RoofIt: 2D Parametric Sketch Extraction from Rough Models for High-quality Roof Reconstruction

Xiaogang Wang
Southwest University
Chongqing, China
wangxiaogang@swu.edu.cn

Yuncong Liu
Southwest University
Chongqing, China
a646023527@email.swu.edu.cn

Yu Zhang*
Beijing Freedo Technology Co., Ltd.
Beijing, China
zhangyu@freedotech.com

Abstract

3D building roof reconstruction remains a challenging task in urban modeling, particularly when processing rough roof models containing noise and structural imperfections. While existing methods have achieved significant progress, they often fail to maintain reconstruction quality when handling such inputs. This paper presents a unified reconstruction pipeline that transforms rough structured roof models into high-quality reconstructions by integrating 2D sketch optimization. Our approach begins by applying statistical priors to segment the input rough roof model into planar primitives with constrained orientations. We then extract 2D roof sketches and establish precise correspondences between sketch elements and planar segments. A structure-aware optimization process enforces geometric constraints to refine these sketches. The optimized parametric sketches guide the correction of planar primitives, which are subsequently fused into an enhanced 3D roof surface representation. Experimental evaluations demonstrate that our method outperforms current baseline approaches, achieving superior reconstruction accuracy while effectively handling noise and structural irregularities inherent in rough roof models. The proposed pipeline proves particularly effective in recovering sharp features and maintaining surface regularity from challenging input data.

CCS Concepts

• **Computing methodologies** → **Mesh geometry models**.

Keywords

roof reconstruction, 3D shape analysis, sketch analysis, planar primitives

ACM Reference Format:
Xiaogang Wang, Yuncong Liu, and Yu Zhang. 2026. RoofIt: 2D Parametric Sketch Extraction from Rough Models for High-quality Roof Reconstruction. In . ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

*Corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
Conference'17, Washington, DC, USA
© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Three-dimensional building reconstruction is a fundamental task in digital city applications such as urban planning, mapping, simulation, and virtual reality. Among building components, roofs remain particularly challenging due to their complex geometries and the noisy, incomplete, and irregularly sampled nature of input data. Recovering clean, regular, and structurally faithful roof models from such observations therefore remains an important problem.

Existing roof reconstruction methods can be broadly grouped into footprint-based, learning-based, and plane-based approaches. Footprint-based methods [27, 29] rely on 2D contours extracted from aerial imagery, but their performance degrades when roof structures are complex or contour extraction is unreliable. Learning-based methods [4, 25] exploit data-driven shape priors and often achieve strong reconstruction fidelity, yet they may fail to preserve the sharp geometric regularity required in practical roof modeling. Plane-based methods [10, 16] generate regular surfaces through planar primitive fitting and intersection, but are highly sensitive to errors in plane detection and geometric intersection, especially under severe noise or missing observations.

A central difficulty in roof reconstruction is the simultaneous need for robustness, structural accuracy, and geometric regularity. In rough 3D roof data, these goals are tightly coupled: noise and non-uniform sampling degrade local plane estimation, incomplete observations destabilize plane intersections, and even small angular deviations can substantially distort characteristic roof features. As a result, direct optimization in 3D is often unstable, since primitive parameters are difficult to estimate reliably and structural relations are hard to infer from corrupted observations alone.

A key observation of this work is that roof structure is reflected not only by 3D planar primitives, but also by the arrangement of projected feature lines. Even when the input geometry is degraded, many roof-defining features remain more stable in suitable 2D projected domains, where ridges, valleys, and boundaries admit compact parametric representations. This suggests that 2D sketches provide a more effective intermediate representation for capturing structural regularities than direct 3D fitting alone.

However, a single 2D projection is usually insufficient for complex roofs, since different components may overlap, degenerate, or become ambiguous under one projection. To address this issue, we construct multi-level parametric sketches over projected domains, so that complementary structural information can be preserved while maintaining explicit correspondence with 3D planar primitives.

Based on this idea, we propose a reconstruction framework with three tightly coupled components: multi-level parametric sketch extraction, explicit correspondence establishment between

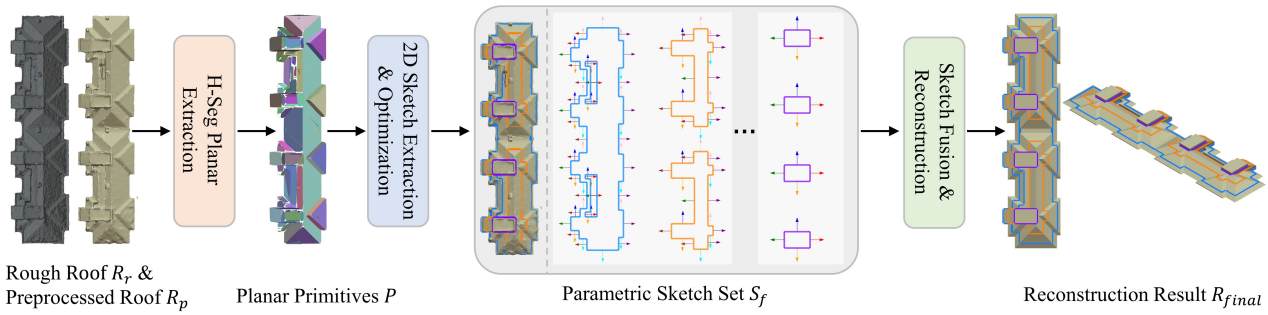

Figure 1: Overview of the proposed method. We first preprocess the input rough roof model R_r by smoothing and removing irrelevant structures. Next, we perform Hierarchical-Segmentation-Based Planar Extraction (H-Seg Planar Extraction) on the preprocessed model R_p to obtain the initial planar primitives P. Then, in 2D Sketch Extraction and Optimization, we extract and optimize leveled 2D sketches along the z-axis from the point cloud associated with P, and obtain the filtered sketch set S_f. Finally, in Sketch Fusion and Reconstruction, we fuse S_f, transfer the 2D sketch corrections back to 3D, convert them into planar primitives, and reconstruct the final high-quality roof model R_{final}.

2D sketches and 3D planar primitives, and bidirectional propagation of structural corrections from 2D constraints to 3D geometry. In our framework, 2D sketches are not merely auxiliary abstractions, but the main structural representation for separating stable roof organization from noisy geometric observations. The optimized 2D relations are then transferred back to 3D to regularize planar primitives and improve reconstruction quality.

Experimental results demonstrate that the proposed representation is effective for noisy roof models with strong structural organization. Compared with existing plane-based methods, our approach is more robust to imperfect input; compared with learning-based methods, it better preserves roof features and geometric regularity.

The main contributions of this work are summarized as follows:

- We propose a structured roof reconstruction framework that uses multi-level 2D parametric sketches as an intermediate representation to refine rough 3D roof geometry.
- We introduce an explicit correspondence mechanism between 2D sketches and 3D planar primitives, enabling reliable cross-domain reasoning under noisy input conditions.
- We develop a bidirectional optimization scheme that propagates structural constraints inferred in 2D back to 3D planar correction, leading to accurate and geometrically regular reconstruction of structured roof models.

2 Related Works

2.1 Roof Reconstruction

Roof reconstruction is a fundamental task in urban 3D modeling, as roof geometry largely determines the fidelity of reconstructed buildings. Early studies mainly relied on geometric reasoning and optimization. For example, Chen et al. [2] reconstructed buildings from airborne LiDAR point clouds through primitive clustering and geometric modeling. Huang et al. [7] adopted a top-down strategy by matching observations to a predefined roof library, while Li et al. [12] represented buildings as combinations of aligned boxes. PolyFit [16] formulated polygonal reconstruction as a binary optimization problem, and Zhou et al. [31] improved reconstruction quality by combining local planar fitting with global regularity constraints. Other works also explored robust boundary extraction and graph-based roof topology modeling [18, 19].

Another line of research incorporates learning-based priors or explicit model selection mechanisms. Henn et al. [6] used supervised learning to infer plausible roof models from handcrafted features. More recently, deep networks have been introduced to directly predict roof structures from measured data. Point2Roof [11] reconstructed 3D roof wireframes from airborne LiDAR point clouds, and Lu et al. [15] improved topology inference by learning frequency-domain geometric features. Planes4LoD2 [20] combined neural detection with vectorized roof reconstruction from images. In parallel, the release of large-scale benchmarks such as Building3D [24] has further promoted learning-based roof reconstruction methods [9, 13, 14].

Despite this progress, existing methods remain limited in our target setting. Real-world roof scans often suffer from severe noise, missing observations, and complex structures. Traditional optimization-based methods are sensitive to these imperfections, while recent learning-based approaches [9, 13, 14] have mainly been evaluated on roofs with relatively simple topologies. In contrast, our method is designed for noisy and structurally complex roof data and aims to produce high-fidelity reconstructions under challenging real-world conditions.

2.2 High-Quality 3D Reconstruction

High-quality 3D reconstruction has mainly developed along two directions: geometry-driven optimization and data-driven neural methods.

Traditional geometry-based approaches recover structured surfaces by fitting primitives to point clouds or meshes. Probabilistic sampling methods such as RANSAC are widely used to detect reliable primitives from noisy data [21], while variational methods improve reconstruction quality through global energy optimization [3, 22]. For polygonal reconstruction, recent studies have further emphasized compact and watertight surface generation through plane arrangement techniques [1, 23].

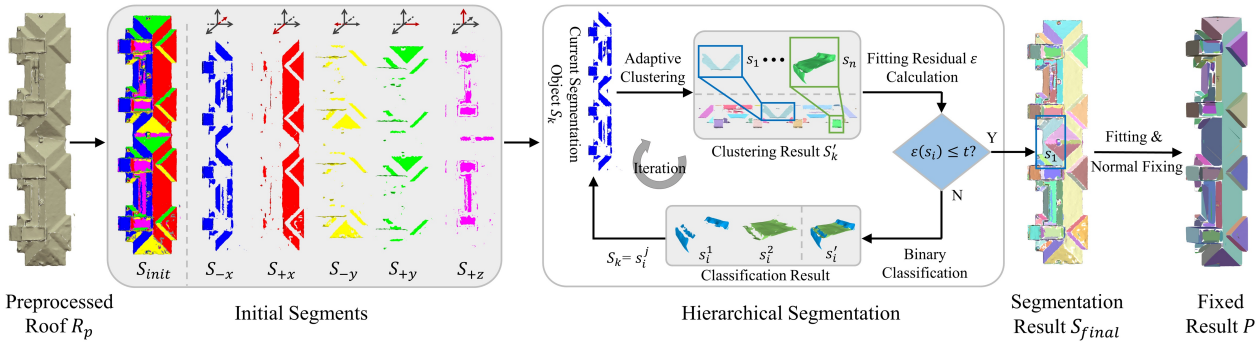

Figure 2: H-seg Planar Extraction: Hierarchical-Segmentation-Based Planar Extraction.

Neural approaches have recently emerged as an important alternative by learning shape priors directly from data. PrimitiveNet [8] and HPNet [26] jointly predict primitive types and parameters for input points. Nerve [32] reconstructs 3D wireframes from piecewise linear curve representations, while NeurCADRecon [4] improves surface smoothness and fidelity with curvature-aware constraints. PolyGen [17] instead generates polygonal models autoregressively by predicting vertices and faces.

Different from these general-purpose reconstruction methods, our goal is high-quality roof modeling from coarse and noisy 3D observations. We therefore introduce a dedicated pipeline that optimizes extracted 2D roof contour sketches, propagates the corrections back to 3D geometry, and combines the refined structure with polygonal reconstruction to produce accurate roof models.

3 Method

Our method mainly consists of three core modules, as shown in Figure 1. We take a rough 3D roof model R_r as input, and obtain R_p through preprocessing before it enters the core modules. First, through the hierarchical-segmentation-based planar extraction(H-seg planar extraction), we perform hierarchical segmentation on R_p and extract the initial set of planar primitives P. Next, we extract multi-level 2D sketches from the sampled points $\mathcal{P}$ of R_p, obtaining the 2D sketch set S. We then denoise and optimize each level of sketches in S, and after filtering, obtain the refined 2D sketch set S_f. Finally, we select a key sketch from S_f and fuse the remaining sketches with it as the foundation. The fusion sketch guide the correction of planar primitive set P, which are subsequently fused into an enhanced 3D roof surface representation.

3.1 Preprocess

For rough 3D roof mesh models with significant noise, preprocessing is required as they are difficult to be segmented directly. The first preprocessing step is smoothing. Additionally, fence-like structures often appear around the roof's bottom edges due to data acquisition artifacts. These structures do not belong to the roof itself and introduce segmentation noise, so we remove them. This preprocessing ensures cleaner input data for subsequent hierarchical segmentation and 2D contour extraction.

3.2 H-seg planar extraction

As illustrated in Figure 2, we extract roof planar primitives from R_p using a hierarchical segmentation strategy. Here, a planar primitive denotes a *bounded planar patch* rather than an infinite plane. Each primitive is represented by a plane equation together with its inlier points, which define its orientation and spatial support for subsequent correspondence and optimization.

Statistical analysis shows that roof surfaces typically align with no more than nine dominant directions (see the supplementary material for details). To exploit this prior and simplify the initial classification, we adopt an axis-aligned orientation assumption and discretize the normal space into a small set of canonical directions. Specifically, each triangular face f_i in R_p is assigned to one of six orientation categories

$$C = \{+x, -x, +y, -y, +z, null\},$$

where the first five categories correspond to the reference direction set

$$V = \{\vec{v}_{+x}, \vec{v}_{-x}, \vec{v}_{+y}, \vec{v}_{-y}, \vec{v}_{+z}\}.$$

For each face f_i, we compute the angle θ_{ij} between its normal $\vec{n}_i$ and each reference direction $\vec{v}_j \in V$, and assign its category by

$$c_i = \begin{cases} C_j, & \text{if } \min_j \theta_{ij} < \theta_{th}, \\ null, & \text{otherwise}, \end{cases} \tag{1}$$

where θ_{th} is an angular threshold. Faces labeled as *null* are excluded from subsequent plane extraction, yielding the initial classification

$$S_{init} = \{S_{+x}, S_{-x}, S_{+y}, S_{-y}, S_{+z}\}.$$

We then refine each orientation class $S_k \in S_{init}$ by Euclidean clustering, producing a set of connected face subsets $S'_k = \{s_i\}_{i=1}^N$. Each subset s_i is approximated as a bounded planar primitive s_i^p. To robustly estimate its supporting plane, we first determine the dominant normal direction of s_i through voting. Specifically, face normals within an angular tolerance are treated as identical, and the dominant normal $\vec{n}_i^p$ is defined as

$$\vec{n}_i^p = \arg\max_{\mathbf{n}} \sum_{k=1}^{N} \mathbb{I}(\mathbf{n}_k \cdot \mathbf{n} \ge \cos\epsilon), \tag{2}$$

where ϵ is the normal similarity threshold.

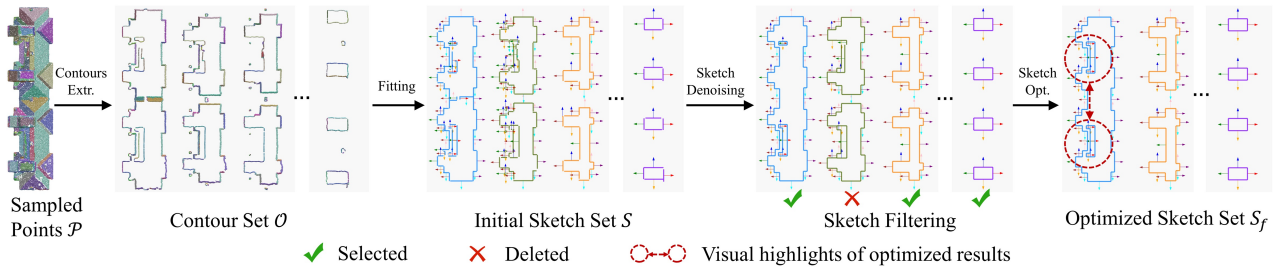

Figure 3: 2D Sketch Extraction and Optimization.

After obtaining $\vec{n}_i^p$, we sample points from s_i to form $\mathcal{P}_i$, and downsample them to $\hat{\mathcal{P}}_i$. Combining $\hat{\mathcal{P}}_i$ with $\vec{n}_i^p$, we construct candidate supporting planes in point-normal form and select the one with the minimum fitting residual.

If the minimum residual $\varepsilon(s_i)$ exceeds a threshold t, s_i is regarded as insufficiently planar and is further subdivided. Specifically, we perform a binary reclassification of the faces in s_i according to whether their orientations are closer to the z-axis or to the x-/y-axes, yielding $s_i' = \{s_i^j \subset s_i\}_{j=1}^2$. The same estimation and verification procedure is then recursively applied to each subset s_i^j until all segments satisfy the planar fitting criterion.

After recursion terminates, we obtain the final segmentation result S_{final}. Since the plane orientations in S_{final} may still be locally inconsistent, we further snap them to a set of at most nine standard directions, producing the final planar primitive set P. Each primitive in P is a bounded planar patch represented by its plane equation and an interior point. This hierarchical strategy extracts structurally meaningful planar primitives while effectively suppressing over-segmentation and fragmented faces.

3.3 Parametric Sketch Extraction and Optimization

As illustrated in Figure 3, we extract a set of 2D parametric sketches from the inlier point set $\mathcal{P}$ derived from the segmented planar primitives. Note that the sketch is not obtained by directly projecting the boundaries of 3D planar primitives. Instead, each primitive contributes through its inlier points, which are voxelized and organized into horizontal contour layers, from which 2D line structures are extracted. The sketch therefore serves as an intermediate representation of the projected structural arrangement of planar primitives.

We first voxelize $\mathcal{P}$ to generate a uniform voxel grid $V = \{v_{x,y,z}\}$. Based on voxel occupancy along the vertical direction, the occupied voxels are organized into contour levels $O = \{o_k\}_{k=1}^M$, where the k-th contour level is defined as

$$o_k = \{v_{x,y,k} \mid x \in [1, X] \cap \mathbb{Z},\ y \in [1, Y] \cap \mathbb{Z}\}. \tag{3}$$

Here, X and Y denote the voxel resolution in the x- and y-directions. Each voxel $v_{x,y,k}$ stores the subset of sampled points falling inside it, $p_{x,y,k} = \{p_i \in \mathcal{P}\}_{i=1}^N$, and thus implicitly aggregates the primitive information inherited from these points.

For each contour level o_k, we project the occupied voxels onto the XOY plane and assign a 2D orientation label to each projected voxel. This orientation is inferred from the planar primitives associated with the contained points rather than from voxel boundaries. Specifically, for each point $p_i \in p_{x,y,k}$, we retrieve the normal vector $\vec{n}_{p_i} = (n_x, n_y, n_z)$ of its corresponding planar primitive and project it onto the XOY plane. Since the corrected roof normals are constrained to the nine standard directions and satisfy $n_x \cdot n_y = 0$, the projected normal can only take one of four axis-aligned directions:

$$\vec{n}_{p_i}^{proj} = \begin{cases} (1,0), & \text{if } n_x > 0, \\ (-1,0), & \text{if } n_x < 0, \\ (0,1), & \text{if } n_y > 0, \\ (0,-1), & \text{if } n_y < 0. \end{cases} \tag{4}$$

These directions are mapped to four 2D orientation labels

$$D = \{up, down, left, right\}.$$

The mapping is defined as

$$label(\vec{n}_{p_i}^{proj}) = \begin{cases} right, & \text{if } \vec{n}_{p_i}^{proj} = (1,0), \\ left, & \text{if } \vec{n}_{p_i}^{proj} = (-1,0), \\ up, & \text{if } \vec{n}_{p_i}^{proj} = (0,1), \\ down, & \text{if } \vec{n}_{p_i}^{proj} = (0,-1). \end{cases} \tag{5}$$

For each projected voxel, the final orientation label is determined by majority voting over all contained points. Since the contour hierarchy is built along the z-axis, primitives with normals aligned with the positive z-direction do not provide directional information in the XOY projection and are excluded from voting. Each occupied voxel is thus assigned a dominant 2D orientation label together with an aggregated planar primitive label.

Within each contour level, adjacent voxels sharing the same orientation label are grouped into connected components, each representing a locally coherent directional pattern in the projected plane. We fit a 2D line segment to each component, obtaining the parametric sketch set $S = \{\hat{s}_k\}_{k=1}^M$. Meanwhile, planar primitive labels are transferred from voxels to line segments through a two-stage voting strategy:

(1) **Voxel-level primitive assignment.** For each voxel, we count the primitive labels of the points in $p_{x,y,k}$ and assign the most frequent one.
(2) **Segment-level primitive assignment.** For each fitted line segment, we collect the primitive labels of its associated voxels and assign the majority label.

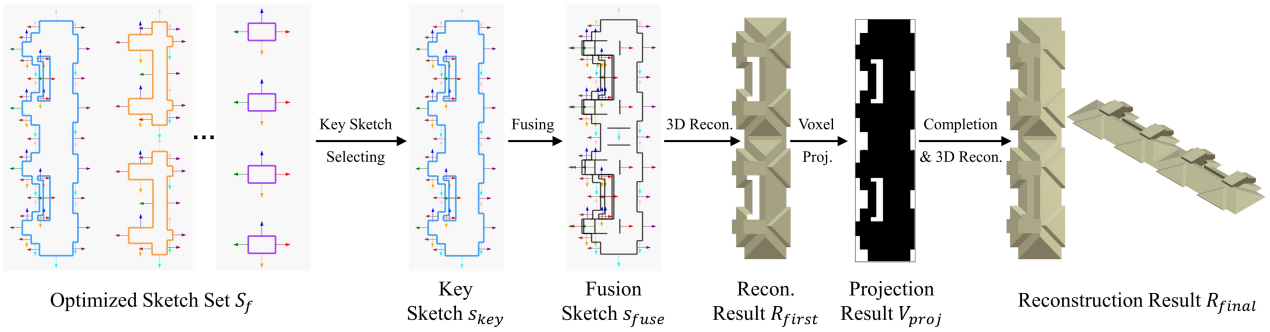

Figure 4: Sketch Fusion and Reconstruction.

This establishes an explicit correspondence between 3D planar primitives and 2D sketch elements.

After line fitting, adjacent collinear segments are merged to simplify topology and improve geometric consistency, while extremely short segments are removed. We further eliminate asymmetric short segments, where symmetry is evaluated with respect to the horizontal center of the sketch bounding box. Next, we construct a graph whose nodes are segment endpoints and whose edges are the line segments, and remove nodes of degree one. These operations mainly suppress unstable structures and isolated noise, producing the denoised sketch set S_d.

We then filter S_d according to closure and structural support. Sketches located at low contour levels but associated with only a few planar primitives are regarded as non-contributory, since they usually correspond to noise or weak structural evidence, and are therefore discarded. This yields the selected sketch set S_{sel}.

For each selected sketch $s^i_{sel} \in S_{sel}$, we further enforce collinearity and symmetry constraints on its line segments. In addition, segments at different contour levels but sharing the same planar primitive label are required to remain consistent with the corresponding corrected plane. We also perform symmetry completion with respect to the horizontal center of the sketch bounding box: each connected component should either have a symmetric counterpart or be self-symmetric. If a component lacks its counterpart, the mirrored component is added. After these corrections, we obtain the optimized sketch set S_f.

The optimized sketches are then used to refine the original planar primitives. Based on the orientation and primitive labels of the line segments, we recompute supporting planes and derive clipping boundaries from the sketch topology. For each line segment l_i, we first compute an initial supporting plane p_i from its two endpoints and the normal vector associated with its planar primitive label. Let the two adjacent line segments of l_i be denoted by l^1_i and l^2_i, with corresponding supporting planes p^1_i and p^2_i. We use p^1_i and p^2_i to clip p_i, yielding an initially bounded plane patch p'_i. According to the original primitive label of l_i, p'_i is then further constrained by the inliers of the corresponding original planar primitive, producing the corrected planar primitive p^i_c. Since plane boundaries in Vertex Group representations are typically constrained by inliers, we sample points on p^i_c and use them as its inliers. In this way, the structural regularity encoded in the 2D sketches is propagated back to 3D, effectively reducing noise in the originally fitted planar primitives.

3.4 Sketch Fusion and Reconstruction

We fuse the optimized sketches to recover a complete roof structure, as illustrated in Figure 4. Rather than directly taking the union of all line segments from different contour levels, sketch fusion aims to construct a geometrically consistent and structurally stable set of planar primitives for subsequent 3D reconstruction. Since multi-level sketches may contain redundant, fragmented, or conflicting segments, we adopt a hierarchical fusion strategy in which a key sketch provides the primary structural scaffold and upper-level sketches are incorporated only as complementary refinements.

We first select a key sketch $s_{key} \in S_f$, defined as the sketch associated with the largest number of planar primitives and typically located at a lower contour level. Lower-level sketches usually preserve the dominant roof footprint and provide more complete structural support, whereas higher-level sketches often correspond to local height transitions, secondary facets, or fragmented details. We therefore use s_{key} as the base topology and iteratively examine line segments from higher contour levels to obtain the fused sketch s_{fuse}.

Each line segment l_j is associated with its contour level, orientation label, and original planar primitive label. During fusion, inter-level conflicts are resolved by prioritizing the current fused scaffold. A segment from an upper-level sketch is retained only if it contributes new and reliable structural information; otherwise, it is discarded as redundant or noisy. Specifically, the following checks are performed:

(1) **Planar primitive compatibility.** Let ρ be the planar primitive corresponding to l_j. If ρ is absent from the fused sketch, l_j is added directly. Otherwise, we compare the inlier distribution I^ρ_j of l_j with the inlier distribution I^ρ of the corresponding fused primitive. If the two differ significantly, l_j is retained, as this usually indicates that the same primitive has been split into multiple spatially distinct roof facets during boundary-aware clipping; otherwise, it is treated as redundant.

(2) **Geometric consistency.** For segments passing the previous test, we compute the one-way Chamfer distance between the corrected surface induced by l_j and the already fused planes.

A small distance indicates geometric consistency and allows l_j to be incorporated, whereas a large discrepancy suggests an unstable or spurious structure and leads to rejection.

(3) **Data support.** We further examine the number of inliers supporting the plane corresponding to l_j. Segments with insufficient support are removed as unreliable.

In this way, conflicts among contour-level sketches are resolved conservatively: the lower-level key sketch determines the principal topology, while upper-level sketches only supplement missing planes, distinct clipped facets, or strongly supported local structures. This preserves the dominant roof organization without being destabilized by fragmented high-level details.

After fusion, we convert the fused sketch into a refined planar primitive set P_f. Although the fusion is performed in 2D sketch space, each line segment in s_{fuse} carries its planar primitive label and is associated with a corrected primitive normal and contour level. The fused sketch therefore provides sufficient information to recover a consistent set of 3D supporting planes together with their clipping relations. Using P_f as input, we perform an initial reconstruction and obtain the intermediate roof model R_{first}.

At this stage, R_{first} may still contain hollow regions caused by missing bottom clipping surfaces. To repair them, we use R_{first} as prior geometric evidence. Specifically, we voxelize R_{first} and project the occupied voxels onto the xy-plane to obtain a binary occupancy map V_{proj}, where occupied cells are assigned 1 and unoccupied cells 0. We then extract connected components from the unoccupied cells using DFS. Components that do not touch the boundary of the projected domain are treated as interior hollow regions and thus as candidates for bottom clipping. For each such region, we lift the centers of its 2D voxels to 3D by assigning appropriate z-coordinates, and use the resulting points as inliers to fit the corresponding bottom clipping surface. These surfaces are added to P_f, producing the updated planar primitive set P'_f.

Finally, we run the reconstruction algorithm again on P'_f to obtain the final roof model R_{final}. Our sketch fusion stage mainly guarantees structural consistency in planar primitive selection and boundary organization, while the final watertightness and global topological closure are achieved by the downstream reconstruction algorithm on the refined primitive set. In other words, our contribution is to provide a cleaner, more complete, and more structurally consistent set of planes and clipping cues, thereby improving the robustness of the final reconstruction.

4 Experiments

In this section, we first present the implementation details of our experiments. We then conduct comparative evaluations against both learning-based and plane-based methods on sloped roof datasets. Finally, we perform ablation studies to validate the contribution of key components in our framework. Additional experimental results are provided in the supplementary materials.

4.1 Experimental Setup

4.1.1 Dataset. The dataset we use consists of real-world scanned data of sloped roofs in the form of 3D mesh models without bottom surfaces. These data which comprises 70 distinct roof models contain sufficiently complex roof structures incorporating various styles of sloped roofs, along with substantial noise structures, presenting significant challenges that can comprehensively test the algorithm's robustness, effectiveness, and generalization capability.

Table 1: Quantitative comparisons with the state-of-the-art methods NeurCAD[4], PrimFit[10], KSR[1], and Compod[23].

Method	NeurCAD	PrimFit	KSR	Compod	Ours
Times(s)	2394.4	5160.5	462.8	134.8	**132.4**
RMS↓	0.0763	0.0974	0.0424	0.0557	**0.0387**
MM↓	0.0690	0.0587	0.0365	0.0446	**0.0218**

4.1.2 Implementation Details. Each roof model is first normalized to $[-1, 1]^3$. Before reconstruction, we apply the preprocessing pipeline in Sec. 3.2, using Dual-DMP [5] for smoothing. The same preprocessing is applied to all compared methods. For plane detection, all plane-based baselines use GoCoPP [28].

For final mesh generation, we use Compod [23], which serves only as the last-stage mesh reconstruction module in our pipeline. Runtime is measured according to the required pipeline of each method. For plane-based methods, runtime is counted from plane detection onward, since plane detection is a necessary prerequisite. For learning-based methods, we report the runtime of the method itself, as no separate plane detection stage is required. All methods are tested under the same hardware conditions and implementation settings for fair comparison.

In the segmentation stage, the classification angle threshold is set to $\theta_{th} = 10^\circ$, which works well for most roof models. The number of sampled points is fixed to 100K.

Following [30], we use two complementary metrics to evaluate geometric accuracy: Root Mean Square Hausdorff Distance (RMS) and Maximum Mean Hausdorff Distance (MM). RMS measures global reconstruction fidelity, while MM captures localized maximum deviation. Together, they provide a balanced evaluation of overall and worst-case performance.

4.2 Comparison

Our method is compared with one learning-based method, NeurCAD [4], and three plane-based methods, PrimFit [10], KSR [1], and Compod [23]. All methods take the preprocessed roof model as input. For computational efficiency, we evaluate the entire reconstruction pipeline, including plane detection and 3D mesh reconstruction.

Quantitative comparisons with state-of-the-art methods are reported in Table 1. The results show that our method achieves a strong balance between efficiency and accuracy. In terms of runtime, our method (132.4s) is comparable to the fastest baseline, Compod (134.8s), and is substantially faster than NeurCAD (2394.4s), PrimFit (5160.5s), and KSR (462.8s). More importantly, our method achieves the best geometric accuracy under both metrics, with the lowest RMS (0.0387) and MM (0.0218), indicating both the highest global fidelity and the smallest local errors.

As shown in the qualitative comparison in Figure 5, our method produces higher-quality reconstructions than the baselines. During segmentation, we enforce normal consistency of planar primitives

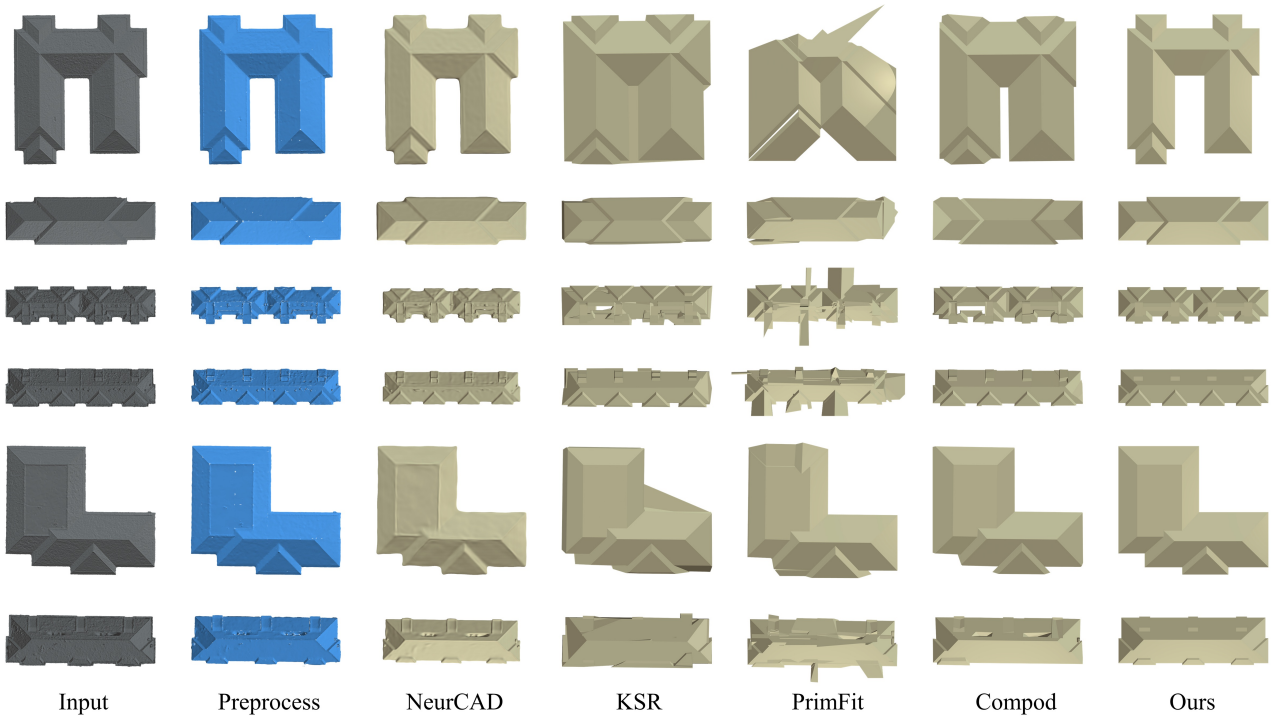

Figure 5: Comparison with state-of-the-art approaches.

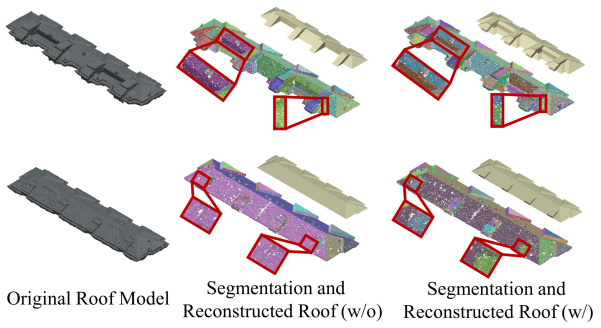

Figure 6: Visualizing the reconstruction result without(w/o) and with(w/) hierarchical segmentation strategy.

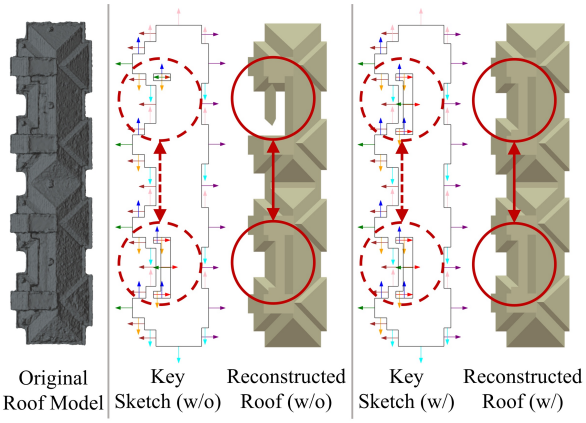

Figure 7: Visualizing the reconstruction result without(w/o) and with(w/) sketch symmetry repair strategy.

using statistical priors, leading to more regular roof structures. We further refine the primitives through sketch-based optimization and remove noisy structures, which improves reconstruction quality. NeurCAD reconstructs complete roofs but does not preserve sharp features well, and its surfaces are often insufficiently planar. Among plane-based methods, KSR and Compod are sensitive to noise, leading to redundant structures and local distortions, while PrimFit performs worse and produces substantial redundant elements. By propagating 2D sketch constraints back to 3D planar correction and suppressing noisy structures, our method reconstructs cleaner and more accurate roof models.

4.3 Ablation Study

4.3.1 Evaluation Of Hierarchical Segmentation. We conducted a comparative experiment by disabling H-Seg, where we only performed clustering on the initial classification results while keeping all other settings unchanged. As shown in the results in Figure 6, the third column displays the segmentation results without hierarchical segmentation, which, when compared to the results with hierarchical segmentation enabled, exhibit errors that lead to missing roof

structures. These issues arise because, after clustering, there still exist clusters containing faces oriented in different directions. If these faces are directly approximated as planes, the segmentation results will be incorrect, causing missing coverage in the segmented planar primitives. Therefore, hierarchical segmentation effectively separates faces with different orientations within clusters, significantly improving the accuracy of the segmentation results.

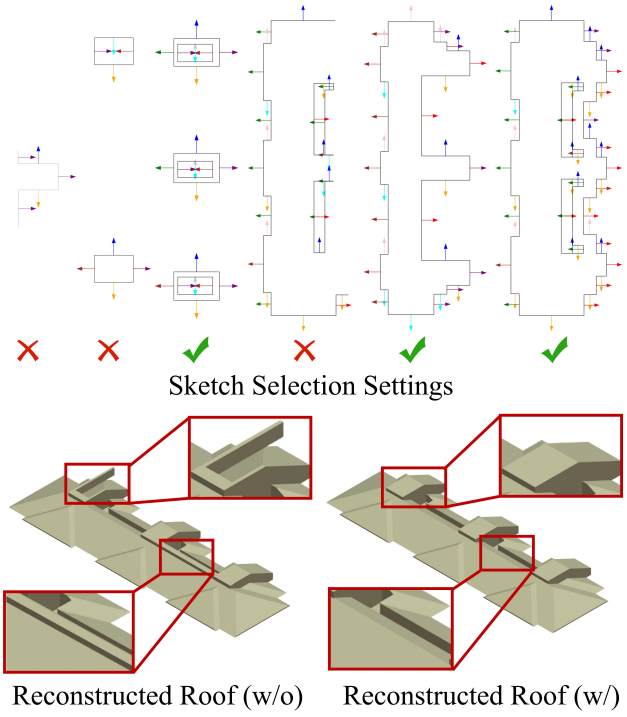

Figure 8: Visualizing the reconstruction result without(w/o) and with(w/) sketch filtering.

4.3.2 Evaluation of Sketch Symmetry Repair. As shown in Figure 7, the reconstruction without symmetric repair clearly shows missing roof structures (red boxes). Since the rough roof contains numerous noise structures, the extracted sketches also include many noise line segments. After denoising, if key roof structures are severely affected by noise, their related line segments are also treated as noise and removed. Without symmetric repair, these major structures would be missing in the final reconstruction. Therefore, symmetric sketch repair can restore those roof structures severely affected by noise, significantly enhancing our algorithm's robustness.

4.3.3 Evaluation of Sketch Filtering. As shown in Figure 8, when all sketches participate in fusion, noise structures appear in the final reconstruction. Moreover, due to the inclusion of noise-related planar primitives, the reconstruction quality deteriorates, resulting in redundant surfaces. These issues occur because disabling sketch filtering allows non-effective planar primitives to be incorporated into the fusion process, compromising reconstruction quality. Sketch filtering effectively removes sketches containing line segments related to non-effective planar primitives, thereby improving final reconstruction quality.

4.3.4 Evaluation of Key Sketch Selecting. We conducted a comparative experiment on the same model by randomly selecting different sketches as the key sketch while keeping other settings unchanged. As shown in Figure 9, selecting the sketch containing the most planar primitives as the key sketch yields better reconstruction results. This is because the key sketch serves as the foundation for the fusion module - when adding line segments to it, noise segments and duplicate segments are filtered based on this foundation. If the key sketch contains fewer associated planar primitives, noise segments may be confused with some non-noise segments, causing valid segments to be incorrectly filtered out and resulting in missing structures in the reconstruction.

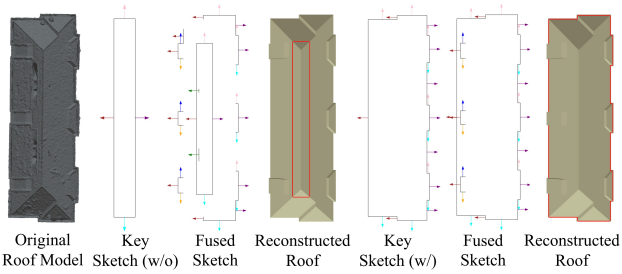

Figure 9: Visualizing the reconstruction result without(w/o) and with(w/) key sketch selecting strategy.

5 Conclusion

This paper presents a unified framework that bridges 2D sketch abstraction and 3D geometric refinement for robust structured roof reconstruction from imperfect data. Unlike methods operating solely in 3D, our approach is based on a cross-domain optimization strategy. We first perform hierarchical segmentation of the rough roof mesh to extract planar primitives under robust orientation constraints. We then construct and optimize multi-level 2D roof contour sketches, which provide a compact and topologically clean representation for enforcing structural regularities. The optimized 2D constraints are subsequently transferred back to 3D to guide planar primitive correction. This dual-domain design enables structural priors and geometric corrections that are difficult to impose directly in 3D. Extensive quantitative and qualitative experiments on real-world datasets with roofs ranging from simple to highly complex demonstrate that our method achieves strong performance while maintaining a good balance between reconstruction quality and computational efficiency.

Despite these advantages, the current framework has several limitations. First, the sketch abstraction stage tends to omit fine-grained details such as small dormers and chimneys. Second, our method is mainly designed for geometric correction and regularization rather than geometry completion, and therefore struggles when large roof regions are missing. In future work, we will explore more advanced frameworks to address these limitations.

Acknowledgments

Acknowledgments. This work was supported in part by the Chongqing Natural Science Foundation (CSTB2024NSCQ-MSX1026), in part by the Fundamental Research Funds for the Central Universities (No. SWU-KT25012), in part by the National Key R&D Program of China (No. 2024YFC3808901).

References

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

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

[3] David Cohen-Steiner, Pierre Alliez, and Mathieu Desbrun. 2004. Variational shape approximation. In *ACM SIGGRAPH 2004 Papers*. 905–914.

[4] Qiujie Dong, Rui Xu, Pengfei Wang, Shuangmin Chen, Shiqing Xin, Xiaohong Jia, Wenping Wang, and Changhe Tu. 2024. NeurCADRecon: Neural Representation for Reconstructing CAD Surfaces by Enforcing Zero Gaussian Curvature. *arXiv preprint arXiv:2404.13420* (2024).

[5] Shota Hattori, Tatsuya Yatagawa, Yutaka Ohtake, and Hiromasa Suzuki. 2022. Learning self-prior for mesh denoising using dual graph convolutional networks. In *European Conference on Computer Vision*. Springer, 363–379.

[6] Andre Henn, Gerhard Gröger, Viktor Stroh, and Lutz Plümer. 2013. Model driven reconstruction of roofs from sparse LIDAR point clouds. *ISPRS Journal of photogrammetry and remote sensing* 76 (2013), 17–29.

[7] Hai Huang, Claus Brenner, and Monika Sester. 2013. A generative statistical approach to automatic 3D building roof reconstruction from laser scanning data. *ISPRS Journal of photogrammetry and remote sensing* 79 (2013), 29–43.

[8] Jingwei Huang, Yanfeng Zhang, and Mingwei Sun. 2021. Primitivenet: Primitive instance segmentation with local primitive embedding under adversarial metric. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 15343–15353.

[9] Shangfeng Huang, Ruisheng Wang, Bo Guo, and Hongxin Yang. 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.

[10] Jingen Jiang, Mingyang Zhao, Shiqing Xin, Yanchao Yang, Hanxiao Wang, Xiaohong Jia, and Dong-Ming Yan. 2023. Structure-aware surface reconstruction via primitive assembly. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 14171–14180.

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

[12] Minglei Li, Peter Wonka, and Liangliang Nan. 2016. Manhattan-world urban reconstruction from point clouds. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*. Springer, 54–69.

[13] Yujun Liu, Ruisheng Wang, Shangfeng Huang, and Guorong Cai. 2025. EdgeDiff: Edge-aware Diffusion Network for Building Reconstruction from Point Clouds. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 17008–17018.

[14] Yuzhou Liu, Lingjie Zhu, Hanqiao Ye, Shangfeng Huang, Xiang Gao, Xianwei Zheng, and Shuhan Shen. 2025. BWFormer: Building Wireframe Reconstruction from Airborne LiDAR Point Cloud with Transformer. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 22215–22224.

[15] Ziqiong Lu, Linxi Huan, Qiyuan Ma, and Xianwei Zheng. 2023. Holistic Geometric Feature Learning for Structured Reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 21807–21817.

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

[17] Charlie Nash, Yaroslav Ganin, SM Ali Eslami, and Peter Battaglia. 2020. Polygen: An autoregressive generative model of 3d meshes. In *International conference on machine learning*. PMLR, 7220–7229.

[18] Charalambos Poullis. 2019. Large-scale urban reconstruction with tensor clustering and global boundary refinement. *IEEE transactions on pattern analysis and machine intelligence* 42, 5 (2019), 1132–1145.

[19] Jing Ren, Biao Zhang, Bojian Wu, Jianqiang Huang, Lubin Fan, Maks Ovsjanikov, and Peter Wonka. 2021. Intuitive and efficient roof modeling for reconstruction and synthesis. *arXiv preprint arXiv:2109.07683* (2021).

[20] Philipp Schuegraf, Jie Shan, and Ksenia Bittner. 2024. PLANES4LOD2: Reconstruction of LoD-2 building models using a depth attention-based fully convolutional neural network. *ISPRS Journal of Photogrammetry and Remote Sensing* 211 (2024), 425–437.

[21] Silvia Sellán, Noam Aigerman, and Alec Jacobson. 2020. Developability of height-fields via rank minimization. *ACM Trans. Graph.* 39, 4 (2020), 109.

[22] Martin Skrodzki, Eric Zimmermann, and Konrad Polthier. 2020. Variational shape approximation of point set surfaces. *Computer Aided Geometric Design* 80 (2020), 101875.

[23] Raphael Sulzer and Florent Lafarge. 2025. Concise Plane Arrangements for Low-Poly Surface and Volume Modelling. In *European Conference on Computer Vision*. Springer, 357–373.

[24] Ruisheng Wang, Shangfeng Huang, and Hongxin Yang. 2023. Building3d: An urban-scale dataset and benchmarks for learning roof structures from point clouds. In 2023 IEEE. In *CVF International Conference on Computer Vision (ICCV)*. 20019–20029.

[25] Zixiong Wang, Yunxiao Zhang, Rui Xu, Fan Zhang, Peng-Shuai Wang, Shuangmin Chen, Shiqing Xin, Wenping Wang, and Changhe Tu. 2023. Neural-Singular-Hessian: Implicit Neural Representation of Unoriented Point Clouds by Enforcing Singular Hessian. *ACM Transactions on Graphics (TOG)* 42, 6 (2023), 1–14.

[26] Siming Yan, Zhenpei Yang, Chongyang Ma, Haibin Huang, Etienne Vouga, and Qixing Huang. 2021. Hpnet: Deep primitive segmentation using hybrid representations. In *Proceedings of the IEEE/CVF international conference on computer vision*. 2753–2762.

[27] Dayu Yu, Fan Ye, Peng Yue, Min Chen, and Filip Biljecki. 2025. RooFormer: Reconstructing detailed 3D roof models from high-resolution remote sensing imagery using transformer. *ISPRS Journal of Photogrammetry and Remote Sensing* 227 (2025), 745–758.

[28] Mulin Yu and Florent Lafarge. 2022. Finding good configurations of planar primitives in unorganized point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6367–6376.

[29] Fuyang Zhang, Nelson Nauata, and Yasutaka Furukawa. 2020. Conv-mpn: Convolutional message passing neural network for structured outdoor architecture reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2798–2807.

[30] Zhenyu Zhang, Mingyang Zhao, Zeyu Shen, Yuqing Wang, Xiaohong Jia, and Dong-Ming Yan. 2024. Interactive reverse engineering of CAD models. *Computer Aided Geometric Design* 111 (2024), 102339.

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

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