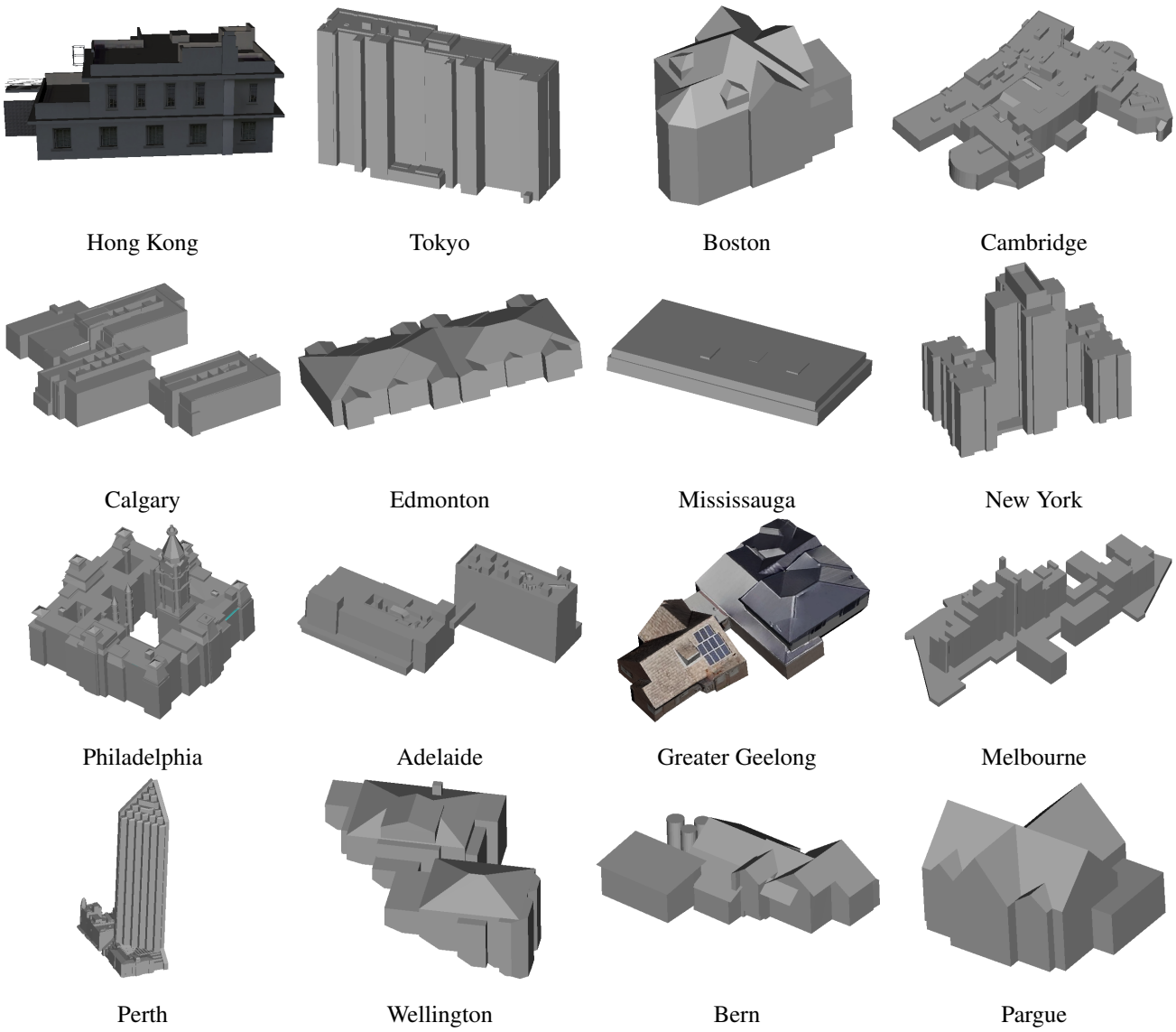

Figure 5: Illustration of BuildingWorld Dataset.

Supplementary Material

Figure 5 presents representative building models from multiple cities included in the BuildingWorld dataset, revealing the wide stylistic and functional diversity of buildings ranging from dense residential neighborhoods and urban commercial districts to industrial zones and leisure-oriented facilities.

Conclusion

In this work, we introduce BuildingWorld, a large-scale, structured dataset tailored for 3D building understanding across diverse geographic and architectural contexts. By aggregating about five million LoD2 building models from 44 cities worldwide, and providing both real and simulated aerial LiDAR point clouds, BuildingWorld offers a unique resource for training and evaluating urban-scale 3D vision systems. The integration of the Cyber City framework further enables the generation of procedurally diverse synthetic scenes. We demonstrate that models trained exclusively on simulated data from BuildingWorld achieve competitive performance on real-world datasets, validating BuildingWorld's effectiveness in bridging the synthetic-to-real gap.

Acknowledgments

This work was supported by the Key Technological Innovation Program of Ningbo City under Grant No. 2024Z297, and by National Key Research and Development Program of China (2022YFB2602105).

References

Achiam, J.; Adler, S.; Agarwal, S.; Ahmad, L.; Akkaya, I.; Aleman, F. L.; Almeida, D.; Altenschmidt, J.; Altman, S.;

Anadkat, S.; et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Bauchet, J.-P.; and Lafarge, F. 2020. Kinetic shape reconstruction. *ACM Transactions on Graphics (TOG)*, 39(5): 1–14.

Chen, D.; Wang, R.; and Peethambaran, J. 2017. Topologically aware building rooftop reconstruction from airborne laser scanning point clouds. *IEEE Transactions on Geoscience and Remote Sensing*, 55(12): 7032–7052.

Chen, M.; Hu, Q.; Yu, Z.; Thomas, H.; Feng, A.; Hou, Y.; McCullough, K.; Ren, F.; and Soibelman, L. 2022. Stpls3d: A large-scale synthetic and real aerial photogrammetry 3d point cloud dataset. *arXiv preprint arXiv:2203.09065*.

Demir Ozbek, E.; Zlatanova, S.; Ates Aydar, S.; and Yomralioglu, T. 2016. 3D Geo-Information requirements for disaster and emergency management. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 41: 101–108.

Deng, T.; Zhang, K.; and Shen, Z.-J. M. 2021. A systematic review of a digital twin city: A new pattern of urban governance toward smart cities. *Journal of management science and engineering*, 6(2): 125–134.

Ernst, F.; Şenol, H. İ.; Akdağ, S.; and Barutcuoglu, Ö. 2021. Virtual Reality for City Planning. *Harran Üniversitesi Mühendislik Dergisi*, 6(3): 150–160.

Gao, W.; Nan, L.; Boom, B.; and Ledoux, H. 2021. SUM: A benchmark dataset of semantic urban meshes. *ISPRS Journal of Photogrammetry and Remote Sensing*, 179: 108–120.

Government, T. M. 2021. Tokyo Digital Twin Project. Accessed: 2025-01-17.

Guo, D.; Yang, D.; Zhang, H.; Song, J.; Zhang, R.; Xu, R.; Zhu, Q.; Ma, S.; Wang, P.; Bi, X.; et al. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Hao, Q.; Han, T.; Liu, Y.; Huang, S.; Zhu, D.; Su, J.; Wu, Y.; and Cai, G. 2025. Edge First: Edge-Guided Geometry for Superior 3D Roof Wireframe Reconstruction. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.

Hao, Q.; Liu, Z.; Zhu, D.; Su, J.; Gong, Z.; Wu, Y.; and Cai, G. 2024. LEAF: Learning Edge Attraction Filed for Precise Roof Wireframe Reconstruction. In *Proceedings of the 2024 8th International Conference on Electronic Information Technology and Computer Engineering*, 612–619.

Hu, Q.; Yang, B.; Khalid, S.; Xiao, W.; Trigoni, N.; and Markham, A. 2022. Sensaturban: Learning semantics from urban-scale photogrammetric point clouds. *International Journal of Computer Vision*, 130(2): 316–343.

Huang, J.; Stoter, J.; Peters, R.; and Nan, L. 2022. City3D: Large-scale building reconstruction from airborne LiDAR point clouds. *Remote Sensing*, 14(9): 2254.

Huang, Q.; Zhang, R.; Liu, K.; Gong, M.; Zhang, H.; and Huang, H. 2025. ArcPro: Architectural Programs for Structured 3D Abstraction of Sparse Points. *arXiv preprint arXiv:2503.02745*.

Huang, S.; Wang, R.; Guo, B.; and Yang, H. 2024. PBWR: Parametric-building-wireframe reconstruction from aerial LiDAR point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 27778–27787.

Jiang, T.; Wang, Y.; Zhang, Z.; Liu, S.; Dai, L.; Yang, Y.; Jin, X.; and Zeng, W. 2023. Extracting 3-D Structural Lines of Building from ALS Point Clouds using Graph Neural Network Embedded with Corner Information. *IEEE Transactions on Geoscience and Remote Sensing*.

Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A. C.; Lo, W.-Y.; et al. 2023. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, 4015–4026.

León-Sánchez, C.; Giannelli, D.; Agugiaro, G.; and Stoter, J. 2021. Testing the new 3D bag dataset for energy demand estimation of residential buildings. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 46: 69–76.

Li, L.; Song, N.; Sun, F.; Liu, X.; Wang, R.; Yao, J.; and Cao, S. 2022. Point2Roof: End-to-end 3D building roof modeling from airborne LiDAR point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 193: 17–28.

Li, Z.; and Shan, J. 2022. RANSAC-based multi primitive building reconstruction from 3D point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 185: 247–260.

Lin, L.; Liu, Y.; Hu, Y.; Yan, X.; Xie, K.; and Huang, H. 2022. Capturing, reconstructing, and simulating: the urbanscene3d dataset. In *European Conference on Computer Vision*, 93–109. Springer.

Liu, A.; Feng, B.; Xue, B.; Wang, B.; Wu, B.; Lu, C.; Zhao, C.; Deng, C.; Zhang, C.; Ruan, C.; et al. 2024a. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*.

Liu, Y.; Obukhov, A.; Wegner, J. D.; and Schindler, K. 2024b. Point2Building: Reconstructing buildings from airborne LiDAR point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 215: 351–368.

Liu, Y.; Wang, R.; Huang, S.; and Cai, G. 2025a. EdgeDiff: Edge-aware Diffusion Network for Building Reconstruction from Point Clouds. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 17008–17018.

Liu, Y.; Zhu, L.; Ye, H.; Huang, S.; Gao, X.; Zheng, X.; and Shen, S. 2025b. BWFormer: Building Wireframe Reconstruction from Airborne LiDAR Point Cloud with Transformer. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 22215–22224.

Luo, H.; Zhang, J.; Liu, X.; Zhang, L.; and Liu, J. 2024. Large-scale 3d reconstruction from multi-view imagery: A comprehensive review. *Remote Sensing*, 16(5): 773.

Luo, Y.; Ren, J.; Zhe, X.; Kang, D.; Xu, Y.; Wonka, P.; and Bao, L. 2022. Learning to construct 3d building wireframes from 3d line clouds. *arXiv preprint arXiv:2208.11948*.

Nan, L.; and Wonka, P. 2017. Polyfit: Polygonal surface reconstruction from point clouds. In *Proceedings of the IEEE international conference on computer vision*, 2353–2361.

Oquab, M.; Darcet, T.; Moutakanni, T.; Vo, H.; Szafraniec, M.; Khalidov, V.; Fernandez, P.; Haziza, D.; Massa, F.; El-Nouby, A.; et al. 2023. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*.

Otsuka, D.; Mae, S.; Yamada, R.; and Kataoka, H. 2025. Pre-training with 3D Synthetic Data: Learning 3D Point Cloud Instance Segmentation from 3D Synthetic Scenes. *arXiv preprint arXiv:2503.24229*.

Pan, Y.; Zhu, M.; Lv, Y.; Yang, Y.; Liang, Y.; Yin, R.; Yang, Y.; Jia, X.; Wang, X.; Zeng, F.; et al. 2023. Building energy simulation and its application for building performance optimization: A review of methods, tools, and case studies. *Advances in Applied Energy*, 10: 100135.

Peralta, D.; Casimiro, J.; Nilles, A. M.; Aguilar, J. A.; Atienza, R.; and Cajote, R. 2020. Next-best view policy for 3d reconstruction. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part IV 16*, 558–573. Springer.

Selvaraju, P.; Nabail, M.; Loizou, M.; Maslioukova, M.; Averkiou, M.; Andreou, A.; Chaudhuri, S.; and Kalogerakis, E. 2021. Buildingnet: Learning to label 3d buildings. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10397–10407.

Song, J.; Xia, S.; Wang, J.; and Chen, D. 2020. Curved buildings reconstruction from airborne LiDAR data by matching and deforming geometric primitives. *IEEE Transactions on Geoscience and Remote Sensing*, 59(2): 1660–1674.

Tack, F.; Buyuksalih, G.; and Goossens, R. 2012. 3D building reconstruction based on given ground plan information and surface models extracted from spaceborne imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67: 52–64.

Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.-A.; Lacroix, T.; Rozière, B.; Goyal, N.; Hambro, E.; Azhar, F.; et al. 2023a. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Touvron, H.; Martin, L.; Stone, K.; Albert, P.; Almahairi, A.; Babaei, Y.; Bashlykov, N.; Batra, S.; Bhargava, P.; Bhosale, S.; et al. 2023b. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

Wang, R.; Huang, S.; and Yang, H. 2023. 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*, 20076–20086.

Wang, S.; Cai, G.; Cheng, M.; Junior, J. M.; Huang, S.; Wang, Z.; Su, S.; and Li, J. 2020. Robust 3D reconstruction of building surfaces from point clouds based on structural and closed constraints. *ISPRS Journal of Photogrammetry and Remote Sensing*, 170: 29–44.

Wichmann, A.; Agoub, A.; and Kada, M. 2018. Roofn3d: Deep learning training data for 3d building reconstruction. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42: 1191–1198.

Winiwarter, L.; Pena, A. M. E.; Weiser, H.; Anders, K.; Sánchez, J. M.; Searle, M.; and Höfle, B. 2022. Virtual laser scanning with HELIOS++: A novel take on ray tracing-based simulation of topographic full-waveform 3D laser scanning. *Remote Sensing of Environment*, 269: 112772.

Xu, R.; Wang, X.; Wang, T.; Chen, Y.; Pang, J.; and Lin, D. 2024. Pointllm: Empowering large language models to understand point clouds. In *European Conference on Computer Vision*, 131–147. Springer.

Yang, H.; Huang, S.; and Wang, R. 2024. A method for roof wireframe reconstruction based on self-supervised pretraining. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10: 239–246.

Yang, H.; Huang, S.; Wang, R.; and Wang, X. 2024. Self-supervised pre-training for 3-D roof reconstruction on LiDAR data. *IEEE Geoscience and Remote Sensing Letters*, 21: 1–5.

Yang, S.; Cai, G.; Du, J.; Chen, P.; Su, J.; Wu, Y.; Wang, Z.; and Li, J. 2022. Connectivity-aware Graph: A planar topology for 3D building surface reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 191: 302–314.

Yu, D.; Ji, S.; Liu, J.; and Wei, S. 2021. Automatic 3D building reconstruction from multi-view aerial images with deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 171: 155–170.

Zang, Y.; Mi, W.; Xiao, X.; Guan, H.; Chen, J.; and Li, D. 2024. Compound 3D building modeling with structure-aware partition and primitive assembly from airborne laser scanning point clouds. *International Journal of Digital Earth*, 17(1): 2375112.

Zhang, C.; Zeng, W.; and Liu, L. 2021. UrbanVR: An immersive analytics system for context-aware urban design. *Computers & Graphics*, 99: 128–138.

Zhang, W.; Li, Z.; and Shan, J. 2021. Optimal model fitting for building reconstruction from point clouds. *IEEE journal of selected topics in applied earth observations and remote sensing*, 14: 9636–9650.

Zhou, Q.-Y.; and Neumann, U. 2010. 2.5 d dual contouring: A robust approach to creating building models from aerial lidar point clouds. In *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part III 11*, 115–128. Springer.

Zhou, Q.-Y.; and Neumann, U. 2011. 2.5 D building modeling with topology control. In *CVPR 2011*, 2489–2496. IEEE.

Zolanvari, S.; Ruano, S.; Rana, A.; Cummins, A.; Da Silva, R. E.; Rahbar, M.; and Smolic, A. 2019. DublinCity: Annotated LiDAR point cloud and its applications. *arXiv preprint arXiv:1909.03613*.