The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-M-9-2025
30th CIPA Symposium "Heritage Conservation from Bits:
From Digital Documentation to Data-driven Heritage Conservation", 25–29 August 2025, Seoul, Republic of Korea

Deep Contour Detection for Enhanced Heritage Sites Visualization by Laser Point Clouds

Tian Zhang[1*], Sagi Filin[1]

[1]Mapping and Geo-Information Engineering, Technion – Israel Institute of Technology, Haifa, Israel -

(tianz, filin)@technion.ac.il (*corresponding author)

Keywords: Point Cloud Denoising, Deep Learning, 3D Contour Detection, Heritage Documentation

Abstract

Laser scanning is a key tool for capturing detailed 3D representations of cultural heritage sites, enabling applications in visualization, conservation, and analysis. However, the resulting point clouds comprise of visually monotonous surfaces with little intuitive value for direct interpretation. Conventional representations—such as collections of surfaces, 3D lines, or boundary-based topological models—often lack the expressiveness or adaptability needed to capture semantically rich, human-intuitive features like salient architectural regions. These salient regions serve as interpretable medium to understand the scan content but remain difficult to extract. Therefore, this paper proposes a framework that learns salient region-aware representations from raw point clouds. Central to the proposed approach is a heat-propagation-driven graph network trained with a new spectral supervision signal. To ensure scalability, our network operates on local patches and introduces a prediction aggregation scheme that efficiently scales to heritage-scale scenes with millions of points. Results show superior performance over existing salient region depiction approaches, producing high-quality results in under two seconds. It is orders of magnitude faster than state-of-the-art methods, even for large-scale noisy terrestrial scans of millions of points.

1. Introduction

Laser scanning has become a ubiquitous and powerful tool for the 3D documentation and analysis of cultural heritage sites. These scans, typically captured using terrestrial laser scanners, provide highly detailed geometric information. However, most campaigns apply a uniform point density across all regions—resulting in dense, semantically unstructured point clouds. While the raw data is valuable, its sheer volume and lack of semantic content make interpretation and visualization a significant challenge. Thus, a suitable representation from the native point form is of great interest for e.g., visual interpretation, analysis, and planning (Hackel et al., 2016; Wang et al., 2020; Arav et al., 2022; Zang et al., 2023). A key challenge in devising this representation lies in selectively emphasizing salient regions, namely important entities of structural complexity that attract attention. In the cultural heritage-related realm, these regions are often tightly linked to architectural significance and historical context. Current representations often fall short of capturing this saliency, where existing methods range from detecting valley- and ridge-lines, delineating object boundaries by 3D lines, and fitting planar surfaces (Judd et al., 2007; Xia and Wang, 2017; Lu et al., 2019; Yu and Lafarge, 2022). While useful, these approaches tend to lack the expressiveness or adaptability required to capture semantically rich architectural features in a form intuitive to humans.

Early works focusing on salient regions, have largely relied on handcrafted geometric traits, e.g., curvature, normal variation, or surface intersections (Judd et al., 2007; Hackel et al., 2016). These methods typically segmented regions of interest and then connected them using such heuristics as shortest paths (Xia and Wang, 2017; Lu et al., 2019). As such, they became dependent on the normal estimation quality and handcrafted thresholds, which are unreliable in the presence of noise and non-uniform densities. In more recent years, neural-based architectures have emerged to overcome the limitations of heuristic-driven pipelines. Wang et al. (2020) proposed a mesh-based neural model that classifies salient points and forms contours, though it is constrained to small, watertight meshes. Himeur et al. (2021) trained a convolutional network using distance-based labels to classify both salient and near-salient regions, boosting recall but still relying on handcrafted geometric features. Zang et al. (2023) utilized dynamic-graph convolutions to detect salient regions and built dense graphs to infer connectivity. In a domain-specific setting, Huang et al. (2024) employed a transformer architecture to jointly detect and connect salient regions in building roof structures. Alternatively, Arav et al. (2025) shifted focus toward learning surface anomalies—areas difficult to fit during surface reconstruction—as indicators of saliency. Despite promising advances, by formulating salient region detection purely as a classification problem—often without accounting for data noise, existing methods tend to produce sub-optimal outputs that require extensive post-processing. The common disregard for noise leads to numerous false responses, introducing irrelevant points that hinder accurate depiction. Furthermore, most architectures have primarily been tested on watertight, low-volume datasets, indicating limited applicability to real-world, large-scale heritage scans.

To address these limitations, we propose in this paper a neural framework that aims to improve the data quality and highlight regions of interest robustly under noise and varying densities. We introduce a new objective in contrast to classification-based approaches, which better reflects the underlying geometry when detecting salient regions. To achieve faithful representation, we focus on generating high confidence-region proposals. To efficiently handle large-scale data, our approach adopts a divide-and-conquer strategy: it processes small, overlapping patches independently, and then aggregates the results into a coherent global representation. This enables our model to process point clouds with millions of points in under two seconds. The proposed model shows significant improvement over existing learning-based solutions on noisy, challenging heritage scans. To the best of our knowledge, this is the first approach that en-

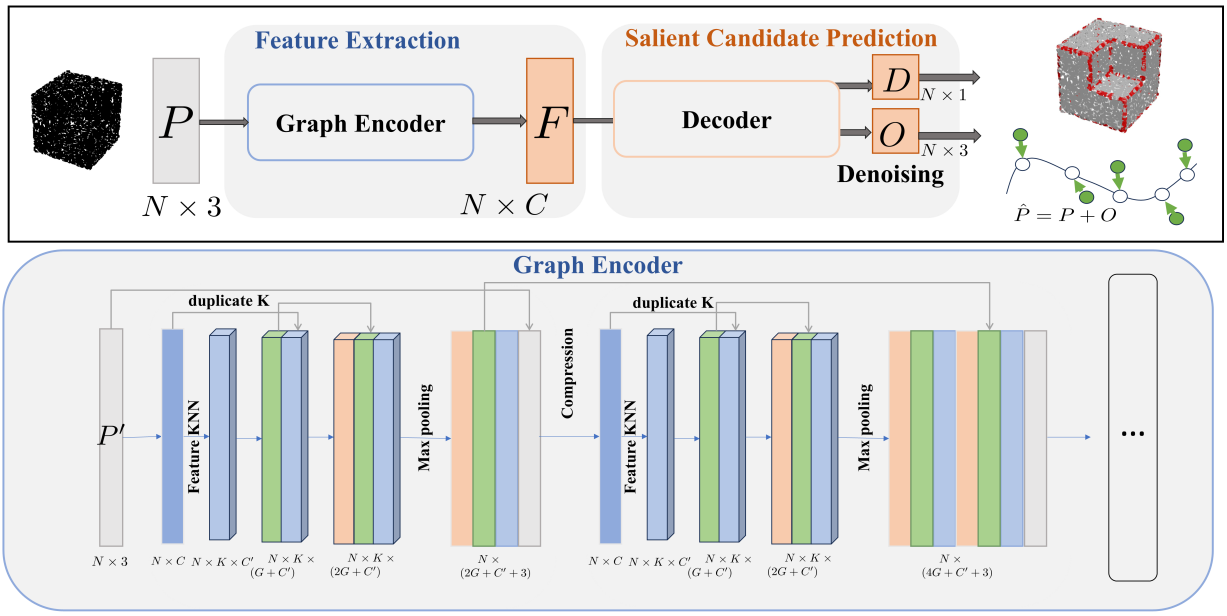

Figure 1. Overview of our deep contour detection network.

ables direct salient region detection from large, noisy data of millions of points.

2. Methodology

Our learning formulation is built upon a dynamic graph convolutional (DGCNN) encoder-decoder architecture designed to robustly capture both local geometric cues and global structural context. Specifically, we employ five layers of densely connected dynamic graph convolutions with skip connections to effectively model hierarchical point relationships. This is followed by a transformer refinement layer that enhances semantic discrimination across surface regions.

Our encoder learns point representations that are both feature-aware and robust to varying noise levels. To supervise this representation, we employ two parallel MLP heads: *i*) a *denoising head*, which predicts per-point offsets to refine raw input points toward the underlying clean surface, and *ii*) a *salient region proposal head*, which predicts a differentiable distance to feature (DDF) value per-point (Fig. 1). To make this representation geometry-aware, we additionally supervise the network to regress a scalar field encoding the normalized distance-to-feature for each point, forming a smooth approximation to a surface-aware importance field. This predicted field forms the basis of a *distance-to-feature representation*. We stitch overlapping patch-level predictions into a global proposal for large-scale scans, enabling coherent extraction of salient structures across complex surfaces. This modular design—graph-based encoding, saliency-aware surface depiction, and patch-based global modeling—allows our method to generalize well across both synthetic and large real-world datasets.

2.1 Network Architecture

Graph Encoder. Our graph encoder is designed to learn representations well-suited for DDF regression. To this end, we adopt a multi-stage architecture with dense blocks and skip connections to ensure rich local aggregation and efficient gradient flow. Each block convolves a point’s feature f_i with its neighbors $f_j \in \mathcal{N}i$ as:

$$e_{ijm} = \mathrm{ReLU}\left(\theta_m(f_j - f_i) + \phi_m f_i\right), \tag{1}$$

where θ_m and ϕ_m are learnable MLPs at layer m. The resulting grouped features are processed through dense MLP stacks, and aggregated as:

$$x_{im} = \max_{j \in \mathcal{N}i} \mathrm{MLP}\left(\mathrm{MLP}(e_{ijm}) | e_{ijm}\right), \tag{2}$$

where $|$ denotes concatenation. Dense intra-block connections and inter-block skip-connections allow deep representation learning with minimal redundancy, while maintaining a compact model size (Fig. 1). We add a transformer layer in the end to refine the features per point and obtain a shape context, F_i, per point.

Decoder. We learn a scalar DDF $\hat{D} \in \mathcal{R}^{N\times 1}$ from the learned shape context F_i through a score-function as detailed in Zhang and Filin (2024). To further improve robustness under noisy input, we use the learned features F_i to concurrently predict per-point offsets $O \in \mathcal{R}^{N\times 3}$ through a lightweight three-layer MLP.

Salient candidate prediction objective. A naïve approach to salient structure modeling may rely on direct edge prediction. However, this strategy often suffers from severe class imbalance and high sensitivity to prediction noise. Instead, we propose to learn a DDF field (Fig. 2), where each point predicts its scalar proximity to the nearest salient feature. This formulation is not only more stable but also aligns naturally with the continuous nature of geometric structures. Rather than relying on simple Euclidean distances (Matveev et al., 2022), we ground our approach in intrinsic geometry and formulate the DDF prediction as a distance field regression problem. Inspired by the *heat method* and its modeling on non-rigid complex forms (Crane et al., 2017; Zhang et al., 2023), we leverage the observation that the gradient of a diffused heat field aligns with the geodesic

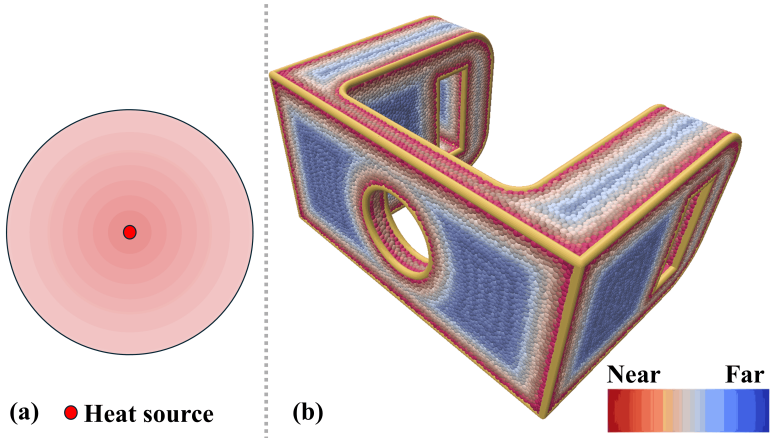

Figure 2. (a) Heat propagation (b) distance to feature D as a smoothing varying scalar field.

paths. Let Δ be the negative semi-definite Laplace-Beltrami operator, heat impulse μ_0 diffused over a brief time t results in a smooth field μ_t given by:

$$(\mathbf{I} - t\Delta)\mu_t = \mu_0, \tag{3}$$

where $\mathbf{I}$ is the identity matrix and μ_0 is a Dirac delta. The resulting vector field $\mathbf{X} = -\nabla\mu_t/|\nabla\mu_t|$ approximates the direction of geodesics. We then solve the Poisson equation:

$$\Delta D = \nabla \cdot \mathbf{X}, \tag{4}$$

to recover the scalar geodesic distance field $D \in \mathcal{R}^{N\times 1}$, which serves as our training target for salient proximity. Importantly, this formulation requires only sparse linear solve, which are efficient and reusable across instances.

Compatibility of architecture and prediction objective The convolutional design of architecture is particularly suited for this objective: graph-based neighborhood aggregation closely mimics the behavior of Laplacian operators Δ. In this sense, our network not only captures the local shape context at each point F_i, but also implicitly aligns with the differential operators used in the heat method. This synergy between architecture and objective enables the network to produce distance fields that are both geometrically coherent and structurally consistent.

2.2 Training Objectives

For the salient candidate prediction, the network is trained to regress a scalar DDF, $\hat{D} \in \mathcal{R}^{N\times 1}$, which estimates the proximity of each point to the nearest salient regions. We supervise this regression with a standard ℓ_1 loss with ground truth computed by the heat method $D \in \mathcal{R}^{N\times 1}$.

$$\mathcal{L}_{\text{DDF}} = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{D}_i - D_i\right|, \tag{5}$$

To further account for the global distribution of predicted distances, we incorporate a histogram loss (Imani and White, 2018; Matveev et al., 2022), which encourages the model’s output distribution to match a target histogram over a predefined interval. In our experiments, we found that using 244 bins yields the best validation performance. Incorporating this loss significantly improves the quality of the regression by enforcing a better global structure in the predicted field. A distinctive aspect of our formulation is the use of a bilateral loss to capture the local surface shape context at each point. This loss differentiates neighboring points on the same surface and those on separate ones, thereby preventing undesired interactions during offset prediction. As a result, our network learns an edge-aware shape context in a straightforward and lightweight manner through a standard back-propagation (Zhang and Filin, 2024).

2.3 Large-volume data inference

Processing large-scale point clouds poses significant computational and memory challenges due to their size and density. To enable inference on these scans, we employ a hierarchical spatial partitioning strategy based on a ball-tree structure (Shtain and Filin, 2020). This allows us to divide the full scan into a collection of non-overlapping patches of fixed scale. To prevent salient regions from being split along patch boundaries—potentially hindering their detection—we vary the leaf size of the ball tree. This generates multi-scale overlapping patches, ensuring that each point appears in multiple contextual neighborhoods. For each patch, our network predicts a per-point scalar field corresponding to the normalized distance to salient features. Since a single point may receive multiple predictions across patches, we aggregate these values to form a consistent global estimate. Among the aggregation strategies we evaluated—we found that computing the *minimum predicted value* across all appearances of a point yielded the most accurate value. Hence, the distance-to-feature prediction for a point p is defined as:

$$d(p) = \min_{d_s \in \mathcal{D}_p} d_s, \tag{6}$$

where $\mathcal{D}_p$ denotes the set of all predictions for p from overlapping patches. In this way, we obtain a coherent DDF estimation that preserves sharp transitions and avoids artifacts introduced by patch boundaries.

3. Results

3.1 Dataset and Implementation

Training Pipeline. Our training is decomposed into two stages to handle imperfect noisy input. First, we train the network solely for the purpose of denoising, to learn the robust representation of the underlying surface from noisy input. Second,

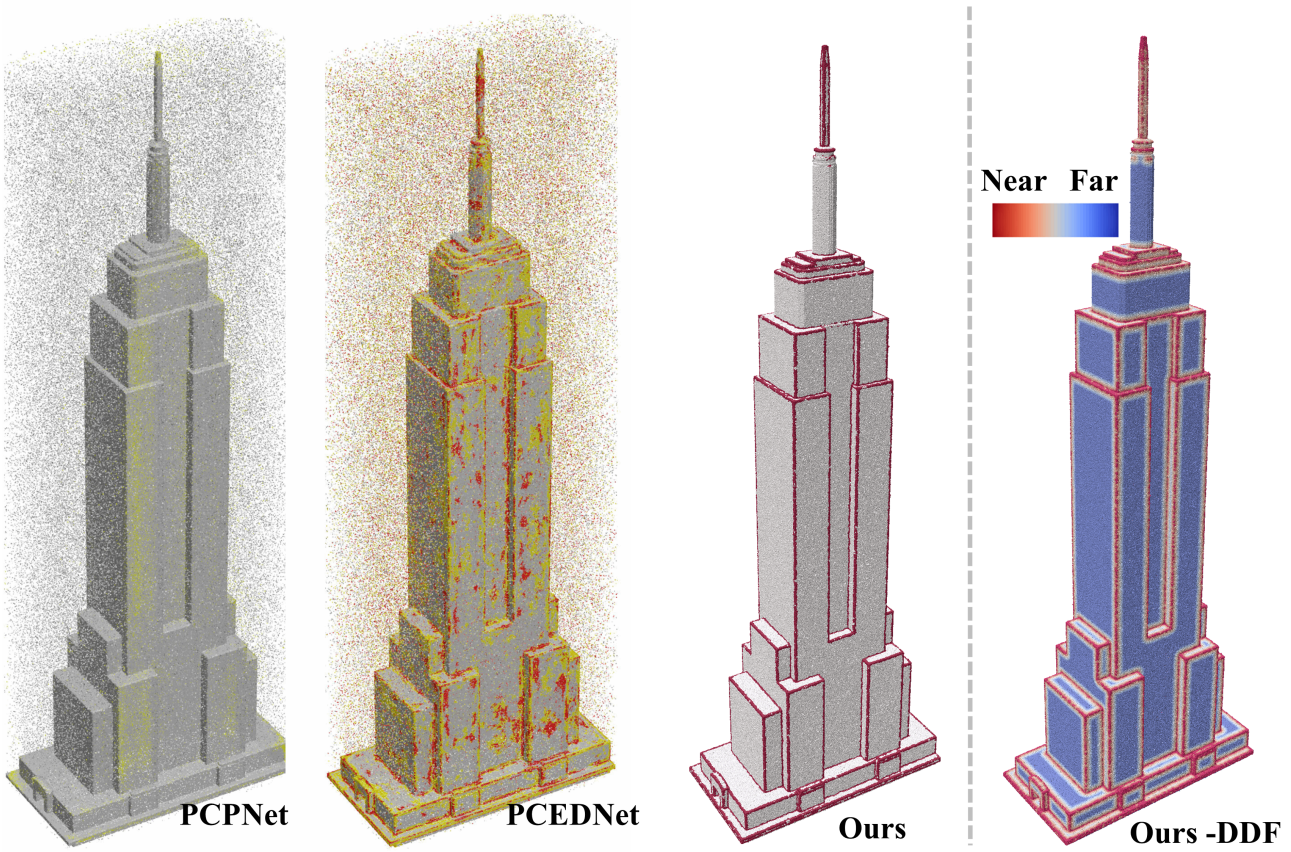

Figure 3. Comparison of salient entity detection (left three columns) on noisy scans (Empire dataset) with 1 million points, PCEDNet results obtained from the supplementary materials at Himeur et al. (2021).

Dataset	Metric	Bilateral	Jet-denoising	MRPCA	GLR	Score-denoise	PCDNF	Ours
PU	CD ($\times 10^4$)	6.304	5.788	5.775	3.839	3.089	4.257	**2.481**
	P2M ($\times 10^4$)	4.73	4.267	4.081	2.707	2.026	3.108	**1.525**
PC	CD ($\times 10^4$)	6.077	5.787	5.57	4.488	3.556	5.126	**3.002**
	P2M ($\times 10^4$)	2.189	2.144	1.976	1.439	1.546	1.879	**1.352**

Table 1. Comparison with state-of-the-art methods (see text for references); CD and P2M are scaled by 10^4.

the network weights are transferred to fine-tune the model to predict salient candidates on CAD data.

Denoising. For quantitative evaluation, we use two benchmark test datasets: the PU-Net (PU, twenty models, Yu et al., 2018) and PointCleanNet (PC, ten models, Rakotosaona et al., 2020). We use the PU training set of twenty shapes, and following Yu et al. (2018), apply Poisson disk sampling to generate point clouds with resolutions ranging from 10K to 50K points. As is customary (e.g., Luo and Hu, 2021; Liu et al., 2023), the point clouds are partitioned into patches of 1K points before being fed to the model. During training, we add Gaussian noise with standard deviations (std.) between 0.5% and 2% of the shape's bounding box dimension. At test time, unseen noise levels in the range of $1\% - 3\%$ are used to evaluate denoising performance and generalization ability.

Salient Candidate Prediction For the second stage, we train on the ABC dataset (Koch et al., 2019), which consists of one million CAD models. Similarly to Matveev et al. (2022), we use the first data chunk for all our experiments, as it provides sufficient diversity and complexity. We compare our method against two baselines on the ABC dataset to quantitatively assess salient region detection quality, namely, PIE-Net (Wang et al., 2020) and DEF (Matveev et al., 2022). However, they are inapplicable to large-scale scans, as their optimization processes fail to converge on such inputs due to memory and runtime constraints. Instead, for large scans, we compare our approach with feature detection models, PCPNet (Guerrero et al., 2018) and PCEDNet (Himeur et al., 2021), as they managed to converge. Specifically, we test on Lans and Empire datasets offered by Himeur et al. (2021), each containing millions of points and are characterized by complex transitions and rich architectural features, offering diverse geometric variations and structural details (Fig. 3 and 4).

Metrics. We assess denoising using Chamfer Distance (CD) and point-to-mesh distance (P2M), with all point clouds normalized to the unit sphere for consistency. For salient region detection, we additionally compute CD and Hausdorff Distance (HD) against ground-truth features sampled from human-annotated curves. These metrics capture both average and worst-case geometric deviations, providing a comprehensive evaluation of feature detection accuracy.

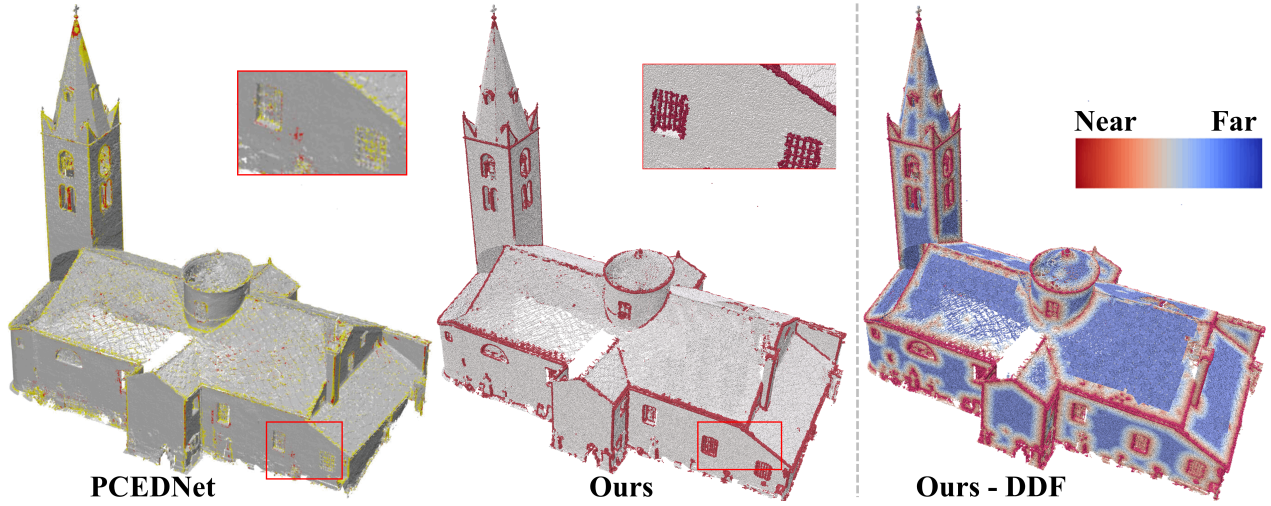

Figure 4. Comparison of salient entities detection on Lans dataset of 1.5 million points.

3.2 Model Analysis

Noise level. The performance of our denoising model consistently outperforms both DL and algorithmic state-of-the-art baselines, at different noise range from $1-3\%$ (whereas training was conducted on $0.5-2\%$) demonstrating robustness even at the unseen noise levels of 3% (Table 1). On the PU dataset, our method achieves CD/P2M scores of $2.481/1.525$, representing improvements of $19.7\%/24.7\%$ over the Score-denoise $(3.089/2.026)$ and $41.7\%/50.9\%$ improvement over PCDNF $(4.257/3.108)$, respectively. Comparable performance gains were also observed on the PC dataset, further validating the effectiveness of our approach.

Quality of salient region detection We evaluate our detected salient regions against baseline methods on a set of 68 representative shapes from the ABC dataset. As the official DEF implementation is not publicly available, we rely on the DEF input data and use the salient point samples shared by the authors. For quantitative comparisons, we report the results of PIE-Net and DEF as presented in the original DEF paper (Table 2). As shown in Table (2), our method consistently outperforms both DEF and PIE-Net. Notably, the Chamfer and Hausdorff distance errors of PIE-Net are largely due to missed or misclassified keypoints—highlighting its limited ability to capture critical geometric features.

	PIE-NET Wang et al. (2020)	DEF Matveev et al. (2022)	Ours(64^3)
CD↓	0.97	0.04	**0.01**
HD↓	2.19	0.55	**0.192**

Table 2. Quantitative comparisons on parametric curve extraction.

3.3 Salient Region Detection on Large TLS Scans

To evaluate the proposed model on large-scale real-world terrestrial laser scans, we performed analysis on the dataset introduced in Himeur et al. (2021). Fig. (3) shows the results on the ”Empire” scene, which consists of over one million points and exhibits substantial noise across large surface areas. PCPNet (Guerrero et al., 2018) demonstrates moderate noise rejection but fails to localize salient geometric features, instead producing scattered predictions. PCEDNet (Himeur et al., 2021), on the other hand, struggles to separate noise from structure and generates visually noisy and inconsistent saliency outputs. In contrast, our method produces sharp and spatially coherent salient region predictions, with significantly fewer false positives. This improvement is attributed to our noise-aware feature learning strategy and the robustness of our distance-to-feature field representation.

Table 3. Inference time comparison on large-scale TLS scans. Our method significantly outperforms prior work in runtime while achieving higher saliency quality.

Method	**Empire (1M pts)**	**Lans (1.5M pts)**
PCEDNet (Himeur et al., 2021)	14.3 min	21.6 min
Ours	**2.0 sec**	**3.1 sec**

In Fig. (4), we apply our method to ”Lans” dataset of a classical architectural church complex with over 1.5 million points. While PCEDNet captures coarse saliency, often producing near-salient predictions near windows and edges, it lacks precise localization. In contrast, our method offers more detailed and structurally aligned saliency maps and transitions around windows, columns, and façade structures. This demonstrates the efficacy of our DDF-based representation in encoding geometric importance with high fidelity in the presence of real-world noise and structural complexity. In visualizations, red indicates salient regions, yellow indicates near-salient regions, and gray represents background points.

Our approach is orders of magnitude faster than prior work, owing to its architectural design that is inherently optimized for large-scale inference. As listed in Table 3, PCEDNet (Himeur et al., 2021) requires over 14.3 minutes for the Empire scan and 21.6 minutes for the Lans scan, both of which are dense terrestrial point clouds. In contrast, our method completes inference in just 2 and 3.1 seconds, respectively. This dramatic speedup stems from our patch-based prediction framework and minimal post-processing, making our approach highly scalable.

4. Conclusions

Neural approaches to heritage interpretation offer promising solutions for enhancing visualization, analysis, and conservation efforts. In this paper, we presented a novel learning-based

salient region representation robust to noise, effectively highlighting content that attracts human attention and reflects architectural complexity. By introducing a Laplacian-aware architecture and our distance-to-feature field formulation, our network learned to assign geometry-aware importance values per point, enabling the retrieval of high-confidence salient region candidates. To handle large-scale scans, we adopted a divide-and-conquer strategy, partitioning the data using a novel hierarchical structure and aggregating predicted salient points. This approach produced highly descriptive predictions even for noisy scans with millions of points. Our results demonstrated significant improvements over state-of-the-art methods.

References

Arav, R., Filin, S., Pfeifer, N., 2022. Content-aware point cloud simplification of natural scenes. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–12.

Arav, R., Wittich, D., Rottensteiner, F., 2025. Evaluating saliency scores in point clouds of natural environments by learning surface anomalies. *ISPRS Journal of Photogrammetry and Remote Sensing*, 224, 235–250.

Crane, K., Weischedel, C., Wardetzky, M., 2017. The heat method for distance computation. *Communications of the ACM*, 60(11), 90–99.

Guerrero, P., Kleiman, Y., Ovsjanikov, M., Mitra, N. J., 2018. Pcpnet learning local shape properties from raw point clouds. *Computer graphics forum*, 37number 2, Wiley Online Library, 75–85.

Hackel, T., Wegner, J. D., Schindler, K., 2016. Contour detection in unstructured 3d point clouds. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1610–1618.

Himeur, C.-E., Lejemble, T., Pellegrini, T., Paulin, M., Barthe, L., Mellado, N., 2021. PCEDNet: A lightweight neural network for fast and interactive edge detection in 3D point clouds. *ACM Transactions on Graphics (TOG)*, 41(1), 1–21.

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

Imani, E., White, M., 2018. Improving regression performance with distributional losses. *International conference on machine learning*, PMLR, 2157–2166.

Judd, T., Durand, F., Adelson, E., 2007. Apparent ridges for line drawing. *ACM transactions on graphics (TOG)*, 26(3), 19–es.

Koch, S., Matveev, A., Jiang, Z., Williams, F., Artemov, A., Burnaev, E., Alexa, M., Zorin, D., Panozzo, D., 2019. Abc: A big cad model dataset for geometric deep learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9601–9611.

Liu, Z., Zhao, Y., Zhan, S., Liu, Y., Chen, R., He, Y., 2023. PCDNF: Revisiting learning-based point cloud denoising via joint normal filtering. *IEEE Transactions on Visualization and Computer Graphics*.

Lu, X., Liu, Y., Li, K., 2019. Fast 3D line segment detection from unorganized point cloud. *arXiv preprint arXiv:1901.02532*.

Luo, S., Hu, W., 2021. Score-based point cloud denoising. *Proc. of the IEEE/CVF ICCV*, 4583–4592.

Matveev, A., Rakhimov, R., Artemov, A., Bobrovskikh, G., Egiazarian, V., Bogomolov, E., Panozzo, D., Zorin, D., Burnaev, E., 2022. Def: Deep estimation of sharp geometric features in 3d shapes. *ACM Transactions on Graphics*, 41(4).

Rakotosaona, M.-J., La Barbera, V., Guerrero, P., Mitra, N. J., Ovsjanikov, M., 2020. Pointcleannet: Learning to denoise and remove outliers from dense point clouds. *Computer graphics forum*, 39number 1, Wiley Online Library, 185–203.

Shtain, Z., Filin, S., 2020. Hierarchical proximity-based oversegmentation of 3-D point clouds for efficient graph feature detection. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2, 273–280.

Wang, X., Xu, Y., Xu, K., Tagliasacchi, A., Zhou, B., Mahdavi-Amiri, A., Zhang, H., 2020. Pie-net: Parametric inference of point cloud edges. *Advances in neural information processing systems*, 33, 20167–20178.

Xia, S., Wang, R., 2017. A fast edge extraction method for mobile LiDAR point clouds. *IEEE Geoscience and Remote Sensing Letters*, 14(8), 1288–1292.

Yu, L., Li, X., Fu, C.-W., Cohen-Or, D., Heng, P.-A., 2018. Punet: Point cloud upsampling network. *Proc. of the IEEE Conference on CVPR*, 2790–2799.

Yu, M., Lafarge, F., 2022. Finding good configurations of planar primitives in unorganized point clouds. *Proc. of the IEEE/CVF Conference on CVPR*, 6367–6376.

Zang, Y., Chen, B., Xia, Y., Guo, H., Yang, Y., Liu, W., Wang, C., Li, J., 2023. LCE-NET: Contour extraction for large-scale 3-D point clouds. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1–13.

Zhang, T., Elnashef, B., Filin, S., 2023. Spatio-temporal registration of plants non-rigid 3-d structure. *ISPRS Journal of Photogrammetry and Remote Sensing*, 205, 263–283.

Zhang, T., Filin, S., 2024. Joint neural denoising and consolidation for portable handheld laser scan. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, 499–505.