ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume X-G-2025
ISPRS Geospatial Week 2025 “Photogrammetry & Remote Sensing for a Better Tomorrow…”, 6–11 April 2025, Dubai, UAE

Edge-aware joint neural denoising and normal estimation for mobile and handheld 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

Keywords: Point cloud denoising, Mobile laser scanners, Handheld laser scanners, Deep learning, Normal estimation

Abstract

Portable laser scanners, handheld and mobile, have gained popularity for their ability to rapidly and economically document scenes. However, the acquired data are characterized by hight levels of noise and by low resolution, both affecting their consequent analysis and 3D modeling. It is customary to enhance their quality by denoising the data by means of point position updates where current approaches independently predict the displacement per point. Such strategies neglect local structural consistency and often yield a non-smooth outcome. To address these shortcomings this paper formulates denoising by local contextual relationships and point assignment to the underlying surface in an end-to-end framework. To extract contextual information it introduces densely packed graph convolution layers and a global attention mechanism. Realizing also that utilization of the conventional L2-norm-driven approaches tends to oversmooth the surface, it introduces a novel bilateral loss that not only mitigates the noise but also preserves sharp geometric features. As a result, our newly developed network learns a shape context representation that measures neighbor similarity and contributes to more accurate surface normals. Performance analysis demonstrates that over 93% of points deviate by ≤ 1 cm – double the percentage achieved by state-of-the-art denoising networks.

1. Introduction

Laser scans acquired by terrestrial platforms provide detailed characterization of 3D object space. Therefore, they are favorable in a plethora of applications, e.g., built environment capture, road map production, city modeling, and forest inventory (Ye et al., 2020; Yu and Lafarge, 2022; Hakula et al., 2023). Though stationary scanners produce the highest quality outcome, the acquisition time they require prohibits the capture of broad scenes. In contrast, portable laser scanners, handheld and mobile (PLS & MLS, respectively) offer broad scene coverage, flexibility, efficiency, and reduced manpower in the acquisition. Nonetheless, due to device instability, inaccurate scanner poses determination, and ranging system limitation, the outcome yields a noisy and sparse scene representation with an excessive rate of outliers that in turn affect downstream applications (Xia et al., 2020; Ren et al., 2022; Li et al., 2024; Conti et al., 2024; Antova, 2024).

Hand-crafted 3D noise attenuation approaches have focused on minimizing the distance towards the actual surface, specifically along its normal direction (Taubin, 1995; Desbrun et al., 2000; Peng et al., 2001). Noise attenuation was in the form of L1-median, graph Laplacian framework, reweighting by distance to the surface, or robust bi-variate polynomial fitting (e.g., Fleishman et al., 2005; Huang et al., 2013; Sun et al., 2015; Mattei and Castrodad, 2017; Dinesh et al., 2020). While successful under some setups, they necessitated heavy manual parameter tuning, yet failed to preserve local geometric details when the noise levels were high and the density low. Current deep learning (DL) point cloud denoising networks aim to predict the point displacement to the surface, either in a Euclidean multi-scale manner, using feature concatenation, or by learning global features (Pistilli et al., 2020; Rakotosaona et al., 2020; Luo and Hu, 2021; Edirimuni et al., 2023b). More recent research has realized the benefit of displacement along the normal direction. Accordingly, normals were introduced as auxiliary information (e.g., via principal component analysis, PCA) for the network to learn both point and normal corrections (Edirimuni et al., 2023a; Liu et al., 2023). Though aiming to become adaptive to different levels of noise (e.g., Pistilli et al., 2020), the learned surface description still struggled to handle high noise levels and outliers, and to preserve sharp shape transitions. Their reliance on L2-norm measures led them to fail when the 3D data density was uneven (e.g., Liu et al., 2023). That is so despite a careful local neighbor selection, advanced feature modeling, and surface normal supervision. Also, their applications were restricted to watertight models and dense scans, leaving the challenging, sparse, and noisy ones unattended.

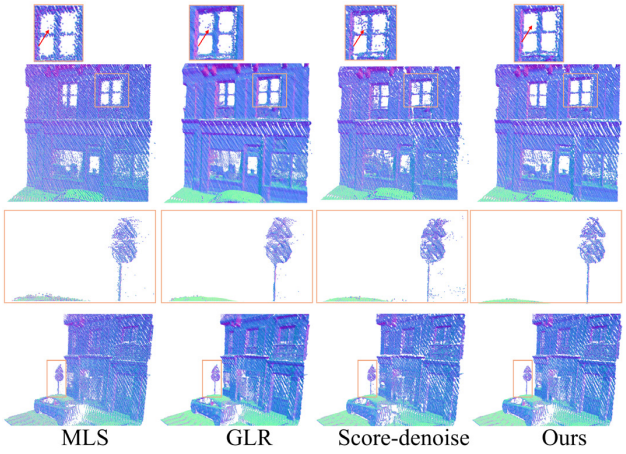

Figure 1. Denoising the RueMadame dataset. Compared to high residual in other approaches, ours recovers clean structures.

To address these challenges, this paper proposes a new neural denoising framework that exhibits robustness to the presence of outliers and preserves underlying surface features. Its contributions are: *i*) an introduction of a new local shape context embedding that captures both local and global surface features; *ii*) a joint treatment of noise and outliers under varying

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-G-2025-1037-2025 | © Author(s) 2025. CC BY 4.0 License.

noise levels; *iii*) the introduction of a novel loss function, which jointly handles point position and surface content; *iv*) a shape-aware surface representation, which leads to a robust surface attribute estimation from denoised points; and *v*) faster training and efficient inference. Analysis shows that our denoising and normal estimation results on benchmarks and real-world scans set state-of-the-art.

2. Related Work

Early point denoising research has focused on weighted averaging over a local neighborhood, using surface subdivision methods for efficient low-pass filtering. Averaging was performed iteratively, gradually smoothing the intersection of planes or high-curvature regions, and therefore required addressing shape shrinkage (Taubin, 1995; Desbrun et al., 2000; Peng et al., 2001). To reflect the underlying surface, moving-polynomial (jet-) fitting facilitated denoising by local projection (Fleishman et al., 2005; Cazals and Pouget, 2005; Guennebaud and Gross, 2007). As local fitting exerted smoothness, Lipman et al. (2007), and Huang et al. (2013) proposed to redistribute points to the local L1-median and introduced repulsion forces to secure coverage of the underlying surface. Nonetheless, their application has led to detail loss. To treat positional updates globally, graph Laplacian-based approaches optimized point positions by reducing the mean curvature (Zeng et al., 2019). Nonetheless, failing at surface intersections, they led to a loss of sharp features (c.f. Sec. 4). In addressing that, bilateral filter-based approaches first estimated surface normals and then denoised the data based on Euclidean proximity and normal differences (Fleishman et al., 2003; Digne and De Franchis, 2017). Though effective at low noise levels, their sensitivity to higher ones and the existence of outliers turned them unreliable in sparse observations. Alternative, robust estimation driven methods aimed to refine the normal direction by minimizing the discrepancy in a local neighborhood and then iteratively updating the points position (Avron et al., 2010; Sun et al., 2015; Mattei and Castrodad, 2017). As they sequentially performed a least-squares adjustment, results were suboptimal and were prone to failure in the presence of large outlier rates, even when incorporating sophisticated graph Laplacian regularization-related terms (Dinesh et al., 2020).

Shifting towards DL approaches, Pistilli et al. (2020) used a dynamic graph convolution (DGCNN) to predict the point's position on a clean surface while applying an L2 loss minimization. This approach resulted in blurring geometric details. A PointNet architecture was applied to detect outliers by Rakotosaona et al. (2020). The noisy points offset predictions were iterative, and an L2-norm loss was used. Later, Luo and Hu (2021) modeled the noise distribution and defined the noise-free surface as the region of the highest density. The network learned the gradient of the distribution during training and iteratively performed a marching process toward dense regions by applying the gradient directions during testing. Due to inexact vector predictions, this iterative process converged very slowly, leading to a lengthy inference time. To secure fast convergence, in Edirimuni et al. (2023b) a progressive data denoising was performed by four replications of their denoising module (the weights per module were learned independently). A DGCNN architecture was used, with L2 norm minimization, and supervision of each intermediate result carried out until becoming noise-free. Edirimuni et al. (2023a) estimated normals and denoised points jointly using a patch-based, PointNet, architecture. Their aim was to learn a common feature under different noise levels for all central points. Then, an MLP predicted the per point denoised coordinates and normal. Here again, an L2 loss was applied. As points were independently processed, the training phase was slow. Liu et al. (2023) also used a patch-based network but added to the input PCA normals, they precomputed, under the premise they may help preserve sharp features and identify relevant neighbors. The network learned to correct the normals along with point offsets. However, this approach was sensitive to the quality of the input normals and was memory-intensive.

By predicting offsets independently, paying less attention to the local context, and using L2-driven cost functions, current approaches tend to yield non-smooth results with a loss of geometric details. The common neglect of outliers' presence also deforms the results further. When surface normals are introduced, their treatment is not adequately reflected in the loss design to ensure consistent predictions. Also, the data on which these architectures were tested are dense and watertight, suggesting that limited attention has been directed toward improving the quality of challenging MLS and PLS data. To address these limitations, we propose a formulation that models local to global shape context, and jointly handles outliers and noise. We introduce a new feature-preserving loss form, that enables the identification of points belonging to geometrically similar features and attenuating the effects of dissimilar points, under varying noise levels.

3. Methodology

Our network is a graph-convolution based, with a point-representation-encoder for shape context extraction followed by a score-estimation-decoder for handling outlier and displacement prediction. Its input is an observed set of N points, P', composed of noisy and outlying returns:

$$P' = \{p_i + d_i\}_{p_i \in P} \cup O, \tag{1}$$

where $p_i \in P$, is the perfect scanned surface sample set, $d_i \in D$ is the ranging noise offsets from the surface; and O is the outlying point set. Their exclusion, $\hat{P} = P' \setminus O$, yields an outlier-free set. $\hat{P}$ is refined by the set of predicted displacements, D, to align the individual points with the true surface.

3.1 Point Representation Encoder

A naive dynamic graph edge convolution does not guarantee a robust description of the local surface structure under varying noise levels and density changes. Also, stacking of dynamic graph convolution layers for multiple levels of local context embedding leads to a vanishing gradient and an overfit to local connections (Pistilli et al., 2020; Liu et al., 2023). Hence, our network introduces densely connected graph-constitutional layers with skip connections. The neighbor proximity in the first layer is set by Euclidean distance while in the following ones, it is set by the closeness of the learned features (in the sense of intrinsic properties). Given a feature vector for the i-th point, f_i, and its neighbors vectors, $f_j \in \mathcal{N}_i$, the dense block convolutions is computed by:

$$e_{ijm} = \text{Relu}\left(\theta_m \cdot (f_j - f_i) + \phi_m \cdot f_i\right), \tag{2}$$

where $\mathcal{N}_i$ is neighborhood, m is the layer index, and θ_m and ϕ_m are MLPs. We further refine each grouped feature, e_{ijm},

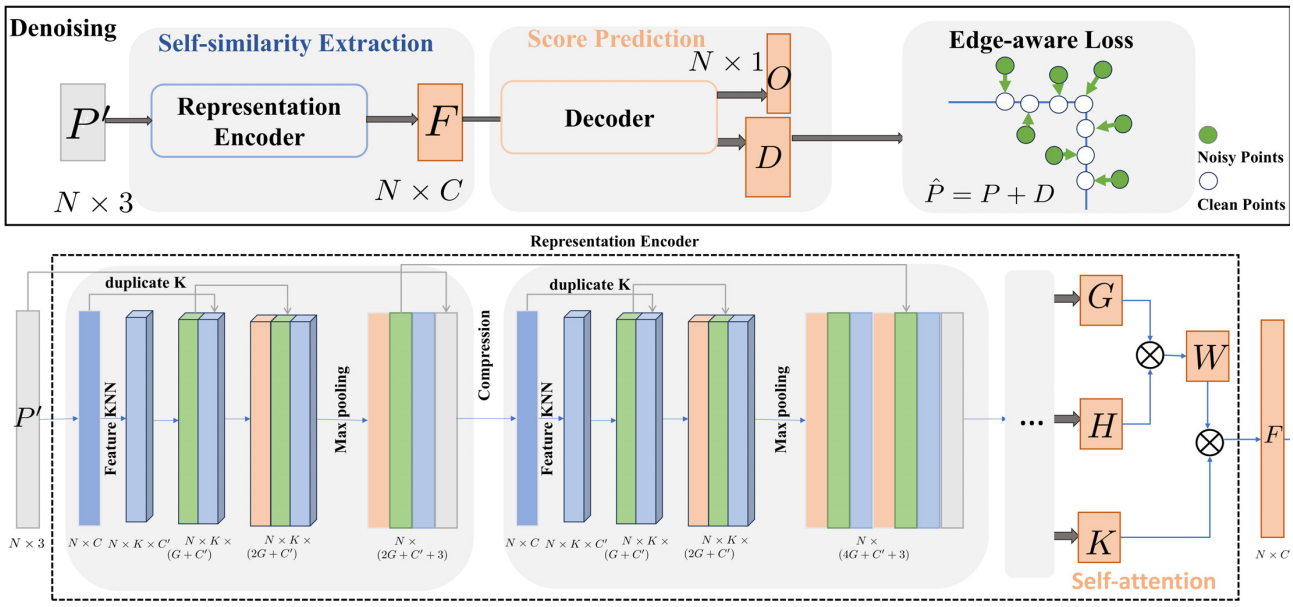

Figure 2. Overview of our denoising network.

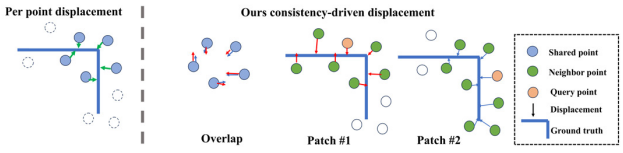

Figure 3. Our consistency driven denoising prediction for shared neighbors, ensuring robustness due to single displacement constraint.

via a chain of densely connected MLPs and the final activation x_{im} of p_i at layer m:

$$x_{im} = \max_{j \in \mathcal{N}_i} \text{MLP}\left(\text{MLP}(e_{ijm}) || e_{ijm}\right), \tag{3}$$

We introduce dense connections both within and between the dense blocks (Fig. 2). Within blocks, the output of each MLP is passed to all subsequent ones. Between blocks, the point features that each block produces are introduced in the form of skip-connections, as an input to the subsequent blocks. This act ensures gradient flow, and contribute to a reduced model size (Fig. 2).

To avoid over-fitting to local cues and secure global context, a self-attention unit (Li et al., 2019) is added. It processes the output of the densely connected blocks, and includes a query, key, and values components, G, H, and K, respectively, where:

$$G = \hat{F}W_G \quad H = \hat{F}W_H \quad K = \hat{F}W_K, \tag{4}$$

$\hat{F}$, is the densely connected blocks output, and W_G, W_H and W_K are MLPs, where the dimensions of G and H remain unchanged. We derive the $N \times N$ weight matrix, W, from the alignment of G and H after the application of a softmax function, providing weights by which K is multiplied to yield the learned per point shape context feature, F_i.

3.2 Decoder, Consistency-driven Score-based Denoising

Prior work denoised point clouds by predicting a per-point offset to the underlying surface using sets of MLPs (Rakotosaona et al., 2020; Edirimuni et al., 2023b). However, they often led to high residuals due to the inability to differentiate outliers. The independent, per-point, prediction also disregarded local neighbor consistency, resulting in a non-smooth output. This is addressed here by using the learned shape context, F_i, which classified each point as an inlier or an outlier by:

$$\text{O}_i = \text{MLP}\left(F_i\right), \tag{5}$$

where O_i is an indicator. We concurrently use F_i to estimate offsets, but instead of predicting offsets per the remaining points, we create dense, overlapping, k-neighbor patches on the point cloud surface and evaluate offsets for its neighbors based on the center point shape context (Fig. 3). To do so, our score estimation module parameterizes traditional point-to-surface distance queries in a high-dimensional space (Fig. 4). It takes, as input, a learnable relative positional encoding of a neighbor $p_j \in \mathcal{N}_i$

$$PosEncoding(p_j) = \text{MLP}\left(p_j - p_i\right), \tag{6}$$

and the shape context, F_i, of p_i. It outputs a predicted displacement $D(p_j)$ for p_j,

$$D(p_j) = \text{Score}\left(PosEncoding(p_j), F_i\right), \tag{7}$$

where $\texttt{Score}(\cdot)$ is an MLP. As p_j and p_i are mutual neighbors, this setting creates overlaps, and accordingly multiple predicted corrections per point, in reference to all the neighborhoods it is part of (Fig. 3). Consequently, we encourage consistent predictions by minimizing discrepancies between ones from different neighborhood sets. Though the central point and its neighbors may lie on different surface elements, our loss function (Sec. 3.3) rules out such an interaction in a straightforward manner. Thereby, we encourage the shape context feature to be surface aware.

3.3 Edge-aware Representation Learning

To enable the filtered point cloud to approximate the surface, a common solution is to apply an L2-norm distance for the loss

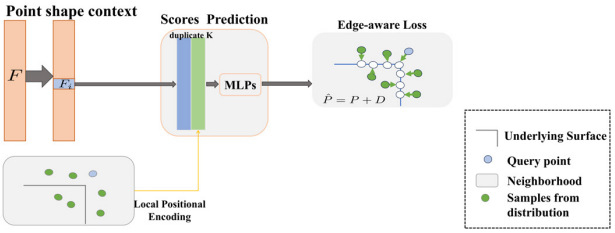

Figure 4. Score prediction module taking the local point shape context and query the offsets to the underlying surface. This query is parameterized by MLPs and shared for all neighbors.

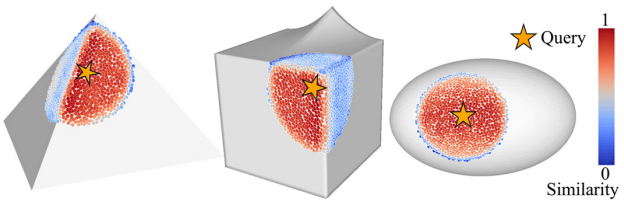

Figure 5. Visualization of the shape context feature similarity to its neighbors for selected PU models. Our learned representation is shape-aware, where the starred point is the query and neighbors' coloring is by cosine similarity of the learned features.

(Rakotosaona et al., 2020; Edirimuni et al., 2023a). However, the L2 loss optimizes displacements towards the closet surface point and leads to the blur of sharp features. Our approach is different. We propose a formulation that reflects the local surface shape and encourages the denoised outcome to minimize a bilateral loss. We define $\hat{p}_i = p_i + D(p_i)$ to be the denoised point and $\bar{p}_j \in \mathcal{N}_i$ a ground-truth point within $\hat{p}_i$ neighborhood, and denote the normal of the ground-truth point closest to $\hat{p}_i$, by $\bar{n}_i$, and by $\bar{n}_j$ that of $\bar{p}_j$. Then, the loss, $L_{\text{bilateral}}$, is given by:

$$L_{\text{bilateral}} = \frac{\sum_{\bar{p}_j \in \mathcal{N}_i} |(\hat{p}_i - \bar{p}_j) \cdot \bar{n}_j| \cdot \phi(\|\hat{p}_i - \bar{p}_j\|) \cdot \theta(\bar{n}_i, \bar{n}_j)}{\sum_{\bar{p}_j \in P_i} \phi(\|\hat{p}_i - \bar{p}_j\|) \cdot \theta(\bar{n}_i, \bar{n}_j)} \tag{8}$$

where $\phi(\|\hat{p}_i - \bar{p}_j\|)$ is a monotonically decreasing weighting function, and $\theta(\bar{n}_i, \bar{n}_j) = \exp\big(-\big(\frac{1-\bar{n}_i \cdot \bar{n}_j}{1-\cos(\gamma)}\big)^2\big)$ decreases with the deviation between $\bar{n}_i$ and $\bar{n}_j$, where γ a bandwidth constant. This loss enforces minimization of difference along the normal and therefore is surface-aware. The term $\theta(\bar{n}_i, \bar{n}_j)$ differentiates neighbors from other surfaces and rules out interactions. Hence it ensures edge-awareness. The surface-aware property in $\hat{p}_i$ is reflected in the shape context F_i through back-propagation.

For consistent performance, we add at the training stage random Gaussian, Laplace, and non-uniform noise to sampled points from the ground truth clean surface.

3.4 Feature-affinity-based Attribute Estimation

Current methods estimate surface features, e.g., normal and curvature values, from noisy scans by applying n-Jet fitting and using the network to assign weights to local neighbors, thereby mitigating noise effect during surface fitting (Ben-Shabat and Gould, 2020). They typically require extensive PCA computations during training, making them computationally intensive and memory-demanding (Liu et al., 2023). Solving linear systems for normal estimation during training iterations further adds to the complexity. Our key contribution here is a shape context feature that inherently captures the similarity of surrounding points and allows robust surface fitting from the denoised points. As Fig. (5) shows, our framework learns edge-aware features that naturally differentiate similar points, e.g., at intersections of planes or singular regions of cubic surfaces. To compute robust attributes from this feature representation, we use the denoised point, $\hat{p}_i$, its local shape context, F_i, and its local neighbors to fit an n-Jet "height function" to any point not in the local tangent space $z(x, y) = \sum_{k=0}^{n} \sum_{j=0}^{k} \phi_{k-j,j} x^{k-j} y^j$, where ϕ is the jet coefficients vector that consists of $N_n = \frac{(n+1)(n+2)}{2}$ terms. In matrix form $\mathbf{M}_{N_p \times N_n} \phi = \mathbf{b}$, where $\mathbf{M}$ is the the Vandermonde matrix, N_p is the number of points, and $\mathbf{b} = \big(z_1, z_2, \cdots, z_{N_p}\big)^T$. To minimize the influence of dissimilar points, we assign weights by the normalized dot product between F_i and $F_j \in \mathcal{N}_i$ (cosine similarities) followed by an application of a softmax function. Setting $\mathbf{W} = \text{diag}(w_1, w_2, \ldots, w_{N_p})$, the least-squares solution is given by $\phi = (\mathbf{M}^T \mathbf{W}\, \mathbf{M})^{-1} \mathbf{M}^T \mathbf{W} \mathbf{b}$. The first two n-jet coefficients yield the estimated normal vector, $\mathbf{n}_i = \frac{(-\phi_1, -\phi_2, 1)}{\|(-\phi_1, -\phi_2, 1)\|}$. The remaining ones represent second-order terms and higher.

4. Results

4.1 Implementation and Dataset

For our quantitative analysis, we utilize two well-established benchmark datasets: PU-Net (PU, 20 models, Yu et al., 2018) and PointCleanNet (PC, 10 models, Rakotosaona et al., 2020). The model is trained using the PU-Net training set, which includes 20 shapes, with point clouds generated from the original meshes through Poisson disk sampling at varying resolutions between 10K and 50K points, in line with the setup in Yu et al. (2018). Adhering to common conventions (e.g., Liu et al., 2023), the point clouds are segmented into local patches of 1K points prior to being processed by the model. To emulate realistic measurement noise during training, Gaussian noise with 0.5% to 2% standard deviation, of the shape's bounding box diameter was added. During testing, to evaluate the model's robustness on previously unseen noise levels, Gaussian noise with std. ranging from 1% to 3% is introduced.

For evaluation on actual MLS scans, we adopt the MLS dataset collected from Paris (Serna et al., 2014), which features complex urban facades along with noise, reflective outliers, and mobile platform motion (Fig. 1). As no ground truth is available, therefore only qualitative analysis is reported, following established protocols (c.f. Luo and Hu, 2021; Liu et al., 2023; Edirimuni et al., 2023b). Additionally, we test our model on point clouds of elaborate architectural structures captured by a GeoSLAM ZEB-REVO handheld LiDAR. These scans exhibit significant noise levels, non-uniform and low-density sampling, and a substantial presence of outliers.

Baselines. We benchmark our method against leading denoising techniques, including classical robust methods such as Jet-denoising (Cazals and Pouget, 2005), bilateral filtering (Bilateral, Digne and De Franchis, 2017), moving robust PCA (MRPCA, Mattei and Castrodad, 2017), and graph Laplacian regularization (GLR, Dinesh et al., 2020). In addition, comparisons are made with recent learning-based methods such as Score-Denoise (Luo and Hu, 2021) and the state-of-the-art PCDNF (Liu et al., 2023). For normal estimation, we compare

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. Performance comparison with state-of-the-art approaches, where both CD and P2M metrics are scaled by a factor of 10^4.

Figure 6. Denoising results for the PU and PC datasets with 3% noise level. Shape dimensions are normalized and coloring is based on the relative normalized distance to the ground truth for uniformity.

against DeepFit's Jet fitting (Jet, Ben-Shabat and Gould, 2020) and PCDNF.

Metrics. To quantitatively assess denoising performance, we employ two widely used metrics: Chamfer Distance (CD) and Point-to-Mesh distance (P2M). Since the point clouds differ in scale and resolution, all denoised outputs are first normalized to fit within a unit sphere prior to evaluation. For quantitative estimation of the derived normals, we measure the per-point angle between the predicted and ground truth normals, and then compute the root-mean-square-error (RMSE) as our metric. We analyze the percentage of points per shape with angular errors lower than $5^\circ$and 10° (pgp5 & pgp10, respectively).

Implementation Details. The training was on an NVIDIA GeForce RTX 2080 GPU (single GPU) using PyTorch 1.9.0 with CUDA 11.1. We train the network for 1M iterations with a learning rate of 1×10^{-4}. For a fair comparison of test times, all learning-based methods were tested on an NVIDIA GeForce RTX 2080 GPU.

4.2 Model Analysis

Noise level. Our denoising approach consistently outperforms both deep learning-based and traditional state-of-the-art methods across a range of noise levels from 1% to 3%—despite being trained only on noise levels between 0.5% and 2%—demonstrating strong generalization to unseen noise conditions, particularly at 3% (Fig. 6 & Table 1). On the PU dataset, our method achieves CD/P2M scores of $2.481/1.525$, corresponding to improvements of $19.7\%/24.7\%$ over Score-denoise ($3.089/2.026$) and $41.7\%/50.9\%$ over PCDNF ($4.257/3.108$). Comparable performance gains are observed on the PC dataset as shown in (Fig. 6).

A visual assessment of challenging curved surfaces and structural shapes—such as the fandisk and icosahedron—reveals that heavy noise severely obscures the original geometry. GLR, Score-denoise, and PCDNF struggle to uncover the true shape, leaving behind considerable residual noise (Fig. 6). We hypothesize that GLR and Score-denoise tend to misinterpret certain noise patterns as genuine surface details, resulting in visibly non-smooth surfaces, particularly in models like the horse and fandisk. Meanwhile, PCDNF's reliance on external normal estimation makes it vulnerable to noise corruption, diminishing its shape reconstruction accuracy under high-noise scenarios. By contrast, our approach is explicitly designed to robustly extract geometric context features across a range of noise intensities, enabling it to better preserve structural details and achieve results that remain faithful to the underlying ground truth.

Runtime. Evaluating our network performance from a runtime perspective, our training time per iteration is only 0.25 sec. compared to 2.38 sec. for the state-of-the-art PCDNF—a lightweight model, an order of magnitude improvement in training efficiency and another merit of our model. Also, as we do not require extensive PCA computations before inference, our denoising (including normal estimation) runtime for 100K points was 20.6 sec. compared to 303.5 sec. of the PCDNF, making our inference 15 times faster.

Normal Estimation. To evaluate the quality of our computed surface normals, we use the PC dataset, where the ground truth normals per shape are available. The point cloud used for evaluation is of low resolution (the number of sampled points is $10\times$ sparser than the ground truth mesh), also featuring noise levels of up to 3%. We compare our normal estimation performance and statistics against the Jet and PCDNF. Results demonstrate how ours outperforms both baseline methods (Fig. 7). As the Jet fitting is a robust polynomial surface fitting designed for low-noise levels, its performance quickly drops compared to ours. As noted, the reliance of the PCDNF on initial normal estimates leads to its failure at high noise levels where the degraded normal quality translates to poor position and normal corrections. Ours, in contrast, demonstrate robustness, benefiting from reliable weighting derived from our shape content features and the denoised points (Fig. 5) . Our results show significant improvement in the estimation accuracy, with a notably

Model	RMSE			pgp10 (%)			pgp5 (%)		
	PCDNF	Jet	Ours	PCDNF	Jet	Ours	PCDNF	Jet	Ours
boxunion	60.3	46.5	40.8	7.2	45.1	60.1	5.9	25.2	54.0
box push	47.1	15.1	9.2	14.6	67.6	89.0	12.4	33.8	70.9
galera	53.7	27.4	21.6	4.5	24.9	39.8	3.5	3.6	7.5
column head	61.5	44.9	36.6	4.9	23.9	35.6	4.3	12.6	24.6
cylinder	48.2	10.3	7.2	6.6	88.5	96.5	5.0	55.4	90.2
icosahedron	46.4	8.7	6.7	4.4	81.6	89.6	3.0	44.3	75.7
netsuke	51.5	26.1	17.0	3.7	26.9	33.7	2.7	8.9	12.4
star smooth	49.9	19.4	11.2	2.6	45.7	75.7	1.9	13.9	34.2
Average	52.3	24.8	**18.8**	6.1	50.5	**65.0**	4.8	24.2	**46.2**

Table 2. Comparison of normal estimation across different models, including RMSE, pgp10, and pgp5 values for PCDNF, Jet, and the updated data. RMSE values are lower-better, while pgp10 and pgp5 values are higher-better.

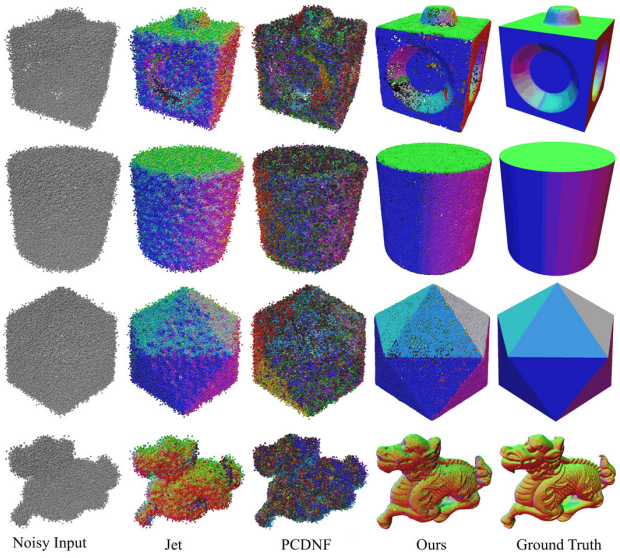

Figure 7. Ours and baseline normals for models with 3% noise level. Color coding is by mapping the per point normal vector to RGB channels.

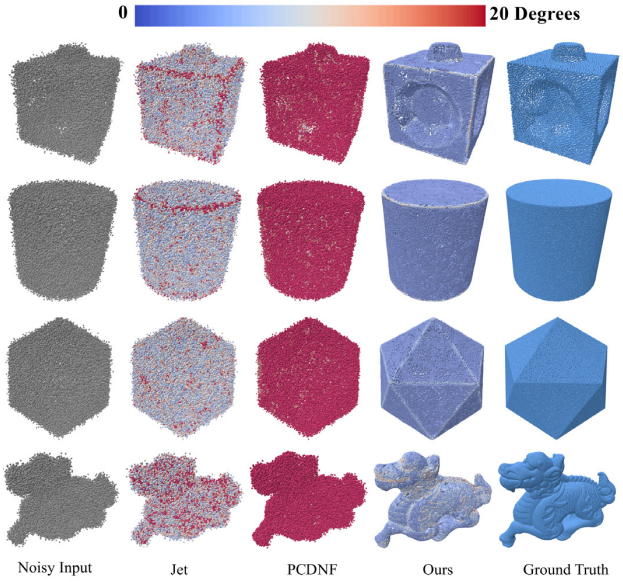

Figure 8. Visualization of ours and baseline normal errors for models with 3% noise level. Color coding is by the angular error.

lower average RMSE (18.8) compared to PCDNF (52.3) and Jet fitting (24.8), indicating improved accuracy in normal estimation (Table 2). Additionally, the percentage of good normal predictions (pgp10 and pgp5) is substantially higher than both methods. The average pgp10 reaches 65.0%, outperforming Jet fitting (50.5%) and PCDNF (6.1%). Similarly, the pgp5 value is 46.2%, surpassing Jet fitting (24.2%) and PCDNF (4.8%).

As Figs. (7 & 8) demonstrate, our estimated normals are edge-preserving, illustrated by the sharp edges on the box push, cylinder, and icosahedron models, where the denoised output accurately reveals the correct structure. These results demonstrate how our learned representation effectively differentiates between surfaces for normal estimation, resulting in low normal errors at singular points (Fig. 8). Our model shows substantial improvement not only on structured objects where state-of-the-art methods collapsed, but also excels on complex, high-curvature models such as netsuke, particularly in transition areas and singular points.

4.3 Quantitative denoising analysis of mobile and handheld LS data

For real-world-data testing, we first apply it to the MLS benchmark. Of focus are two representative scenes and their zoom-in view, with many outlying points due to the reflective metal surface and erroneous returns (Fig. 1). Results show how our proposed approach generates outlier-free, clean surface structures near the window frame and the metal traffic signs, while denoising by the baseline methods exhibits high residual error and distort the surface structure. These results demonstrate our outlier detection capabilities, compared to others who are attempting to denoise these points yielding further distortions. We further evaluated the proposed model on a PLS dataset, which is characterized by low density, high level of noise, and excessive rate of outliers (Fig. 9). On subset #I, the GLR distorted shapes due to its over-smoothing tendencies, and the Score-denoise neglected outliers, distorted sharp features, and yielded high residual noise on the surface (Fig. 10). In contrast, ours effectively preserved sharp features at the intersections of two planes. This preservation of features is due to the shape context feature we have learned with the support of the bilateral loss. It is further illustrated by the normal-coded color visualization. On subset #II, the groin vault, which is a combination of the complex curved surface and their intersections, our proposed method removed noise and preserved the curved structure, compared to the baseline methods, which either over-smoothed the data or created artifacts and voids (Fig. 11). Turning to subset #III (Fig. 12) that features a supporting column base, our model produced clean, surface-adherent, results on complex architectural

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-G-2025-1037-2025 | © Author(s) 2025. CC BY 4.0 License.

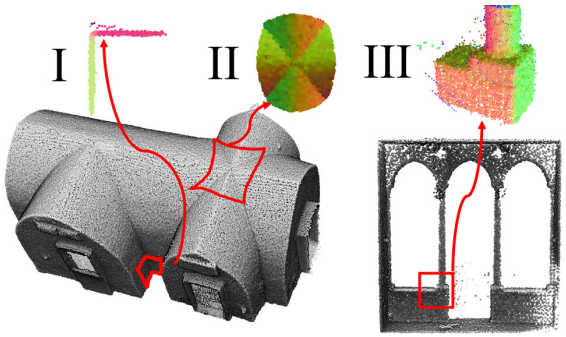

Figure 9. Handheld LS data with normals coloring of the study subsets.

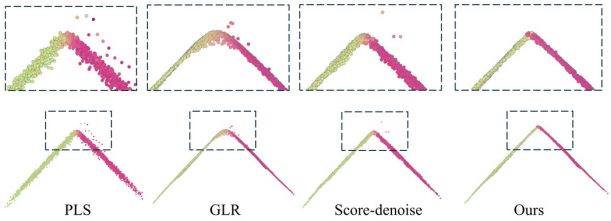

Figure 10. Our denoising result on PLS model #I preserves sharp features while others fail.

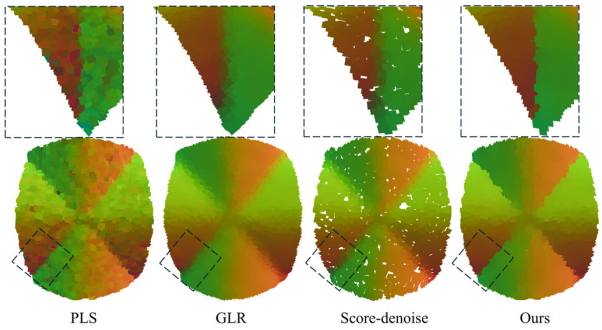

Figure 11. Denoising of a groin vault PLS data (model #II), points are rendered by normals. Ours is smooth and preserves sharp features. The GLR and Score-denoise yields artifacts.

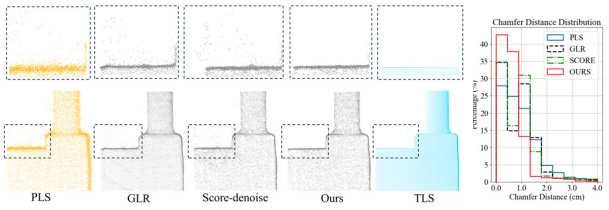

Figure 12. Denoising of a supporting column PLS data (model #III). Our results are cleanest also handling high outliers rate.

features where other methods struggled with noise and outliers. Our model is designed to preserve features and promote piecewise smoothness, hence excelling in all cases. With the availability of TLS data for this subset (Fig. 12.left), we can also evaluate performance quantitatively. Using the CD to the ground truth, our proposed model achieves the lowest residual statistics Fig. (12). It yields a compact residual distribution, with over 93% of points $\leq$ 1 cm, double the percentage achieved by the GLR and Score-denoise.

5. Conclusions

Neural approaches to point cloud denoising offer promising solutions for mitigating noise in cost-effective scanners. Here, a learning-based denoising architecture that effectively removes outliers, adjusts points for noise reduction, and preserves sharp features, was presented. Through densely connected dynamic graph convolution layers, our network has learned to capture the underlying shape context of each point. This shape context is further refined using global attention, enabling the network to leverage both local and global information. Our outlier filtering allows predicting precise offsets for the remaining point set and ensures they are noise-free. By integrating a consistency-based prediction framework and a bilateral form, we achieved accurate displacement predictions and edge-aware denoising output. Our learned shape context encapsulates rich geometric information, facilitating robust weighting of similar points, which also enhances normal estimation. Extensive experiments demonstrate the effectiveness of our approach in both denoising and normal estimation on standard benchmarks. The results show significant improvements in the quality of mobile and portable laser scans, effectively removing outliers and preserving fine features. The outputs produced by our method closely align with those obtained from high-precision terrestrial scanners, with over 93% of points deviating by $\leq$ 1 cm—double the percentage achieved by state-of-the-art methods.

References

Antova, G., 2024. Portable laser scanning solutions for 3D modelling of large buildings. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, 13–19.

Avron, H., Sharf, A., Greif, C., Cohen-Or, D., 2010. L1-sparse reconstruction of sharp point set surfaces. *ACM Transactions on Graphics (TOG)*, 29(5), 1–12.

Ben-Shabat, Y., Gould, S., 2020. Deepfit: 3d surface fitting via neural network weighted least squares. *Computer Vision–ECCV 2020: 16th European Conference*, Springer, 20–34.

Cazals, F., Pouget, M., 2005. Estimating differential quantities using polynomial fitting of osculating jets. *Computer aided geometric design*, 22(2), 121–146.

Conti, A., Pagliaricci, G., Bonora, V., Tucci, G., 2024. A comparison between terrestrial laser scanning and hand-held mobile mapping for the documentation of built heritage. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, 141–147.

Desbrun, M., Meyer, M., Schröder, P., Barr, A. H., 2000. Anisotropic feature-preserving denoising of height fields and bivariate data. *Graphics interface*, 11number 10, Citeseer, 145–152.

Digne, J., De Franchis, C., 2017. The bilateral filter for point clouds. *Image Processing On Line*, 7, 278–287.

Dinesh, C., Cheung, G., Bajić, I. V., 2020. Point cloud denoising via feature graph laplacian regularization. *IEEE Transactions on Image Processing*, 29, 4143–4158.

Edirimuni, D., Lu, X., Li, G., Robles-Kelly, A., 2023a. Contrastive learning for joint normal estimation and point cloud filtering. *IEEE Transactions on Visualization and Computer Graphics*.

Edirimuni, D., Lu, X., Shao, Z., Li, G., Robles-Kelly, A., He, Y., 2023b. Iterativepfn: True iterative point cloud filtering. *Proc. of the IEEE/CVF Conference on CVPR*, 13530–13539.

Fleishman, S., Cohen-Or, D., Silva, C. T., 2005. Robust moving least-squares fitting with sharp features. *ACM transactions on graphics (TOG)*, 24(3), 544–552.

Fleishman, S., Drori, I., Cohen-Or, D., 2003. Bilateral mesh denoising. *ACM SIGGRAPH 2003 Papers*, 950–953.

Guennebaud, G., Gross, M., 2007. Algebraic point set surfaces. *ACM siggraph 2007 papers*, 23–es.

Hakula, A., Ruoppa, L., Lehtomäki, M., Yu, X., Kukko, A., Kaartinen, H., Taher, J., Matikainen, L., Hyyppä, E., Luoma, V. et al., 2023. Individual tree segmentation and species classification using high-density close-range multispectral laser scanning data. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 9, 100039.

Huang, H., Wu, S., Gong, M., Cohen-Or, D., Ascher, U., Zhang, H., 2013. Edge-aware point set resampling. *ACM transactions on graphics (TOG)*, 32(1), 1–12.

Li, R., Li, X., Fu, C.-W., Cohen-Or, D., Heng, P.-A., 2019. Pu-gan: a point cloud upsampling adversarial network. *Proc. of the ICCV*, 7203–7212.

Li, X., Lu, J., Ding, H., Sun, C., Zhou, J. T., Chee, Y. M., 2024. Pointcvar: Risk-optimized outlier removal for robust 3d point cloud classification. *Proc. of the AAAI Conference on Artificial Intelligence*, 38number 19, 21340–21348.

Lipman, Y., Cohen-Or, D., Levin, D., Tal-Ezer, H., 2007. Parameterization-free projection for geometry reconstruction. *ACM Transactions on Graphics (ToG)*, 26(3), 22–es.

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*.

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

Mattei, E., Castrodad, A., 2017. Point cloud denoising via moving rpca. *Computer Graphics Forum*, 36number 8, Wiley Online Library, 123–137.

Peng, J., Strela, V., Zorin, D., 2001. A simple algorithm for surface denoising. *Proc. Visualization, 2001. VIS'01.*, IEEE, 107–548.

Pistilli, F., Fracastoro, G., Valsesia, D., Magli, E., 2020. Learning graph-convolutional representations for point cloud denoising. *ECCV*, Springer, 103–118.

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.

Ren, J., Pan, L., Liu, Z., 2022. Benchmarking and analyzing point cloud classification under corruptions. *International Conference on Machine Learning*, PMLR, 18559–18575.

Serna, A., Marcotegui, B., Goulette, F., Deschaud, J.-E., 2014. Paris-rue-madame database: A 3d mobile laser scanner dataset for benchmarking urban detection, segmentation and classification methods. *4th international conference on pattern recognition, applications and methods ICPRAM 2014*.

Sun, Y., Schaefer, S., Wang, W., 2015. Denoising point sets via L0 minimization. *Computer Aided Geometric Design*, 35, 2–15.

Taubin, G., 1995. A signal processing approach to fair surface design. *Proc. of the 22nd annual conference on Computer graphics and interactive techniques*, 351–358.

Xia, S., Chen, D., Wang, R., Li, J., Zhang, X., 2020. Geometric primitives in LiDAR point clouds: A review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 685–707.

Ye, C., Zhao, H., Ma, L., Jiang, H., Li, H., Wang, R., Chapman, M. A., Junior, J. M., Li, J., 2020. Robust lane extraction from MLS point clouds towards HD maps especially in curve road. *IEEE Transactions on Intelligent Transportation Systems*, 23(2), 1505–1518.

Yu, L., Li, X., Fu, C.-W., Cohen-Or, D., Heng, P.-A., 2018. Pu-net: 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.

Zeng, J., Cheung, G., Ng, M., Pang, J., Yang, C., 2019. 3D point cloud denoising using graph Laplacian regularization of a low dimensional manifold model. *IEEE Transactions on Image Processing*, 29, 3474–3489.