This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

EdgeDiff: Edge-aware Diffusion Network for Building Reconstruction from Point Clouds

Yujun Liu[1] Ruisheng Wang[1,†] Shangfeng Huang[2] Guorong Cai[3]
[1]Shenzhen University [2]University of Calgary [3]Jimei University
yujunliu2024@mails.szu.edu.cn; ruiswang@szu.edu.cn;
shangfeng.huang@ucalgary.ca; guorongcai@jmu.edu.cn

Abstract

*Building reconstruction is a challenging problem at the intersection of computer vision, photogrammetry and computer graphics. 3D wireframe presents a compelling representation for building modeling through its compact structure. Existing wireframe reconstruction methods employing vertex detection and edge regression have achieved promising results. In this paper, we develop an **Edge**-aware **Diff**usion network, dubbed **EdgeDiff**. As a novel paradigm for wireframe reconstruction, the EdgeDiff generates wireframe models from noise using a conditional diffusion model. During the training process, the ground truth wireframes firstly are formulated as a set of parameterized edges and then diffused into a random noise distribution. EdgeDiff learns both the noise reversal process and the network structure simultaneously. During inference, EdgeDiff iteratively refines the generated edge distribution using the denoising diffusion implicit model, enabling flexible single- or multi-step denoising and dynamic adaptation to buildings of varying complexity. Additionally, given the unique structure of wireframes, we introduce an edge attention module to extract point-wise attention from point features, using it as auxiliary information to facilitate learning of edge cues and guide the network toward improved edge awareness. Extensive experiments on the real-world Building3D dataset demonstrate that our approach achieves state-of-the-art performance.*

1. Introduction

Building reconstruction has emerged as a central challenge in computer vision and photogrammetry [9, 16], with far-reaching implications across domains including smart cities, autonomous navigation, and augmented and virtual reality. Among various geometric representations for building modeling, like point clouds [7], meshes [8], and signed distance functions [4, 36], 3D wireframes [31, 32, 44] exhibit unique advantages through its compact yet expressive point-line structure. However, reconstructing accurate wireframes from real-world point cloud data remains a challenging problem, primarily due to noise, incompleteness, and irregular sampling in the input measurements.

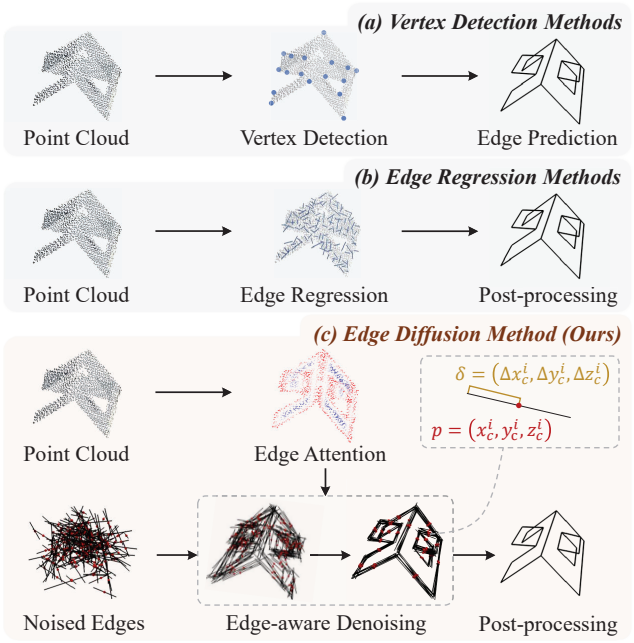


Figure 1. Comparison of building wireframe reconstruction paradigms. Different from existing methods based on vertex detection (a) and edge regression (b), EdgeDiff (c) proposes a novel noise-to-wireframe paradigm that iteratively refines edge parameters through edge-aware denoising.

Deep learning-based methods [14, 23, 26, 55] have shown tremendous potential in building wireframe reconstruction. Existing approaches can be broadly categorized into two paradigms: vertex detection methods [23, 26] and edge regression methods [14, 55]. The former first detects

†Corresponding author

vertices from point clouds and then predicts the topology of edges (See Fig. 1 (a)). While this decomposition appears intuitive, it inherently leads to error accumulation across stages. This insight has driven the emergence of edge regression methods [14, 55], which employ learnable queries to directly regress parameterized edges (See Fig. 1 (b)). Edge regression methods demonstrate superior stability and accuracy by effectively addressing the aforementioned limitations. This paradigm evolution reveals that wireframe reconstruction strategies are shifting from cue search inference methods toward regression-based methods for edge distributions. Note that popular diffusion models, as a unique distribution regression strategy, have demonstrated stronger distribution regression capabilities on images compared to methods based on standard transformers. Motivated by this, we first propose a wireframe reconstruction paradigm based on diffusion models, which iteratively refines the edge distribution from a noisy edge distribution, as shown in Fig. 1 (c).

The popular diffusion models [11, 35, 41] have emerged as a groundbreaking framework for generative modeling, demonstrating remarkable capabilities through their progressive denoising process. The exceptional performance of diffusion model in high-fidelity image synthesis [19, 39] has catalyzed their adoption across diverse 2D vision tasks, notably in object detection [2, 28, 48] and dense prediction [17, 22]. Motivated by these advances, researchers have extended diffusion-based approaches to 3D vision tasks, where the inherent geometric complexity of point cloud representation introduces unique challenges and opportunities. Pioneering works in point cloud completion [21, 29] and shape generation [24, 42] have demonstrated the effectiveness of noise-guided structure modeling in 3D space, highlighting the potential of diffusion models for complex geometric tasks. These successes can be attributed to two fundamental strengths of diffusion models: their capacity to model complex data distributions through controlled noise perturbation, and their iterative refinement mechanism that inherently aligns with progressive geometric understanding.

Inspired by these observations, the EdgeDiff based on conditional diffusion models is proposed for wireframe reconstruction (See Fig. 1 (c)). Our key insight is to reformulate wireframe reconstruction as a structured generative process that progressively evolves from random noise to precise geometric representations. We first parameterize wireframe using six-dimensional edge representations that include the midpoints of the edges and their offsets. During training, these parameterized edges undergo controlled diffusion into gaussian distributions. EdgeDiff then learns to reverse this process by iteratively refining noisy edges into accurate geometric structures, guided by point cloud. At inference stage, EdgeDiff enables flexible single- or multi-step denoising and dynamic adaptation to varying edge numbers, generating final wireframes from random initializations. Additionally, considering the crucial role of conditioning signals in diffusion models [39, 52], we leverage edge points as geometric cues [18, 47] and introduce an edge attention generation module that incorporates point-wise attention for edge-aware denoising. Extensive experiments on the Building3D dataset demonstrate that our approach achieves state-of-the-art performance, establishing diffusion-based approaches as a promising direction for wireframe reconstruction.

Our contributions are summarized in the following:

- We introduce EdgeDiff, the first noise-to-wireframe paradigm that leverages diffusion models for progressive wireframe reconstruction through iterative structure refinement.
- We develop a novel edge-aware denoising module incorporating an edge attention generation module, which enhances the denoising process by capturing fine-grained edge cues from point clouds.
- Comprehensive experiments conducted on the Building3D dataset demonstrate that EdgeDiff achieves state-of-the-art performance, showing the effectiveness of our noise-to-wireframe paradigm.

2. Related Work

Building Reconstruction. Traditional building reconstruction methods have predominantly adopted mesh-based representations as their output format [15, 45]. Representative approaches like 2.5D Dual Contouring [54] extend classical dual contouring algorithm [20], while Topology Aware [1] incorporates Manhattan World constraints for dense triangle surface mesh reconstruction. Despite the capability of capturing fine details, these methods incur substantial spatial complexity. City3D [13] advances PolyFit [34] by leveraging optimization-based techniques to generate polygon mesh models with controlled detail preservation. However, achieving desired results requires manual tuning of hyperparameters through empirical trial-and-error processes.

Recent advances in deep learning and large-scale datasets have sparked significant interest in wireframe reconstruction. Compared to mesh-based methods, wireframe models provide a more efficient representation by encoding structural information with minimal faces, particularly advantageous for large-scale scenarios. Early attempts [10, 12, 51] primarily focused on edge point detection for point cloud contour tracing, yet failed to generate complete wireframe representations. Several methods adopt a two-stage pipeline that first detects vertices followed by edge prediction, as demonstrated in PC2WF [26] and Point2Roof [23]. However, these approaches suffer from inaccurate corner and edge point detection, leading to degraded reconstruction quality. Recent studies have shifted towards direct edge regression [14, 55], eliminating the need for explicit

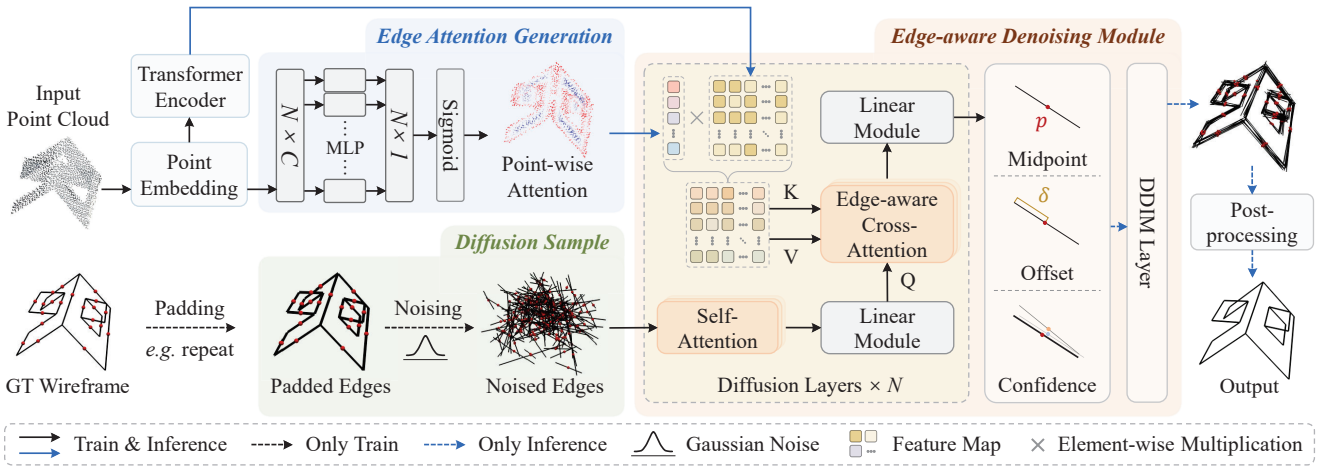


Figure 2. Overview of our proposed EdgeDiff. During training, ground truth wireframes are first padded and progressively diffused into gaussian noise. At inference, EdgeDiff transforms random gaussian into precise wireframe structures through edge-aware denoising, enabling flexible sampling steps and dynamic edge number adaptation.

vertex detection. PBWR [14] introduces a transformer-based architecture with learnable queries for direct edge regression from point clouds, achieving superior performance on the Building3D [44] dataset. In contrast to existing methods, we propose a novel noise-to-wireframe generative paradigm that iteratively refines noisy edge parameters into precise geometric structures via diffusion modeling with point cloud guidance.

Diffusion Model for Regression Tasks. Diffusion models have emerged as a powerful framework in generative tasks [19, 25, 40, 41, 49, 50], demonstrating exceptional capabilities in noise handling, flexible distribution modeling, and progressive refinement. Their success has recently extended to regression tasks, attracting substantial research interest [2, 17, 28, 30, 48]. Notable examples include DDP [17], which introduces a conditional diffusion process for dense pixel prediction under image guidance. DiffusionDet [2] reformulates object detection as a generative process, leveraging diffusion models to transform random boxes into precise detections. Similarly, DiffusionTrack [48] advances object tracking by representing targets through point sets, recasting tracking as a point set diffusion process. While diffusion-based approaches have demonstrated promising results in 3D tasks, including point cloud completion [21, 29] and shape generation [24, 42], their potential in wireframe reconstruction remains unexplored. Our work bridges this gap by introducing a noise-to-wireframe paradigm that enables novel possibilities for diffusion-based 3D geometric understanding.

3. Method

We first introduce preliminaries and problem formulation in Sec. 3.1. The proposed noise-to-wireframe paradigm and key components of EdgeDiff are then elaborated in Sec. 3.2, followed by the training and inference procedures detailed in Sec. 3.3 and Sec. 3.4.

3.1. Preliminaries

Problem Formulation. Our goal is to reconstruct a wireframe $\mathcal{W}$ from an input point cloud $P \in \mathbb{R}^{N\times 7}$, where each point contains coordinates, RGB values, and reflection intensity. We represent the wireframe $\mathcal{W}$ as a collection of edges $\mathcal{W} \in \mathbb{R}^{M\times 6}$, where M denotes the number of edges. Each edge $e^i \in \{\mathcal{W}|i = 1, \ldots, M\}$ is parameterized as a six-dimensional vector $e_i = (x_c^i, y_c^i, z_c^i, \Delta x_c^i, \Delta y_c^i, \Delta z_c^i)$. The first three components (x_c^i, y_c^i, z_c^i) specify the edge midpoint coordinates for accurate localization, while the remaining three components $(\Delta x_c^i, \Delta y_c^i, \Delta z_c^i)$ define the offset from the midpoint to an endpoint of the edge, determining both edge length and orientation.

To learn this representation, we introduce a novel diffusion-based learning framework that progressively refines noisy edges into geometrically precise structures. Our method first applies controlled noise injection to ground truth wireframes $\mathcal{W} = {e^i}_{i=1}^M$ during training, and then learns to reverse this process through iterative denoising. Unlike existing vertex detection or edge regression methods, we reformulate wireframe reconstruction as a novel noise-to-wireframe diffusion process.

Conditional Diffusion Model. Diffusion models [11, 41] constitute a powerful class of generative models characterized by forward noising and reverse denoising processes. The conditional variant [39, 52] extends this framework by incorporating a conditioning signal x. In the forward process, a target sample z_0 is progressively corrupted with

gaussian noise $\epsilon \sim \mathcal{N}(0, \mathbf{I})$ over T timesteps, defined as:

$$q(z_t|z_0) = \mathcal{N}(z_t; \sqrt{\bar{\alpha}_t}z_0, (1-\bar{\alpha}_t)\mathbf{I}) \quad (1)$$

where $\bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s)$, and $\beta_t \in (0,1)$ is a predefined variance schedule [35]. This process systematically transforms z_0 into increasingly noisy latents z_t at each timestep. The reverse process learns to reconstruct clean samples from noisy latents z_t conditioned on the input x, which is formulated as:

$$p_\theta(z_{t-1}|z_t, x) = \mathcal{N}(z_{t-1}; \mu_\theta(z_t, x, t), \Sigma_\theta(z_t, x, t)) \quad (2)$$

where μ_θ and Σ_θ are learnable mean and variance functions. To optimize this process, a neural network f_θ is trained to predict the clean sample z_0 from noisy observations z_t by minimizing the following objective:

$$L_{\text{train}} = \mathbb{E}_{z_0, \epsilon\sim\mathcal{N}(0,\mathbf{I})}\left[||f_\theta(z_t, x, t) - z_0||^2\right] \quad (3)$$

We formulate wireframe reconstruction within a conditional diffusion framework, where the ground truth wireframe z_0 serves as the data sample and building point cloud x as the condition. A neural network f_θ learns to recover z_0 from noisy latents z_t given x, iteratively refining parameterized edges into precise wireframe.

3.2. EdgeDiff Framework

The architecture of EdgeDiff comprises three key components: point encoder, edge attention generation, and edge-aware denoising module, as illustrated in Fig. 2. EdgeDiff takes as input point clouds and extracts conditional features, while the edge attention generator computes point-wise attention weights. These features and attention weights are then utilized by the edge-aware denoising module to progressively refine noisy edges into accurate wireframe representations. We elaborate on the training and inference procedures in Sec. 3.3 and Sec. 3.4, respectively.

Point Encoder. Drawing inspiration from point transformer architectures [33, 43, 53], point encoder first employs a point embedding module to extract initial features F_{embed}. These embeddings are then processed by a transformer encoder that leverages self-attention mechanisms to capture global structural patterns within the point cloud, yielding encoded features F_{enc}.

Edge Attention Generation. Prior works [39, 52] on conditional diffusion models reveal that the quality of conditioning signals directly affects generation accuracy. Furthermore, we observe that edge points in point clouds inherently encode crucial geometric information [18, 38, 47] for wireframe reconstruction. To leverage these edge cues, we propose an edge attention generation module for capturing edge-specific features. The module incorporates a lightweight MLP predictor that processes point cloud embedding features F_{embed} to extract edge information. The extracted features are normalized into attention weights $A \in (0,1)$ through sigmoid activation to guide the denoising process. We leverage ground truth (GT) edge attention weights derived from GT wireframes following Algorithm 1 for supervision. These GT weights encode geometric priors by integrating vertex distances, edge distances, and edge projection positions. The module optimization employs a weighted mean squared error loss for end-to-end network training.

Algorithm 1 Ground Truth Edge Attention Generation

Input: Input point cloud set P; The vertex set and edge set of the ground truth wireframe $\{V, E\}$.
Output: Point-wise attention weight A.
Hyperparameter: Balance ratio $r \in (0,1)$; Default projection factor $\alpha \in [0,1]$
Initialization: Distance factor list D; Projection factor list $F \leftarrow \alpha$

for all $P_i \in P$ **do**
  Search P_i's nearest vertex and nearest edge $\{V_i, E_i\}$.
  Compute the Euclidean distances d_{vi} and d_{ei}.
  if $d_{vi} > d_{ei}$ **then**
    Update the factor F with the distance f_i from the P_i's projection to the edge endpoint. $F_i \leftarrow f_i$
  Select the minimum distance d_i from d_{vi} and d_{ei}.
  Add d_i to list D. $D_i \leftarrow d_i$
Normalized distance factor. $D \leftarrow 1 - (D/max(D))$
Calculate weights by ratio r. $A \leftarrow D \cdot ((1-r) + r \cdot |2F - 1|)$
return A.

Edge-aware Denoising Module. The denoising Module comprises stacked diffusion layers, each integrating three core components: a linear module, self-attention mechanism, and edge-aware cross attention. Each layer leverages a self-attention module to process noisy edge queries for modeling global context, where queries (*Q*), keys (*K*), and values (*V*) are computed from noised edges. The linear module employs a lightweight feedforward network with normalization layers. We devise an edge-aware cross-attention mechanism that derives queries from self-attention outputs and computes *K* and *V* as weighted combinations of encoder features F_{enc} and edge attention A. These attention weights, obtained from our edge attention generation module, guide the denoising process to focus on edge-relevant features. This edge-aware architecture enhances the model's ability to capture fine-grained structural details and wireframe topology, particularly for complex architectural features and edge coherence. After multiple diffusion layers, MLP-based prediction heads generate final outputs including edge confidence scores, midpoint coordinates, and offset parameters.

3.3. Training

During the training process, we first diffuse the ground truth (GT) wireframe into a distribution of random gaussian noise. The model is then trained to reverse this process, allowing it to learn the wireframe structure of the point clouds.

GT Wireframe Padding. Following the formulation in Sec. 3.1, we represent wireframes as a collection of edges, where each edge is parameterized in a six-dimensional space. To accommodate the varying number of edges across different wireframes, we normalize the representation by padding each ground truth wireframe to a fixed number N_{train}, which acts as a predefined hyperparameter. Motivated by recent advances in diffusion-based regression [2, 28], we investigate multiple padding strategies for parameterized edges. Our empirical studies demonstrate that repeat padding achieves superior performance, as analyzed in Sec. 4.4.

Wireframe Diffusion Sample. The diffusion process injects gaussian noise into padded ground truth wireframes, where noise magnitude is controlled by α_t (defined in Eq. (1)). Following [35], we adopt a cosine schedule in which α_t decreases monotonically with timestep t. Additionally, we introduce a scaling mechanism for the GT wireframes, motivated by the observation that edge denoising requires less precise details compared to image generation tasks [3, 5]. This modification enables us to simplify the learning objective by enhancing the signal-to-noise ratio. A detailed analysis of this approach is presented in Sec. 4.4.

Training Loss. The network is trained under the supervision of ground truth wireframes and their corresponding point cloud edge attention weights. We adopt the geometric loss terms from PBWR [14], supplemented with a simple mean squared error loss to supervise the training of point cloud edge attention weights.

3.4. Inference

The inference process of EdgeDiff involves a denoising sampling procedure that transforms noisy edges into wireframes. The process initializes with a set of random edges sampled from a gaussian distribution $X_T \sim \mathcal{N}(0, \mathbf{I})$, and iteratively refines them to obtain the final wireframe X_0.

Sample Step. During inference, we employ a stepwise diffusion sampling procedure for wireframe generation. Starting from timestep T, the process iteratively generates edge sets with interval Δt until timestep zero. At each step, the model predicts edge parameters conditioned on current noise level and derives subsequent noise input following DDIM [41]. We enable flexible sampling configurations, supporting both single- and multi-step sampling to balance reconstruction quality and computational cost.

Dynamic Inference. EdgeDiff exhibits remarkable inference flexibility through its diffusion-based framework. While the model is trained using N_{train} edges, inference can be performed with N_{eval} edges without model retraining. In contrast to previous approaches [14] that employ fixed learnable queries, our method initializes from random noise and performs iterative denoising. This dynamic inference mechanism proves essential for wireframe reconstruction, where building structures involve varying geometric complexities demanding adaptive edge representations.

Post-processing. To ensure topological completeness,we follow the post-processing scheme of [14], which includes confidence-based non-maximum suppression (NMS) and DBSCAN clustering [6] to merge adjacent endpoints, preserving wireframe continuity and structural integrity.

4. Experiments

4.1. Datasets and Evaluation Metrics

Datasets. We evaluate our method on Building3D [44], a large-scale real-world dataset specifically designed for learning wireframe from point clouds. Building3D contains two subsets: Entry-level and Tallinn. The Entry-level subset, a subset of Tallinn with simple buildings, comprises 5,698 training samples and 583 testing samples, while the Tallinn subset includes 32,618 training samples and 3,472 testing samples, representing the full scale and complexity of urban environments.

Metrics. We adopt the evaluation metrics established by the Building3D [44] dataset to assess our approach. The evaluation encompasses both corner and edge accuracy in the reconstructed wireframe. Specifically, we employ the Average Corner Offset (ACO), computed as the mean Euclidean distance between predicted and ground-truth corners, and Corner Precision (CP), Corner Recall (CR), and their harmonic mean Corner F1-score (CF1). For edge-related evaluation, we use Edge Precision (EP), Edge Recall (ER), and Edge F1-score (EF1), where EF1 provides a balanced measure of precision and recall in edge reconstruction.

4.2. Implementation Details

Our framework is implemented in PyTorch [37] and all experiments are conducted on a single NVIDIA A100 GPU. The training process employs the AdamW [27] optimizer with parameters set to $\beta_1 = 0.9$, $\beta_2 = 0.999$, and a learning rate $\alpha = 5 \times 10^{-4}$. The diffusion timestep was set to 1000 with a cosine variance scheduler. The transformer encoder comprise 3-layer with 256-dimensional hidden units and 128-dimensional feed-forward networks. The diffusion layer has 8 layers, each with 256-dimensional representations for both hidden states and feed-forward networks. Both components utilize 4-head attention and a 0.1 dropout rate. For post-processing, we follow the PBWR [14] with distance, confidence, and NMS thresholds set to 0.2, 0.5, 0.2 in training and 0.1, 0.7, 0.15 in testing.

Type	Method	Distance (m) ACO	Corner CP	Corner CR	Corner CF1	Edge EP	Edge ER	Edge EF1
Entry-level dataset								
Building3D benchmark [44]	PointMAE*	0.36	0.71	0.50	0.59	0.81	0.26	0.39
	PointNet++*	0.34	0.79	0.52	0.63	0.84	0.33	0.47
	Stratified Transformer*	0.38	0.72	0.51	0.62	0.75	0.22	0.34
	Supervised	0.26	0.89	0.66	0.76	0.91	0.46	0.61
Vertex detection	PC2WF [26]	0.45	0.24	0.13	0.16	0.02	0.14	0.04
	Point2Roof [23]	0.30	0.66	0.48	0.56	0.71	0.26	0.38
Edge regression	PBWR [14]	**0.22**	0.97	**0.85**	**0.91**	0.91	**0.82**	0.86
Edge diffusion	EdgeDiff (Ours)	**0.22**	**0.99**	**0.85**	**0.91**	**0.96**	**0.82**	**0.88**
Tallinn dataset								
Building3D benchmark [44]	PointMAE*	0.33	0.75	0.47	0.58	0.52	0.12	0.20
	PointM2AE*	0.32	0.79	0.58	0.67	0.50	0.07	0.12
	Linear self-supervised	0.35	0.70	0.60	0.65	0.67	0.16	0.25
	Supervised	0.29	0.90	0.53	0.66	0.88	0.23	0.36
Vertex detection	PC2WF [26]	0.52	0.18	0.67	0.28	0.02	0.15	0.01
	Point2Roof [23]	0.39	0.65	0.30	0.41	0.66	0.08	0.14
Edge regression	PBWR [14]	**0.22**	0.96	0.68	0.80	0.91	0.65	0.76
Edge diffusion	EdgeDiff (Ours)	**0.22**	**0.98**	**0.74**	**0.84**	**0.92**	**0.70**	**0.79**

Table 1. Quantitative comparisons on the Building3D dataset [44]. Through the novel noise-to-wireframe paradigm, EdgeDiff achieves state-of-the-art performance on both subsets. $*$ indicates the method serves as the feature extractor in the baseline network of Building3D benchmark [44]. We highlight the **best** and our results.

4.3. Comparison with the State-of-the-Art Methods

Quantitative results. We conduct extensive evaluations of EdgeDiff on two subsets of the Building3D benchmark [44]. As shown in Tab. 1, EdgeDiff outperforms existing methods on the Entry-level dataset, achieving corner precision and recall scores of 0.99 and 0.85, respectively, substantially outperforming corner detection methods like Point2Roof [23] (0.66 and 0.48, respectively). For edge evaluation, our method attains precision and recall of 0.96 and 0.82, surpassing the previous state-of-the-art edge regression method PBWR [14].

On the more challenging Tallinn dataset, vertex detection methods suffer significant performance degradation, with Point2Roof [23] only achieving CR/ER of 0.30/0.08, revealing inherent limitations in complex architectural modeling. EdgeDiff maintains robust performance with CR/ER of 0.74/0.70, outperforming the prior state-of-the-art PBWR [14]. Through the integration of diffusion-based generative modeling, our approach achieves superior performance with CF1 of 0.84 and EF1 of 0.79, validating the effectiveness of diffusion model in wireframe reconstruction.

Qualitative results. We present extensive qualitative comparisons against existing methods on the Building3D dataset [44]. As shown in Fig. 3, existing deep learning approaches such as EC-Net [51] and PIE-Net [46] only extract partial edge points and fail to reconstruct complete wireframe. While vertex detection methods such as PC2WF [26] and Point2Roof [23] generate wireframe outputs, their inherent reliance on vertex leads to geometric inconsistencies and incorrect edge connectivity in complex structures. In contrast, EdgeDiff produces accurate wireframes with preserved building topology through robust feature learning enabled by generative noise injection.

For comprehensive evaluation, we demonstrate EdgeDiff reconstruction performance on large-scale scenes in Fig. 4 and present detailed comparisons against traditional mesh reconstruction methods in Fig. 5. While methods like PolyFit [34], Topology Aware [1], and 2.5D Dual [54] generate complete meshes, they introduce redundant facets that complicate building structures. EdgeDiff addresses this limitation by producing precise wireframe models that preserve structural features while reducing model complexity.

4.4. Ablation Studies

Parameterized Strategy. Our midpoint-offset parametrization significantly outperforms endpoint representation across all metrics, achieving improvements of 57% and

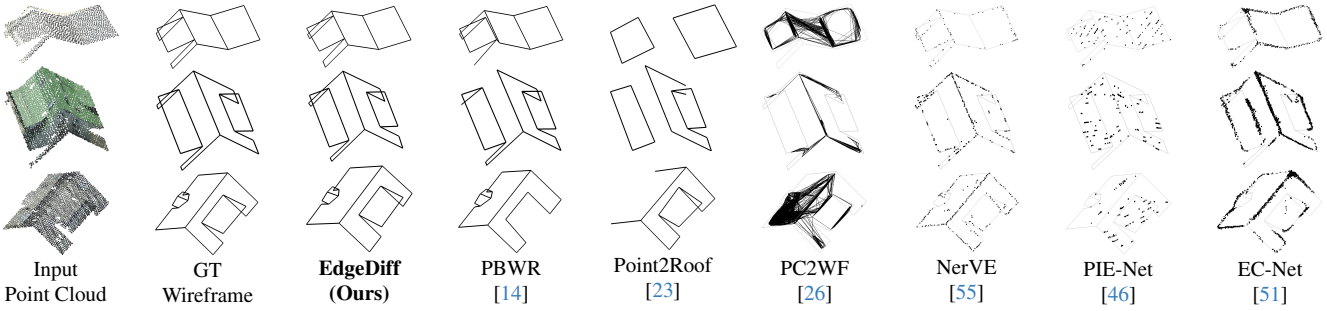


Figure 3. Qualitative comparisons on the Building3D dataset [44]. Ground truth wireframes are rendered in lighter shades to facilitate visual comparison across different methods. Additional visualizations in the *supplementary materials* validate generalization.

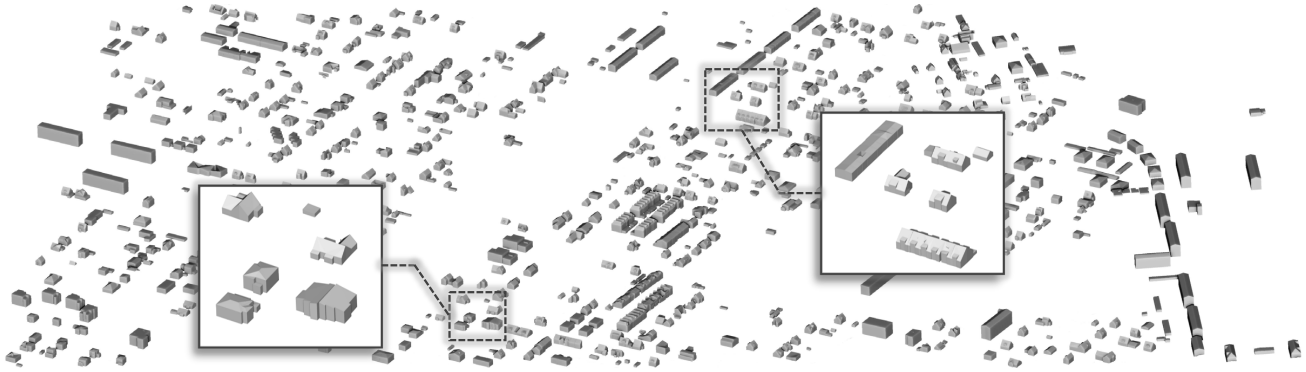


Figure 4. Reconstruction of Tallinn City with EdgeDiff. The wireframe reconstruction transformed into integrated roof and facade meshes, spanning a 1000 $\times$ 2000 m region in central Tallinn. The full city reconstruction is provided in the *supplementary materials*.

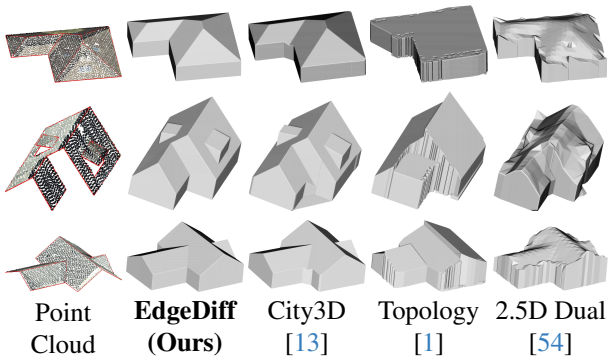


Figure 5. Qualitative comparisons between our EdgeDiff and building mesh reconstruction methods. EdgeDiff achieves more compact representations while preserving geometric fidelity.

69% in CF1 and EF1 scores respectively, as shown in Tab. 2. These gains arise from decomposing edge into stable center anchors and explicit direction-length encodings via decoupled parametrization. As illustrated in Fig. 6 (b), endpoint representation produces scattered predictions that hinder post-processing, while midpoint-offset parametrization maintains spatial consistency. This decomposition naturally aligns with the progressive optimization of diffusion models, where midpoint refinement during denoising establishes robust foundations for offset optimization.

Padding Type. Our evaluation of three wireframe padding types reveals insightful results, as shown in Tab. 2. The experiments indicate that simple edge repetition achieves better performance compared to alternative approaches. This superiority stems from building wireframes' inherent geometric properties, where similar edge patterns frequently recur within structures. Edge repetition preserves these structural regularities during training, facilitating more effective noise-to-wireframe mapping. Random sampling approaches introduce unnecessary geometric uncertainty, potentially hindering the learning of precise structural alignments crucial for architectural reconstruction.

Signal Scale. We analyze various scaling factors for normalized wireframe parameters during noise injection and inverse scaling. As indicated in Tab. 2, a scaling factor of 2.0 yields optimal performance for wireframe reconstruction. A factor of 1.0 results in insufficient signal strength against noise, hindering feature learning, while 3.0 leads to distortion through excessive amplification. The selected factor of 2.0 preserves geometric fidelity by maintaining an optimal signal-to-noise ratio, balancing structural preservation and noise robustness.

Edge Attention. As shown in Tab. 2, our model demonstrates strong performance even without edge attention, validating our noise-to-wireframe paradigm. While initial edge

Abaltion Study	Method	Distance (m) ACO	Corner CP	Corner CR	Corner CF1	Edge EP	Edge ER	Edge EF1
Entry-level dataset								
Parameterized Strategy	Endpoint	0.52	0.26	0.50	0.34	0.13	0.35	0.19
	Midpoint-Offset	**0.22**	**0.99**	**0.85**	**0.91**	**0.96**	**0.82**	**0.88**
Padding Type	Gaussian	0.23	0.98	0.81	0.89	0.94	0.77	0.84
	Uniform	0.23	0.98	0.82	0.89	0.94	0.78	0.85
	Repeat	**0.22**	**0.99**	**0.85**	**0.91**	**0.96**	**0.82**	**0.88**
Signal Scale	1.0	0.25	0.98	0.78	0.87	0.94	0.74	0.82
	2.0	**0.22**	**0.99**	**0.85**	**0.91**	**0.96**	**0.82**	**0.88**
	3.0	0.23	0.98	0.82	0.89	0.94	0.78	0.85
Edge Attention	w/o Edge Attention	0.29	0.96	0.83	0.89	0.88	0.80	0.83
	w/ Edge Attention	0.25	0.97	0.60	0.74	0.92	0.66	0.77
	w/ Edge Attention + Supervised	**0.22**	0.99	0.85	0.91	**0.96**	0.82	0.88
	w/ GT Edge Attention	**0.22**	**0.99**	**0.88**	**0.93**	0.94	**0.87**	**0.90**

Table 2. Ablation of different architecture designs. Default settings of EdgeDiff are marked in gray, with best results in **bold**.

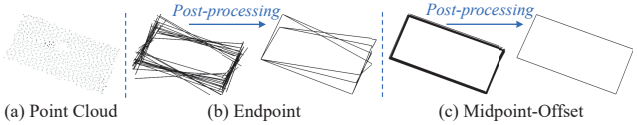


Figure 6. Comparison of wireframe reconstruction results with different parameterization strategies. (a) Input point cloud. (b) Endpoint parameterization represents edges. (c) Our proposed parameterization strategy using midpoint and offset.

attention integration shows marginal degradation due to insufficient supervision, incorporating a dedicated attention loss yields substantial improvements of 25% and 16% in CR and ER, respectively. Experiments with ground truth edge attention demonstrate optimal performance, validating the effectiveness of edge attention and revealing potential for further enhancement in attention prediction.

Matching between N_{train} and N_{eval} . We conduct extensive experiments to analyze how the number of training and inference edges influences model performance. As illustrated in Fig. 7 left, we first evaluate matching configurations where $N_{train} = N_{eval} \in \{100, 200, 300, 500\}$. The results reveal that higher edge numbers consistently yield better EF1 scores, demonstrating enhanced reconstruction quality with richer edge representations. We further investigate different inference settings in Fig. 7 right by fixing $N_{train} = 300$ and varying N_{eval} from 100 to 1000. While increasing N_{eval} consistently improves performance, it also leads to longer inference time. Our analysis suggests that $N_{train} = N_{eval} = 300$ achieves an optimal balance between reconstruction quality and computational efficiency, while confirming EdgeDiff enables inference with arbitrary edge counts.

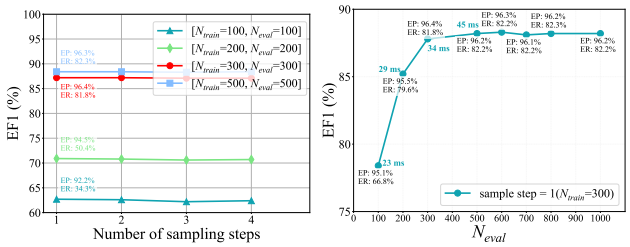


Figure 7. Left: impact of matched training and inference edge numbers on EF1 (%). Right: variation with fixed training edge number and differing inference edge number on EF1 (%).

5. Conclusion

In this work, we propose EdgeDiff, the first diffusion-based framework that reformulates wireframe reconstruction as a noise-to-wireframe generation process, which is a novel wireframe reconstruction paradigm. Our approach exhibits several appealing properties, including edge attention guidance for accurate structure learning and dynamic inference. Extensive experiments on the Building3D dataset demonstrate that EdgeDiff significantly outperforms existing methods while maintaining geometric fidelity, validating the effectiveness of diffusion model for structured 3D reconstruction tasks. Future research will explore more concise frameworks that directly generate global wireframes without relying on post-processing steps.

Acknowledgement. This work is supported by the National Key Research and Development Program of China (2022YFB2602105), the Key Technological Innovation Program of Ningbo City under Grant No. 2024Z297, and the Technical Service Project for Real-Scene 3D Guangxi Construction under Grant No. KWBD5C2024025.

References

[1] Dong Chen, Ruisheng Wang, and Jiju Peethambaran. Topologically aware building rooftop reconstruction from airborne laser scanning point clouds. *IEEE Transactions on Geoscience and Remote Sensing*, 55(12):7032–7052, 2017. 2, 6, 7

[2] Shoufa Chen, Peize Sun, Yibing Song, and Ping Luo. Diffusiondet: Diffusion model for object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 19830–19843, 2023. 2, 3, 5

[3] Ting Chen, Lala Li, Saurabh Saxena, Geoffrey Hinton, and David J Fleet. A generalist framework for panoptic segmentation of images and videos. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 909–919, 2023. 5

[4] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5939–5948, 2019. 1

[5] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021. 5

[6] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, pages 226–231, 1996. 5

[7] Haoqiang Fan, Hao Su, and Leonidas J Guibas. A point set generation network for 3d object reconstruction from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 605–613, 2017. 1

[8] Thibault Groueix, Matthew Fisher, Vladimir G Kim, Bryan C Russell, and Mathieu Aubry. A papier-mâché approach to learning 3d surface generation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 216–224, 2018. 1

[9] Jianwei Guo, Haobo Qin, Yinchang Zhou, Xin Chen, Liangliang Nan, and Hui Huang. Fast building instance proxy reconstruction for large urban scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(11):7267–7282, 2024. 1

[10] Chems-Eddine Himeur, Thibault Lejemble, Thomas Pellegrini, Mathias Paulin, Loic Barthe, and Nicolas Mellado. Pcednet: A lightweight neural network for fast and interactive edge detection in 3d point clouds. *ACM Transactions on Graphics (TOG)*, 41(1):1–21, 2021. 2

[11] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 2, 3

[12] Zeyu Hu, Mingmin Zhen, Xuyang Bai, Hongbo Fu, and Chiew-lan Tai. Jsenet: Joint semantic segmentation and edge detection network for 3d point clouds. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16*, pages 222–239. Springer, 2020. 2

[13] 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, 7

[14] 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, 3, 5, 6, 7

[15] ZhangJin Huang, Yuxin Wen, ZiHao Wang, Jinjuan Ren, and Kui Jia. Surface reconstruction from point clouds: A survey and a benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):9727–9748, 2024. 2

[16] Christian Häne, Christopher Zach, Andrea Cohen, and Marc Pollefeys. Dense semantic 3d reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9): 1730–1743, 2017. 1

[17] Yuanfeng Ji, Zhe Chen, Enze Xie, Lanqing Hong, Xihui Liu, Zhaoqiang Liu, Tong Lu, Zhenguo Li, and Ping Luo. Ddp: Diffusion model for dense visual prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 21741–21752, 2023. 2, 3

[18] Li Jiang, Hengshuang Zhao, Shu Liu, Xiaoyong Shen, Chi-Wing Fu, and Jiaya Jia. Hierarchical point-edge interaction network for point cloud semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10433–10441, 2019. 2, 4

[19] Yuming Jiang, Tianxing Wu, Shuai Yang, Chenyang Si, Dahua Lin, Yu Qiao, Chen Change Loy, and Ziwei Liu. Videobooth: Diffusion-based video generation with image prompts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6689–6700, 2024. 2, 3

[20] Tao Ju, Frank Losasso, Scott Schaefer, and Joe Warren. Dual contouring of hermite data. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 339–346, 2002. 2

[21] Yoni Kasten, Ohad Rahamim, and Gal Chechik. Point cloud completion with pretrained text-to-image diffusion models. *Advances in Neural Information Processing Systems*, 36, 2024. 2, 3

[22] Hsin-Ying Lee, Hung-Yu Tseng, and Ming-Hsuan Yang. Exploiting diffusion prior for generalizable dense prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7861–7871, 2024. 2

[23] 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. 1, 2, 6, 7

[24] Muheng Li, Yueqi Duan, Jie Zhou, and Jiwen Lu. Diffusion-sdf: Text-to-shape via voxelized diffusion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12642–12651, 2023. 2, 3

[25] Xiaoming Li, Xinyu Hou, and Chen Change Loy. When stylegan meets stable diffusion: a w+ adapter for personalized image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2187–2196, 2024. 3

[26] Yujia Liu, Stefano D'Aronco, Konrad Schindler, and Jan D Wegner. Pc2wf: 3d wireframe reconstruction from raw point

clouds. In *International Conference on Learning Representations (ICLR)*, 2021. 1, 2, 6, 7

[27] I Loshchilov. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 5

[28] Run Luo, Zikai Song, Lintao Ma, Jinlin Wei, Wei Yang, and Min Yang. Diffusiontrack: Diffusion model for multi-object tracking. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 3991–3999, 2024. 2, 3, 5

[29] Shitong Luo and Wei Hu. Diffusion probabilistic models for 3d point cloud generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2837–2845, 2021. 2, 3

[30] Weiyi Lv, Yuhang Huang, Ning Zhang, Ruei-Sung Lin, Mei Han, and Dan Zeng. Diffmot: A real-time diffusion-based multiple object tracker with non-linear prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19321–19330, 2024. 3

[31] Xueqi Ma, Yilin Liu, Wenjun Zhou, Ruowei Wang, and Hui Huang. Generating 3d house wireframes with semantics. In *European Conference on Computer Vision*, pages 223–240. Springer, 2025. 1

[32] D Man and A Vision. A computational investigation into the human representation and processing of visual information. *WH San Francisco: Freeman and Company, San Francisco*, 1:1, 1982. 1

[33] Ishan Misra, Rohit Girdhar, and Armand Joulin. An end-to-end transformer model for 3d object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 2906–2917, 2021. 4

[34] Liangliang Nan and Peter Wonka. Polyfit: Polygonal surface reconstruction from point clouds. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2353–2361, 2017. 2, 6

[35] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pages 8162–8171. PMLR, 2021. 2, 4, 5

[36] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 165–174, 2019. 1

[37] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 5

[38] Wonseok Roh, Hwanhee Jung, Giljoo Nam, Jinseop Yeom, Hyunje Park, Sang Ho Yoon, and Sangpil Kim. Edge-aware 3d instance segmentation network with intelligent semantic prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20644–20653, 2024. 4

[39] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 2, 3, 4

[40] Ludan Ruan, Yiyang Ma, Huan Yang, Huiguo He, Bei Liu, Jianlong Fu, Nicholas Jing Yuan, Qin Jin, and Baining Guo. Mm-diffusion: Learning multi-modal diffusion models for joint audio and video generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10219–10228, 2023. 3

[41] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising Diffusion Implicit Models. In *ICLR*, 2021. 2, 3, 5

[42] Arash Vahdat, Francis Williams, Zan Gojcic, Or Litany, Sanja Fidler, Karsten Kreis, et al. Lion: Latent point diffusion models for 3d shape generation. *Advances in Neural Information Processing Systems*, 35:10021–10039, 2022. 2, 3

[43] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017. 4

[44] 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. 1, 3, 5, 6, 7

[45] Tao Wang, Jing Wu, Ze Ji, and Yu-Kun Lai. Sparse convolutional networks for surface reconstruction from noisy point clouds. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3212–3221, 2024. 2

[46] Xiaogang Wang, Yuelang Xu, Kai Xu, Andrea Tagliasacchi, Bin Zhou, Ali Mahdavi-Amiri, and Hao Zhang. Pie-net: Parametric inference of point cloud edges. *Advances in neural information processing systems*, 33:20167–20178, 2020. 6, 7

[47] Chengzhi Wu, Junwei Zheng, Julius Pfrommer, and Jürgen Beyerer. Attention-based point cloud edge sampling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5333–5343, 2023. 2, 4

[48] Fei Xie, Zhongdao Wang, and Chao Ma. Diffusiontrack: Point set diffusion model for visual object tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19113–19124, 2024. 2, 3

[49] Yazhou Xing, Yingqing He, Zeyue Tian, Xintao Wang, and Qifeng Chen. Seeing and hearing: Open-domain visual-audio generation with diffusion latent aligners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7151–7161, 2024. 3

[50] Yanwu Xu, Yang Zhao, Zhisheng Xiao, and Tingbo Hou. Ufogen: You forward once large scale text-to-image generation via diffusion gans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8196–8206, 2024. 3

[51] Lequan Yu, Xianzhi Li, Chi-Wing Fu, Daniel Cohen-Or, and Pheng-Ann Heng. Ec-net: an edge-aware point set consolidation network. In *Proceedings of the European conference on computer vision (ECCV)*, pages 386–402, 2018. 2, 6, 7

[52] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3836–3847, 2023. 2, 3, 4

[53] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip HS Torr, and Vladlen Koltun. Point transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 16259–16268, 2021. 4

[54] Qian-Yi Zhou and Ulrich Neumann. 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*, pages 115–128. Springer, 2010. 2, 6, 7

[55] Xiangyu Zhu, Dong Du, Weikai Chen, Zhiyou Zhao, Yinyu Nie, and Xiaoguang Han. Nerve: Neural volumetric edges for parametric curve extraction from point cloud. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13601–13610, 2023. 1, 2, 7