

PointDGRWKV: Generalizing RWKV-like Architecture to Unseen Domains for Point Cloud Classification

Hao Yang^{1*}, Qianyu Zhou^{2*}, Haijia Sun³, Xiangtai Li⁴, Xuequan Lu⁵,
Lizhuang Ma^{1†}, Shuicheng YAN⁶

¹School of Computer Science, Shanghai Jiao Tong University, Shanghai, China

²College of Computer Science and Technology, Jilin University, Jilin, China

³School of Information Management, Nanjing University, Nanjing, China

⁴College of Computing and Data Science, Nanyang Technological University, Singapore

⁵Department of Computer Science and Software Engineering, The University of Western Australia, Australia

⁶School of Computing, National University of Singapore, Singapore

Abstract

Domain Generalization (DG) has been recently explored to enhance the generalizability of Point Cloud Classification (PCC) models toward unseen domains. Prior works are based on convolutional networks, Transformer or Mamba architectures, either suffering from limited receptive fields or high computational cost, or insufficient long-range dependency modeling. RWKV, as an emerging architecture, possesses superior linear complexity, global receptive fields, and long-range dependency. In this paper, we present the first work that studies the generalizability of RWKV models in DG PCC. We find that directly applying RWKV to DG PCC encounters two significant challenges: RWKV's fixed direction token shift methods, like Q-Shift, introduce spatial distortions when applied to unstructured point clouds, weakening local geometric modeling and reducing robustness. In addition, the Bi-WKV attention in RWKV amplifies slight cross-domain differences in key distributions through exponential weighting, leading to attention shifts and degraded generalization. To this end, we propose PointDGRWKV, the first RWKV-based framework tailored for DG PCC. It introduces two core modules to enhance spatial modeling and cross-domain robustness, while maintaining RWKV's linear efficiency. In particular, we present Adaptive Geometric Token Shift to model local neighborhood structures to improve geometric context awareness. In addition, Cross-Domain key feature Distribution Alignment is designed to mitigate attention drift by aligning key feature distributions across domains. Extensive experiments on multiple benchmarks demonstrate that PointDGRWKV achieves state-of-the-art performance on DG PCC.

Code — <https://github.com/yxltya/PointDGRWKV>

Introduction

3D point clouds play a crucial role in various applications, such as autonomous driving, augmented reality, and robotics (Caesar et al. 2020; Billinghamurst et al. 2015; Thrun et al. 2019). Recently, point cloud classification (PCC)

*These authors contributed equally.

†Corresponding author.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

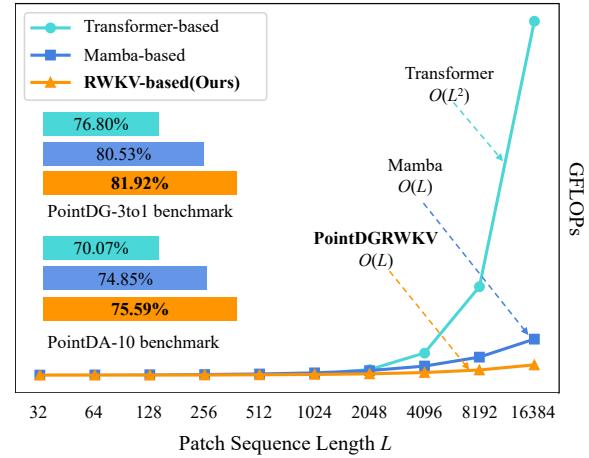


Figure 1: Accuracy-speed tradeoff in DG PCC. (Left) Overall accuracy of methods with different architectures (Right) GFLOPs increase with sequence length. Our PointDGRWKV achieves superior performance with its linear complexity.

tasks (Qi et al. 2017a,b; Phan et al. 2018) have made significant progress in understanding the local geometry and global shapes. However, most of them usually assume that the training and testing data share the same distribution. When the model is applied to unknown domains, the performance often drops significantly due to domain shifts induced by different sensors, environments, scanning angles, etc.

To address this issue, domain generalization (DG) (Long et al. 2025; Yang et al. 2025) has been introduced into point cloud analysis, aiming to train models solely on source domains and generalize well in unknown domains. The mainstream DG PCC methods tend to learn domain-invariant features via data augmentation (Xiao et al. 2023), adversarial training (Lechner et al. 2022), and consistency learning (Kim et al. 2023). Nonetheless, most of them are based on CNNs, and suffer from a limited receptive field, making it challenging to capture global structural information and harm the generalizability. Subsequently, point Transformer (Han et al.

2022) was introduced to enhance global modeling capabilities in DG PCC. However, its inherent attention involves high computational complexity, which limits its efficiency in practical applications. Point Cloud Mamba has recently shown the potential of sequence modeling in DG PCC (Yang et al. 2025). Nevertheless, due to its fixed state space size, it is difficult to fully capture long-range dependencies, especially under long sequence lengths.

Recently, Reception Weighted Key Value (RWKV) (Peng et al. 2023; Chen et al. 2025) has demonstrated excellent capabilities in long-range dependency modeling and capturing global information in NLP and vision tasks. Moreover, the core WKV attention mechanism exhibits a linear computational complexity, which significantly reduces the computational overhead of traditional self-attention. They demonstrate strong scalability in various vision tasks and even in point cloud analysis. Despite its gratifying progress, enhancing the generalizability of RWKV-like models in unseen domains for point cloud analysis remains an open problem, as directly applying RWKV to DG PCC tasks is non-trivial.

In this paper, we aim to improve the generalizability of RWKV-like architectures toward unseen domains in point cloud classification. Our motivations mainly lie in two aspects. *Firstly*, RWKV’s fixed direction token shift, e.g., Q-Shift, would inevitably introduce spatial distortions to unstructured point clouds due to the inconsistent order of token arrangement and spatial proximity, weakening the model’s ability to model local geometry and thus affecting robustness in unseen domains. Secondly, the Bi-WKV attention mechanism in RWKV is highly sensitive to slight discrepancies in *key* distribution between the source domain and the unseen domain. The nonlinear amplification characteristics, i.e., the exponential function, can easily amplify the shift in the focus of attention, undermining the generalization performance of the model in unknown domains.

Motivated by the aforementioned analysis, we propose PointDGRWKV, a novel RWKV-based framework for domain generalized point cloud classification. PointDGRWKV excels in strong generalizability, linear complexity, and capabilities in modeling long-range dependency and global structure information. Our proposed method has two key modules. **Firstly**, we design a lightweight, parameter-free Adaptive Geometric Token Shift mechanism (AGT-Shift) based on the inherent spatial characteristics of point clouds. It constructs local neighborhoods through spatial partitioning and dynamically integrates structural features to enhance the model’s ability to model geometric contexts. This mechanism is specifically designed based on the characteristics of point clouds. **Secondly**, we propose a Cross-Domain *Key* feature Distribution Alignment module (CD-KDA) to address the nonlinear amplification effect of *key* vectors on weight calculation in the Bi-WKV attention mechanism. By aligning the *key* distributions between source domains at the mean and covariance levels, we explicitly alleviate the cross-domain shift of attention and improve the generalization performance of the model in unseen domains. As shown in Fig. 1, PointDGRWKV achieves superior performance with less computational overhead compared to existing Transformer-based and Mamba-based methods on mul-

iple DG benchmarks. Our contributions are three-fold:

- We propose PointDGRWKV, a novel RWKV-based framework for domain generalizable point cloud classification that excels in strong generalizability toward unseen domains, global receptive fields, linear complexity, and long-range dependency.
- We design Adaptive Geometric Token Shift (AGT-Shift) and Cross-Domain *key* feature Distribution Alignment (CD-KDA) to enhance RWKV’s geometry perception ability and the generalizability toward unseen domains.
- Extensive experiments on multiple DG benchmarks verify the superiority and effectiveness of PointDGRWKV compared to state-of-the-art approaches.

Related Work

Point Cloud Classification (PCC) aims to accurately categorize 3D point cloud data. Early works such as PointNet (Qi et al. 2017a) and PointNet++(Qi et al. 2017b) pioneered the use of MLP-based architectures to directly learn features from raw point clouds. Subsequent research expanded on this by incorporating Convolutional Neural Networks (CNNs)(Li et al. 2018; Wang et al. 2019) to better capture local geometric patterns. Nevertheless, CNN-based approaches often struggle with limited receptive fields, especially in deeper networks. To address this, Vision Transformers (ViTs)(Zhao et al. 2021; Fang et al. 2024; Deng et al. 2024) have recently been adopted in PCC, offering enhanced global context modeling capabilities. Methods like PCT(Guo et al. 2021) and Point Transformer (Zhao et al. 2021) leverage self-attention mechanisms to capture long-range dependencies across points. Recently, Point Mamba and Point Cloud Mamba (Liang et al. 2024; Zhang et al. 2025) have introduced Mamba-like models into point cloud analysis, and achieved a global receptive field with linear complexity. While these models achieve impressive results on standard benchmarks, their generalization to novel or unseen domains remains a significant challenge.

Domain Generalized Point Cloud Classification (DG PCC) DG PCC aims to enhance the PCC model’s generalizability to previously unseen domains. Existing DG methods primarily focus on learning domain-invariant representations through meta learning (Huang et al. 2021), contrastive learning (Wei, Gu, and Sun 2022) consistency regularization (Kim et al. 2023), and data augmentation (Lehner et al. 2022; Xiao et al. 2023). While these approaches have shown promising results, many are built on CNN-based backbones, whose inherently limited receptive fields constrain their ability to capture global structural information critical for robust generalization. Subsequently, Huang et al. (Huang et al. 2023) proposed Transformers-based subdomain alignment and domain-aware attention mechanisms, while suffer from the quadratic computational costs. Recently, PointDGMamba (Yang et al. 2025) introduced Mamba-based architectures (Wang et al. 2025) in DG PCC to improve generalization to unseen domains. Although Mamba offers linear inference efficiency and long-sequence modeling capabilities, its fixed-size state space constrains its ability to capture

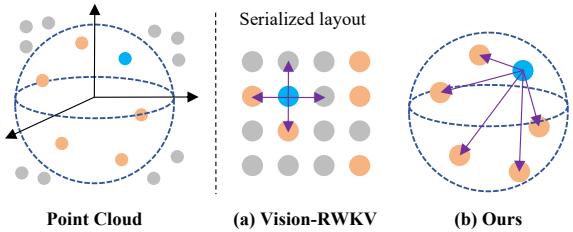


Figure 2: Comparison of local structure modeling between Vision-RWKV and our PointDGRWKV. (a) Vision-RWKV uses fixed four-directional token shifts, which may distort spatial relationships in unstructured point clouds. (b) In contrast, our Adaptive Geometric Token Shift dynamically aggregates local information based on geometric neighborhood, making it more suitable for point cloud data.

long-range spatial dependencies in point clouds. This highlights the need for novel architectures in DG PCC that can simultaneously support global context modeling and maintain better long-range dependencies on long sequence length.

Reception Weighted Key Value (RWKV). RWKV (Yuan et al. 2024; Chen et al. 2025) has garnered increasing attention due to its significant advantages in global receptive fields, computational complexity, and advantages in long sequence modeling. The core innovation is its linear WKV attention mechanism and spatial mixing and channel mixing, balancing local features and global dependencies through gating and recursion mechanisms, and supporting parallel training and efficient inference. Recently, Point-RWKV introduce RWKV in point cloud analysis, but did not really open-source their implementations. Regarding the unstructured and sparse nature of point clouds, as well as cross-domain differences such as sensor or scene changes, pose new challenges to RWKV (Yin, Li, and Dong 2024). To our knowledge, this is the first work that studies the generalizability of RWKV-based models toward unseen domains in point cloud tasks. This paper uses the popular vision-RWKV (Yuan et al. 2024) as the baseline.

Method

Revisiting the RWKV

RWKV (Duan et al. 2024) incorporates a token shift function, e.g., Q-Shift, which introduces interactions among nearby positions along the channel dimension, enriching local context without increasing the computational cost:

$$\text{Q-Shift}_S(X) = X + (1 - \mu_S)X^* \quad (1)$$

$$X^* = \text{Concat}(X_1, X_2, X_3, X_4),$$

where X^* denotes a sliced vector of X , capturing tokens from positions adjacent to the current location in the channel dimension. However, when applied to point clouds, this operation may distort the underlying spatial structure (Fig.2).

Moreover, in the attention mechanism adopted by Bi-WKV (Duan et al. 2024), the attention weight of each

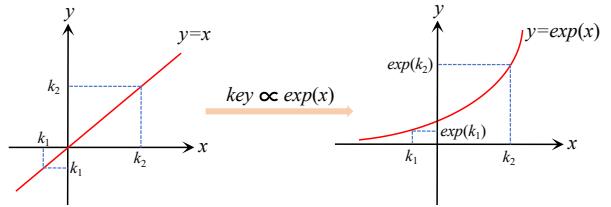


Figure 3: Illustration of the exponential function’s amplification effect on *key* differences: for example, $k_1 = -0.3$ and $k_2 = 1.0$ differ by only 1.3, but after exponentiation, they become $e^{-0.3} \approx 0.74$ and $e^{1.0} \approx 2.72$, showing a magnified gap. This indicates that in Bi-WKV attention, small *key* differences can be significantly amplified, potentially causing biased attention and harming the model’s generalizability.

token is formulated as follows:

$$\text{wkv}_t = \text{Bi-WKV}(K, V)_t$$

$$= \frac{\sum_{i=0, i \neq t}^{T-1} e^{-\frac{|t-i|-1}{T} \cdot \mathbf{w} + \mathbf{k}_i} \cdot \mathbf{v}_i + e^{\mathbf{u} + \mathbf{k}_t} \cdot \mathbf{v}_t}{\sum_{i=0, i \neq t}^{T-1} e^{-\frac{|t-i|-1}{T} \cdot \mathbf{w} + \mathbf{k}_i} + e^{\mathbf{u} + \mathbf{k}_t}}, \quad (2)$$

where \mathbf{k}_i and \mathbf{v}_i represent the *key* and *value* of the i -th token, respectively, and \mathbf{w} is the learnable distance decay parameter, and \mathbf{u} is the learnable bias term. Since the *key* \mathbf{k} appears directly in the exponential function, its distribution has an exponential amplification effect on attention results (Fig.3), which can lead to attention drift and degrade the model’s generalization performance.

To address these limitations in the context of point cloud domain generalization, we propose two modules, as illustrated in Fig. 4: Adaptive Geometric Token Shift (AGT-Shift), which enhances local structure modeling via spatial partitioning, and Cross-Domain Key Distribution Alignment (CD-KDA), which improves the robustness of attention by aligning *key* feature distributions across domains.

Adaptive Geometric Token-Shift

When adapting the token shift mechanism of RWKV to the point cloud domain, two key challenges arise. Firstly, point clouds inherently lack regular topological structures, making it difficult to establish consistent spatial directions such as “up,” “down,” “left,” and “right” as in image data. Secondly, point cloud datasets are typically large-scale, and conventional operations like KNN search or graph construction introduce substantial computational and memory overhead, thereby limiting scalability. Consequently, a central challenge lies in achieving an effective balance between computational efficiency and the ability to model spatial structures.

To address this issue, we propose Adaptive Geometric Token Shift (AGT-Shift). AGT-Shift efficiently constructs the nearest neighborhood through spatial partitioning and introduces a weighted feature aggregation scheme among neighboring points to enable token shifting and enhance structural awareness. By avoiding the explicit computation of pairwise distance matrices, the method circumvents the quadratic complexity typically found in KNN-based approaches.

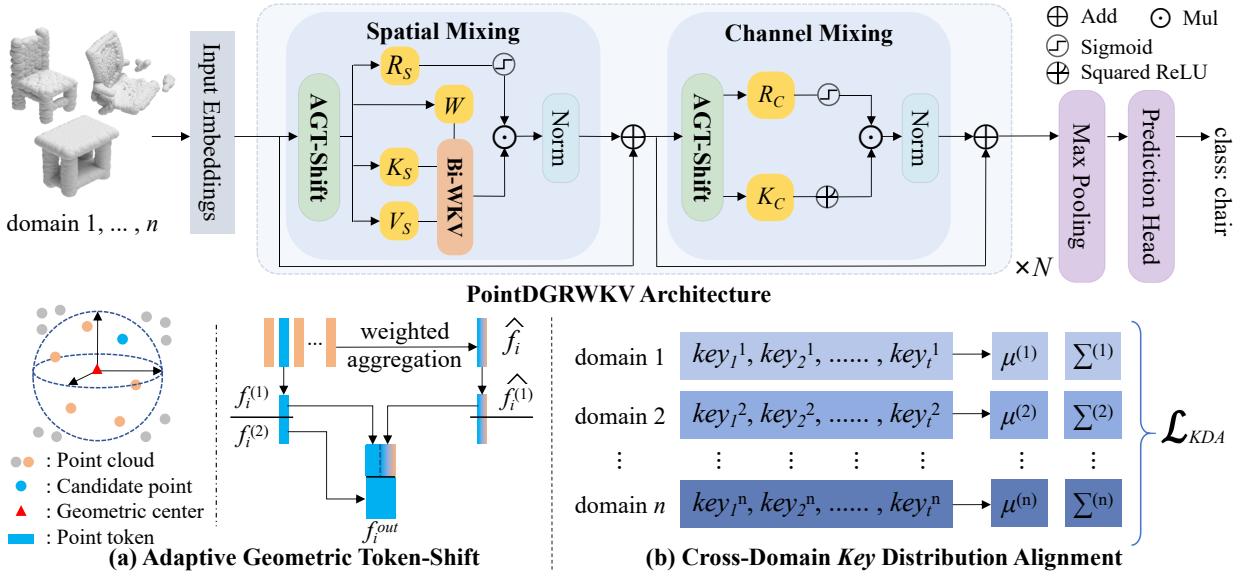


Figure 4: The architecture of PointDGRWKV consists of two key components: (a) **Adaptive Geometric Token-Shift**: it introduces an adaptive offset strategy based on geometric relations to fuse features from neighboring points in a weighted manner. It enables structure-aware feature shift and enhancement, improving the model’s ability to capture local geometric structures. (b) **Cross-Domain Key Distribution Alignment**: By minimizing the distribution differences of *key* across different source domains, CD-KDA reduces inconsistencies in cross-domain attention and enhances the model’s generalizability across domains.

Concretely, AGT-Shift partitions the 3D space into a set of spatial sub-regions using a spatial hashing technique with fixed step sizes. Points within the same sub-region are treated as forming a local context block, capturing localized geometric structures. For each point token, partial feature fusion is then conducted by computing a weighted average of the feature tokens within its corresponding region, facilitating efficient context-aware feature enhancement.

Let the point cloud features be denoted as $F \in \mathbb{R}^{B \times N \times C}$ and the corresponding point coordinates as $X \in \mathbb{R}^{B \times N \times 3}$, where B, N, C , represent the batch size, number of points, and feature dimensions, respectively. The 3D space is discretized into a set of spatial grids \mathcal{G}_i and each point is assigned to a grid cell based on its coordinates. For a given point $i \in \mathcal{G}_i$, its token shift feature is defined as:

$$\hat{f}_i = \sum_{j \in \mathcal{G}_i} w_{ij} f_j, \quad w_{ij} = \frac{\exp(-\|x_j - \mu_i\|)}{\sum_{k \in \mathcal{G}_i} \exp(-\|x_k - \mu_i\|)}, \quad (3)$$

where $\mu_i = \frac{1}{|\mathcal{G}_i|} \sum_{j \in \mathcal{G}_i} x_j$ denotes the geometric center of subregion \mathcal{G}_i , and w_{ij} is the contribution of point j to the representation of point i , with higher weights assigned to closer points. To preserve the discriminability of the original features and avoid excessive perturbation, we selectively perturb only a subset of channels and introduce a residual fusion mechanism to ensure stable feature refinement:

$$f_i^{\text{out}} = [\lambda f_i^{(1)} + (1 - \lambda) \hat{f}_i^{(1)} \| f_i^{(2)}], \quad (4)$$

where $f_i^{(1)}$ represents the first C' channels (used for disturbance), $f_i^{(2)}$ is the remaining channel, and $\lambda \in (0, 1)$ controls the degree of the disturbance in token shift.

Remark. Note that our presented AGT-Shift module does not rely on KNN or explicit adjacency graph construction, nor does it depend on additional parameter learning. All aggregation processes can be completed through tensor operations, and the overall computational complexity is $\mathcal{O}(N)$.

Cross-Domain Key Distribution Alignment

We observe that there are significant differences in the distribution of k across different domains, such as high or low mean values of k features and significant differences in variance in some domains. These distribution shifts will cause significant bias at e^k level, leading to shifts of attention focus position within the domain, severely undermining the model’s generalizability to unseen domains.

To address this issue, we propose Cross-Domain Key Feature Distribution Alignment(CD-KDA) to enhance the modeling stability and structural generalization ability of attention mechanisms in cross-domain scenes. Regarding the nature of point cloud data, the *key* vector k encodes the relative importance of each point within its local neighborhood or in relation to the global context. Specifically, k captures the spatial selection tendencies of the points: its mean μ reflects the global attention focus, while its variance Σ characterizes the semantic or geometric diversity among points. Consequently, *if different source domains exhibit distinct distributions of k due to geometric discrepancies, it will lead to domain shifts in the aggregation of point cloud structures governed by attention mechanisms*. In contrast, although the *value* vector v also contributes to attention computation, it only serves as the weighted content to be aggregated. It does not influence the generation of attention weights directly, nor does it appear within the exponential function of the atten-

Method	Setting	Venue	Backbone	PointDA-10 Benchmark					PointDG-3to1 Benchmark					Avg.
				M,S*→S	M,S→S*	S,S*→M	Avg.	ABC→D	ABD→C	ACD→B	BCD→A	Avg.	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		
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		
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	PCM	81.02	46.83	83.92	70.59	72.27	91.24	57.28	87.54	77.08		
PointDGMamba	DG	AAAI'2025	PCM	84.33	52.83	87.38	74.85	74.20	95.51	61.71	90.68	80.53		
V-RWKV'	DG	ICLR'2025	V-RWKV	81.90	49.52	85.49	72.24	73.42	92.12	57.88	88.18	77.90		
PointDGRWKV	DG	-	V-RWKV	84.39	54.10	88.49	75.66	76.37	95.99	63.92	91.38	81.92		

Table 1: Performance comparison between the proposed method and the state-of-the-art point cloud classification methods on the PointDA-10 and PointDG-3to1 benchmarks. The metric used is overall classification accuracy (%), and **Avg.** indicates the mean accuracy across all target domain scenarios. The highest result in each benchmark is marked in **bold**.

tion formulation. As a result, its effect on generalization is less immediate and critical than that of \mathbf{k} .

Additionally, the spatial decay parameter \mathbf{w} and bias \mathbf{u} in Bi-WKV are shared model parameters that encode the model’s inherent sensitivity to spatial distance and positional priors. These parameters should be learned jointly across multiple source domains to capture a unified inductive bias, and therefore should not be forcibly aligned across domains.

Based on these observations, we argue that *aligning the dynamic input key representations \mathbf{k} across source domains is the most critical and effective strategy to enhance generalization*. Let the set of source domains be denoted as $\{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_n\}$, with the corresponding key features extracted from each domain represented by $\mathbf{k}^{(i)} \in \mathbb{R}^{T \times C}$. We define the following objective function for alignment:

$$\mathcal{L}_{\text{CD-KDA}} = \frac{1}{|\mathcal{P}|} \sum_{(i,j) \in \mathcal{P}} \left\| \mu^{(i)} - \mu^{(j)} \right\|_2^2 + \left\| \Sigma^{(i)} - \Sigma^{(j)} \right\|_F^2, \quad (5)$$

where $\mu^{(i)} = \frac{1}{T} \sum_t \mathbf{k}_t^{(i)}$ is the *key* mean of the i -th domain, $\Sigma^{(i)}$ is the corresponding covariance matrix, \mathcal{P} is the set of unordered domain pairs between all source domains, and $\|\cdot\|_F$ is the Frobenius norm. As such, by minimizing the distribution of \mathbf{k} representations between source domains, the cross-domain stability of the attention mechanism is enhanced. The introduced CD-KDA explicitly aligns the distribution of \mathbf{k} representations in different source domains on a source domain basis, thereby improving the consistency of the model’s attention distribution to the unseen domain.

Training and Inference

During the training phase, the total loss of the model includes the classification loss and the cross-domain feature alignment loss, as follows:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{cls}} + \lambda_2 \mathcal{L}_{\text{CD-KDA}}, \quad (6)$$

where \mathcal{L}_{cls} is the cross-entropy loss used to supervise the model’s ability to recognize point cloud classes. Hyperparameters λ_1 and λ_2 , respectively, adjust the weights of these two loss terms during training. During the inference stage, only the trained feature extractor and classifier are used. The model no longer requires access to any source domain data and directly predicts the class labels of the target domain samples. This allows for efficient and scalable deployment in unseen domains without additional adaptation.

Experiments

Experiments Setup

Implementation Details. Our model was trained on an NVIDIA RTX 4090 GPU. The optimizer AdamW (Loshchilov and Hutter 2017) is used, with an initial learning rate of 1×10^{-4} , a cosine annealing scheduling strategy, and a weight decay of 1×10^{-4} . During the training process, preprocessing and enhancement operations such as scaling, normalization, and random jitter were applied to the input point cloud data. The model adopts a four-stage hierarchical structure for gradually extracting and aggregating multi-scale point cloud features. The number of RWKV blocks included in each stage is 1, 1, 2 and 2, respectively. In all experiments, $\lambda_1 = 1$ and $\lambda_2 = 0.3$ by default.

Benchmarks. To evaluate the generalizability of our method in DG PCC, we conduct experiments on two benchmarks. The first is PointDA-10 (Qin et al. 2019; Dai et al. 2017; Wu et al. 2015), which includes ModelNet-10 (M), ShapeNet-10 (S), and ScanNet-10 (S*) with 10 shared categories. ModelNet and ShapeNet contain clean point clouds generated from synthetic 3D models, while ScanNet captures real-world scenes with frequent missing regions due to occlusion. It defines three cross-domain settings: M, S*→S; M, S→S*; and S, S*→M. The second benchmark is PointDG-3to1 (Yang et al. 2025), including ModelNet-5 (A), ScanNet-5 (B),

AGTS	KDA	M,S*→S	M,S→S*	S,S*→M	Avg.	Gain
✓	✓	81.70	49.52	85.49	72.24	-
		83.39	51.10	86.21	73.57	1.33
		82.50	53.48	86.86	74.28	2.04
✓	✓	84.39	54.10	88.49	75.66	3.42

Table 2: Ablation study on the AGT-Shift (AGTS) and CD-KDA (KDA) modules on the PointDA-10 benchmark.

ShapeNet-5 (C), and 3D-FUTURE-Completion (D) (Liu et al. 2024; Fu et al. 2021), sharing five classes. It adopts a “leave-one-out” setting to form four settings: ABC→D, ABD→C, ACD→B, and BCD→A. Following common DG practices, the training uses only source samples, and the evaluation is performed on the target domain’s testing set.

Comparison Results

Comparison Methods. To comprehensively evaluate the proposed method, we compare it with representative point cloud classification models, including CNN-based methods such as PointDAN (Qin et al. 2019), DefRec (Achituv, Maron, and Chechik 2021), GAST (Zou et al. 2021), PDG (Wei, Gu, and Sun 2022), MetaSets (Huang et al. 2021), and PointNeXt (Qian et al. 2022), as well as Transformer-based approaches like SUG (Huang et al. 2023), PCT (Guo et al. 2021), and GBNet (Qiu, Anwar, and Barnes 2021). We also include Mamba-based methods PCM (Zhang et al. 2025) and PointDGMamba (Yang et al. 2025). Additionally, we evaluate V-RWKV, a modified Vision-RWKV (Duan et al. 2024) variant with a different number of blocks. Due to the unavailability of training code, PointRWKV (He et al. 2025) is excluded.

Benchmark Results. We conduct a comprehensive evaluation of the proposed PointDGRWKV method on two widely used multi-domain point cloud generalization benchmarks: PointDA-10 and PointDG-3to1. The results are presented in Table 1. PointDGRWKV consistently outperforms existing methods in terms of average overall accuracy across both benchmarks. Specifically, on the three domain generalization tasks of PointDA-10, PointDGRWKV achieves better performance than the state-of-the-art PointDGMamba across all DG tasks, with an average accuracy of 75.66%. Notably, on the PointDG-3to1 benchmark, PointDGRWKV achieves an average accuracy of 81.92% across four domain shifts, outperforming PointDGMamba by a significant margin of 1.39%. The improvement is particularly pronounced in the most challenging ACD→B setting.

Analysis of Improvements. Overall, PointDGRWKV demonstrates high performance across different domains, indicating strong cross-domain stability. We attribute this consistent improvement to the AGT-Shift mechanism, which better captures the local geometric structures inherent to unstructured point cloud data, effectively mitigating the information mismatch caused by the “pseudo-local receptive field” issue in the original RWKV. Additionally, the CD-KDA module alleviates attention misalignment in Bi-WKV caused by domain-specific variations in *key* distributions, en-

Shift	M,S*→S	M,S→S*	S,S*→M	Avg.
KNN-RandOne	83.39	52.85	87.38	74.54
KNN-Avg	83.87	53.36	85.51	74.25
KNN-WAvg	83.35	53.31	86.80	74.82
AGT-Shift (Ours)	84.39	54.10	88.49	75.66

Table 3: Ablations on different shifting strategies on the PointDA-10 benchmark.

abling the model to learn more consistent structural perception across source domains and thereby enhancing generalization to unseen domains. It is worth noting that while the proposed method shows clear improvements on average, its advantage is relatively modest in certain simpler settings, suggesting that the primary gains come from improved robustness and stability under more complex domain shifts.

Ablation Study

Effectiveness of AGT-Shift and CD-KDA. To further validate the specific roles of the proposed components in the model, we conducted ablation experiments on the PointDA-10 benchmark, and the results are presented in Table 2. Firstly, we constructed a basic version of V-RWKV’ without AGT-Shift and CD-KDA modules. Compared with the basic model, the introduction of AGT-Shift improved the overall performance in all three tasks, indicating that this module has a positive effect on modeling the local geometric structure of point clouds. Furthermore, we separately introduced the CD-KDA module for evaluation. We observed stable performance improvements, indicating that the cross-domain *key* feature alignment mechanism has a certain effect in alleviating attention bias and improving generalization ability. Finally, when both modules are introduced simultaneously, the model achieves better performance on all transfer tasks, indicating that AGT-Shift and CD-KDA are complementary, jointly promoting the overall performance.

Effects of Different Shifting Strategies. To evaluate the effectiveness of our proposed AGT-Shift module, we compare it with three different token shifting strategies: (1) KNN-Random Replacement (KNN-RandOne): For each point, its K nearest neighbors are first identified using KNN search. Then, one neighbor is randomly selected to replace the original point’s feature. (2) KNN-Mean Aggregation (KNN-Avg): After obtaining the K nearest neighbors, the output feature is computed as the average of all neighbor features, replacing the original. (3) KNN-Weighted Aggregation (KNN-WAvg): Different from KNN-Avg, a soft weighting scheme is applied based on spatial distance, where closer neighbors contribute more. Table 3 shows that the performances of these strategies are 74.54%, 74.25%, 74.82% for KNN-RandOne, KNN-Avg, KNN-WAvg, respectively. Notably, KNN-RandOne performs competitively despite its simplicity, suggesting that introducing randomness can help alleviate overfitting to local patterns. However, all three variants suffer from quadratic computational complexity due to KNN. In contrast, our AGT-Shift achieves better performance while maintaining linear complexity and avoiding

Setting	M,S*→S	M,S→S*	S,S*→M	Avg.
None	83.39	51.10	86.21	73.57
Only v	83.67	51.89	86.68	74.08
k and v	84.63	54.07	88.34	75.68
Only k (Ours)	84.39	54.10	88.49	75.66

Table 4: Ablation study comparing different alignment settings for key (k) and value (v) in the CD-KDA module.

Scale	M,S*→S	M,S→S*	S,S*→M	Avg.
Ours-Base	83.99	53.83	87.62	75.15
Ours-Standard	84.39	54.10	88.49	75.66
Ours-Large	84.63	54.61	89.14	76.13

Table 5: Generalization results of PointDGRWKV across varying network scales.

pairwise distance computation, highlighting its efficiency and robustness in large-scale domain generalization tasks.

Impact of Key and Value Alignment in CD-KDA. To further investigate the roles of different components in the attention mechanism, we conduct an ablation study isolating the effects of the *key* (k) and *value* (v) features in our proposed CD-KDA module. Specifically, we design the following variants: (1) None: no alignment is performed, serving as a baseline; (2) Only v: alignment is applied solely on the v features; (3) k and v: both k and v are aligned simultaneously; (4) Only k (Ours): alignment is applied solely on the *key* representations k, as proposed in our method. The results in Table 4 show that aligning only the value vector v brings limited performance improvement, suggesting that despite contributing to feature aggregation, v has a relatively minor influence on cross-domain generalization. In contrast, aligning both the key k and value v vectors leads to a more noticeable performance gain, indicating that promoting feature consistency does benefit generalization. Interestingly, the best performance is achieved when alignment is applied solely to k, confirming our hypothesis that the key vector, which directly influences attention weights via the exponential function, plays a more critical role in guiding spatial focus and structural understanding. Therefore, aligning k across domains significantly stabilizes the attention mechanism and enhances generalization performance.

Visualization and Analysis

T-SNE Feature Visualization. To investigate the effectiveness of each proposed module, we visualize the features of target distributions under four different configurations using t-SNE, as illustrated in Fig. 5. Specifically, (I) shows the baseline without AGT-Shift and CD-KDA, (II) removes only the AGT-Shift module, and (III) removes only the CD-KDA module, while (IV) represents our complete model. The visualization is conducted on the test set of the ShapeNet-5 (C) dataset under the PointDG-3to1 benchmark, with different colors denoting different classes. Among the first three ones, the baseline (I) exhibits the lowest intra-class compactness,

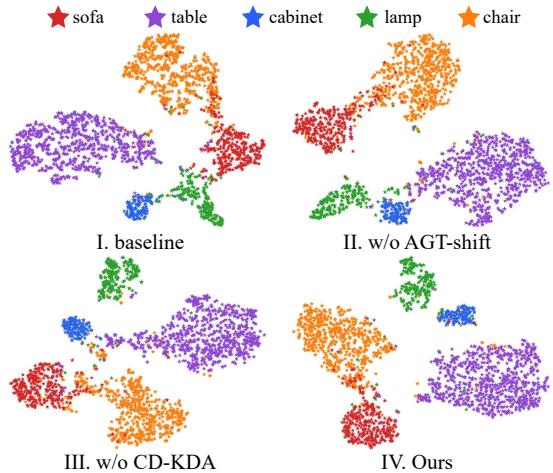


Figure 5: T-SNE visualization results of target domain feature distribution of PointDGRWKV under different module configurations. Different colors represent different classes.

indicating limited discriminability. Removing either AGT-Shift (II) or CD-KDA (III) leads to moderate improvements, but both still show less compact clusters and more ambiguous boundaries compared to the full model. Notably, these differences are especially clear in categories such as cabinet (blue) and table (purple). The complete model (IV), in contrast, achieves the most compact intra-class distributions and the clearest inter-class boundaries, demonstrating superior feature separability and confirming the essential roles of AGT-Shift and CD-KDA in DG.

Effect of Model Scale. To examine the impact of model capacity on generalization, we design three variants of our PointDGMamba: Ours-Base, Ours-Standard, and Ours-Large. As summarized in Table 5, the Standard version follows the default configuration used in our main experiments. Ours-Base reduces the number of network blocks by half, resulting in a shallower architecture. In contrast, Ours-Large increases the feature dimension and performs denser point sampling during processing. Among these, Ours-Large achieves the best average accuracy, indicating that increased representational power and finer geometric granularity benefit generalization. Meanwhile, the Base model still performs competitively to state-of-the-art methods, indicating the proposed method is effective even at lower computational cost.

Conclusion

We propose PointDGRWKV, the first RWKV-based framework for DG PCC. It enhances spatial perception and generalization while retaining RWKV’s efficient sequence modeling and linear complexity. To address RWKV’s limitations on 3D data, we introduce AGT-Shift for improved local geometric modeling and CD-KDA to reduce attention drift by aligning key distributions across domains. Extensive experiments on PointDA-10 and PointDG-3to1 benchmarks confirm that our method achieves state-of-the-art performance with a strong balance between efficiency and robustness.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62502178, 62472282, 72192821) and the Fundamental Research Funds for the Central Universities (No. YG2023QNA35).

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.
- Billinghurst, M.; Clark, A.; Lee, G.; et al. 2015. A survey of augmented reality. *Foundations and Trends® in Human-Computer Interaction*, 8(2-3): 73–272.
- Caesar, H.; Bankiti, V.; Lang, A. H.; Vora, S.; Liong, V. E.; Xu, Q.; Krishnan, A.; Pan, Y.; Baldan, G.; and Beijbom, O. 2020. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11621–11631.
- Chen, T.; Zhou, X.; Tan, Z.; Wu, Y.; Wang, Z.; Ye, Z.; Gong, T.; Chu, Q.; Yu, N.; and Lu, L. 2025. Zig-rir: Zigzag rwkv-in-rwkv for efficient medical image segmentation. *IEEE Transactions on Medical Imaging*.
- Dai, A.; Chang, A. X.; Savva, M.; Halber, M.; Funkhouser, T.; and Nießner, M. 2017. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5828–5839.
- 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*.
- Duan, Y.; Wang, W.; Chen, Z.; Zhu, X.; Lu, L.; Lu, T.; Qiao, Y.; Li, H.; Dai, J.; and Wang, W. 2024. Vision-rwkv: Efficient and scalable visual perception with rwkv-like architectures. *arXiv preprint arXiv:2403.02308*.
- 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.
- 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.
- He, Q.; Zhang, J.; Peng, J.; He, H.; Li, X.; Wang, Y.; and Wang, C. 2025. Pointrwkv: Efficient rwkv-like model for hierarchical point cloud learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 3410–3418.
- 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.
- 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.
- Lechner, 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. Pointtencn: Convolution on x-transformed points. *Advances in Neural Information Processing Systems*, 31.
- 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. 2024. Cloudmix: Dual mixup consistency for unpaired point cloud completion. *IEEE Transactions on Visualization and Computer Graphics*.
- Long, S.; Zhou, Q.; Ying, C.; Ma, L.; and Luo, Y. 2025. Diverse target and contribution scheduling for domain generalization. *IEEE Transactions on Image Processing*.
- Loshchilov, I.; and Hutter, F. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.
- Peng, B.; Alcaide, E.; Anthony, Q.; Albalak, A.; Arcadinho, S.; Biderman, S.; Cao, H.; Cheng, X.; Chung, M.; Grella, M.; et al. 2023. Rwkv: Reinventing rnns for the transformer era. *arXiv preprint arXiv:2305.13048*.
- 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.

Thuruthel, T. G.; Shih, B.; Laschi, C.; and Tolley, M. T. 2019. Soft robot perception using embedded soft sensors and recurrent neural networks. *Science Robotics*, 4(26): eaav1488.

Wang, H.; He, Q.; Peng, J.; Yang, H.; Chi, M.; and Wang, Y. 2025. Mamba-yolo-world: marrying yolo-world with mamba for open-vocabulary detection. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.

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.

Wu, Z.; Song, S.; Khosla, A.; Yu, F.; Zhang, L.; Tang, X.; and Xiao, J. 2015. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1912–1920.

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.

Yang, H.; Zhou, Q.; Sun, H.; Li, X.; Liu, F.; Lu, X.; Ma, L.; and Yan, S. 2025. Pointdgmamba: Domain generalization of point cloud classification via generalized state space model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 9193–9201.

Yin, Z.; Li, C.; and Dong, X. 2024. Video rwkv: Video action recognition based rwkv. *arXiv preprint arXiv:2411.05636*.

Yuan, H.; Li, X.; Qi, L.; Zhang, T.; Yang, M.-H.; Yan, S.; and Loy, C. C. 2024. Mamba or rwkv: Exploring high-quality and high-efficiency segment anything model. *arXiv preprint arXiv:2406.19369*.

Zhang, T.; Yuan, H.; Qi, L.; Zhang, J.; Zhou, Q.; Ji, S.; Yan, S.; and Li, X. 2025. Point cloud mamba: Point cloud learning via state space model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 10121–10130.

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.

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.