

PointDGMamba: Domain Generalization of Point Cloud Classification via Generalized State Space Model

Hao Yang^{1*}, Qianyu Zhou^{1*}, Haijia Sun^{2*}, Xiangtai Li^{3,5}, Fengqi Liu¹, Xuequan Lu⁴,
Lizhuang Ma^{1†}, Shuicheng Yan^{3,5}

¹Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China

²School of Information Management, Nanjing University, Nanjing, China

³Skywork AI, Singapore

⁴Department of Computer Science and Software Engineering, The University of Western Australia, Australia

⁵Nanyang Technological University, Singapore

{yanghao.cs, zhouqianyu, liufengqi, lzma}@sjtu.edu.cn, 522023140081@smail.nju.edu.cn

Abstract

Domain Generalization (DG) has been recently explored to improve the generalizability of point cloud classification (PCC) models toward unseen domains. However, they often suffer from limited receptive fields or quadratic complexity due to using convolution neural networks or vision Transformers. In this paper, we present the first work that studies the generalizability of state space models (SSMs) in DG PCC and find that directly applying SSMs into DG PCC will encounter several challenges: the inherent topology of the point cloud tends to be disrupted and leads to noise accumulation during the serialization stage. Besides, the lack of designs in domain-agnostic feature learning and data scanning will introduce unanticipated domain-specific information into the 3D sequence data. To this end, we propose a novel framework, *PointDGMamba*, that excels in strong generalizability toward unseen domains and has the advantages of global receptive fields and efficient linear complexity. PointDGMamba consists of three innovative components: Masked Sequence Denoising (MSD), Sequence-wise Cross-domain Feature Aggregation (SCFA), and Dual-level Domain Scanning (DDS). In particular, MSD selectively masks out the noised point tokens of the point cloud sequences, SCFA introduces cross-domain but same-class point cloud features to encourage the model to learn how to extract more generalized features. DDS includes intra-domain scanning and cross-domain scanning to facilitate information exchange between features. In addition, we propose a new and more challenging benchmark PointDG-3to1 for multi-domain generalization. Extensive experiments demonstrate the effectiveness and state-of-the-art performance of PointDGMamba.

Code — <https://github.com/yxlytya/PointDGMamba>

Introduction

3D point cloud learning is a fundamental 3D vision task that has applications in many fields, such as autonomous driving, robot navigation, virtual reality, and urban modeling.

*These authors contributed equally.

†Corresponding author.

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Some methods such as PointNet (Qi et al. 2017a), PointNet++ (Qi et al. 2017b), and DGCNN (Phan et al. 2018) have made great progress in point cloud classification. However, existing methods typically excel only on seen datasets and may encounter performance degradation on unseen domains. This is mainly due to domain shifts, *e.g.*, differences caused by sensor types, scanning angles, and environmental conditions, across different domains.

To address this issue, domain generalization (DG) techniques have been recently introduced into point cloud classification (PCC) (Huang et al. 2021) to learn domain-invariant features, thereby improving the model’s generalizability. The mainstream methods include data augmentation (Xiao et al. 2023), adversarial training (Lehner et al. 2022), and consistency learning (Kim et al. 2023). Nevertheless, most of them are based on CNNs and inherently suffer from a limited receptive field, making it difficult to capture global information. As a result, the model may struggle to fully understand the overall structure of the data, leading to sub-optimal classification performance. One solution is to use a Vision Transformer (Han et al. 2022), but its internal attention layer inevitably introduces higher quadratic computational complexity with increased input points. Therefore, how to effectively model global information while keeping low computational complexity is a key issue in DG PCC.

Recently, Mamba (Gu and Dao 2023), as an emerging State Space Model (SSM) model, has been demonstrated to be effective in capturing long-range dependencies and global information (Xing et al. 2024; Zhang et al. 2024). More importantly, they can accomplish some tasks with linear complexity, which solves the problem of a limited receptive field and avoids quadratic complexity. The representative work, Point Mamba (Liang et al. 2024) and Point Cloud Mamba (PCM) (Zhang et al. 2025) propose several serialization methods to scan 3D point cloud data into specific sequences. Nonetheless, these methods often produce less desirable performance in unseen domains because they do not adequately account for domain shifts or do not include any tailored designs. As shown in Figure 1, there exist noticeable performance gaps between PCM and state-of-the-art DG PCC methods. Thus, and understanding the barriers

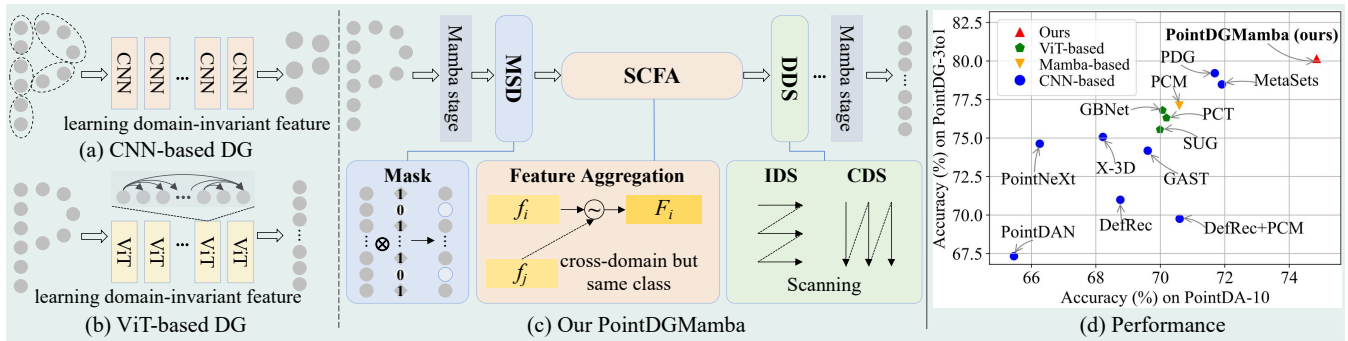


Figure 1: **Left:** Comparisons between previous works and our PointDGMamba. Previous domain generalization (DG)-based point cloud classification (PCC) methods often rely on Convolution Neural Networks (CNNs) or Vision Transformers (ViTs) to learn domain-invariant features, while often suffering from limited receptive fields with local perception (a) or high quadratic complexity with global perception (b). **Middle:** In contrast, we propose a novel framework, PointDGMamba (c), that excels in strong generalizability toward unseen domains and has the advantages of global receptive fields and efficient linear complexity. Our PointDGMamba consists of Masked Sequence Denoising (MSD), Sequence-wise Cross-domain Feature Aggregation (SCFA), and Dual-level Domain Scanning (DDS). **Right:** In the widely-used PointDA-10 benchmark and our proposed PointDG-3to1 benchmark (d), our PointDGMamba demonstrates superior accuracy against state-of-the-art methods.

preventing Mamba-based models from effectively handling distribution shifts in DG and enhancing the generalizability of SSMs remain critical challenges for the point cloud field.

In this paper, we aim to boost the generalizability of Mamba-like models toward unseen domains in point cloud classification. Our motivations mainly lie in *three aspects*. *Firstly*, we observe that the inherent topology of the point cloud tends to be disrupted during Mamba’s serialization process, and even generates some unexpected noises unrelated to the current state. Such noise would accumulate during the training, subsequently affecting the model’s performance when unseen data is used as input. *Secondly*, we observe that existing blocks of Mamba-like models are usually hand-crafted and over-heuristic, overlook the designs of learning domain-agnostic features, and tend to overfit specific domains. This may introduce unanticipated domain-specific information into the sequence data, thereby weakening Mamba’s effectiveness in handling distribution shifts. *Thirdly*, an unresolved issue is how to effectively convert 3D point cloud data into 1D sequence data suitable for Mamba in DG PCC. Though recent studies have investigated different scanning methods for point cloud tasks, these rigid and fixed scanning approaches inevitably introduce human biases and largely ignore domain-agnostic considerations. Additionally, they are highly susceptible to varying conditions, posing challenges to their application in unseen domains.

Motivated by the above facts, we propose PointDG-Mamba, a novel State Space Model-based framework for domain generalizable point cloud classification. PointDG-Mamba excels in strong generalizability toward unseen domains and has the advantages of global receptive fields and efficient linear complexity. It mainly includes three core components: Masked Sequence Denoising (MSD), Sequence-wise Cross-domain Feature Aggregation (SCFA), and Dual-level Domain Scanning (DDS). Specifically, MSD selectively masks the noised point tokens of the point cloud

sequences and uses the purified features for classification, mitigating the adverse effects of noise accumulation. Then, SCFA is designed to aggregate cross-domain but same-class point cloud features to prompt the model to extract more domain-generalized features. Finally, DDS, including two scanning methods, intra-domain scanning and cross-domain scanning, is proposed to facilitate sufficient information interaction between different parts of the features. As such, it converts 3D point cloud data into 1D sequence data suitable for Mamba-like models in varying unseen domains. Provided that the number of source domains in the existing DG PCC benchmark is limited, we also present a new multi-domain generalization benchmark, PointDG-3to1, which is more diverse, practical, and challenging. It includes 4 variants of leave-one-out settings, with 3 domains used as source domains and the remaining one as the unseen domain.

Our contributions can be summarized as follows:

- We propose PointDGMamba, a novel State Space Model-based framework for domain generalizable point cloud classification that shows strong generalizability toward unseen domains and has the advantages of global receptive fields and efficient linear complexity.
- We design Masked Sequence Denoising (MSD), Cross-domain Feature Aggregation (SCFA), and Dual-level Domain Scanning(DDS) to improve the generalizability of the SSM-based model in DG PCC.
- We propose a more challenging multi-domain generalization benchmark PointDG-3to1. Extensive experiments on PointDA-10 and our PointDG-3to1 benchmarks demonstrate the effectiveness and superiority of our PointDGMamba against state-of-the-art competitors.

Related Work

Point Cloud Classification (PCC) makes an accurate classification of the given 3D point cloud data. Pioneering

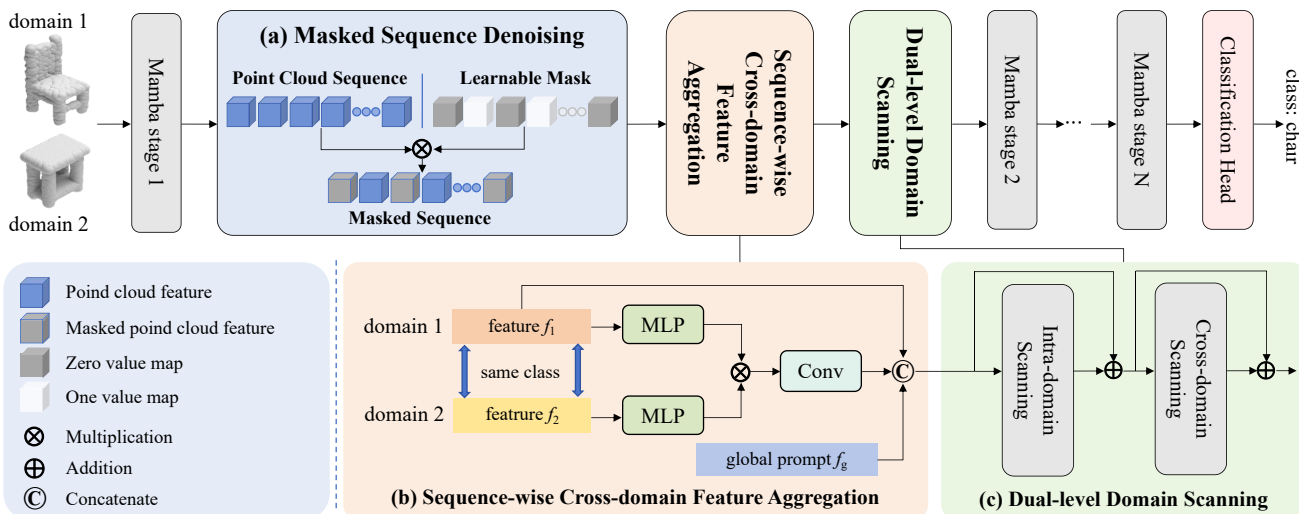


Figure 2: The framework of PointDGMamba. It consists of three components: (a) Masked Sequence Denoising (MSD) is presented to mask out noised point patches in the sequence and thus mitigate adverse effects of noise accumulation during the serialization stage; (b) Sequence-wise Cross-domain Feature Aggregation (SCFA) is introduced to aggregate cross-domain but same-class point cloud features with the global prompt to extract more generalized features, thereby strengthening Mamba’s effectiveness in handling distribution shifts. (c) Dual-level Domain Scanning, including intra-domain scanning and cross-domain scanning, is proposed to facilitate sufficient information interaction between different parts of the features.

works (Qi et al. 2017a,b) use MLP-like architecture to learn the representation on the point cloud directly. Researchers have explored many frameworks (Li et al. 2018; Wang et al. 2019) based on Convolutional Neural Networks (CNNs) to enhance the understanding of point cloud geometric structures. However, they suffer from a limited receptive field when stacking deep layers. Recently, vision Transformers (Zhao et al. 2021; Fang et al. 2024; Deng et al. 2024) have been introduced to PCC due to the merit of global receptive fields inherent in ViT. Some representative methods, such as PCT (Guo et al. 2021) and Point Transformer (Zhao et al. 2021), unveil the potential of self-attention layers and propose Transformer-based architectures to model global context and dependencies within point clouds. Point-BERT (Yu et al. 2022), and Point-MAE (Pang et al. 2022) introduce the Masked Point Modeling (MPM) as a pre-text task for reconstructing masked point clouds. They use masked autoencoders for BERT-style pre-training or self-supervised learning to achieve stronger representations. Despite their gratifying progress, most of these methods perform well on seen datasets and may encounter significant performance degradation when generalizing to unseen, novel domains.

Domain Generalized Point Cloud Classification. Although domain adaptation techniques have been explored in point cloud areas (Liu et al. 2024a; Jiang et al. 2024; Li et al. 2025) to narrow the domain shifts, the target data is not always accessible in real scenarios, which might fail these methods. Domain generalization has recently been introduced into PCC to improve the generalizability toward unseen domains. The mainstream DG PCC models aim to learn domain-invariant features and are mainly categorized as adversarial learning, consistency learning (Kim et al. 2023),

data augmentation (Lehner et al. 2022; Xiao et al. 2023), *etc.* Despite their remarkable progress in DG, the lack of global receptive fields hinders further developments of CNN-based models for boosting generalization performance. To address this issue, Huang et al. (Huang et al. 2023) propose subdomain alignment and domain-aware attention to achieve DG with Transformers. Nonetheless, it suffers from quadratic complexity to the resolution rising from the attention mechanism, leading to extra computation and memory overhead. Moreover, it is limited to single-source domain generalization. Thus, it is necessary to investigate the multi-domain generalization that excels in global information modeling and low computational complexity in PCC.

Mamba. Mamba (Gu and Dao 2023), as well as the state space model (SSM), has garnered increasing attention due to its significant advantages in global receptive fields and computational complexity. VMamba (Liu et al. 2024b) and Vim (Zhu et al. 2024) propose visual SSMs to deploy Mamba for vision tasks. In the area of point cloud understanding, PointMamba (Liang et al. 2024) introduced a reordering strategy to scan data in a specific sequence to capture point cloud structures. Besides, Mamba3D (Han et al. 2024) used local norm pooling blocks to extract local geometric features. Similarly, Zhang et al. presented PCM (Zhang et al. 2025), which converts point clouds into one-dimensional point sequences using a consistent traversal serialization method, ensuring that adjacent points in the sequence remain adjacent in space. Unfortunately, no research investigates Mamba’s generalizability in point cloud tasks. To our knowledge, this is the first work that studies the generalizability of SSM-based models toward unseen domains in point cloud tasks. This paper uses the popular

PCM (Zhang et al. 2025) as the baseline.

Method

In this section, we present PointDGMamba, a novel State Space Model-based framework for DG point cloud classification that excels in strong generalizability toward unseen domains and has the advantages of global receptive fields and efficient linear complexity. As shown in Figure 2, it includes three core components: Masked Sequence Denoising (MSD), Sequence-wise Cross-domain Feature Aggregation (SCFA), and Dual-level Domain Scanning(DDS). Concretely, MSD is presented to selectively mask out noised point patches in the sequence and thus alleviate the noise accumulation during the serialization stage. Besides, SCFA is introduced to aggregate cross-domain but same-class point cloud features with the global prompt to extract more generalized features, thereby strengthening Mamba’s effectiveness in handling distribution shifts. Finally, DDS, including two scanning methods: intra-domain scanning and cross-domain scanning, is proposed to facilitate sufficient information interaction between different parts of the features. All modules are inserted after the first Mamba stage.

Masked Sequence Denoising

We have noticed that during Mamba’s serialization process, the inherent topology of the point cloud often gets disrupted when 3D point cloud data is converted into 1D sequences, leading to the generation of unexpected noise that is unrelated to the current state. This noise can be accumulated during the training, which could negatively impact the model’s performance when it encounters unseen data.

To address this issue, we propose Masked Sequence Denoising (MSD) to selectively mask the noised point tokens of the point cloud sequences and use the purified features for classification, mitigating the adverse effects of noise accumulation. This process not only preserves the basic feature of the point cloud but also ensures that the denoised sequence can highly represent the original structure. Specifically, we define the point cloud feature as f and the mask as m . The masked feature map F can be represented as:

$$F = f \otimes m, \quad (1)$$

Ideally, an element of 0 in the mask sequence indicates that the features are to be masked, while an element of 1 indicates that the features are to be preserved. However, since this mask consists of only 0 and 1 values, the gradient cannot be back-propagated during the training. We can solve this problem by using the Gumbel-Softmax method (Jang, Gu, and Poole 2016), where the mask m at a certain position can be represented as:

$$m = \frac{\exp\left(\frac{g_1 + \log p_1}{\tau}\right)}{\sum_{i=0}^1 \exp\left(\frac{g_i + \log p_i}{\tau}\right)} \quad (2)$$

where g_i represents Gumbel noise, τ is the temperature parameter, and p_i represents the probability that the current position is 0 or 1. The higher the probability of p_i becomes,

the closer the probability value of the current position is to a value of 1, otherwise it is closer to a value of 0. This indicates that while features are preserved, noise is greatly suppressed. As shown in Figure 2(a), the input sequence will be multiplied by a learnable mask to obtain the masked sequence. Then, the denoised and purified sequence features will be forwarded for classification.

Remark. Unlike existing Masked Point Modeling (MPM)-based methods, we intend to filter out the noised tokens selectively rather than reconstruct the masked tokens in the point cloud sequence. Besides, as for the technical designs, our MSD involves addressing the challenge of learning a discrete number (0 or 1) during the back-propagation, setting it apart from other MPM frameworks.

Sequence-wise Cross-domain Feature Aggregation

We observe that the current blocks of Mamba models are manually designed and over-heuristic. They neglect the designs of domain-agnostic features and tend to overfit specific domains. This can introduce unexpected domain-specific information into the sequence data, reducing Mamba’s effectiveness in dealing with distribution shifts.

To address this, we propose a Sequence-wise Cross-domain Feature Aggregation (SCFA) to aggregate cross-domain but same-class point cloud features to prompt the model to extract more generalized features. Specifically, as shown in Figure 2(b), the denoised point cloud features f_1 are fed into this module. Then, the randomly selected same-class point cloud feature from other domains f_2 is aggregated with f_1 to get the Cross-Domain Feature f' :

$$f' = \text{Conv}(\text{MLP}(f_1) \otimes \text{MLP}(f_2)), \quad (3)$$

where \otimes represents the element-wise multiplication, MLP represents multi-layer perceptron and Conv represents a convolutional layer.

In addition, we also introduce Global Prompt f_g to capture global information of the entire source domain, which consists of a set of learnable vectors. It is similar to a system message in large language models. Then, we aggregate the global prompt f_g with f_1 and f' to avoid unanticipated domain-specific information in the sequence data. During the training, point cloud features will be re-aggregated as:

$$F = \text{Concat}(f_1, f', f_g), \quad (4)$$

where F is the aggregated point cloud features.

Dual-level Domain Scanning

A key challenge is converting 3D point cloud data into 1D sequence data suitable for Mamba in DG PCC. While recent studies have explored various scanning methods for point cloud tasks, these rigid and fixed approaches often introduce human biases and largely overlook domain-agnostic factors. Moreover, they are susceptible to varying conditions, making it difficult to apply them in unseen domains.

In order to facilitate the interaction of different feature information for generalization, we design Dual-level Domain Scanning, including Intra-domain Scanning (IDS) and Cross-domain Scanning (CDS). As shown in Figure 3, the

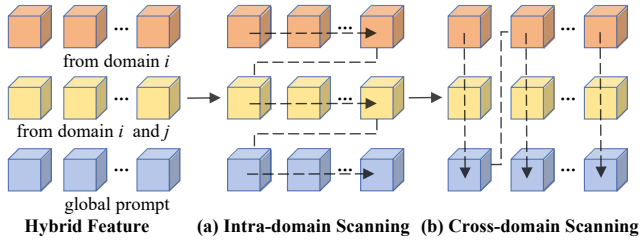


Figure 3: Dual-level Domain Scanning (DDS) comprises Intra-domain Scanning and Cross-domain Scanning.

cubes of different colors in the figure represent different features. IDS treats features as three unrelated sequences, scanning them one after another in order. Only after scanning the current feature can the next feature be scanned. CDS treats features as three related sequences. After scanning a data point of the first feature, the data points at the same position for the other two features will be scanned sequentially. This can promote the interaction between each feature. We use arrows in the figure to indicate the order in which each data is scanned. The scanning process could be formulated as:

$$F_{out} = CDS(IDS(F)), \quad (5)$$

where F_{out} is the output feature after two types of scanning. Then, the output features will be sent to the later Mamba stages and classification head. It is worth noting that the DDS module processes aggregated features, so it needs to be used together with the SCFA module.

Training and Inference

Training. During the training, we use Cross Entropy loss to measure the difference between the model’s prediction and the ground truths. Specifically, for given input samples x and its labels y , Cross Entropy loss \mathcal{L}_{class} can be formulated as:

$$\mathcal{L}_{class} = - \sum_i y_i \log(\hat{y}_i), \quad (6)$$

where \hat{y}_i is the probability that the sample x_i predicted by the model belongs to a certain class.

In addition, our PointDGMamba model requires a set of point clouds with the same class but from different source domains during the training phase for each experiment. We employ random resampling techniques during the data loading to ensure that the number of same-class point clouds in different source domains is consistent.

Inference. During the inference phase, only the point cloud data from a single domain is input into the model. It is worth noting that each point cloud does not need to interact with features from other domains except the Global Prompt. We replace it with each point cloud’s feature itself in SCFA.

Experiments

Experiment Settings

Implementation. In the training process, we used the AdamW (Loshchilov and Hutter 2017) optimizer with an

initial learning rate of $1e^{-4}$, a cosine decay schedule, and a weight decay of $1e^{-4}$. The number of epochs was set to 200. During the first 5 epochs, we employed a warmup mechanism to gradually increase the learning rate, reducing initial instability. Subsequently, the learning rate decreased following a cosine function, maintaining a relatively low learning rate of $1e^{-5}$ in the later stages of training. During the training phase, we used the PointMix (Chen et al. 2020) method for data augmentation to obtain more training samples.

Benchmark. To evaluate our method, we use the widely-used PointDA-10 (Qin et al. 2019) benchmark, which consists of ModelNet-10(M), ShapeNet-10(S), and ScanNet-10(S*), and contain 10 shared categories. Among them, ModelNet and ShapeNet were obtained from synthetic 3D models. ScanNet is sampled from the real world. Compared to ModelNet and ShapeNet, point clouds in ScanNet often have certain missing parts due to object occlusion. In this benchmark, we randomly select two domains as the source domain and the remaining domain as the target domain for each experiment, including 3 DG scenarios: (1) M, S*→S; (2) M, S→S* and (3) S, S*→M. In all experiments, both the training and testing sets of the source domains are used, while the target domain only uses the testing set.

PointDG-3to1 Benchmark

The existing DG PCC benchmarks usually include a limited number of source domains, *e.g.*, the number of source domains is only two, resulting in a lack of sufficient object diversities in the training data. This is not conducive to the model’s generalization in the unseen domain. In this paper, we propose a multi-domain generalization benchmark named PointDG-3to1 consisting of four sub-datasets, which is more diverse, practical, and challenging.

The PointDG-3to1 benchmark includes four sub-datasets: ModelNet-5 (A), ScanNet-5(B), ShapeNet-5 (C), and 3D-FUTURE-Completion (D). There are 5 shared classes in each dataset, including “cabinet”, “chair”, “lamp”, “sofa”, and “table”. The 3D-FUTURE-Completion (Liu et al. 2024a) was generated from the 3D-FUTURE (Fu et al. 2021) dataset, with each point cloud consisting of 2048 points. When performing DG, it includes 4 variants of leave-one-out (LOO) settings: (1) ABC→D; (2) ABD→C; (3) ACD→B, and (4) BCD→A, where 3 domains are used as source domains and the remaining one as the unseen domain. Following the common practice of DG, whole source domains are used for training (the training and testing sets), and the testing set is used for evaluation.

Comparisons to the State-of-the-art Methods

Comparison Methods. To evaluate the effectiveness of our proposed method, we compare several state-of-the-art approaches in PCC. These include CNN-based methods such as PointDAN (Qin et al. 2019), DefRec (Achituve, Maron, and Chechik 2021), GAST (Zou et al. 2021), PDG (Wei, Gu, and Sun 2022), MetaSets (Huang et al. 2021), PointNeXt (Qian et al. 2022) and X-3D (Sun et al. 2024), and Transformer-based methods such as SUG (Huang et al. 2023), PCT (Guo et al. 2021) and GBNet (Qiu, Anwar, and Barnes 2021), and Mamba-based methods such as

Method	Setting	Venue	Backbone	PointDA-10 Benchmark				PointDG-3to1 Benchmark				
				M,S*→S	M,S→S*	S,S*→M	Avg.	ABC→D	ABD→C	ACD→B	BCD→A	Avg.
PointDAN	DA	NeurIPS'2019	PointNet	77.38	40.32	78.69	65.46	58.85	81.66	48.86	79.95	67.33
DefRec	DA	WACV'2021	DGCNN	77.23	44.28	84.77	68.76	72.76	79.97	43.29	87.94	70.99
DefRec+PCM	DA	WACV'2021	DGCNN	80.02	48.39	83.39	70.60	68.81	82.90	44.00	83.33	69.76
GAST	DA	ICCV'2021	DGCNN	79.43	47.69	81.72	69.61	71.78	86.43	52.31	86.21	74.18
MetaSets	DG	CVPR'2021	PointNet	81.39	50.86	83.48	71.91	73.24	92.41	60.97	87.28	78.48
PDG	DG	NeurIPS'2022	PointNet	79.82	51.73	83.51	71.69	73.38	92.98	60.57	89.90	79.21
PointNeXt	DG	NeurIPS'2022	PointNet	77.31	43.32	78.16	66.26	71.47	91.70	46.39	88.95	74.63
X-3D	DG	CVPR'2024	PointNet	78.06	46.91	79.69	68.22	71.58	91.89	48.34	88.45	75.07
PCT	DG	CVM'2021	PointTrans	80.23	48.29	81.91	70.14	71.43	87.43	58.43	88.34	76.41
GBNet	DG	TMM'2021	PointTrans	79.94	48.92	81.34	70.07	72.78	87.83	57.76	88.82	76.80
SUG	DG	MM'2023	PointTrans	78.34	49.59	82.03	69.99	71.58	89.62	54.66	86.35	75.55
PCM	DG	AAAI'2025	Mamba	81.02	46.83	83.92	70.59	72.27	91.24	57.28	87.54	77.08
PointDGMamba	DG	-	Mamba	84.33	52.83	87.38	74.85	74.20	95.51	61.71	90.68	80.53

Table 1: Comparison results with state-of-the-art PCC methods on the PointDA-10 and our proposed PointDG-3to1 benchmark. **Avg.** denotes the average classification accuracy across all target domains. We have highlighted the best result in **black**.

SCFA		MSD	M,S*→S	M,S→S*	S,S*→M	Avg.
CDF	GP					
			82.45	48.19	81.49	70.71
✓			83.35	50.07	84.83	72.75
✓	✓		83.51	50.72	85.46	73.23
✓	✓	✓	84.33	52.83	87.38	74.85

Table 2: Ablation studies on SCFA and MSD modules on the PointDA-10 benchmark. CDF and GP represent Cross-Domain Features and Global Prompt, respectively.

PCM (Zhang et al. 2025). Since some Transformer-based and Mamba-based methods were not originally designed for domain generalization, we made appropriate modifications to enable their application to DG PCC settings. *More details could be referred to the supplementary.*

Results on PointDA-10 benchmark. As shown in Table 1, we report the comparison results on the PointDA-10 benchmark, indicating that our PointDGMamba achieves the best generalization performance on all these domains. Specifically, Our PointDGMamba outperforms the existing state-of-the-art methods, *e.g.*, PointDGMamba is superior to MetaSets by 2.94% in average generalization performance, demonstrating the effectiveness of our PointDGMamba in boosting the generalizability. The main reason is that these DG methods heavily rely on CNNs and ViTs, either suffering from limited receptive fields or quadratic complexity. In contrast, our SSM-based method, PointDGMamba, achieves global receptive fields and linear complexity and makes tailored designs for learning domain-invariant features in SSM.

Results on PointDG-3to1 benchmark. As shown on the right of Table 1, we report the generalization performance on the PointDG-3to1 benchmark, and our proposed PointDGMamba achieves superior generalization performance in all scenarios. In particular, the average generalization per-

DDS		M,S*→S	M,S→S*	S,S*→M	Avg.
IDS	CDS				
		82.99	50.92	84.65	72.85
✓		83.70	51.89	85.82	73.80
	✓	84.15	51.72	85.60	73.82
✓	✓	84.33	52.83	87.38	74.85

Table 3: Ablation studies on the DDS module on the PointDA-10 benchmark, where IDS and CDS are Intra-domain Scanning and Cross-domain Scanning, respectively.

formance of PointDGMamba exceeds the existing state-of-the-art method PDG by 1.32%. On the challenging 3D Feature and ScanNet target domains, our method outperforms the second-best method by 0.82% and 0.74%, respectively, which also demonstrates the superiority of our PointDG-Mamba to CNN-based and ViT-based methods.

Ablation Study

In this section, we provide ablation studies to verify the contribution of each proposed component. For simplicity, all experiments are conducted using the PointDA-10 benchmark.

Effectiveness of SCFA and MSD. Table 2 demonstrates the contribution of SCFA and MSD while keeping DDS unchanged. As aforementioned, SCFA leverages three parts of point cloud features: source domain features, cross-domain features (CDF), and global prompt (GP). The first row means just using the source domain features without any aggregation. After adding CDF, the generalization performance is improved in all scenarios, proving the effectiveness of CDF in promoting the extraction of more generalized features in the model. After using GP, the performance has also been improved, indicating that the model can extract generalization features on a global scale. Finally, after further adding MSD, we achieved the best generalization perfor-

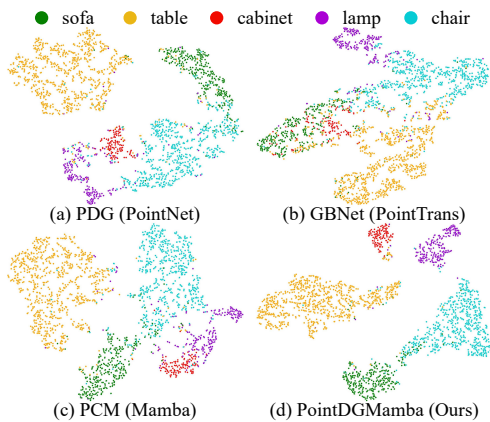


Figure 4: Visualization of our PointDGMamba and other state-of-the-art methods using t-SNE, where they are tested on the ShapeNet-5(C) dataset of the PointDG-3to1 benchmark. Different colors represent different classes.

mance, indicating that MSD has significant advantages in removing noise that is not conducive to generalization.

Effect of IDS and CDS. Table 3 shows the impact of Intra-Domain Scanning (IDS) and Cross-Domain Scanning (CDS) of our DDS module. The baseline means the model without any scanning, and the accuracy is less desired, indicating that scanning operations are crucial for feature processing and information fusion. Specifically, when using IDS only, the model can achieve certain performance improvements because it can capture the relationships between features to some extent. However, due to the lack of cross-domain information interaction, the classification accuracy is still not perfect. When using only CDS, we can also observe improvements in performance since the model can better fuse feature information from different domains, while the performance is not the best compared to using both scanning methods simultaneously. This indicates that during the feature scanning process, we should focus on not only the feature information within the domain but also cross-domain feature interaction, to enhance the model’s generalizability.

Visualization and Analysis

Feature Visualization. To more intuitively demonstrate the impact of PointDGMamba, we used t-SNE to visualize the features. We chose to visualize the testing set of ShapeNet-5(C) dataset on the PointDG-3to1 benchmark. This is because the testing sets of others have too few point clouds, or the number of point clouds of different classes varies by order of magnitude, which is not intuitive enough to verify our motivation. Figure 4 shows the feature visualization of the PointNet-based method PDG, Point Transformer-based method GBNet, Mamba-based method PCM, and our PointDGMamba. Specifically, PointDGMamba has achieved excellent characterization, manifested as stronger intra-class compactness and inter-class discrimination ability. Especially on the “cabinet (blue)” and “lamp (green)”, the difference between PointDGMamba is more obvious, with stronger compactness. These findings confirm the superior-

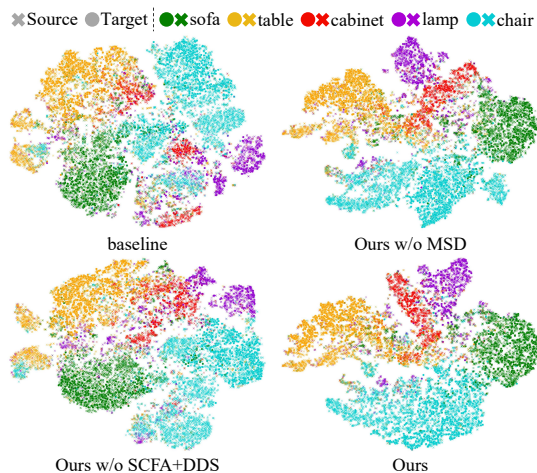


Figure 5: Visualization on the distributions of ablations of our PointDGMamba. We also visualize the source and target domains, with marker “x” representing the entire source domain and circles representing the target domain.

ity of PointDGMamba in improving generalizability.

In addition, we also visualized the impact of different modules on PointDGMamba. The SCFA and DDS modules have a causal relationship and must be used together. As shown in Figure 5, we visualize the point cloud features of the target and the source domains. From the perspective of the target domain, the model has poor inter-class discrimination ability when no modules are used. In the absence of MSD or SCFA+DDS, the inter-class discriminative ability of the model has been improved to some extent, but the compactness of some classes is poor, such as the “lamp (purple)”. When all modules are used together, our method can achieve optimal performance. From the source domain perspective, incomplete point clouds in ScanNet-5 make it difficult for the model to fully align point clouds of the same class in the feature space. After adding all modules, our method still allows point clouds of the same class in the source and target domains to stay closer in the feature space, such as “table (orange)” and “chair (Cyan)”.

Conclusion

We propose PointDGMamba, a novel State Space Model-based framework for domain generalizable point cloud classification. It excels in strong generalizability toward unseen domains and has the advantages of global receptive fields and efficient linear complexity. Specifically, Masked Sequence Denoising is presented to mitigate the adverse effects of noise accumulation during the serialization stage. Cross-domain Feature Aggregation and Dual-level Domain Scanning are designed to strengthen Mamba’s effectiveness in learning domain-invariant features and avoiding unanticipated domain-specific information in the sequence data. In addition, we also proposed a benchmark PointDG-3to1 that includes more domains. Extensive experiments with analyses demonstrate the effectiveness and superiority of our PDGMmaba against state-of-the-art competitors.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 72192821), the National Natural Science Foundation of China (No. 62472282), the Fundamental Research Funds for the Central Universities (No. YG2023QNA35) and YuCaiKe [2023] Project Number (231111310300).

References

- Achituve, I.; Maron, H.; and Chechik, G. 2021. Self-supervised learning for domain adaptation on point clouds. In *Proceedings of Winter Conference on Applications of Computer Vision*, 123–133.
- Chen, Y.; Hu, V. T.; Gavves, E.; Mensink, T.; Mettes, P.; Yang, P.; and Snoek, C. G. 2020. Pointmixup: Augmentation for point clouds. In *Proceedings of European Conference Computer Vision*, 330–345.
- Deng, Z.; Li, X.; Li, X.; Tong, Y.; Zhao, S.; and Liu, M. 2024. VG4D: Vision-Language Model Goes 4D Video Recognition. *arXiv preprint arXiv:2404.11605*.
- Fang, Z.; Li, X.; Li, X.; Buhmann, J. M.; Loy, C. C.; and Liu, M. 2024. Explore in-context learning for 3d point cloud understanding. *Advances in Neural Information Processing Systems*, 36.
- Fu, H.; Jia, R.; Gao, L.; Gong, M.; Zhao, B.; Maybank, S.; and Tao, D. 2021. 3d-future: 3d furniture shape with texture. *International Journal of Computer Vision*, 129: 3313–3337.
- Gu, A.; and Dao, T. 2023. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*.
- Guo, M.-H.; Cai, J.-X.; Liu, Z.-N.; Mu, T.-J.; Martin, R. R.; and Hu, S.-M. 2021. Pct: Point cloud transformer. *Computational Visual Media*, 7: 187–199.
- Han, K.; Wang, Y.; Chen, H.; Chen, X.; Guo, J.; Liu, Z.; Tang, Y.; Xiao, A.; Xu, C.; Xu, Y.; et al. 2022. A survey on vision transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1): 87–110.
- Han, X.; Tang, Y.; Wang, Z.; and Li, X. 2024. Mamba3d: Enhancing local features for 3d point cloud analysis via state space model. *arXiv preprint arXiv:2404.14966*.
- Huang, C.; Cao, Z.; Wang, Y.; Wang, J.; and Long, M. 2021. Metasets: Meta-learning on point sets for generalizable representations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8863–8872.
- Huang, S.; Zhang, B.; Shi, B.; Li, H.; Li, Y.; and Gao, P. 2023. Sug: Single-dataset unified generalization for 3d point cloud classification. In *Proceedings of the ACM International Conference on Multimedia*, 8644–8652.
- Jang, E.; Gu, S.; and Poole, B. 2016. Categorical reparameterization with gumbel-softmax. *arXiv preprint arXiv:1611.01144*.
- Jiang, J.; Zhou, Q.; Li, Y.; Zhao, X.; Wang, M.; Ma, L.; Chang, J.; Zhang, J. J.; and Lu, X. 2024. PCoTTA: Continual Test-Time Adaptation for Multi-Task Point Cloud Understanding. *Advances in Neural Information Processing Systems*.
- Kim, H.; Kang, Y.; Oh, C.; and Yoon, K.-J. 2023. Single domain generalization for lidar semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 17587–17598.
- Lehner, A.; Gasperini, S.; Marcos-Ramiro, A.; Schmidt, M.; Mahani, M.-A. N.; Navab, N.; Busam, B.; and Tombari, F. 2022. 3d-vfield: Adversarial augmentation of point clouds for domain generalization in 3d object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 17295–17304.
- Li, Y.; Bu, R.; Sun, M.; Wu, W.; Di, X.; and Chen, B. 2018. Pointcnn: Convolution on x-transformed points. *Advances in Neural Information Processing Systems*, 31.
- Li, Y.; Zhou, Q.; Gong, J.; Zhu, Y.; Dazeley, R.; Zhao, X.; and Lu, X. 2025. DAPoinTr: Domain Adaptive Point Transformer for Point Cloud Completion. *AAAI Conference on Artificial Intelligence*.
- Liang, D.; Zhou, X.; Wang, X.; Zhu, X.; Xu, W.; Zou, Z.; Ye, X.; and Bai, X. 2024. Pointmamba: A simple state space model for point cloud analysis. *arXiv preprint arXiv:2402.10739*.
- Liu, F.; Gong, J.; Zhou, Q.; Lu, X.; Yi, R.; Xie, Y.; and Ma, L. 2024a. Cloudmix: Dual mixup consistency for unpaired point cloud completion. *IEEE Transactions on Visualization and Computer Graphics*.
- Liu, Y.; Tian, Y.; Zhao, Y.; Yu, H.; Xie, L.; Wang, Y.; Ye, Q.; and Liu, Y. 2024b. Vmamba: Visual state space model. *arXiv preprint arXiv:2401.10166*.
- Loshchilov, I.; and Hutter, F. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.
- Pang, Y.; Wang, W.; Tay, F. E.; Liu, W.; Tian, Y.; and Yuan, L. 2022. Masked autoencoders for point cloud self-supervised learning. In *European Conference on Computer Vision*, 604–621.
- Phan, A. V.; Le Nguyen, M.; Nguyen, Y. L. H.; and Bui, L. T. 2018. Dgcnn: A convolutional neural network over large-scale labeled graphs. *Neural Networks*, 108: 533–543.
- Qi, C. R.; Su, H.; Mo, K.; and Guibas, L. J. 2017a. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 652–660.
- Qi, C. R.; Yi, L.; Su, H.; and Guibas, L. J. 2017b. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in Neural Information Processing Systems*, 30.
- Qian, G.; Li, Y.; Peng, H.; Mai, J.; Hammoud, H.; Elhoseiny, M.; and Ghanem, B. 2022. Pointnext: Revisiting pointnet++ with improved training and scaling strategies. *Advances in Neural Information Processing Systems*, 35: 23192–23204.
- Qin, C.; You, H.; Wang, L.; Kuo, C.-C. J.; and Fu, Y. 2019. Pointdan: A multi-scale 3d domain adaption network for point cloud representation. *Advances in Neural Information Processing Systems*, 32.
- Qiu, S.; Anwar, S.; and Barnes, N. 2021. Geometric back-projection network for point cloud classification. *IEEE Transactions on Multimedia*, 24: 1943–1955.

- Sun, S.; Rao, Y.; Lu, J.; and Yan, H. 2024. X-3D: Explicit 3D Structure Modeling for Point Cloud Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5074–5083.
- Wang, Y.; Sun, Y.; Liu, Z.; Sarma, S. E.; Bronstein, M. M.; and Solomon, J. M. 2019. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics*, 38(5): 1–12.
- Wei, X.; Gu, X.; and Sun, J. 2022. Learning generalizable part-based feature representation for 3D point clouds. *Advances in Neural Information Processing Systems*, 35: 29305–29318.
- Xiao, A.; Huang, J.; Xuan, W.; Ren, R.; Liu, K.; Guan, D.; El Saddik, A.; Lu, S.; and Xing, E. P. 2023. 3d semantic segmentation in the wild: Learning generalized models for adverse-condition point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9382–9392.
- Xing, Z.; Ye, T.; Yang, Y.; Liu, G.; and Zhu, L. 2024. Segmamba: Long-range sequential modeling mamba for 3d medical image segmentation. *arXiv preprint arXiv:2401.13560*.
- Yu, X.; Tang, L.; Rao, Y.; Huang, T.; Zhou, J.; and Lu, J. 2022. 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*, 19313–19322.
- Zhang, T.; Yuan, H.; Qi, L.; Zhang, J.; Zhou, Q.; Ji, S.; Yan, S.; and Li, X. 2025. Point could mamba: Point cloud learning via state space model. *AAAI Conference on Artificial Intelligence*.
- Zhang, Z.; Liu, A.; Reid, I.; Hartley, R.; Zhuang, B.; and Tang, H. 2024. Motion mamba: Efficient and long sequence motion generation with hierarchical and bidirectional selective ssm. *arXiv preprint arXiv:2403.07487*.
- Zhao, H.; Jiang, L.; Jia, J.; Torr, P. H.; and Koltun, V. 2021. Point transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 16259–16268.
- Zhu, L.; Liao, B.; Zhang, Q.; Wang, X.; Liu, W.; and Wang, X. 2024. Vision mamba: Efficient visual representation learning with bidirectional state space model. *arXiv preprint arXiv:2401.09417*.
- Zou, L.; Tang, H.; Chen, K.; and Jia, K. 2021. Geometry-aware self-training for unsupervised domain adaptation on object point clouds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 6403–6412.