SFD-Net: Sharp Feature Detection Network Based on Local Geometric Features

Inyoung Oh[1,2] and Kwang Hee Ko[1]*

[1] Department of Mechanical and Robotics Engineering, Gwangju Institute of Science and Technology (GIST), 123 Cheomdangwagi-ro, Buk-gu, Gwangju 61005, Republic of Korea
[2] Center for AI Research, Korea Institute of Science and Technology (KIST), Hwarang-ro 14-gil 5, Seongbuk-gu, Seoul 02792, Republic of Korea
deanoh90@kist.re.kr, khko@gist.ac.kr

Abstract. Reliable detection of sharp features in point clouds—loci where the surface-normal field is discontinuous—remains challenging under increased local density variation and noise. SFD-Net couples a compact multi-scale Local Geometric Descriptor (LGD) with an enhanced PointNet++ backbone. LGD aggregates second-moment statistics of surface normal differences at three nested neighborhood scales, yielding rigid-motion- and scale-invariant cues that remain stable under joint noise and spacing variability. On the ABC benchmark, SFD-Net achieves state-of-the-art F1-score and competitive False Positive Rate (FPR) across four noise conditions under a consistent evaluation setup. LGD further improves competitive detectors in a plug-and-play manner without architectural changes, and transfer to Stanford Large-Scale 3D Indoor Spaces (S3DIS) confirms generalization to real scans. Downstream applications potentially benefit from more reliable feature localization. Our source code and pretrained models are publicly available at `https://github.com/inyoungoh-cde/SFD-Net`.

Keywords: Point cloud · Sharp feature detection · Local geometric descriptor · Density variation · Multi-scale normal statistics

1 Introduction

Point clouds are the native output of depth cameras, LiDAR scanners, and structured-light systems, and they underpin a broad range of 3D perception tasks, including semantic segmentation [25, 26, 32], object detection [23, 42], surface reconstruction [16, 33], wireframe generation [5, 40, 41], and autonomous driving [17, 30]. Unlike meshes or voxels, point sets carry no explicit topology; thus, any local geometric computation—normal estimation, curvature, feature detection—depends entirely on neighborhood quality. When neighborhoods become irregular due to noise or increased local spacing variability, even simple differential quantities can become unreliable.

* Corresponding author.

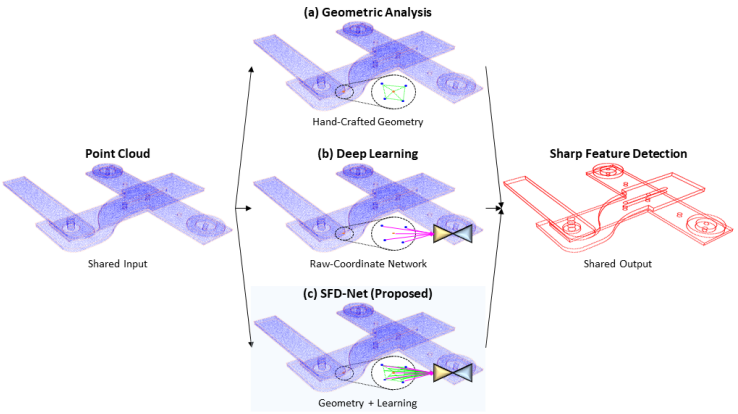

Fig. 1: Comparison of sharp feature detection paradigms. All three paradigms map a shared input point cloud to the same sharp-feature-detection task, differing only in the intermediate stage. Unlike hand-crafted geometric analysis (a) and raw-coordinate deep learning (b), the proposed SFD-Net (c) adopts geometric-guided learning that, from the *same* local neighborhood, jointly extracts multi-scale local geometric descriptors (green) and raw point features (pink) and feeds them into a deep learning backbone, combining the stability of classical geometry with the discriminative power of learned representations.

Among all geometric structures on a surface, *sharp features*, as shown in Figure 1, are particularly important. A sharp feature is a locus where the surface-normal field is discontinuous; equivalently, G^1 continuity fails and the tangent plane changes abruptly [29]. Creases at patch intersections and corners where several patches meet are canonical examples. Reliable identification of these structures in point clouds is a prerequisite for robust separation of faces, edges, and corners, and it directly improves downstream estimation of differential quantities [7], mesh reconstruction [20, 43], and boundary-aware semantic segmentation [8, 9, 28]. Despite its importance, sharp feature detection remains an open problem whenever inputs deviate from the clean, dense regime assumed by most prior methods.

Learning-based detectors routinely achieve high accuracy on clean, densely sampled benchmarks; however, two correlated perturbations expose a fundamental weakness. First, measurement noise corrupts local normal estimates, destabilizing the differential cues on which detection depends. Second, point subsampling increases the spread of local k-NN spacing. Although uniform random thinning is used to reduce the point count and the selection is uniform, the resulting point set exhibits a broader distribution of local k-NN spacing, thereby increasing local density variation. This spacing variability renders neighborhood-based statistics inconsistent: fine-scale features may fall entirely outside small neighborhoods, while large neighborhoods over-smooth the discontinuities that define sharp features. Both perturbations co-occur in practical scanning pipelines, yet existing methods rarely address their joint effect. Additionally, sharp features

constitute a small minority class, so standard objectives bias predictions toward the dominant non-sharp class, resulting in low recall for these rare but critical points.

We present SFD-Net, a two-stage framework that addresses these gaps through a compact Local Geometric Descriptor (LGD) paired with an enhanced PointNet++ [26] backbone. LGD estimates Principal Component Analysis (PCA) [12] normals at three nested k-NN scales and aggregates second-moment statistics of normal differences into a 3D isotropy index per scale. By operating on normal differences rather than raw coordinates or normals alone, LGD is invariant to rigid motion and global scale, insensitive to normal sign ambiguity, and stabilized across spacing variability through its multi-scale structure. The descriptor is concatenated with raw coordinates and processed by a PointNet++ augmented with two lightweight Transformer encoder blocks and a composite Focal Tversky [1]–Dice loss [31] tailored for severe class imbalance. Critically, LGD is architecture-agnostic: it can be prepended to any backbone that accepts per-point feature channels, enabling plug-and-play improvement of existing detectors.

Extensive evaluations on the ABC benchmark [18] demonstrate that SFD-Net achieves state-of-the-art F1-score across four noise conditions with density variation. LGD provides consistent F1 and False Positive Rate (FPR) improvements when prepended to PointNet++, RepSurf [27], and EDWG [41] without architectural modifications, and models trained on ABC CAD data generalize effectively to real S3DIS [2] scenes.

Contributions. We highlight our main contributions below:

- **Robustness under density variation and noise.** LGD maintains stable feature discrimination under increased local k-NN spacing variability induced by uniform random thinning and under additive Gaussian noise across multiple severity levels.
- **Plug-and-play improvement.** Prepending LGD to PointNet++, RepSurf, and EDWG improves Average F1 and FPR in all cases without modifying backbone architectures.
- **Fair comparison under a consistent evaluation setup.** All baselines share the same dataset split, the same uniform random thinning and noise procedure, and are retrained from scratch, ensuring evaluation differences reflect model quality rather than experimental conditions.
- **Downstream applicability.** The improved localization accuracy of sharp features benefits tasks such as wireframe generation [41], registration [34], and boundary-aware semantic segmentation [8].

2 Related Work

Sharp-feature detection on point clouds intersects three research threads: classical geometry-based feature extraction, deep learning on unstructured point sets, and learning-based edge and boundary detection. We survey each in turn, emphasizing how robustness to density variation and noise has been treated.

Classical geometry-based detection. Early methods derive feature indicators from local differential quantities computed over k-NN or radius neighborhoods. Surface variation—the ratio of the smallest PCA eigenvalue to the eigenvalue sum—is an efficient and widely used descriptor [24], but its zero breakdown point under the sample mean makes it sensitive to outliers and neighborhood irregularities [14]. Subsequent variants analyzed eigenvalue distributions under varying noise [3], used normal curvature from quadratic fitting [44], improved normal estimation via candidate-set optimization [45], and replaced fragile moments with robust location estimators [19]. Local quadric fitting provides access to Gaussian and mean curvatures via fundamental forms [22], while normal-statistics thresholding has been combined with learning in hybrid pipelines [13]. Beyond purely local statistics, graph-spectral methods resample point clouds via high-pass graph filtering to emphasize contours and other high-frequency structures, extracting sharpness cues without explicit normal estimation [6]. Despite their efficiency, these methods rely on manually tuned thresholds and fragile normal estimates, and their reliability degrades when sampling density varies spatially alongside noise.

Deep learning on unstructured point clouds. Representative architectures are surveyed along the progression from point- to edge- to surface-level geometric reasoning, and further to multilayer perceptron (MLP)- and attention-based global designs. PointNet pioneered permutation-invariant processing of raw coordinates via shared MLPs and global pooling [25], and PointNet++ extended it with hierarchical multi-scale grouping that improves robustness to moderate density variation. DGCNN dynamically constructs k-NN graphs in feature space and applies EdgeConv for local relational encoding, improving boundary delineation at the cost of sensitivity to noisy or uneven neighborhoods [36]. RepSurf enriches inputs with lightweight triangular and umbrella surface priors around each point, boosting accuracy with small overhead while inheriting k-NN vulnerability to sampling irregularities. PointMLP replaces explicit geometric extractors with residual MLP stacks and a geometric affine normalization module, yielding fast inference but limited performance on tasks requiring precise curvature or topology reasoning [21]. Transformer-based designs, including Point Transformer [46], extend context modeling beyond local neighborhoods, providing strong backbone baselines for dense prediction. More recent designs inject geometric structure into learning: CurveNet aggregates features along learned point curves to capture local continuity [37], while GDANet disentangles each shape into sharp and gentle geometric components through attention for complementary understanding [39]. Despite this progress, sharp-feature detection remains challenging because target regions are sparse and neighborhood-derived cues can collapse to single-class predictions under joint density variation and noise.

Learning-based edge and sharp feature detection. Several methods specifically target edges, corners, or boundaries in point clouds. EC-Net extends PointNet++ with an edge-aware loss to improve edge localization [43]. PIE-Net uses

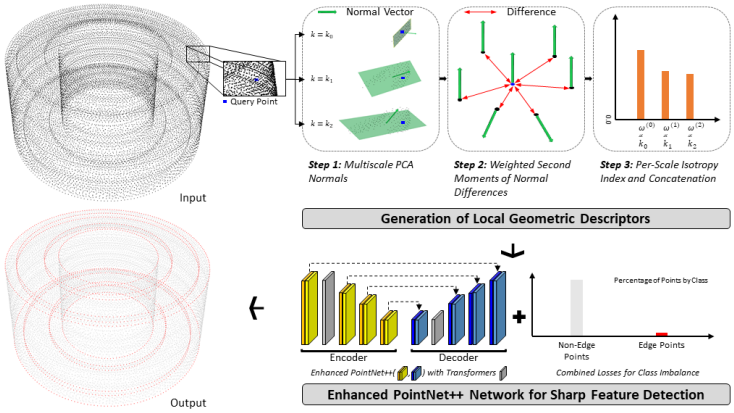

Fig. 2: Overview of SFD-Net. Stage 1 builds a 3D descriptor $\boldsymbol{\phi}_i$ per query point via (Step 1) PCA normals at three nested k-NN scales, (Step 2) weighted second moments of normal differences $\boldsymbol{\Sigma}_i^{(s)}$, and (Step 3) per-scale isotropy indices concatenated into $\boldsymbol{\phi}_i$. Stage 2 concatenates $\boldsymbol{\phi}_i$ with raw coordinates and feeds an enhanced PointNet++ trained with a class-imbalance loss for point-wise sharp feature detection.

a two-stage PointNet++ pipeline with focal loss for sparse edge detection [35]. PCEDNet encodes multi-scale geometric cues into a compact representation for efficient edge classification [11]. NerVE learns volumetric edge representations combining PointNet++, convolutional neural networks (CNNs), and graph edge convolution [48]. BoundED uses lightweight geometric-statistics-based features for efficient boundary and edge detection [4]. SFC-Net refines sharp feature localization through displacement-based iterative consolidation [47]. MSL-Net uses normal-angle descriptors for sharp feature detection over k-NN neighborhoods with a lightweight MLP [15]. EdgeFormer combines handcrafted geometric features with Transformer-based attention [38]. EDWG integrates multi-scale feature fusion and attention for edge detection and wireframe reconstruction.

Summary and gap. Despite substantial progress, two gaps remain. First, most learning-based detectors rely on neighborhood-derived normals or coordinates without explicit stabilization against joint density variation and noise, causing degradation when spacing becomes irregular. Second, descriptor-level improvements are rarely evaluated in a plug-and-play fashion across multiple architectures. Our LGD is designed to fill both gaps: it stabilizes multi-scale second-moment statistics of normal differences and is architecture-agnostic, enabling consistent improvement across different backbones under a consistent evaluation setup.

3 Method

Figure 2 outlines the two-stage SFD-Net pipeline. Stage 1 computes a compact LGD $\boldsymbol{\phi}_i \in \mathbb{R}^3$ for every point from multi-scale PCA normal statistics. Stage 2 concatenates LGD with raw coordinates and feeds the result to an enhanced PointNet++ backbone trained with an imbalance-aware objective. LGD is architecture-agnostic: it attaches to any backbone that accepts per-point feature channels, requiring no architectural modification.

3.1 Problem Setup

Let $\mathcal{P} = \{\mathbf{p}_i\}_{i=1}^{N} \subset \mathbb{R}^3$ be an unordered point set. Each point $\mathbf{p}_i$ carries a binary label $y_i \in \{0, 1\}$ indicating sharp (y_i=1) or non-sharp (y_i=0). Following geometric modeling conventions, a sharp feature is a locus where the surface-normal field is discontinuous (G^1 continuity fails); in discrete point clouds this manifests as abrupt orientation changes in neighborhood-based normal estimates. The goal is to predict $\hat{y}_i \in \{0, 1\}$ for all points under noisy measurements and degraded neighborhood regularity caused by increased local spacing variability.

3.2 LGD: Multi-Scale Normal-Difference Isotropy

Three empirical obstacles motivate LGD. First, purely coordinate-based covariances conflate positional scale and sampling density, so curvature surrogates fluctuate when neighborhoods are sparse or irregular. Second, a single neighborhood size forces a brittle trade-off between responsiveness to thin features and stability on coarse or noisy regions. Third, explicit normal orientation alignment introduces branch-dependent behavior near sharp features. LGD addresses all three by estimating normals at three nested supports, aggregating differences of these normals, and summarizing the resulting distribution via an isotropy index with a clear geometric interpretation.

Neighborhoods and nesting. Let k_2 be the largest neighborhood size, with $k_0 < k_1 < k_2$. For each point $\mathbf{p}_i$, a single k_2-NN query returns neighbors ordered by increasing Euclidean distance, and nested subsets are extracted as prefixes:

$$\mathcal{N}_s(i) = \mathrm{kNN}(\mathbf{p}_i;\, k_s), \quad s \in \{0, 1, 2\}, \tag{1}$$

guaranteeing $\mathcal{N}_0(i) \subset \mathcal{N}_1(i) \subset \mathcal{N}_2(i)$ without additional queries.

Step 1: PCA normals at three scales. For scale s and point $\mathbf{p}_i$, center the neighborhood: $\bar{\mathbf{p}}_i^{(s)} = \frac{1}{k_s} \sum_{j \in \mathcal{N}_s(i)} \mathbf{p}_j$. The local covariance is

$$\mathbf{C}_i^{(s)} = \sum_{j \in \mathcal{N}_s(i)} (\mathbf{p}_j - \bar{\mathbf{p}}_i^{(s)})(\mathbf{p}_j - \bar{\mathbf{p}}_i^{(s)})^\top = \mathbf{Q}_i^{(s)} \,\mathrm{diag}(\lambda_{i,1}^{(s)}, \lambda_{i,2}^{(s)}, \lambda_{i,3}^{(s)})\, \mathbf{Q}_i^{(s)\top}, \tag{2}$$

with ordered eigenvalues $\lambda_{i,1}^{(s)} \le \lambda_{i,2}^{(s)} \le \lambda_{i,3}^{(s)}$. The unit PCA normal is the minimum-eigenvalue eigenvector: $\mathbf{n}_i^{(s)} = \mathbf{q}_{i,1}^{(s)}$. No orientation propagation is applied; only unit directions are retained, enabling flip invariance.

Step 2: Weighted second moments of normal differences. Sharp features manifest as directional changes of local surface orientation, so LGD encodes pairwise normal differences

$$\Delta\mathbf{n}_{ij}^{(s)} = \mathbf{n}_j^{(s)} - \mathbf{n}_i^{(s)} \in \mathbb{R}^3, \quad j \in \mathcal{N}_s(i), \tag{3}$$

with magnitude-squared weights $w_{ij}^{(s)} = \|\Delta\mathbf{n}_{ij}^{(s)}\|_2^2 + \varepsilon_w$ ($\varepsilon_w > 0$) to emphasize informative differences while remaining well-defined when $\Delta\mathbf{n}_{ij}^{(s)} = \mathbf{0}$. The normalized weighted second-moment tensor is

$$\boldsymbol{\Sigma}_i^{(s)} = \frac{1}{Z_i^{(s)}} \sum_{j \in \mathcal{N}_s(i)} w_{ij}^{(s)}\, \Delta\mathbf{n}_{ij}^{(s)} \Delta\mathbf{n}_{ij}^{(s)\top}, \qquad Z_i^{(s)} = \sum_{j \in \mathcal{N}_s(i)} w_{ij}^{(s)} + \varepsilon_z, \tag{4}$$

with $\varepsilon_z > 0$ preventing division by zero. By construction, $\boldsymbol{\Sigma}_i^{(s)}$ is positive semi-definite.

Invariance. Consider a similarity transform $T(\mathbf{x}) = c\mathbf{R}\mathbf{x} + \mathbf{t}$ ($c > 0$, $\mathbf{R} \in SO(3)$). Centering in Eq. (2) removes $\mathbf{t}$, and the covariance maps to $\mathbf{C}_i'^{(s)} = c^2\mathbf{R}\mathbf{C}_i^{(s)}\mathbf{R}^\top$. Scaling all eigenvalues equally by c^2 leaves the eigenvectors unchanged up to rotation, so the minimum-eigenvalue normal becomes $\mathbf{n}_i'^{(s)} = \pm\mathbf{R}\mathbf{n}_i^{(s)}$. Hence $\Delta\mathbf{n}_{ij}'^{(s)} = \pm\mathbf{R}\Delta\mathbf{n}_{ij}^{(s)}$ and, since the unit normals carry no scale, $\boldsymbol{\Sigma}_i'^{(s)} = \mathbf{R}\boldsymbol{\Sigma}_i^{(s)}\mathbf{R}^\top$, whose eigenvalues—and thus each $\omega_i^{(s)}$ and the descriptor $\boldsymbol{\phi}_i$—are unchanged. The $\pm$ cancels in the outer product $\Delta\mathbf{n}\Delta\mathbf{n}^\top$, yielding scale, rigid-motion, and flip invariance; we confirm this empirically in the supplementary.

Step 3: Per-scale isotropy index and concatenation. Let $\mu_{i,1}^{(s)} \le \mu_{i,2}^{(s)} \le \mu_{i,3}^{(s)}$ be the eigenvalues of $\boldsymbol{\Sigma}_i^{(s)}$. On a locally planar patch, normal differences are nearly collinear ($\boldsymbol{\Sigma}_i^{(s)}$ has one dominant eigenvalue), whereas near a crease or corner the difference vectors span multiple directions and the spectrum becomes more isotropic. The isotropy index

$$\omega_i^{(s)} = 1 - \frac{\mu_{i,3}^{(s)}}{\mu_{i,1}^{(s)} + \mu_{i,2}^{(s)} + \mu_{i,3}^{(s)}} \in [0, 1) \tag{5}$$

increases as directional spread grows. The three scale-specific responses are concatenated into the final descriptor:

$$\boldsymbol{\phi}_i = \left[\omega_i^{(0)},\, \omega_i^{(1)},\, \omega_i^{(2)}\right] \in \mathbb{R}^3. \tag{6}$$

The three scales play complementary roles: the smallest support (k_0) provides high spatial acuity for narrow creases; the intermediate (k_1) balances localization and stability; the largest (k_2) suppresses variance from noise or local undersampling. Together they supply discriminative cues that remain stable as local k-NN spacing variability increases.

3.3 Enhanced PointNet++ Backbone

The per-point input is formed by concatenating coordinates and LGD:

$$\mathbf{u}_i = [\mathbf{p}_i \parallel \boldsymbol{\phi}_i] \in \mathbb{R}^6. \tag{7}$$

Any detector can ingest $\mathbf{u}_i$ without modification. For the main SFD-Net instantiation, a PointNet++ encoder-decoder is used with two lightweight Transformer encoder blocks inserted to capture non-local dependencies that bounded-neighborhood PointNet blocks cannot model: (i) after the first Set Abstraction stage (d_{model}=96, 6 heads, 2 layers, sequence length N_1=1024) for cross-patch aggregation before aggressive downsampling, and (ii) after the first Feature Propagation stage (d_{model}=256, 8 heads, 2 layers, resolution N_3=64) to reinforce cross-scale consistency during upsampling. Each block follows the standard multi-head attention → feed-forward → residual → layer-norm structure with feed-forward width $4d_{\text{model}}$ and dropout 0.1. The dual Transformer placement jointly improves sharp-feature sensitivity by allowing information flow across distant regions in the coarse hierarchy and refining boundary evidence as features are propagated to dense points.

3.4 Imbalance-Aware Composite Objective

Sharp feature points are a sparse minority (typically 1–10% of all points), so standard cross-entropy biases predictions toward the dominant non-sharp class. A composite of Focal Tversky and Dice losses is adopted to explicitly balance false positives and false negatives while stabilizing overlap optimization. Let $p_i \in [0,1]$ and $y_i \in \{0,1\}$ denote the predicted probability and ground-truth label for the sharp class. Define soft counts:

$$\mathrm{TP} = \sum_i p_i y_i, \quad \mathrm{FP} = \sum_i p_i(1-y_i), \quad \mathrm{FN} = \sum_i (1-p_i) y_i. \tag{8}$$

The Tversky index and Focal Tversky loss are

$$\mathrm{TI} = \frac{\mathrm{TP} + \epsilon}{\mathrm{TP} + \alpha\,\mathrm{FP} + \beta\,\mathrm{FN} + \epsilon}, \qquad \mathcal{L}_{\mathrm{FT}} = (1 - \mathrm{TI})^{\gamma}, \tag{9}$$

where $\alpha, \beta > 0$ weight false positives and false negatives, $\gamma > 0$ focuses on hard examples, and $\epsilon > 0$ is a numerical stabilizer. The Dice loss is

$$\mathcal{L}_{\mathrm{Dice}} = 1 - \frac{2\sum_i p_i y_i + \epsilon}{\sum_i p_i + \sum_i y_i + \epsilon}. \tag{10}$$

The total objective is

$$\mathcal{L} = \lambda_1\,\mathcal{L}_{\mathrm{FT}} + \lambda_2\,\mathcal{L}_{\mathrm{Dice}}, \quad \lambda_1, \lambda_2 > 0, \tag{11}$$

with all hyperparameters $(\lambda_1, \lambda_2, \alpha, \beta, \gamma, \epsilon)$ reported in the supplementary material. The Focal Tversky term explicitly controls the FP–FN balance under severe class imbalance, while the Dice term directly maximizes overlap; together they improve both sharp-feature recall and boundary precision.

4 Experiments

4.1 Datasets and Evaluation Metrics

Synthetic Dataset: ABC. All experiments use the ABC dataset [18], a large-scale benchmark of parametric CAD objects with point-wise ground-truth sharp feature labels. From `chunk_0000` (7,168 models), 1,000 models are selected following a scheme similar to MSL-Net and partitioned into 400 (train), 90 (validation), and 510 (test).

For each model, 10,000 points are retained via uniform random sampling without replacement—matching the point count used in MSL-Net and EDWG. Although the selection is spatially unbiased, the resulting point set exhibits increased local k-NN spacing variability, introducing realistic density variation without the spatial artifacts of voxel or grid filters. Spacing-distribution statistics are provided in the Supplementary. Gaussian noise is then added following the PCPNet [10] data-generation procedure: per-point, per-axis perturbations

$$\tilde{\mathbf{p}}_i = \mathbf{p}_i + \boldsymbol{\eta}_i, \quad \boldsymbol{\eta}_i \sim \mathcal{N}(\mathbf{0},\, (\sigma_{\text{rel}} L)^2 \mathbf{I}_3), \tag{12}$$

where L is the bounding-box diagonal and $\sigma_{\text{rel}} \in \{0.00125, 0.006, 0.012\}$ defines small ($\approx$0.12%), medium (0.6%), and large (1.2%) noise levels relative to L. This yields four evaluation conditions: ABC_{none} (density variation only) and $\text{ABC}_{0.12\%}$, $\text{ABC}_{0.6\%}$, $\text{ABC}_{1.2\%}$ (with progressively increasing noise).

Real Dataset: S3DIS. The S3DIS benchmark [2] comprises point clouds of large-scale indoor scenes annotated with 13 semantic categories. Since point-wise sharp feature labels are not available, pseudo-labels are derived following [8]: a point is labeled sharp if at least one of its four nearest neighbors (k=4) belongs to a different semantic category, and non-sharp otherwise. This dataset is used exclusively for zero-shot generalization evaluation; no S3DIS data is used during training.

Metrics. Primary metrics are Average F1-score (↑) and FPR (↓). In all tables, **bold** = best and underline = second best.

4.2 Implementation Details

All experiments are implemented in PyTorch on an NVIDIA RTX 3090 GPU. Each density-varied cloud (10K-point cap; Sec. 4.1) is subsampled to only 500 points per forward pass; following the PointNet++ setup, this deliberately small budget stress-tests sharp-feature recovery from sparse input, a low-density regime in which prior detectors degrade [15]. All methods use this budget except RepSurf, whose architecture takes no fixed input size and follows its original design. We train for 200 epochs (batch size 16, Adam, learning rate 10^{-3}, cosine annealing warm restarts with T_0=10, T_{mult}=2, minimum 10^{-5}). Baselines are trained from scratch on the same data following their original papers, with batch size varying by hardware and all other hyperparameters fixed; for MSL-Net, unspecified hyperparameters match SFD-Net.

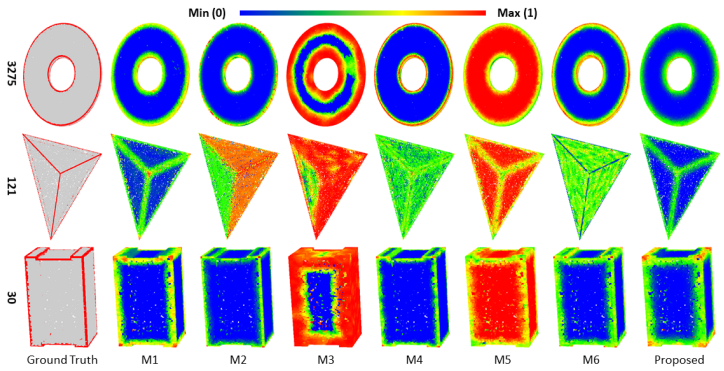

Fig. 3: Per-point geometric feature responses on $\mathrm{ABC}_{\mathrm{none}}$ for three representative models (3275, 121, and 30). Columns: Ground Truth, M1–M6, and the proposed LGD. Color encodes normalized response in $[0, 1]$ (blue $= 0$; red $= 1$). M1: PCA-based [24]; M2: normal-curvature [44]; M3: edge-statistics [13]; M4: robustness-oriented [19]; M5: MSL-Net [15]; M6: EdgeFormer [38]. Additional visualizations are provided in the supplementary material.

4.3 Main Comparison on ABC

Figure 3 visualizes per-point geometric feature responses on $\mathrm{ABC}_{\mathrm{none}}$ for three representative models, comparing LGD against four alternative descriptors (M1–M4) and two learning-based methods (M5, M6). Color encodes normalized response magnitude (blue $= 0$, red $= 1$). The target behavior is twofold: suppressed responses on non-sharp regions and concentrated high responses along sharp feature bands. Among all compared methods, only M1 and the proposed LGD consistently achieve both criteria across all three models. M2 and M3 exhibit responses leaking into flat faces; M5 produces an inverted trend near sharp features; M4 and M6 fail to align with ground-truth sharp bands on model 121. LGD achieves the strongest combination of sharp-feature concentration and non-sharp suppression, providing discriminative geometric cues prior to any learning stage.

We conduct experiments on the ABC dataset under four conditions combining density variation and progressively increasing noise levels. As shown in Table 1, SFD-Net achieves the best F1-score in all four settings, ranging from 61.33% ($\mathrm{ABC}_{\mathrm{none}}$) to 57.05% ($\mathrm{ABC}_{1.2\%}$). The second-best method in every condition is MSL-Net [15], which SFD-Net surpasses by 1.92, 1.31, 2.51, and 4.31 percentage points, with the largest margin (4.31 pp) occurring at the highest noise level ($\mathrm{ABC}_{1.2\%}$).

Generic backbones—PointNet++, DGCNN, and PointMLP—achieve near-random FPR (≈50%), reflecting a strong bias toward the majority non-sharp class under class imbalance. RepSurf performs competitively under clean and low-noise conditions (F1: 55.23 and 57.16) but degrades noticeably at higher perturbation levels (48.17 at $\mathrm{ABC}_{1.2\%}$), suggesting instability when local geometric statistics become unreliable.

Table 1: Comparison on ABC dataset (%). Average F1↑ and FPR↓ across four noise conditions (density variation applied to all); Average = arithmetic mean of Non-sharp and Sharp class metrics. **Bold**: best; underline: second best. Full per-class results are provided in the supplementary material.

Method	ABC_{none}		$ABC_{0.12\%}$		$ABC_{0.6\%}$		$ABC_{1.2\%}$	
	F1↑	FPR↓	F1↑	FPR↓	F1↑	FPR↓	F1↑	FPR↓
PointNet++ [26]	50.89	48.57	55.02	45.88	49.03	49.35	50.21	49.01
DGCNN [36]	49.89	48.90	50.65	48.51	48.44	49.58	47.53	49.99
RepSurf [27]	55.23	45.96	57.16	43.06	54.21	46.55	48.17	49.73
PointMLP [21]	54.59	46.30	47.53	50.00	47.52	50.01	47.65	49.94
PIE-Net [35]	52.55	38.50	52.44	37.35	47.61	40.54	49.56	**39.33**
BoundED [4]	48.12	49.88	47.98	49.94	47.97	49.90	47.70	49.98
SFC-Net [47]	47.92	49.84	48.64	49.53	48.27	49.69	48.03	49.80
MSL-Net [15]	59.41	**29.83**	59.11	**30.79**	56.11	**36.23**	52.74	47.53
EdgeFormer [38]	52.04	31.24	48.66	33.95	34.99	45.16	28.13	48.55
EDWG [41]	54.95	46.06	54.93	46.07	48.20	49.81	47.76	49.90
SFD-Net (ours)	**61.33**	36.93	**60.42**	39.76	**58.62**	40.88	**57.05**	41.94

Among task-specific methods, PIE-Net maintains consistently low FPR (37.35–40.54%) yet yields only mid-range F1-scores (47–52%), indicating that its conservative predictions suppress false positives at the cost of sharp-feature recall. EdgeFormer attains a low FPR under clean conditions (31.24%) but collapses under moderate-to-large noise (F1: 34.99 and 28.13 at 0.6% and 1.2%), revealing a critical fragility to joint density variation and noise. MSL-Net attains the second-best F1-score across all four conditions (59.41/59.11/56.11/52.74) while keeping FPR low under clean and low-noise settings, though its FPR rises markedly at the largest noise level (47.53 at $ABC_{1.2\%}$) as detection becomes unreliable.

SFD-Net keeps FPR competitive while consistently leading in F1-score, demonstrating that stabilizing local geometric descriptors under k-NN spacing variability effectively improves robustness without sacrificing sharp-feature recall.

This behavior is further confirmed in the qualitative comparison of Fig. 4. On model 4654 (circular sharp boundaries), generic backbones such as PointNet++, DGCNN, and PointMLP detect almost no sharp features, reflecting a strong bias toward the majority non-sharp class. Task-specific methods such as PIE-Net and BoundED improve recall but generate excessive false positives across smooth surfaces, failing to distinguish sharp boundaries from non-sharp regions. On model 5251 (rectilinear sharp boundaries), the same pattern holds: generic backbones produce sparse or empty detections, while over-detecting methods lack spatial selectivity. In contrast, SFD-Net recovers the circular boundaries of model 4654 and the straight edges of model 5251 with spatial coherence closest to ground truth, while suppressing spurious responses on non-sharp surfaces.

4.4 Plug-and-Play LGD Evaluation

LGD improves both Average F1 and FPR across all three backbones on ABC_{none} without any architectural modification (Table 2). Gains in F1 range from +4.05 pp (RepSurf) to +7.00 pp (PointNet++), while FPR decreases by 2.62–9.08 pp, confirming that LGD provides complementary geometric cues that generalize across

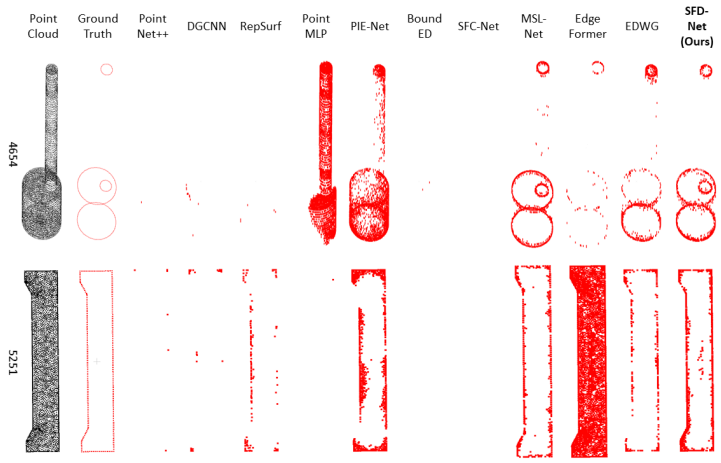

Fig. 4: Qualitative comparison on $\mathrm{ABC}_{\mathrm{none}}$ for two representative models (4654: curved boundary object; 5251: rectilinear boundary object). Ground truth shows sharp feature points only; red points indicate predicted sharp features. Additional visualizations in the supplementary materials validate generalization.

backbone architectures. Notably, LGD + EDWG approaches SFD-Net performance (60.80 vs. 61.33 F1), suggesting that the multi-scale isotropy descriptor is the primary driver of SFD-Net's gains over the baseline EDWG.

4.5 Ablation Study

Table 3 ablates three design axes of SFD-Net on $\mathrm{ABC}_{\mathrm{none}}$.

Removing LGD entirely (a) incurs the largest single F1 drop of -2.78 pp, confirming that geometric descriptors are the primary contribution. Using a smaller scale set (b) or replacing per-scale concatenation with averaging (c) both degrade F1, validating the choice of nested scales $k \in \{20, 40, 80\}$ with concatenation.

Among Transformer placements, the dual SA + FP configuration outperforms both single-position variants (e, f) and the Transformer-free baseline (d) in F1, confirming that cross-patch aggregation and cross-scale consistency are complementary.

For the loss, Cross-Entropy (i) yields the highest FPR (40.85%), while Dice alone (h) achieves the lowest FPR (36.80%) but reduces F1 to 59.16%. The composite Focal Tversky–Dice objective attains the best F1 (61.33%) with near-optimal FPR (36.93%), demonstrating that jointly optimizing overlap and false positive control is essential under severe class imbalance.

4.6 Generalization to Real Scans

Figure 5 shows zero-shot sharp feature detection on `Area_1_hallway_4` (363K points), where the model trained on $\mathrm{ABC}_{\mathrm{none}}$ is applied directly without re-

Table 2: Plug-and-play evaluation of LGD on ABC_{none} (%). LGD is prepended to existing backbones without architectural modification. ΔF1 and ΔFPR report absolute change over the corresponding baseline (↑ improvement; ↓ degradation). **Bold**: best; underline: second best. Full per-class results are provided in the supplementary material.

Method	F1↑	FPR↓	ΔF1	ΔFPR
PointNet++	50.89	48.57	—	—
LGD + PointNet++	57.89	43.70	↑7.00	↓4.87
RepSurf	55.23	45.96	—	—
LGD + RepSurf	59.28	43.34	↑4.05	↓2.62
EDWG	54.95	46.06	—	—
LGD + EDWG	60.80	36.98	↑5.85	↓9.08
SFD-Net (ours)	**61.33**	**36.93**	—	—

Table 3: Ablation study on ABC_{none} (%). Each row isolates one design choice while keeping all others fixed to the full SFD-Net configuration. Input dimensionality is noted in parentheses (ch = channels). **Bold**: best; underline: second best. Loss-term weights and all other hyperparameters are reported in the supplementary material.

Category	Module	Variant	FPR↓	F1↑
Geom. Prior	Descriptor	(a) Coordinates only (3ch)	41.08	58.55
		(b) $k \in \{10, 20, 40\}$ (6ch)	38.87	61.16
		(c) Averaged descriptor (4ch)	38.87	60.05
Network	Transformer	(d) No Transformer	37.60	61.01
		(e) FP only	38.32	60.02
		(f) SA only	39.91	60.18
	Loss	(g) Focal Tversky only	38.14	59.93
		(h) Dice only	**36.80**	59.16
		(i) Cross-Entropy	40.85	60.58
SFD-Net (full)			36.93	**61.33**

training. Because point-wise sharp labels are unavailable for real scans, we report pseudo-label-based metrics with appropriate caveats: as the pseudo-labels are derived from semantic-category boundaries, they capture only coarse object-level edges and tend to penalize fine geometric detections as false positives, so the absolute scores understate true detection quality. Under this proxy, SFD-Net attains the best F1 and FPR among the compared methods (Table 4). Quantitative metrics are reported on the standard Area 5 test split, following the conventional S3DIS protocol; since no S3DIS data is used for training, the qualitative scenes shown here and in the supplementary material (drawn from Areas 1 and 2) serve only for visual illustration and are independent of the Area 5 evaluation fold. Qualitatively, most baselines fall into one of two failure modes: generic backbones such as RepSurf, PointMLP, and SFC-Net detect almost no sharp structures, while DGCNN, PIE-Net, and EDWG generate pervasive false positives across flat surfaces. SFD-Net avoids both failure modes, recovering structural boundaries with spatial coherence comparable to the ground truth while suppressing spurious activations. These results confirm that local geometric relations learned from 10K-point CAD instances transfer reliably to real scenes

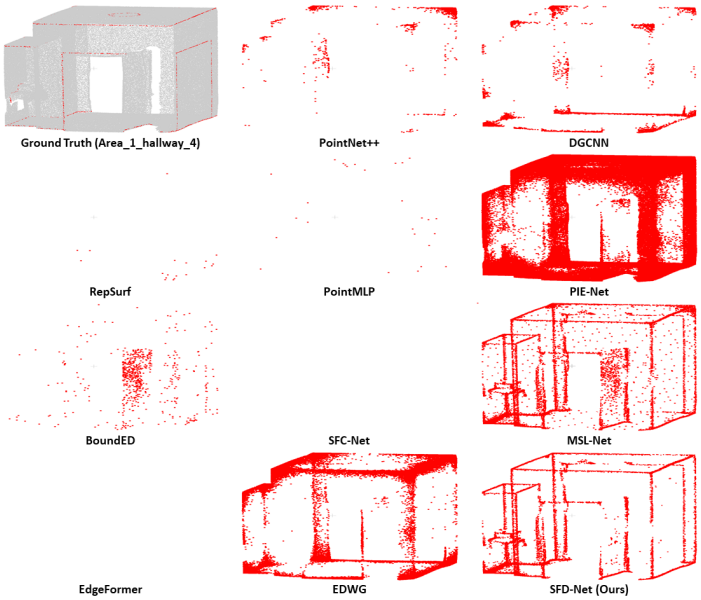

Fig. 5: Zero-shot sharp feature detection on S3DIS (`Area_1_hallway_4`, 363K points); quantitative results in Table 4. Most baselines miss the majority of sharp structures or over-detect on flat surfaces, whereas SFD-Net produces spatially coherent boundaries with suppressed spurious responses. Additional scenes are in the supplementary material.

whose point counts are more than an order of magnitude larger ($\sim$36$\times$; 363K vs. 10K points).

Efficiency. Although SFD-Net builds on PointNet++ and therefore inherits its higher inference cost, LGD is architecture-agnostic and benefits faster backbones as well: as shown in Tables 2 and 4, LGD + EDWG reaches 60.80 F1—within 0.53 pp of SFD-Net—while running over 60$\times$ faster (2.65 vs. 170.7 ms per 500-point cloud). This indicates that the accuracy gains of LGD are not tied to a heavy backbone and can be obtained at substantially lower latency when deployment efficiency is critical. Note that LGD itself adds only a one-time CPU preprocessing cost (35.9 ± 1.9 ms per 10K-point cloud, linear in N)—separate from the network inference times above—which we analyze in the supplementary material.

5 Conclusion

We proposed SFD-Net, a two-stage sharp feature detection framework that couples a compact multi-scale Local Geometric Descriptor (LGD) with an enhanced PointNet++ backbone. LGD provides rigid-motion and scale-invariant geometric cues that remain stable under increased k-NN spacing variability and additive

Table 4: Zero-shot quantitative results on S3DIS Area 5 (%) and computational cost (RTX 3090, batch 1, 500 points, mean±std over 100 runs). Pseudo-label metrics understate true detection quality. FLOPs for RepSurf are omitted owing to its custom surface-construction operators. **Bold**: best.

Method	S3DIS Area 5		Efficiency			
	F1↑	FPR↓	Time (ms)↓	Params (M)↓	FLOPs↓	Mem. (MB)↓
PointNet++	50.48	49.59	163.0±5.7	1.88	1.31	53.4
RepSurf	50.01	48.19	5.81±0.38	**0.98**	—	**18.5**
EDWG	50.88	45.40	**2.65±0.04**	1.20	**1.28**	33.8
SFD-Net (ours)	**53.20**	**41.63**	170.7±8.9	4.59	2.05	113.7

noise, addressing the joint density-variation and noise challenge that prior methods rarely tackle. On the ABC benchmark, SFD-Net achieves state-of-the-art Average F1-score across all four noise conditions. LGD further improves Average F1 and FPR in a plug-and-play manner across three independent backbones without architectural modification, and zero-shot transfer to S3DIS confirms generalization to real scans. These results suggest that reliable sharp-feature localization can benefit downstream tasks such as wireframe generation, registration, and boundary-aware semantic segmentation.

Several directions remain for future work. Replacing PCA-based normal estimation with a learned estimator could improve robustness under heavy noise and severe density imbalance. Beyond binary sharp/non-sharp detection, distinguishing finer geometric categories such as creases, corners, and boundaries would broaden the applicability of the method to richer downstream tasks. Extending evaluation to real-scene corpora with reliable annotations would motivate semi-supervised strategies that combine synthetic CAD labels with proxy supervision from real scans. Finally, moving toward a fully end-to-end differentiable design could improve feature alignment and training efficiency.

Acknowledgements

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2021-0-00315, Development of Techniques for Space-Aware Projection and Contents to Transform an Indoor Space to an XR environment, 60%) and (No. RS-2026-25516746, Development of On-Device VLA-Based Real-Time Action Intelligence Technology for Task Execution of Autonomous Agents, 40%).

References

1. Abraham, N., Khan, N.M.: A novel focal tversky loss function with improved attention u-net for lesion segmentation. In: 2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019). pp. 683–687. IEEE (2019)

2. Armeni, I., Sener, O., Zamir, A.R., Jiang, H., Brilakis, I., Fischer, M., Savarese, S.: 3d semantic parsing of large-scale indoor spaces. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1534–1543 (2016)
3. Bazazian, D., Casas, J.R., Ruiz-Hidalgo, J.: Fast and robust edge extraction in unorganized point clouds. In: 2015 international conference on digital image computing: techniques and applications (DICTA). pp. 1–8. IEEE (2015)
4. Bode, L., Weinmann, M., Klein, R.: Bounded: Neural boundary and edge detection in 3d point clouds via local neighborhood statistics. ISPRS Journal of Photogrammetry and Remote Sensing **205**, 334–351 (2023)
5. Cao, L., Xu, Y., Guo, J., Liu, X.: Wireframenet: A novel method for wireframe generation from point cloud. Computers & graphics **115**, 226–235 (2023)
6. Chen, S., Tian, D., Feng, C., Vetro, A., Kovačević, J.: Fast resampling of three-dimensional point clouds via graphs. IEEE Transactions on Signal Processing **66**(3), 666–681 (2017)
7. Chen, S., Fang, Z., Wan, S., Zhou, T., Chen, C., Wang, M., Li, Q.: Geometrically aware transformer for point cloud analysis. Scientific Reports **15**(1), 16545 (2025)
8. Du, S., Ibrahimli, N., Stoter, J., Kooij, J., Nan, L.: Push-the-boundary: Boundary-aware feature propagation for semantic segmentation of 3d point clouds. In: 2022 International Conference on 3D Vision (3DV). pp. 1–10. IEEE (2022)
9. Gong, J., Xu, J., Tan, X., Zhou, J., Qu, Y., Xie, Y., Ma, L.: Boundary-aware geometric encoding for semantic segmentation of point clouds. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 1424–1432 (2021)
10. Guerrero, P., Kleiman, Y., Ovsjanikov, M., Mitra, N.J.: Pcpnet learning local shape properties from raw point clouds. In: Computer graphics forum. vol. 37, pp. 75–85. Wiley Online Library (2018)
11. Himeur, C.E., Lejemble, T., Pellegrini, T., Paulin, M., Barthe, L., Mellado, N.: 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)
12. Hoppe, H., DeRose, T., Duchamp, T., McDonald, J., Stuetzle, W.: Surface reconstruction from unorganized points. In: Proceedings of the 19th annual conference on computer graphics and interactive techniques. pp. 71–78 (1992)
13. Huang, X., Han, B., Ning, Y., Cao, J., Bi, Y.: Edge-based feature extraction module for 3d point cloud shape classification. Computers & graphics **112**, 31–39 (2023)
14. Huber, P.J.: Robust statistics. In: International encyclopedia of statistical science, pp. 1248–1251. Springer (2011)
15. Jiao, X., Lv, C., Yi, R., Zhao, J., Pan, Z., Wu, Z., Liu, Y.J.: Msl-net: Sharp feature detection network for 3d point clouds. IEEE Transactions on Visualization and Computer Graphics **30**(9), 6433–6446 (2023)
16. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G., et al.: 3d gaussian splatting for real-time radiance field rendering. ACM Trans. Graph. **42**(4), 139–1 (2023)
17. Kim, M., Oh, I., Yun, D., Ko, K.: Improved semantic segmentation network using normal vector guidance for lidar point clouds. Journal of Computational Design and Engineering **10**(6), 2332–2344 (2023)
18. Koch, S., Matveev, A., Jiang, Z., Williams, F., Artemov, A., Burnaev, E., Alexa, M., Zorin, D., Panozzo, D.: Abc: A big cad model dataset for geometric deep learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9601–9611 (2019)
19. Liu, G., Li, X., Sun, S., Yi, W.: Robust and fast normal mollification via consistent neighborhood reconstruction for unorganized point clouds. Sensors **23**(6), 3292 (2023)

20. Liu, Y., Wang, R., Huang, S., Cai, G.: Edgediff: Edge-aware diffusion network for building reconstruction from point clouds. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 17008–17018 (2025)
21. Ma, X., Qin, C., You, H., Ran, H., Fu, Y.: Rethinking network design and local geometry in point cloud: A simple residual mlp framework. arXiv preprint arXiv:2202.07123 (2022)
22. Oh, I., Ko, K.H.: Automated recognition of 3d pipelines from point clouds. The Visual Computer **37**(6), 1385–1400 (2021)
23. Pan, X., Xia, Z., Song, S., Li, L.E., Huang, G.: 3d object detection with pointformer. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 7463–7472 (2021)
24. Pauly, M., Gross, M., Kobbelt, L.P.: Efficient simplification of point-sampled surfaces. In: IEEE Visualization, 2002. VIS 2002. pp. 163–170. IEEE (2002)
25. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 652–660 (2017)
26. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Advances in neural information processing systems **30** (2017)
27. Ran, H., Liu, J., Wang, C.: Surface representation for point clouds. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 18942–18952 (2022)
28. Roh, W., Jung, H., Nam, G., Yeom, J., Park, H., Yoon, S.H., Kim, S.: Edge-aware 3d instance segmentation network with intelligent semantic prior. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20644–20653 (2024)
29. Slamu, W., Cao, J., Yao, X.: Sharp features extraction from point clouds. International Journal of Image and Graphics **12**(04), 1250029 (2012)
30. Song, Z., Liu, L., Jia, F., Luo, Y., Jia, C., Zhang, G., Yang, L., Wang, L.: Robustness-aware 3d object detection in autonomous driving: A review and outlook. IEEE Transactions on Intelligent Transportation Systems **25**(11), 15407–15436 (2024)
31. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: International Workshop on Deep Learning in Medical Image Analysis. pp. 240–248. Springer (2017)
32. Thomas, H., Qi, C.R., Deschaud, J.E., Marcotegui, B., Goulette, F., Guibas, L.J.: Kpconv: Flexible and deformable convolution for point clouds. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 6411–6420 (2019)
33. Wang, J., Chen, M., Karaev, N., Vedaldi, A., Rupprecht, C., Novotny, D.: Vggt: Visual geometry grounded transformer. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 5294–5306 (2025)
34. Wang, W., Zhang, Y., Ge, G., Yang, H., Wang, Y.: A new approach toward corner detection for use in point cloud registration. Remote Sensing **15**(13), 3375 (2023)
35. Wang, X., Xu, Y., Xu, K., Tagliasacchi, A., Zhou, B., Mahdavi-Amiri, A., Zhang, H.: Pie-net: Parametric inference of point cloud edges. Advances in neural information processing systems **33**, 20167–20178 (2020)
36. Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M.: Dynamic graph cnn for learning on point clouds. ACM Transactions on Graphics (tog) **38**(5), 1–12 (2019)

37. Xiang, T., Zhang, C., Song, Y., Yu, J., Cai, W.: Walk in the cloud: Learning curves for point clouds shape analysis. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 915–924 (2021)
38. Xie, Y., Tu, Z., Yang, T., Zhang, Y., Zhou, X.: Edgeformer: local patch-based edge detection transformer on point clouds. Pattern Analysis and Applications **28**(1), 11 (2025)
39. Xu, M., Zhang, J., Zhou, Z., Xu, M., Qi, X., Qiao, Y.: Learning geometry-disentangled representation for complementary understanding of 3d object point cloud. In: Proceedings of the AAAI conference on artificial intelligence. vol. 35, pp. 3056–3064 (2021)
40. Xu, Y., Cao, L., Geng, S., Liu, X.: Rr-net: 3d roof reconstruction from airborne lidar point clouds via edge segmentation and wireframe generation. IEEE Transactions on Geoscience and Remote Sensing (2025)
41. Xu, Y., Guo, J., Cao, L., Liu, X.: Edwg: Efficient edge detection and wireframe generation from point clouds. IEEE Transactions on Instrumentation and Measurement (2025)
42. Yang, B., Luo, W., Urtasun, R.: Pixor: Real-time 3d object detection from point clouds. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp. 7652–7660 (2018)
43. Yu, L., Li, X., Fu, C.W., Cohen-Or, D., Heng, P.A.: Ec-net: an edge-aware point set consolidation network. In: Proceedings of the European conference on computer vision (ECCV). pp. 386–402 (2018)
44. Yun, D., Kim, S., Heo, H., Ko, K.H.: Automated registration of multi-view point clouds using sphere targets. Advanced Engineering Informatics **29**(4), 930–939 (2015)
45. Zhang, J., Cao, J., Liu, X., Chen, H., Li, B., Liu, L.: Multi-normal estimation via pair consistency voting. IEEE transactions on visualization and computer graphics **25**(4), 1693–1706 (2018)
46. Zhao, H., Jiang, L., Jia, J., Torr, P.H., Koltun, V.: Point transformer. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 16259–16268 (2021)
47. Zhao, T., Yu, M., Alliez, P., Lafarge, F.: Sharp feature consolidation from raw 3d point clouds via displacement learning. Computer Aided Geometric Design **103**, 102204 (2023)
48. Zhu, X., Du, D., Chen, W., Zhao, Z., Nie, Y., Han, X.: Nerve: Neural volumetric edges for parametric curve extraction from point cloud. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13601–13610 (2023)