www.cad-journal.net
Computer-Aided Design AND APPLICATIONS

Adaptive Point Cloud Denoising via Attention-Based Hard Masking and Variable-Length Feature Fusion

Wen-Peng Ge[1], Li-Yong Shen[2], Bao-Wen Xu[3] and Ji-Xing Li[4]

[1]University of Chinese Academy of Sciences, gewenpeng22@mails.ucas.ac.cn
[2]University of Chinese Academy of Sciences, lyshen@ucas.ac.cn
[3]AVIC-Digital, Tsing Hua University, xubw001@avic.com
[4]AVIC-Digital, li_jixing@buaa.edu.cn

Corresponding author: Wen-Peng Ge, gewenpeng22@mails.ucas.ac.cn
Bao-Wen Xu, xubw001@avic.com

Abstract. This paper proposes an adaptive point cloud denoising framework based on attention mechanisms, addressing the critical balance between noise suppression and geometric preservation through hard feature masking and dynamic feature fusion. To overcome the limitations of conventional deep learning methods, which are susceptible to outlier noise contamination in feature encoding and constrained by fixed architectures that ignore variations in noise intensity and geometric complexity, our method introduces two innovative modules: First, the Hard Feature Masking module, built upon spatial attention mechanisms, dynamically suppresses noise-contaminated feature channels through point-wise activation analysis, effectively blocking noise propagation pathways in neural networks. Second, the Binary Gated Fusion module enables dynamic adjustment of feature integration depth through learnable gating mechanisms, adaptively selecting feature aggregation paths according to local noise levels and geometric complexity. Extensive experiments on the Synth-A synthetic dataset and real-world scanned data demonstrate that our method outperforms state-of-the-art models in both quantitative metrics (CD, RMSE) and visual quality. Notably, in sparse point cloud scenarios, it achieves a 5.0% improvement in CD accuracy compared to the leading existing model while better preserving sharp geometric features. Ablation studies further confirm the synergistic effects of the dual modules, validating the framework's comprehensive advantages in noise robustness and geometric fidelity.

Keywords: Adaptive Point Cloud Denoising, Attention Mechanism, Hard Feature Masking, Binary Gated Fusion, Dynamic Feature Integration.

DOI: https://doi.org/10.14733/cadaps.2026.790-802

1 INTRODUCTION

As a discrete 3D geometric representation, point clouds have become a core data carrier in computer vision and graphics due to their millimeter-level spatial resolution and sub-surface detail characterization capabilities. They are widely applied in critical fields such as reverse engineering, digital twins, autonomous driving high-definition map construction, and industrial inspection. However, point cloud data acquired via LiDAR scanning, TOF cameras, or multi-view stereo matching are often contaminated by unstructured outliers caused by multipath effects, photon scattering noise, and sensor quantization errors. Such noise not only induces surface normal estimation deviations and topological distortions but also severely degrades the reliability of downstream tasks like point cloud registration and SLAM mapping, highlighting the urgent need for robust adaptive denoising algorithms.

As shown in Figure 1, this work aims to restore noisy point clouds (marked by circles) to their ideal surfaces (solid lines). Traditional denoising methods [1–3], while effective in specific scenarios, heavily rely on prior noise assumptions and manual parameter tuning. Recent deep learning approaches have achieved significant progress: PointCleanNet [4] regresses displacement vectors through residual field construction, PCDNF [5] enhances performance by incorporating normal vector information, and PathNet [6] innovatively introduces reinforcement learning for dynamic denoising path selection. Nevertheless, existing methods still face two critical limitations: (1) Insufficient utilization of neighborhood feature information during encoding, leading to loss of local geometric details; (2) Static network architectures struggle to adapt to heterogeneous noise distributions and geometric complexity, often requiring specialized models for different scenarios. Notably, although PathNet [6] achieves dynamic path selection via geometric feature labels, its dependence on precomputed labels introduces feature extraction error propagation and computational redundancy.

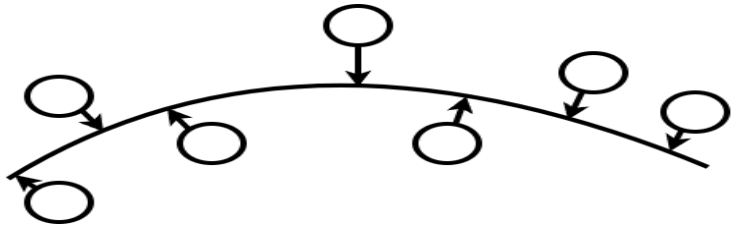


Figure 1: Restore noisy point clouds to their ideal surfaces.

To address these challenges, we propose a dynamic point cloud denoising framework based on attention mechanisms, with two key innovations:

1. **Hard Feature Masking Module:** Cleanses noise-contaminated features by masking outlier characteristics.

2. **Variable-Length Feature Fusion Module:** Dynamically selects fusion depth based on the geometric properties and noise complexity of local point cloud patches.

This framework directly achieves adaptive denoising through end-to-end learning. Experimental results demonstrate that our method autonomously identifies and removes outliers while preserving geometric continuity at sharp features, outperforming state-of-the-art approaches in both quantitative metrics and visual quality.

2 RELATED WORK

This section systematically reviews existing research on point cloud denoising from both traditional methodology and deep learning perspectives.

Traditional Point Cloud Denoising Methods: Traditional approaches for point cloud denoising focus on reducing noise while preserving geometric information. To eliminate noise while retaining shape details and sharp edges, Fleishman et al. proposed a bilateral filtering framework inspired by grayscale image denoising principles. This method accounts for both spatial distances between points and their neighborhoods as well as distances along normal directions, effectively preserving sharp edges [7]. Marc Alexa introduced a shape representation method using point sets, defining smooth manifold surfaces through local mappings and Moving Least Squares (MLS) approximation. This approach enables local computation suitable for large-scale point clouds and controls approximation errors via point density adjustment [8]. Yaron Lipman developed a novel Locally Optimal Projection (LOP) operator for surface approximation of point set data. Independent of local normal estimation or plane fitting, LOP handles noisy and complex geometries, particularly excelling in scenarios with ambiguous orientations or intricate surface folds. Defined through global minimization yet remaining local, LOP provides second-order smooth surface approximations with strong noise and outlier robustness [9]. To further enhance denoising, norm regularization methods have been employed. These techniques suppress noise through regularization constraints while preserving sharp features, though their performance may degrade in smooth regions [10,11]. Xuequan Lu proposed a low-rank matrix approximation-based normal estimation method applicable to both point clouds and meshes. By computing local structures and identifying similar non-local counterparts, this approach robustly estimates normals and introduces a novel filtering framework for point cloud smoothing. Experiments demonstrate its superior accuracy and visual quality in tasks including point cloud filtering, surface reconstruction, and mesh denoising [12]. While achieving satisfactory results, these traditional methods often require extensive parameter tuning and exhibit inconsistent performance across diverse point cloud datasets.

Deep Learning-Based Point Cloud Denoising Methods: The rapid advancement of deep learning has opened new opportunities for point cloud denoising. Deep neural networks automatically learn complex patterns in point cloud data, significantly improving denoising effectiveness. Notable contributions include PCPNet, a deep architecture that first removes outliers before denoising by learning noise patterns and geometric features [4]. Some studies leverage the intrinsic connection between denoising and normal filtering tasks, introducing auxiliary normal filtering objectives to enhance performance [5]. Reinforcement learning has also been explored to train adaptive networks that select denoising strategies based on point cloud characteristics, integrating prior geometric knowledge for flexible noise handling [6]. Score-based denoising networks assess noise levels through score models to achieve precise denoising, demonstrating strong generalization across datasets [13]. The PointASNL framework introduces an Adaptive Sampling (AS) module that reweights neighborhood points and adjusts sampling locations, coupled with a Local-NonLocal (L-NL) module capturing long-range dependencies to improve noise robustness [14]. Addressing the common limitation of position-based filtering methods in preserving sharp features, [15] proposes a feature-preserving approach without normal dependency. Its core innovation lies in searching non-local similar patches through customized similarity metrics, aggregating information in standardized spaces before remapping results to original domains. Existing methods primarily achieve point cloud consolidation through feature aggregation but lack explicit constraints on edges, leading to blurred reconstructed surfaces. Lequan Yu proposed the first edge-aware deep learning framework that jointly predicts point coordinates and point-to-edge distances. By employing an edge-sensitive loss to preserve sharp features, the method surpasses existing techniques on both synthetic and real-world data [16]. Riccardo Roveri et al. proposed PointProNets, a generative convolutional neural network framework. By constructing an end-to-end encoder-decoder architecture for point cloud processing, the method implements feature extraction on localized geometric patches, effectively resolving surface ambiguities while producing detail-preserving dense point clouds. Experimental validations reveal its significant advantages in

both geometric feature conservation and reconstruction accuracy compared to conventional approaches dependent on geometric priors [17]. Recent work by Francesca Pistilli et al. introduced a dynamic graph-convolutional architecture that addresses noise sensitivity through learnable feature-space neighborhood graphs. The framework employs fully-convolutional operations with hierarchical feature aggregation, effectively mitigating high-noise vulnerabilities inherent in conventional methods. Particularly, its adaptive neighborhood modeling mechanism demonstrates enhanced structured noise suppression while preserving geometric fidelity. Empirical validation confirms the framework's superior robustness under extreme noise conditions compared to existing techniques [18].

To address the limitations of existing methods—specifically, the insufficient exploitation of neighborhood feature information during encoding and the inability to adapt denoising depth based on local noise complexity—we propose two key modules: Hard Feature Masking Module, which identifies and suppresses high-noise outlier features to prevent contamination during the feature fusion process; and Variable-Length Feature Fusion Module, which dynamically adjusts the fusion depth according to the geometric structure and noise level of each local region, enabling a more flexible and accurate denoising strategy.

3 METHOD

This chapter introduces an adaptive point cloud denoising method based on attention mechanisms, hard masking, and variable-length feature fusion. Due to the characteristics of acquisition methods, point cloud data are often affected by noise, which adversely impacts subsequent analysis and processing tasks. In Section 3.1, we detail the core principles of the method, including how noise displacement vectors are predicted to restore noisy point clouds, and how hard feature masking and binary gated fusion adaptively adjust denoising operations. Section 3.2 presents the complete network architecture, elaborating on the roles of the encoder, masked sampling module, variable-length feature fusion module, and decoder. In Section 3.3, we introduced the loss functions used in our approach.

3.1 Problem Formulation

Given a noisy point cloud $\hat{P}$, our method first predicts a noise displacement vector. The objective is to recover the point cloud $\hat{P}-\epsilon$ so that it closely approximates the clean point cloud P, where P accurately represents the true 3D geometry of the object while preserving sharp features at the edges. The recovery process can be formulated as:

$$\hat{P} = P + \epsilon$$

Specifically, for each point $\hat{p}_i$ in the point cloud $\hat{P}$, we search for the nearest N points to form a local point cloud patch (in our experiments, $N = 128$). This allows us to decompose the entire point cloud into multiple local patches of size $N \times 3$ for denoising, enabling the processing of point clouds with varying densities. Since neighboring points in a point cloud patch exhibit intrinsic structural relationships, we design an attention-based feature fusion module to fully exploit the structural information of nearby points. However, we recognize that some outlier points in the point cloud patch may contaminate the features of other points during the fusion process. To address this issue, we introduce an attention-based hard feature masking mechanism that automatically removes the features of high-noise points.

Moreover, considering that different patches exhibit varying geometric complexity and noise levels, we propose a binary gating mechanism to adaptively control the denoising intensity by determining the number of iterations for feature fusion. Our network architecture is illustrated in Figure 2.

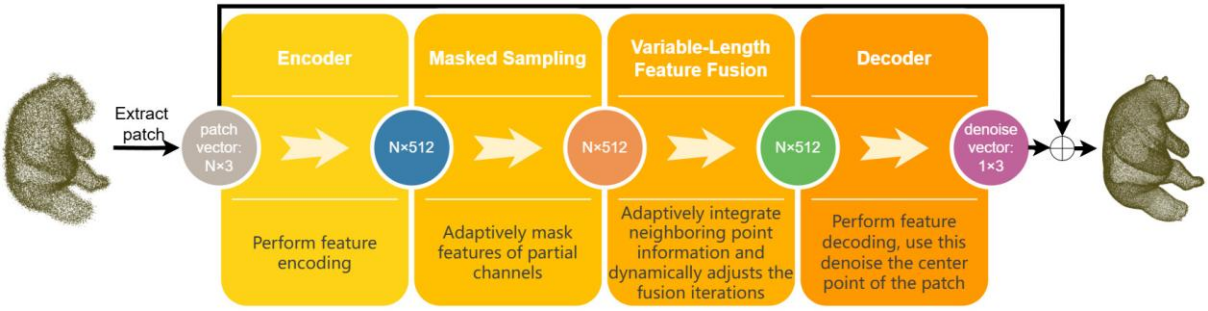


Figure 2: Overview of network architecture .

3.2 Point Cloud Denoising Network

Figure 2 illustrates our complete network architecture. The input consists of a local patch formed by the nearest N neighbors centered at point p_i, and the output is the noise displacement at point p_i, which is used to correct the position of point p_i to make it closer to the true surface. Our network mainly consists of four parts: Encoder, Masked Sampling Module, Variable-Length Feature Fusion Module, and Decoder. The encoder encodes the $N\times3$ dimensional point cloud patch into a high-dimensional feature vector for subsequent processing. The masked sampling module adaptively masks out outliers with high noise. The variable-length feature fusion module adaptively determines the number of feature fusion steps based on the information in the input patch. The decoder reduces the high-dimensional features learned by the network into a three-dimensional noise displacement vector, which is then used to correct the position of the noisy point p_i to bring it closer to the real surface.

3.2.1 Encoder module

The encoder module first centers the input point cloud patch. It consists of four fully connected layers (MLP). Each layer is followed by batch normalization and a ReLU activation function. This architecture transforms the $N\times3$ point cloud patch into a high-dimensional feature representation of $N\times512$, facilitating efficient processing by subsequent deep learning modules.

3.2.2 Masked Sampling Module

To mitigate the interference of high-noise outliers during feature fusion in the point cloud denoising process, we propose a Masked Sampling Module (MSM) to adaptively suppress outlier contamination. As shown in Figure 3, the module operates as follows:

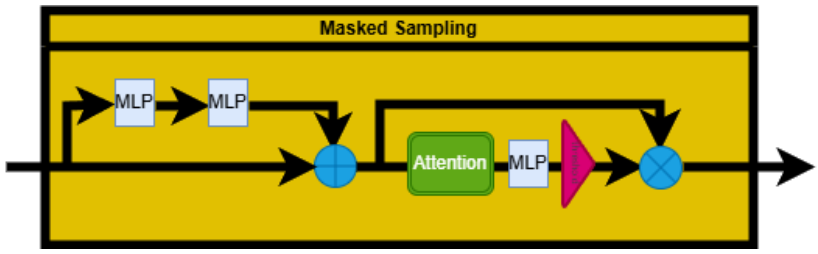


Figure 3: Masked Sampling Module Architecture Visualization.

(1) **Residual MLP Feature Refinement:** We first pass the input feature embedding (N,C), where N is the number of points and C the feature dimension, through a residual MLP. This structure helps maintain gradient flow and enhances features by capturing both local and global information.

(2) Attention-Based Feature Correlation Learning: To quantify feature dependencies among points within a point cloud patch, we introduce an attention-based scoring mechanism, where values range from [0,1]. Higher scores indicate stronger correlations with the underlying surface structure, whereas lower scores correspond to potential outliers.

(3) Binary Mask Generation and Hard Masking: By applying a threshold $\tau = 0.5$, the saliency vector f_s is binarized into a mask $f_t \in \{0,1\}^N$. This hard masking operation zeros out features of outlier points ($f_t = 0$), preventing them from influencing later stages. Figure 4 illustrates the effectiveness of this approach.

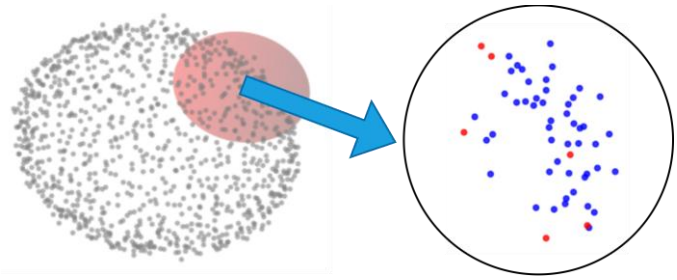


Figure 4: Masking Effect Visualization.

3.2.3 Variable-Length Feature Fusion Module

To adaptively aggregate multi-scale structural features and dynamically adjust the depth of feature fusion based on local geometric complexity and noise distribution, we propose a Variable-Length Feature Fusion Module (VLFFM). As shown in Figure 5, the module operates as follows:

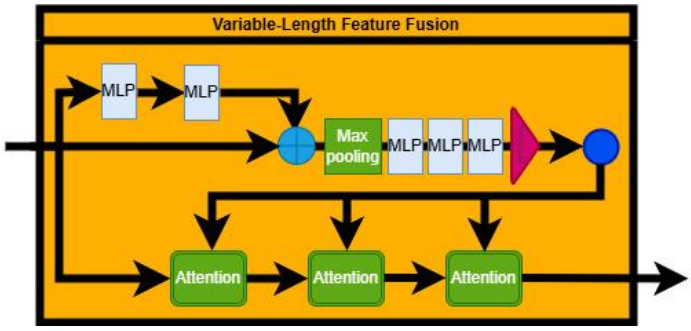


Figure 5: Variable-Length Feature Fusion Module Visualization.

(1) Feature Preprocessing via Residual MLP: The input feature $f \in \mathbb{R}^{N\times C}$ is first processed through a residual MLP subnetwork to enhance its adaptability to subsequent operations. This step

ensures stable gradient propagation while refining local geometric details and global contextual relationships.

(2) Multi-Scale Feature Encoding: A max pooling operation is applied to the refined features to generate a condensed global descriptor, which encapsulates the main structural characteristics of the patch.

(3) Dynamic Fusion Depth Control: After passing through three fully connected layers followed by thresholding, we generate a vector $m \in \{0,1\}^3$ that determines the fusion depth, where 0 indicates no fusion and 1 indicates fusion.

(4) Attention-Based Feature Fusion: Here, we introduce an attention mechanism to effectively fuse information from neighboring points, thereby uncovering the underlying geometric features.

3.2.4 Decoder Module

The decoder module aims to reconstruct denoised coordinates by predicting and compensating for noise-induced displacements. Through progressive dimensionality reduction, we ultimately obtain a 3-dimensional noise displacement vector $\Delta \in \mathbb{R}^3$. The denoised coordinates p' are computed by subtracting the predicted displacement Δ from the original noisy coordinates p:

$$p' = p - \Delta$$

3.3 Loss Function

The proposed loss function is defined as a weighted combination of two complementary objectives to jointly optimize geometric consistency and spatial distribution regularity. Specifically, the total loss L is defined as:

$$L = L_{\text{recon}} + \gamma L_{\text{reg}}$$

The reconstruction loss L_{recon} enforces pointwise geometric consistency between the denoised coordinates and the ground truth surface through an L2 norm penalty:

$$L_{\text{recon}} = \frac{1}{T}\sum_{i=1}^{T} \min_{j \in K(i)} \| \hat{p}_i - p_j^{gt} \|^2$$

where $\hat{p}_i$ and p_i^{gt} represent the denoised and ground truth coordinates of the i-th point, respectively, and $K(i)$ denotes the k-nearest neighbors of i-th point.

The regularization term L_{reg} imposes a uniformity constraint on the denoised point distribution to reduce clustering artifacts and maintain surface continuity. This constraint is implemented using a repulsion-based metric that penalizes abnormally dense regions:

$$L_{\text{reg}} = \frac{1}{T}\sum_{i=1}^{T} \max_{j \in K(i)} \| \hat{p}_i - p_j^{gt} \|^2$$

4 EXPERIMENTAL RESULTS

To rigorously evaluate the robustness of our method, we conducted comparative experiments with four state-of-the-art point cloud denoising methods: Bilateral Filtering[7], POINTCLEANNET [4], PCDNF [5], and PathNet [6]. These methods represent leading algorithms in the field of point cloud denoising and have been widely applied and validated in the relevant literature. To ensure fairness in our experiments, we performed strict comparisons on the same dataset.

We adopted the dataset Synth-A [6] (available at https://github.com/ZeyongWei/PathNet), derived from the test dataset of PU-GAN [19], as the benchmark for our experiments. The Synth-A dataset contains 3D models of various shapes and complexities, allowing for a comprehensive evaluation of denoising algorithms under different conditions. For each model in the Synth-A

dataset, point clouds containing 10K, 20K, and 50K points are generated, and Gaussian noise is applied with standard deviations set to 0.5%, 1.0%, and 1.5% relative to the diagonal of the model's bounding box. Here, the noise intensity is defined as the standard deviation of the Gaussian noise normalized by the bounding box diagonal length, which quantifies the magnitude of perturbations relative to the object scale and controls the severity of noise applied to the point cloud. Our experiments cover varying noise intensities and point cloud densities to validate the generalization ability of our method. Here, the noise intensity is defined as the product of the standard deviation of the Gaussian noise and the diagonal length of the model's bounding box. This quantifies the magnitude of perturbations relative to the object scale and controls the level of noise applied to the point cloud.

4.1 Quantitative Comparison

In these tests, we focused on evaluation metrics based on Chamfer Distance (CD) and Root Mean Squared Error (RMSE), which reflect the accuracy of geometric recovery in the denoising results.

Chamfer Distance (CD) is a commonly used metric for measuring geometric differences between point clouds. It computes the average sum of the shortest distances between two point clouds. Specifically, the Chamfer Distance is defined as:

$$\mathrm{CD}(A,B)=\frac{1}{|A|}\sum_{p\in A}\min_{\hat{p}\in B}\|p-\hat{p}\|^2+\frac{1}{|B|}\sum_{\hat{p}\in B}\min_{p\in A}\|p-\hat{p}\|^2$$

where A and B are two point clouds, $\|p-\hat{p}\|$ represents the Euclidean distance between points p and $\hat{p}$, and $|A|$ and $|B|$ denote the number of points in point clouds A and B, respectively. The first term in Chamfer Distance represents the average squared minimum distance from each point in point cloud A to its nearest neighbor in point cloud B, while the second term represents the average squared minimum distance from each point in point cloud B to its nearest neighbor in point cloud A. A lower Chamfer Distance value indicates a higher geometric similarity between the two point clouds, implying that the denoised point cloud has minimal geometric deviation from the original. In the following experiments, the unit is set to 10^{-5}.

To comprehensively evaluate the denoising effectiveness of our method, in addition to Chamfer Distance (CD) as an evaluation metric, we also introduced the Root Mean Squared Error (RMSE) metric. RMSE is a common measure for evaluating point cloud denoising performance, quantifying the deviation between the denoised and original point clouds. Specifically, RMSE calculates the square root of the average squared difference between each point's position before and after denoising, given by:

$$\mathrm{MSE}=\sqrt{\frac{1}{T}\sum_{i=1}^{T}\mathbf{p}_i-\hat{\mathbf{p}}_i^{\,2}}$$

where T is the number of points in the point cloud, $\hat{\mathbf{p}}_i$ is the position of the i-th point after denoising, and p_i is the position of the nearest point to $\hat{\mathbf{p}}_i$ in the original point cloud. A lower RMSE value indicates a smaller error between the denoised and original point clouds, implying that the denoising algorithm better preserves the original point cloud structure. In the following experiments, the unit is set to 10^{-3}.

To verify the generalization capability of our method, we designed multiple controlled experiments. In these experiments, we varied the noise intensity $\delta=0.5\%$ to 1.5% of the bounding box diagonal length) and point cloud density (ranging from 10K to 50K points per model). The purpose of these controlled experiments was to assess the performance of our method under different noise conditions and point cloud densities.

From the quantitative results, as shown in Table 1, our method set a new performance record in the Chamfer Distance (CD) evaluation metric, achieving significant advantages over other

methods. Notably, the performance advantage of our model is particularly evident when the dataset size is smaller. On the 10K dataset, the average Chamfer Distance result was 36.048, which is 5.0% lower than the second-best method (PCDNF) at 37.950. On the 20K dataset, our method reduced the average Chamfer Distance error to 20.636, approximately 3.5% lower than the closest competitor (PCDNF) at 21.378. These results demonstrate that our method consistently improves performance across different data sizes and noise levels. Overall, thanks to the refined feature fusion and noise suppression mechanisms, our model set a new performance record in Chamfer Distance, reducing the error by 3.6% compared to the second-best method, demonstrating its strong potential in point cloud denoising.

		Synth-A									
		10K			20K			50K			AVE
		0.5%	1%	1.5%	0.5%	1%	1.5%	0.5%	1%	1.5%	
CD	Bilateral Filtering	27.454	54.498	91.046	16.865	32.744	63.476	8.336	17.524	42.096	39.338
	POINTCLEANNET	29.761	42.887	49.873	17.778	22.895	27.940	7.516	10.206	14.750	24.845
	PCDNF	27.167	39.779	46.904	16.176	21.638	26.319	7.072	9.677	**14.495**	23.247
	PathNet	27.997	41.171	48.033	16.955	21.933	27.055	7.155	9.876	14.500	23.853
	Ours	**25.098**	**37.865**	**45.183**	**15.267**	**20.949**	**25.693**	**6.858**	**9.499**	15.690	**22.456**
RMSE	Bilateral Filtering	3.137	3.664	4.295	2.750	3.251	3.982	2.311	2.857	3.696	3.327
	POINTCLEANNET	3.238	3.501	3.627	2.815	2.973	3.133	2.258	2.438	2.699	2.996
	PCDNF	3.154	3.414	3.546	2.745	2.916	3.065	2.218	2.393	**2.672**	2.932
	PathNet	3.167	3.438	3.561	2.763	2.920	3.087	2.216	2.406	2.676	2.946
	Ours	**3.100**	**3.379**	**3.523**	**2.713**	**2.897**	**3.059**	**2.203**	**2.391**	2.758	**2.921**

Table 1: Comparison of CD and RMSE Performance for Different Models on the Synth-A Dataset.

In this experiment, the RMSE metric was also applied to various test scenarios in the Synth-A dataset to further validate the robustness and accuracy of our method. According to the experimental results in Table 1, our method also performed exceptionally well in the RMSE metric. On the 10K dataset, our method achieved an average RMSE of 3.334, which is about 1.1% lower than the second-best method PCDNF at 3.371. On the 20K dataset, our method achieved an average RMSE of 2.893, which is approximately 0.6% lower than the closest competitor PCDNF at 2.911. These results indicate that our method demonstrates significant advantages, particularly in cases where the point cloud data is relatively sparse.

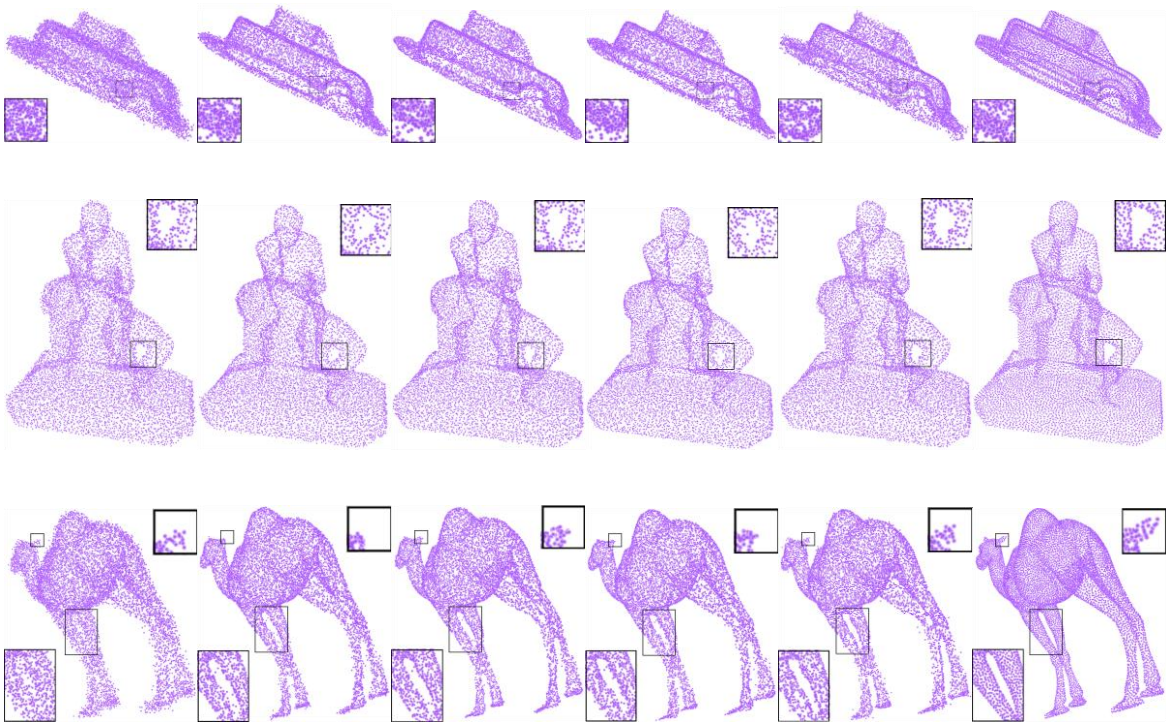
Overall, based on the quantitative results of both CD and RMSE metrics, our method exhibits a clear advantage in denoising performance. While maintaining a low Chamfer Distance, our method also excels in the RMSE metric, proving its dual advantage in both accuracy and stability. This indicates that our proposed denoising model can consistently deliver high-quality denoising results when handling different types of point cloud data.

4.2 Qualitative Comparison

To intuitively demonstrate the denoising performance of our model, we present a comparative visualization of the denoised results. Figure 6 illustrates the visual performance of our method in denoising. The results indicate that, compared to other point cloud denoising models, our method excels in preserving fine details. For instance, in the lower part of the car model, other methods tend to collapse the bottom into a single-layer surface, whereas our approach better retains its structural hierarchy and details, a feature that existing models fail to achieve. Through the local zoom-in views, we can observe that the denoising results produced by our method are closer to the true distribution of the point cloud. However, due to the model's high sensitivity to local details, certain noise along the edges remains insufficiently removed. This suggests that in edge structures, there is still room for improvement. Future work could focus on further enhancing the model's ability to suppress local noise, striking a better balance between detail preservation and noise removal. Additionally, we conducted a detailed analysis on the human statue and camel models. The leg gap of the human statue is more accurately restored using our method, rendering

a more realistic triangular shape, while other methods fail to achieve this precision. In the camel model, our approach outperforms others in restoring ear details and denoising the gap between the front legs. These results demonstrate that our method effectively captures and preserves essential geometric features when handling complex structural details.

To verify the effectiveness of our method in practical applications, we conducted comparative experiments on real scanned point cloud data. As shown in Figure 7, we selected two typical complex scenes: (a) an aircraft model obtained from multi-station point cloud fusion using the Leica RTC360 terrestrial laser scanning system, and (b) an urban vehicle point cloud collected by a vehicle-mounted LiDAR system. In the aircraft point cloud experiment (first column of Figure 7), our method exhibits significant advantages in the wingtip region, producing a more natural wing shape. For the vehicle point cloud captured by vehicle-mounted LiDAR (second column of Figure 7), our method excels in suppressing outliers in the window region, rendering the side window shape more consistent with reality. The results from both experiments indicate that our algorithm demonstrates strong generalization capabilities on real-world data. The effectiveness of our method primarily stems from the collaborative design of the Masked Sampling Module and the Variable-Length Feature Fusion Module. The former employs a feature masking mechanism to effectively suppress outlier disturbances with white noise characteristics, while the latter adaptively captures local geometric context through a variable-length feature fusion strategy, enabling more accurate reconstruction of the underlying clean point cloud and effective removal of Gaussian noise. Since real-world point cloud noise is typically a mixture of white and Gaussian noise, our method demonstrates strong adaptability and robustness when applied to real scanned data.



(a) Input (b) POINTCLEANNET (c) PCDNF (d) PathNet (e) Ours (f) Truth

Figure 6: Qualitative Comparison of Denoising Effects.

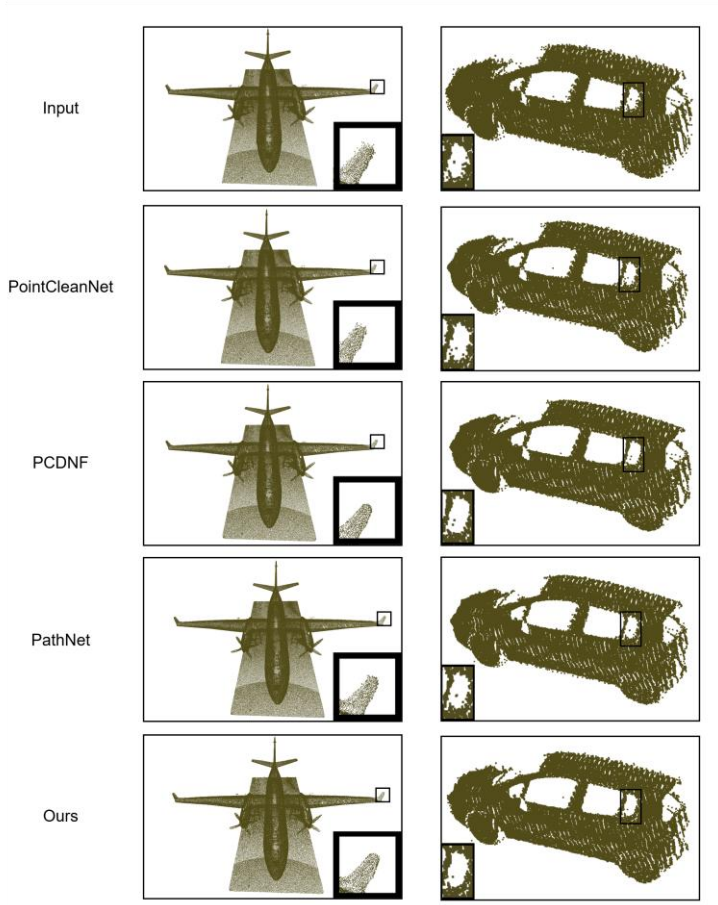


Figure 7: Comparison of Denoising Effects on Real-World Scanned Point Cloud Data.

4.3 Ablation Study

To further validate the effectiveness of different modules in our model, we designed and conducted ablation experiments, specifically testing the impact of removing the mask sampling module and replacing the feature fusion module with a fixed-length version. The primary objective of this experiment is to analyze the contribution of each module and their impact on model performance across datasets of varying scales.

Table 2 presents the experimental results on the Synth-A dataset. We evaluated the impact of removing the mask sampling module and using a fixed-length feature fusion module under different point cloud densities (10K, 20K, 50K). The evaluation results include different noise levels (0.5% , 1% , 1.5%), providing further insights into model performance under various noise conditions.

		Synth-A									
		10K			20K			50K			AVE
		0.5%	1%	1.5%	0.5%	1%	1.5%	0.5%	1%	1.5%	
CD	Without masked sampling	27.548	40.583	47.630	16.702	21.869	26.750	7.175	9.837	14.459	23.617
	Fixed—Length	27.160	40.432	47.594	16.289	21.840	26.539	7.139	9.701	**14.230**	23.436

	Complete	**25.098**	**37.865**	**45.183**	**15.267**	**20.949**	**25.693**	**6.858**	**9.499**	15.690	**22.456**
RMSE	Without masked sampling	3.165	3.433	3.565	2.760	2.924	3.084	2.224	2.409	2.684	2.947
	Fixed—Length	3.154	3.432	3.564	2.751	2.925	3.077	2.224	2.399	**2.670**	2.941
	Complete	**3.100**	**3.380**	**3.524**	**2.713**	**2.897**	**3.059**	**2.203**	**2.391**	2.758	**2.920**

Table 2: The ablation experiment on the Synth-A dataset.

From the results, it is evident that removing the mask sampling module leads to performance degradation across all point densities and noise levels. Specifically, for the 10K dataset, the average performance after removing the mask sampling module is 23.617, which is worse than the complete model's result 22.456. A similar trend is observed in the 20K and 50K datasets, indicating that the mask sampling module plays a crucial role in improving model robustness and reducing noise interference.

Furthermore, replacing the feature fusion module with a fixed-length version also results in noticeable performance degradation. The performance gap is more pronounced in the 10K and 20K datasets, while it becomes relatively smaller in the 50K dataset. Overall, the average performance with the fixed-length module is 23.436, compared to 22.456 for the complete model, indicating a slight performance drop.

In summary, the ablation study demonstrates that both the mask sampling module and the adaptive feature fusion strategy are essential for enhancing model performance. Their complementary effects are particularly evident when dealing with small-scale datasets.

5 CONCLUSIONS

This work presents a robust point cloud denoising framework that addresses key limitations of existing methods through two novel contributions: the Masked Sampling Module for reduce the interference of outliers and the Variable-Length Feature Fusion Module fully leverage the latent structural information of neighboring points and it can adaptively determine the fusion depth based on the complexity of the point cloud. Extensive experiments on synthetic datasets demonstrate state-of-the-art performance, with our method outperforming POINTCLEANNET, PCDNF, and PathNet by 3.6% in Chamfer Distance under varying noise levels (0.5%-1.5%) and point densities (10K-50K points). The proposed dynamic depth control mechanism proves particularly effective in preserving fine structures under sparse sampling conditions, as validated by visual and quantitative analyses.

Wen-Peng Ge, https://orcid.org/0009-0002-2536-6060
Li-Yong Shen, https://orcid.org/0000-0001-5769-4814
Bao-Wen Xu, https://orcid.org/0009-0002-2717-3377
Ji-Xing Li, https://orcid.org/0000-0001-7136-2234

REFERENCES

[1] Sun, Y.; Chen, H.; Qin, J.: Reliable Rolling-guided Point Normal Filtering for Surface Texture Removal, Computer graphics forum, 38(7), 2019, 721-732. https://doi.org/10.1111/cgf.13874

[2] Zhang, S.; Cui, S.; Ding, Z.: Hypergraph spectral analysis and processing in 3D point cloud, IEEE Transactions on Image Processing, 30, 2020, 1193-1206. https://doi.org/10.1109/TIP.2020.3042088

[3] Digne, J.; Valette, S.; Chaine, R.: Sparse geometric representation through local shape probing, IEEE transactions on visualization and computer graphics, 24(7), 2017, 2238-2250. https://doi.org/10.1109/TVCG.2017.2719024
[4] Rakotosaona, M J.; La Barbera, V.; Guerrero, P.: Pointcleannet: Learning to denoise and remove outliers from dense point clouds, Computer graphics forum, 39(1), 2020, 185-203. https://doi.org/10.1111/cgf.13753
[5] Liu, Z.; Zhao, Y.; Zhan, S.: PCDNF: Revisiting learning-based point cloud denoising via joint normal filtering, IEEE Transactions on Visualization and Computer Graphics, 2023. https://doi.org/10.1109/TVCG.2023.3292464
[6] Wei, Z.; Chen, H.; Nan, L.: PathNet: Path-selective point cloud denoising, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024. https://doi.org/10.1109/TPAMI.2024.3355988
[7] Digne ,J.; De Franchis, C.: The bilateral filter for point clouds, Image Processing On Line, 7, 2017, 278-287. https://dx.doi.org/10.5201/ipol.2017.179
[8] Alexa, M.; Behr, J.; Cohen-Or, D.: Computing and rendering point set surfaces, IEEE Transactions on visualization and computer graphics, 9(1), 2003, 3-15. https://doi.org/10.1109/TVCG.2003.1175093
[9] Lipman, Y.; Cohen-Or, D.; Levin, D.: Parameterization-free projection for geometry reconstruction, ACM Transactions on Graphics (ToG), 26(3), 2007, 22-es. https://doi.org/10.1145/1276377.1276405
[10] Sun, Y.; Schaefer, S.; Wang, W.: Denoising point sets via L0 minimization, Computer Aided Geometric Design, 35, 2015, 2-15. https://doi.org/10.1016/j.cagd.2015.03.011
[11] Avron, H.; Sharf, A.; Greif, C.: ℓ1-sparse reconstruction of sharp point set surfaces, ACM Transactions on Graphics (TOG), 29(5), 2010, 1-12. https://doi.org/10.1145/1857907.1857911
[12] Lu, X.; Schaefer, S.; Luo, J.: Low rank matrix approximation for 3D geometry filtering, IEEE transactions on visualization and computer graphics, 28(4), 2020, 1835-1847. https://doi.org/10.1109/TVCG.2020.3026785
[13] Luo, S.; Hu, W.: Score-based point cloud denoising, Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, 4583-4592. https://doi.org/10.48550/arXiv.2107.10981
[14] Yan, X.; Zheng, C.; Li, Z.: Pointasnl: Robust point clouds processing using nonlocal neural networks with adaptive sampling, Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, 5589-5598. https://doi.org/10.48550/arXiv.2003.00492
[15] Wang, J.; Jiang, J.; Lu, X.: Rethinking point cloud filtering: A non-local position-based approach, Computer-Aided Design, 144, 2022, 103162. https://doi.org/10.1016/j.cad.2021.103162
[16] Yu, L.; Li, X.; Fu, C. W.: Ec-net: an edge-aware point set consolidation network, Proceedings of the European conference on computer vision (ECCV), 2018, 386-402. https://doi.org/10.48550/arXiv.1807.06010
[17] Roveri, R.; Öztireli, A. C.; Pandele, I.: Pointpronets: Consolidation of point clouds with convolutional neural networks, Computer Graphics Forum, 37(2), 2018, 87-99. https://doi.org/10.1111/cgf.13344
[18] Pistilli, F.; Fracastoro G.; Valsesia D.: Learning robust graph-convolutional representations for point cloud denoising, IEEE Journal of Selected Topics in Signal Processing, 15(2), 2020, 402-414. https://doi.org/10.1109/JSTSP.2020.3047471
[19] Li, R; Li, X; Fu, C. W.: Pu-gan: a point cloud upsampling adversarial network, Proceedings of the IEEE/CVF international conference on computer vision, 2019, 7203-7212. https://doi.org/10.48550/arXiv.1907.10844