GeoMamba: Dynamic Step-Size Adaptive Modulation for Mamba-Based Point Cloud Classification

Hanxiao Liu[1]and Li Cui[2*]

[1]Yanbian University, Yanji 133002, Jilin Province, China
2Yanbian University, Yanji 133002, Jilin Province, China

*Corresponding author:cuili@ybu.edu.cn

Abstract. Most existing Mamba-based point cloud classification models adopt implicit step-size adaptation, which lacks explicit awareness of 3D geometric structures. This leads to limited classification accuracy and poor robustness against background noise in complex real-world scenarios. To address this issue, we propose GeoMamba, a lightweight geometry-semantics dual-driven dynamic step-size modulation framework. We design a multidimensional geometric encoder to extract fine-grained geometric features, including relative positions, distance variations and local neighborhood statistics, paired with a semantic gating module for context-aware modulation, which constrains the step-size adjustment range to [0.2, 5.0]. With only 0.9M additional parameters, GeoMamba achieves 86.57% overall accuracy on the most challenging ScanObjectNN PB-T50-RS subset without pretraining, outperforming PointMamba by 4.09 percentage points, and reaches 89.21% accuracy with pretraining. Extensive ablation studies and visualization analysis validate the effectiveness and interpretability of our method.

Keywords: Point cloud classification, Mamba, Dynamic step-size modulation, Lightweight model.

1 Introduction

Point cloud classification, a core task in 3D computer vision, serves as a technological foundation for applications such as robot navigation, autonomous driving, and industrial inspection [1-2]. In recent years, Transformers have achieved significant progress in this field due to their global modeling capabilities [3-4]. However, the $O(n^2)$ computational complexity of their attention mechanisms poses challenges for large-scale point cloud processing and edge deployment. State Space Models (SSMs), such as Mamba [5], offer a novel approach to point cloud classification through sequence modeling with linear complexity. PointMamba [6] pioneered the application of Mamba to point cloud classification by employing space-filling curves for efficient feature extraction. However, this method retains the implicit step-size strategy of standard SSMs, resulting in a lack of explicit awareness of 3D geometric structures.

In SSMs, the step size Δ controls the temporal scale of state transitions and serves as the core parameter balancing local and global information. However, existing methods predominantly employ implicit step-size adaptations or implicit learning strategies. This results in models treating all point cloud features equally, failing to effectively distinguish foreground targets from background noise, and exhibiting insufficient robustness in real-world scanning scenarios with occlusions and interference.

Recent efforts have attempted to refine step-size design (e.g., encoding coordinate differences as step sizes [7]); however, these approaches remain constrained by scalar geometric representations and lack fine-grained geometric details and semantic contextual constraints, thereby hindering scene-content-based adaptive feature selection. To address these limitations, this paper introduces GeoMamba, which achieves adaptive dynamic modulation of SSM step sizes through a dual geometry-semantics-driven mechanism. As illustrated in Figure 1 (detailed in Section 4), we design a lightweight geometric encoder based on the PointMamba [6] architecture to extract multidimensional geometric features, including relative positions, distance variations, and local neighborhood statistics. A semantic gating module aggregates global context based on model hidden states, followed by element-wise fusion to generate context-aware modulation signals. This mechanism constrains the dynamic step-size adjustment range to [0.2, 5.0], significantly enhancing the model's robustness in complex real-world scenarios while increasing model parameters by only 0.9M.

The core contributions of this paper are as follows:

We design a lightweight multidimensional geometric encoder that fuses relative positions, distance changes, and local neighborhood statistics to construct a fine-grained isometry-invariant geometric representation. This multidimensional geometric encoder overcomes the limitations of simple geometric modeling relying solely on scalar coordinate differences, as in STREAM [7], providing interpretable structural priors for subsequent step-size modulation.

We introduce a context-aware dynamic step-size modulation paradigm. By aggregating global latent state information and geometric features through a semantic gating module for adaptive element-wise fusion (Hadamard product), our approach intelligently selects between local detail enhancement and background noise suppression, addressing the fundamental flaw of standard Mamba's implicit step-size adaptation, which treats all spatial positions identically.

We construct a lightweight modulation framework that synergizes geometric and semantic information, achieving significant performance gains with only a 0.9M parameter increase.

2 Related Work

In recent years, point cloud classification methods based on SSMs have rapidly gained traction. PointMamba pioneered the application of Mamba to point cloud classification, achieving linear-time modeling through Hilbert curve serialization. However, its adoption of an implicit step-size strategy and lack of geometric awareness result in insufficient robustness in real-world scenarios. STREAM [7] achieved preliminary geometric awareness by encoding coordinate differences as step-size values; however, it remains constrained by scalar representations and lacks fine-grained geometric details and semantic contextual constraints.

ZigzagPointMamba [8] optimized the sequencing path through zigzag scanning but failed to address the fundamental step-size calculation, thereby preventing adaptive adjustments based on point cloud content. Contemporary work further explores enhancement strategies across different dimensions: PointDGMamba [9] proposes Masked Sequence Denoising (MSD) and Dual-level Domain Scanning (DDS) to enhance domain generalization; Spectral Informed Mamba [10] utilizes the Laplacian spectrum to define

geometric traversal order, demonstrating a frequency-domain geometric perspective; CloudMamba [11] introduces Grouped Selective State Space (GS6) to mitigate S6 overfitting through parameter sharing while enhancing geometric awareness via multi-axis scanning; and Efficient Spiking Point Mamba [12] explores the integration of spiking neural networks with Mamba to achieve ultra-low-power inference.

The aforementioned methods exhibit certain limitations, as they fail to account for the impact of the step size—a core parameter—on model performance when processing geometric data. For instance, PointMamba completely disregards 3D structure; STREAM employs only scalar geometry; ZigzagPointMamba, PointDGMamba, and CloudMamba neglect step-size mechanisms; Spectral Informed Mamba relies on computationally complex spectral decomposition; and Spiking Mamba focuses on computational paradigms rather than feature modulation.

This paper introduces GeoMamba, which presents a novel integration of multidimensional geometric features (relative positions, distance changes, and local neighborhood statistics) with global semantic context into SSM step-size calculation. Through a lightweight geometric encoder and semantic gating module, GeoMamba achieves a paradigm shift from an implicit step-size adaptation to a geometry-semantics-driven dynamic step size, significantly enhancing robustness in complex real-world scenarios.

3 Preliminaries

3.1 Selective State Space Model

The core of Mamba is the Selective State Space Model (S6), which achieves efficient sequence feature modeling by discretizing continuous state-space equations. Its core formulation is:

$$h_k = \bar{A} \cdot h_{k-1} + \bar{B}_k \cdot x_k \, , y_k = C_k \cdot h_k + D \cdot x_k \tag{1}$$

Where $\bar{A}_k = exp(A\Delta)\ and\ \bar{B}_k = (A\Delta)^{-1}(exp(A\Delta) - I) \cdot \Delta$Brepresent the state transition matrices, and the step size Δ is the core parameter controlling the rate of state transition: a large Δ enhances the model's ability to capture long-range dependencies, whereas a small Δ enables the model to focus on extracting local information. In standard Mamba, the step size is generated solely from the current input features, calculated as:

$$\Delta = \text{softplus}\left(\text{Linear}(x_k) + b_\Delta\right) \tag{2}$$

This approach fails to account for the inherent structural characteristics of sequential data, making it difficult to adapt to three-dimensional data with strong geometric structures, such as point clouds. Consequently, the model's performance is constrained in complex scenarios.

3.2 Step-Size Limitations of PointMamba and Variants

PointMamba [6] directly adopts the standard Mamba step-size computation method for point cloud classification tasks. Since this model generates step sizes solely based on

the input feature $x_i$of the point cloud, completely disregarding its 3D geometric structure, it treats all point cloud features identically and fails to effectively distinguish foreground targets from background noise.

STREAM [7], the first model to incorporate geometric information into step-size design, uses point cloud coordinate differences as the basis for step-size generation. Although this approach achieves geometric awareness in step-size generation, it relies solely on scalar coordinate differences to represent geometric information. Furthermore, it lacks fine-grained geometric features and a semantic modulation mechanism, preventing the adjustment of geometric feature importance based on global semantic context.

ZigzagPointMamba [8] does not improve the step-size calculation method but instead enhances performance by optimizing sequential path generation and masking strategies. Consequently, it fails to address the core flaw of implicit step-size adaptations, leaving room for improvement in robustness under complex scenes. To tackle these issues, this paper proposes a geometry-semantics-driven step-size calculation method to achieve adaptive step-size modulation. The core formula is:

$$\Delta = f_{\text{geo}}(x_i, N) \odot g_{\text{sem}}(h_{i-1}) \tag{3}$$

Here, f_{geo}represents multidimensional geometric features extracted by the geometric encoder, g_{sem} denotes semantic context features generated by semantic gating, and $\odot$ denotes element-wise multiplication. This operation enables the step size to simultaneously fuse geometric structure and semantic context information.

4 Proposed Method

4.1 Architecture Overview

As shown in Figure 1, GeoMamba follows the concise architectural design of PointMamba [6]. First, n keypoints are obtained via Farthest Point Sampling (FPS). Hilbert curves and transposed Hilbert curves are then used to serialize these keypoints, generating an ordered point sequence. KNN is subsequently applied to construct local point cloud clusters for each serialized keypoint. Local features are extracted via a PointNet-style encoder to obtain serialized tokens, which are subsequently fed into the modified GeoMamba block for global feature extraction. Its core innovation lies in the geometry-semantics-driven step-size modulation mechanism within the GeoMamba block.

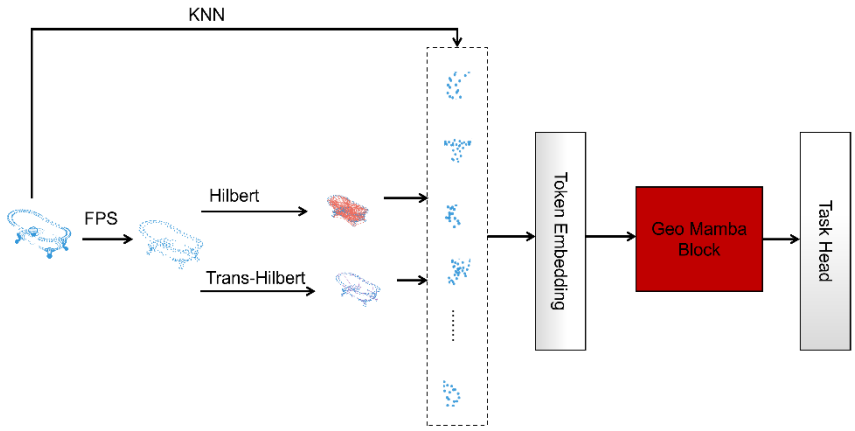

Fig. 1. GeoMamba Overall Architecture Overview. The framework consists of FPS sampling, Hilbert curve serialization, local feature extraction, and the proposed GeoMamba blocks with dual-driven step-size modulation

4.2 Geometric-Semantic Dual-Driven Step-Size Modulation

As shown in Figure 2, this mechanism takes local point cloud groups as input. It extracts multidimensional geometric features f_{geo},including relative positions, distance variations, and local neighborhood statistics, via a geometric encoder, while simultaneously generating , g_{sem} through semantic gating that aggregates global context. Element-wise multiplication of these two outputs enables selective modulation of geometric features by semantic context, yielding the fused feature $f_{fused} = f_{geo} \odot g_{sem}$,where $f_{fused} \in R^{L\times 128}$.

The fused feature undergoes lightweight MLP projection onto the internal dimensions of the SSM, generating an initial modulation signal δ_{mod}. A learnable scalar global gate γ is introduced and soft-weighted via Sigmoid activation, yielding$\tilde{\delta}_{mod} = \sigma(\gamma) \cdot \delta_{mod}$. This ensures that the modulation module exerts only a negligible influence during the early training phases, thereby safeguarding training stability. The weighted modulation signal is mapped to the [0,1] range via the Sigmoid function and subsequently constrained to [0.2,5.0] through linear transformation. This value is multiplied by the original step size to obtain the adaptive step size, as expressed in the following equation:

$$\Delta_{\text{final}} = \Delta_{\text{original}} \cdot \left(0.2 + 4.8 \cdot \sigma(\delta_{\text{mod}})\right) \tag{4}$$

This mechanism determines "where to modulate" through geometric encoding and "how to modulate" through semantic gating, enabling context-aware adaptive adjustment of step size.

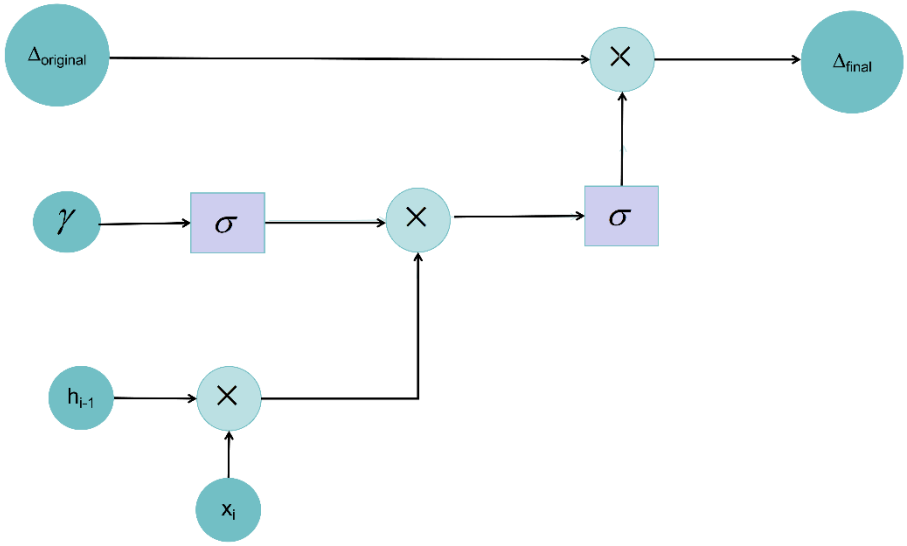

Fig. 2. Geometric-Semantic Fusion and Step-Size Modulation Module Flowchart.

5 Experiments

To thoroughly validate GeoMamba's performance, module effectiveness, and interpretability, we conduct extensive experiments on two classic point cloud classification datasets: ScanObjectNN and ModelNet40. We evaluate the model's overall performance by comparing it with mainstream point cloud classification models, validate the effectiveness of each core module through ablation studies, and reveal the underlying mechanism of dynamic step-size modulation through visual analysis of the step-size modulation factor.

5.1 Experimental Setup

Datasets. ScanObjectNN [2]: A real-world scanned point cloud dataset comprising 15 categories of everyday objects, totaling approximately 15K point cloud samples. It is a highly challenging dataset for point cloud classification. This dataset offers three difficulty levels: OBJ-BG (original samples with background), OBJ-ONLY (samples with background removed), and PB-T50-RS (the most challeng-ing subset with 50% background points added and random rotations, denoted as scan_hardest). We use PB-T50-RS as the core evaluation subset to validate the model's robustness in complex real-world scenarios.

ModelNet40 [13]: A synthetic CAD point cloud dataset comprising 40 classes of man-made objects with 12K point cloud samples. Featuring clean scenes and well-de-

fined object structures, it lacks background interference and occlusion issues. This dataset validates model performance in simple, idealized scenarios and compares the effectiveness of constant versus dynamic step-size scales.

Training Configuration. All experiments were implemented and refined based on the open-source PointMamba codebase, uniformly using Python 3.8, PyTorch 2.0, and CUDA 11.7 on NVIDIA RTX 3090 GPUs. The optimizer used is AdamW with a cosine annealing learning rate schedule. Training is conducted for 300 epochs with a batch size of 32. Different learning rates are used for different datasets: $5 \times 10^{-4}$for ScanObjectNN and $3 \times 10^{-4}$for ModelNet40.

Evaluation Metrics. Overall accuracy (OA) is adopted as the core performance metric, while parameter efficiency (PE = OA(%) / Number of Parameters(M)) is introduced to evaluate model lightweightness. Higher parameter efficiency indicates better performance per unit of parameter count. For ScanObjectNN, parameter efficiency is calculated based on the PB-T50-RS subset; for ModelNet40, it is calculated based on overall dataset accuracy.

5.2 Key Results

5.2.1 ScanObjectNN Dataset.

Without Pretraining. As shown in Table 1, GeoMamba achieves 86.57% OA on the challenging PB-T50-RS subset, surpassing PointMamba by 4.09 percentage points and the parameter-matched PointMamba by 0.21%. Compared to STREAM [7] which uses scalar geometry, GeoMamba improves by 1.25% , validating the effectiveness of multidimensional geometric encoding combined with semantic gating. The model achieves 6.56% parameter efficiency (PE), delivering superior accuracy with approximately 2.6× fewer parameters than Transformer-based methods like Point-MAE [3].

With Pretraining. GeoMamba achieves 89.21% on PB-T50-RS (Table 2), out-performing ZigzagPointMamba [8] by 0.56% and PointMamba by 0.49%. This consistent superiority across both settings demonstrates that step-size modula-tion provides fundamental advantages over path optimization or scalar geometric encoding for real-world scanned data with noise and occlusion.

Table 1. Comparison of Performance Without Pre-training on the ScanObjectNN Dataset. Note: Version 13.2M was adjusted to maintain the same parameter count as GeoMamba.

Method	param.(M)	OBJ-BG(%)	OBJ-ONLY (%)	PB-T50-RS(%)	PE (%/M)
PointMLP [15]	13.2	85.4	—	85.4	6.47
Point-MAE [3]	22.1	86.75	86.92	80.78	3.65
PCM [16]	34.2	—	—	88.1	2.58
PointMamba [6]	12.3	88.30	87.78	82.48	6.71
STREAM [7]	12.3	90.02	88.64	85.32	**6.94**
PointMamba (13.2M)	13.2	90.70	**88.81**	86.36	6.54

GeoMamba (Ours)	13.2	**90.88**	88.30	**86.57**	6.56

Table 2. Comparison of Pre-trained Performance on the ScanObjectNN Dataset.

Method	param.(M)	OBJ-BG(%)	OBJ-ONLY(%)	PB-T50-RS(%)	PE (%/M)
Point-BERT [4]	22.1	87.30	88.12	83.07	3.95
MaskPoint [17]	22.1	89.70	89.30	84.60	4.06
PointMAE [3]	22.1	90.02	88.29	84.60	4.07
ACT [18]	22.1	93.29	91.91	88.21	4.22
Zigzag-PointMamba [8]	12.3	**94.15**	92.10	88.65	**7.65**
PointMamba [6]	12.3	93.80	92.25	88.72	7.63
GeoMamba (Ours)	13.2	93.98	**92.77**	**89.21**	7.12

ModelNet40 Dataset. GeoMamba shows robust performance on the synthetic ModelNet40 dataset, an ideal benchmark for testing basic feature learning capabilities of point cloud models. Without pretraining, it achieves 92.22% overall accuracy, outperforming the 13.2M parameter PointMamba baseline by 0.41 percentage points; with pretrained weights, accuracy rises to 92.59%, slightly exceeding the original 12.3M PointMamba with only a minor 0.9M parameter increase. This proves the dual-driven step-size modulation mechanism effectively enhances feature extraction without extra parameters or sacrificing simple-scenario performance.

Its performance gain here is far smaller than on ScanObjectNN, perfectly matching our core design motivation: ModelNet40's clean, regular CAD point clouds with complete structures make the traditional implicit step-size strategy sufficiently effective, leaving limited optimization space. While ScanObjectNN's noisy real-scanned data with background interference lets GeoMamba's dynamic modulation fully exert its advantages in accurately distinguishing foreground and background via geometric encoding and semantic gating. This validates the mechanism is a targeted upgrade with strong adaptability across both ideal and complex real-world scenarios.

Table 3. Performance Comparison on ModelNet40 Dataset (Without Pre-training).Note: The 13.2M version was adjusted to maintain the same parameter count as GeoMamba.

Method	param.(M)	OA (%)	PE (%/M)
PointNet [14]	3.5	89.2	25.49
PointNet++ [1]	1.5	90.7	**60.47**
Point-MAE [3]	22.1	92.3	4.18
PCM [16]	34.2	**93.4**	2.73
PointMamba [6]	12.3	92.4	7.51
STREAM [7]	12.3	92.7	7.54

PointMamba (13.2M)	13.2	91.81	6.95
GeoMamba (Ours)	13.2	92.22	6.98

Table 4. Performance Comparison on ModelNet40 Dataset (Including Pre-training).

Method	Param. (M)	OA (%)	PE (%/M)
Point-BERT [4]	22.1	92.70	4.19
MaskPoint [17]	22.1	92.60	4.19
PointMAE [3]	15.3	93.20	6.09
ACT [18]	22.1	**93.60**	4.24
ZigzagPointMamba [8]	12.3	93.15	7.57
PointMamba [6]	12.3	92.51	**7.60**
GeoMamba (Ours)	13.2	92.59	7.09

5.3 Ablation Studies

To rigorously validate the effectiveness and synergistic contributions of Ge-oMamba's core components—including local neighborhood features and dis-tance features from the geometric encoder, as well as the semantic gating mecha-nism—comprehensive ablation experiments without pretraining were conducted on the most challenging subset PB-T50-RS of the ScanObjectNN dataset.

Ablation Experiment Design and Quantitative Results. Centered on the core dual-driven geometry-semantics architecture, this abla-tion study designs seven comparative experiments. The results are summarized in Table 5.

Table 5. GeoMamba Ablation Experiment Results (ScanObjectNN PB-T50-RS, NoPre-training).

Model Configuration	param.(M)	OA (%)	Change (%)
Geometric Constants + Semantic Gating	13.2	86.12	-0.45
No Distance Features	13.2	86.15	-0.42
No local neighborhood features	13.1	86.25	-0.32
Geometry only, no gating	12.5	86.36	-0.21
PointMamba (13.2M)	13.2	86.36	-0.21
PointMamba	12.3	82.48	-4.09
Complete GeoMamba	13.2	**86.57**	—

Ablation experiments (Table 5) validate the irreplaceability of each module. The "geometric constant + semantic gating" configuration exhibits the lowest performance, demonstrating that semantic modulation detached from geometric priors blindly suppresses information. Removing distance features causes greater performance loss than removing local neighborhood features, indicating that scalar distance features are more

fundamental than high-dimensional local fea-tures. The "geometry-only without gating" configuration performs on par with PointMamba, proving that performance breakthroughs require semantic gating to provide global contextual selection. The complete model achieves a 4.09% im-provement over the original version, demonstrating the synergistic gain effect of geometric encoding and semantic gating.

Visualization and Interpretability Analysis of Step-Size Modulation Mechanism. To deeply uncover the intrinsic working mechanism of the dual-driven geometry-semantics approach, we conduct a visualization analysis of the step-size modulation factor$\Delta_{final}/\Delta_{original}$ .By comparing the modulation patterns between the full GeoMamba and ablation variants, we validate the physical interpretability of dynamic step size through spatial distribution structures

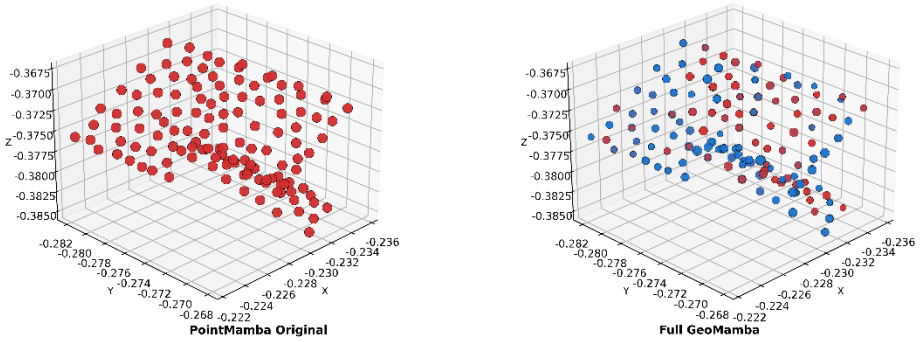

Fig. 3. Comparison of step-size modulation patterns between PointMamba and GeoMamba.

As shown in Figure 3, the original PointMamba model exhibits a uniform red distribution ($\Delta \approx 1.0$), indicating that it applies an implicit step-size adaptation across all spatial locations without distinguishing foreground from background. This "one-size-fits-all" approach manifests in the PB-T50-RS subset, where the model struggles to differentiate object edges from background noise, resulting in insufficient robustness when processing real scan data containing occlusions and artifacts.

In contrast, GeoMamba exhibits distinct structural differentiation between red and blue regions: critical geometric regions such as object edges and corners appear red, whereas smooth surfaces and background noise regions appear blue. This selective modulation strongly couples with physical geometric structures, validating the effectiveness of the dual-driver mechanism.

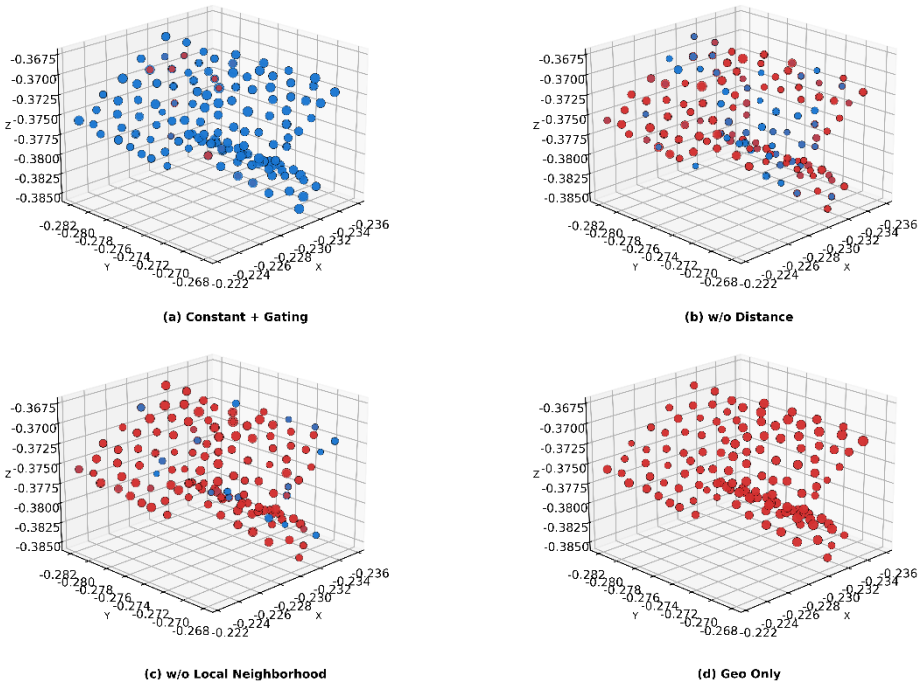

Fig. 4. Spatial distribution of modulation factors across four key ablation configurations.

Figure 4 illustrates the spatial distribution of modulation factors across four key ablation configurations, and the visualized geometric patterns are highly consistent with the quantitative performance results in Table 5, further revealing the core function of each module in the geometry-semantics dual-driven framework and their indispensable synergistic effect.

Fig.4a: Semantic gating detached from geometric guidance exhibits a discrete, spotty blue distribution lacking continuous spatial structure. Without fine-grained geometric priors from the multidimensional encoder, the semantic gating module cannot identify the spatial structure of point clouds or distinguish foreground and background, leading to blind and random feature suppression that fails to target noise regions or enhance key geometric areas, resulting in the lowest performance of this configuration.

Fig.4b: After removing scalar distance features, the visualization reveals local discontinuities and an inability to effectively recognize the overall spatial layout. As the most fundamental geometric feature of point clouds, scalar distance reflects the relative spatial distribution of adjacent points. Its removal causes the geometric encoder to rely only on high-dimensional neighborhood features, leading to fragmented enhanced regions on object edges and corners, and the model cannot form a complete spatial perception of 3D objects for coherent step-size modulation.

Fig.4c: Extensive red coverage indicates that the model overemphasizes global structure while neglecting local details. Removing local neighborhood features leaves only scalar distance features, which provide basic global spatial relationships but fail to capture fine-grained local details such as object concavities and corners. The model thus cannot distinguish key enhanced regions and suppressible smooth surfaces, leading to global feature enhancement that ignores local detail differentiation and weak noise suppression.

Fig.4d : Geometric encoding without semantic constraints produces global red coverage, indicating that all locations are enhanced. The complete geometric encoder can extract valid geometric priors, but the absence of semantic gating removes global contextual selection. The model blindly enhances all geometric regions—including both foreground key structures and background geometric noise—causing valid and invalid features to overlap, and performance remains equivalent to the PointMamba baseline without substantive breakthroughs.

Visual analysis fully confirms the geometry-semantics dual-driver paradigm: geometric encoding extracts fine-grained features to accurately determine where to modulate, while semantic gating aggregates global context to intelligently decide how to modulate; the two modules are complementary and jointly enable context-aware adaptive step-size modulation.

6 Discussion

6.1 Comparison with Related Works

Existing Mamba-based point cloud methods mainly focus on three dimensions: scanning path optimization, geometric encoding, and model architecture scaling. Different from these works, our method focuses on the core parameter of SSMs (step size Δ), which is the key to balancing local and global information. Experimental results show that optimizing only the scanning path (ZigzagPointMamba) or relying on scalar geometric encoding (STREAM) cannot break through the performance bottleneck in complex real-world scenarios, while our geometry-semantics dual-driven step-size modulation achieves significant performance gains with negligible parameter overhead. Compared with transformer-based methods, our Mamba-based framework achieves superior performance and parameter efficiency, demonstrating that Mamba with geometric-aware step-size design has a more suitable architectural inductive bias for point cloud processing.

6.2 Limitations and Future Directions

There are still some limitations in this work. First, the current vector-level modulation with semantic gating has not yet achieved channel-level fine-grained control. In future work, we will introduce efficient channel attention mechanisms to assign adaptive weights to different channels of geometric features. Second, we only validate our method on point cloud classification tasks, and the generalizability to other 3D vision tasks needs further verification. We will extend our method to point cloud detection, segmentation, and denoising tasks in future work. Third, we only conduct experiments on two indoor point cloud datasets, and we will validate the performance on large-scale outdoor datasets in the future. Finally, we will integrate model pruning and quantization techniques to further reduce the computational complexity, enabling efficient deployment on edge devices.

7 Conclusion

In this paper, we propose GeoMamba, a geometry-semantics dual-driven dynamic step-size modulation model for Mamba-based point cloud classification, which addresses the core limitation of existing Mamba-based methods: lack of geometric awareness and semantic context fusion in step-size design. By leveraging a lightweight multidimensional geometric encoder to extract fine-grained geometric features, and a semantic gating module to aggregate global context, our method achieves adaptive dynamic modulation of SSM step sizes. With only 0.9M additional parameters, GeoMamba achieves significant performance improvement on the challenging real-world ScanObjectNN dataset, and outperforms state-of-the-art Mamba-based and transformer-based baselines. Extensive ablation studies and visualization analysis validate the effectiveness and interpretability of our method. We believe this work can provide useful insights for geometric-aware design of SSMs for 3D vision tasks.

Acknowledgments. This research was supported by the Jilin Provincial Natural Science Foundation project (No. YDZJ202601ZYTS079)

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 30, pp. 5099–5108 (2017)
2. Uy, Q.H., Pham, D.T., Hua, B.S., Nguyen, M.K., Yeung, S.K.: Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 1588–1597 (2019)
3. Pang, Y., Wang, W., Tay, F.E., Liu, W., Tian, Y., Yuan, L.: Masked autoencoders for point cloud self-supervised learning. World Scientific Annual Review of Artificial Intelligence. vol. 1, pp. 2440001 (2023)
4. Yu, X., Tang, L., Rao, Y., Huang, T., Zhou, J., Lu, J.: Point-bert: Pre-training 3d point cloud transformers with masked point modeling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 19313–19322 (2022)
5. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. In: First Conference on Language Modeling (COLM) (2024)
6. Liang, D., Zhou, X., Wang, X., Zhu, X., Zeng, Z., Zhang, X., Li, J., Zhang, Z., Xu, X., et al.: Pointmamba: A simple state space model for point cloud analysis. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 37, pp. 32653–32677 (2024)
7. Schöne, F., Beck, J., Nazemi, M., Barber, Q., Fuchs, F., Wörtche, P., Zwönitzer, K., Biedenkapp, A., Hutter, F., Lindauer, M.: Stream: A universal state-space model for sparse geometric data. arXiv preprint arXiv:2411.12603 (2024)
8. Diao, H., Zhang, Y., Xu, X., Zhou, X., Zhu, Q., Wang, Z., Fan, J., Wang, J., Wang, J.: ZigzagPointMamba: Spatial-semantic mamba for point cloud understanding. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 38 (2025)

9. Yang, J., Li, B., Liu, Z., Jiang, Z., Guo, S., Zhang, D.: Pointdgmamba: Domain generalization of point cloud classification via generalized state space model. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, no. 9, pp. 9193–9201 (2025)
10. Bahri, M., Vinyes-Moya, F., Berman, M., Bernardis, G., Avrithis, Y.: Spectral informed mamba for robust point cloud processing. In: Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR). pp. 11799–11809 (2025)
11. Qu, X., Yu, Z., Yuan, C., Li, S.Z.: CloudMamba: Grouped Selective State Spaces for Point Cloud Analysis. arXiv preprint arXiv:2511.07823 (2025)
12. Wu, K., Wen, X., Liu, Z., Liu, Y., Liu, C., Liu, Y., Wang, Z., Li, B.: Efficient spiking point mamba for point cloud analysis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 26393–26403 (2025)