Shaping Without Tearing: Controllable Diffeomorphic Deformations for Topology-Preserving 3D Point Cloud Augmentation

Jian Bi, Qianliang Wu, Jianjun Qian, Lei Luo*, Jian Yang*

PCA Lab, Key Lab of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education
School of Computer Science and Engineering, Nanjing University of Science and Technology, China
{jianbi, wuqianliang, csjqian, cslluo, csjyang}@njust.edu.cn

Abstract

Point cloud data augmentation is critical to improving the generalization of 3D deep learning models. However, existing methods often fail to preserve the underlying manifold structure, leading to semantic distortion or topology violation. This causes models to learn untrustworthy features, thereby limiting the representational ability of the model. To overcome these limitations, we propose ManiPoint, a novel point cloud augmentation framework based on diffeomorphism that explicitly preserves manifold structure during deformation. ManiPoint constructs diffeomorphic transformations via continuous differentiable mappings, ensuring topological consistency and geometric continuity between original and augmented data. To prevent excessive distortion and ensure semantic consistency, we introduce a controllable deformation mechanism that quantitatively constrains the augmentation magnitude and enables fine-grained control over the deformation space. We further provide theoretical analysis, indicating that, compared with topologically inconsistent methods, ManiPoint reduces empirical and vicinal risks by generating diverse and structurally reliable samples. Extensive experiments and visualizations on object-level datasets demonstrate that ManiPoint produces high-quality augmentations and consistently improves model robustness over existing baselines. Meanwhile, the scalability of our method was further verified on the scene-level datasets.

Introduction

The growing demand for high-quality data in autonomous driving (Chen et al. 2017; Hu et al. 2025), robotics (Navarro-Serment, Mertz, and Hebert 2010; Wu et al. 2023, 2024; Lian et al. 2024, 2025), and 3D computer vision (Yang et al. 2024, 2025; An et al. 2025; Zheng et al. 2023; Wang et al. 2025; Wang, Yan, and Yang 2024) has highlighted the challenges associated with acquiring sufficient, diverse, and high-fidelity point cloud datasets. However, acquiring large-scale, high-quality point cloud data presents significant resource challenges due to sensor limitations, environmental constraints, and the inherent complexity of target objects. This research gap has motivated the development of advanced data augmentation techniques, which serve as practical solutions to data scarcity while enhancing the diversity and quality of training point cloud datasets for machine learning applications.

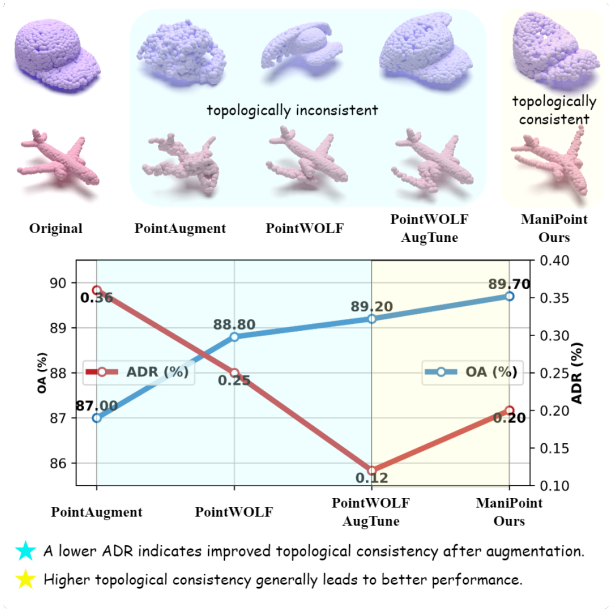

Figure 1: In topologically inconsistent augmentation methods, reducing the average deformation rate (ADR) enhances topological consistency with the original data, thereby improving the performance of the model. However, our method strictly maintains topological consistency and has controllable deformation rate, thereby outperforming all methods with topological inconsistencies in terms of performance.

Conventional data augmentation methods (Qi et al. 2017a; Wang et al. 2019), including random rotation, scaling, shifting, and noise injection, rely on simple linear transformations or random perturbations. The increased randomness in these methods makes it challenging to generate high-quality and diverse training samples. Mixing-based augmentation methods (Chen et al. 2020; Lee et al. 2022, 2021; Zhang et al. 2022a; Ren, Pan, and Liu 2022; Hong, Zhang, and Ma 2023) have shown great potential for point clouds. Although the mixing-based augmentation can improve the performance of the model in some tasks, it has significant shortcomings in interpretability and topology. The mix

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

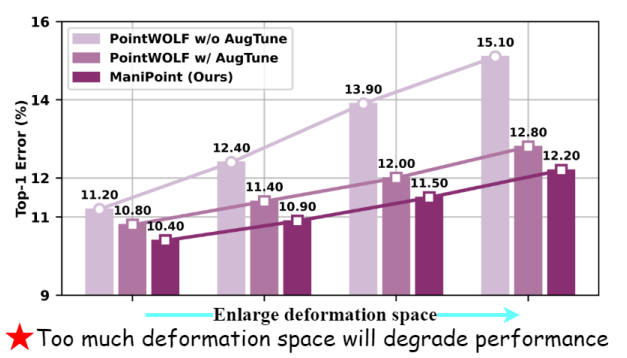

Figure 2: The larger the parameter space of deformation, the more semantically inconsistent samples will be, which will lead to the degradation of model performance.

method generates new data by mixing different samples, which makes it difficult for the model to understand and explain its decision process and reduces the interpretability of the model. In addition, the mixing-based method does not effectively consider the topological structure of the data, ignores the spatial relationship between samples, and destroys the intrinsic geometric structure of point clouds.

Other methods aim to generate different training samples through shape perturbations at different scales (Bi et al. 2025b,a). The fine-grained method PointAugment (Li et al. 2020) employs neural networks to generate augmented samples. However, the extensive search space makes it difficult to ensure realistic shapes in the generated samples, thereby limiting their effectiveness. In contrast, PointWOLF (Kim et al. 2021) uses a coarse-grained approach by dividing point cloud into patches and applying transformations like rotation, translation, and scaling to each patch. Although this method allows explicit control over sample generation and reduces the parameter space of deformation, it may cause fragmentation between patches. They can lead to data distortion, semantic inconsistencies, and disruptions in geometric structure and topological consistency (Figure 1 top). However, as shown in Figure 1 (bottom), when PointWOLF is fine-tuned, the topological consistency is improved and the accuracy of the model is further increased. Moreover, we can see from Figure 2, the larger the parameter space of deformation, the more semantically inconsistent samples will be, which will also lead to the degradation of model performance. Although these methods demonstrate some improvement in model performance, they are often overly engineered, lack explainability, and fail to ensure that the generated sample maintains a manifold structure with the original sample. These limitations restrict their applicability in point cloud understanding (Bi et al. 2025c) tasks.

To address these challenges, we propose ManiPoint, a novel point cloud augmentation framework based on diffeomorphism that explicitly preserves manifold structure during deformation. The core idea involves applying explicit, continuous, and differentiable functions to perform nonlinear deformations on point clouds, effectively maintaining their intrinsic geometric continuity and topological consistency during augmentation. Unlike other methods, our ManiPoint leverages manifold learning theory to ensure local diffeomorphism and topological consistency. Additionally, to prevent excessive distortion and ensure semantic consistency, we introduce a controllable deformation mechanism that quantitatively constrains the augmentation magnitude and enables fine-grained control over the deformation space. Moreover, we provide theoretical analysis, indicating that, compared with topologically inconsistent methods, ManiPoint reduces empirical and vicinal risks by generating diverse and structurally reliable samples. Finally, extensive experiments and visualizations on object-level datasets demonstrate that ManiPoint produces high-quality augmentations and consistently improves model robustness over existing baselines. Meanwhile, the scalability of our method was further verified on the scene-level datasets.

Our main contributions are summarized as follows:

- We propose the first explicit data augmentation framework for point cloud based on diffeomorphism. It guarantees topological consistency and enhances both the diversity and geometric authenticity of generated samples.
- We design a deformation rate to control the intensity of geometric deformations, thereby narrowing the deformation space. This prevents excessive distortion and unnatural augmentation, ensuring high-quality augmented data.
- Theoretical analysis demonstrates that ManiPoint attains lower empirical risk than topologically inconsistent augmentations and generates samples distributed around the true data distribution, minimizing vicinal risk. Extensive experiments have shown that ManiPoint improves the generalization and robustness of various models.

Method

Topological inconsistency can disrupt the underlying manifold structure of point clouds, while an unconstrained deformation space may lead to semantic inconsistency. To address these issues, we propose a novel plug-and-play point cloud augmentation framework that enables fine-grained shape perturbations while preserving topological consistency. First, we establish theoretical foundations through the definition of diffeomorphism and **Theorem 1**. Second, to overcome topological inconsistencies, we introduce a diffeomorphic deformation framework that generates smoothly deformed samples while maintaining topological consistency. Third, to address the issue of excessive deformation space leading to semantic inconsistency, we propose a controllable deformation mechanism that constrains the magnitude of transformations. By adjusting the average deformation rate, the deformation space can be reduced and the authenticity of generated samples can be increased. Finally, we provide theoretical proofs of both necessity and analyze the effectiveness from the perspective of empirical risk and vicinal risk minimization.

Preliminaries

A manifold is a low-dimensional space embedded within a higher-dimensional space. One important concept related to

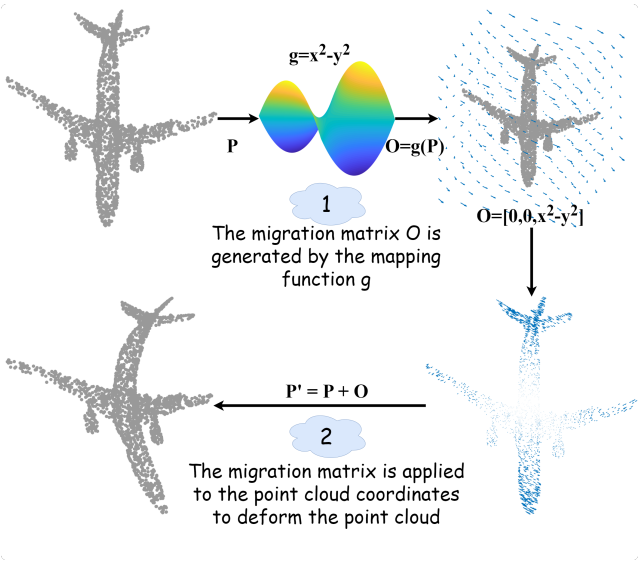

Figure 3: An overview of ManiPoint. The deformation matrix is generated by a continuous and differentiable function to make the point cloud produce smooth deformation.

manifolds is diffeomorphism, which describes a bijective relationship between two manifolds through a smooth, invertible mapping. A diffeomorphism is defined as follows:

Definition 1. (Diffeomorphism) Let $\mathcal{M}$ and $\mathcal{N}$ be two smooth manifolds, and $f : \mathcal{M} \to \mathcal{N}$ is bijective, and f and f^{-1} are both smooth mappings, then f is a Diffeomorphism, denoted as $f \in Diff(\mathcal{M}, \mathcal{N})$.

The key property of a diffeomorphism is that it establishes a smooth and invertible correspondence between two manifolds, thereby maintaining the continuity of their local geometric structures and ensuring global topological equivalence. Specifically, if a mapping $f : \mathcal{M} \to \mathcal{N}$ is a diffeomorphism, then for any point $p \in \mathcal{M}$, the local neighborhood around p is smoothly mapped to a corresponding neighborhood in $\mathcal{N}$ without tearing or folding. This guarantees that not only the local manifold structure is retained, but also that the manifolds share the same topological type.

In the context of data augmentation, a particularly important subclass of diffeomorphisms is volume-preserving, which maintain not only the differential and topological structure, but also the volume measure induced by the manifold. This motivates the following theorem:

Theorem 1. (Volume-Preserving of Diffeomorphism) Let $\mathcal{M}$ and $\mathcal{N}$ be oriented smooth manifolds equipped with volume forms $\omega_{\mathcal{M}}$ and $\omega_{\mathcal{N}}$ respectively. Suppose $f : \mathcal{M} \to \mathcal{N}$ is a diffeomorphism such that the Jacobian determinant of f satisfies $det(J_f(p)) = 1, \forall p \in \mathcal{M}$. Then f is a *volume-preserving diffeomorphism*, *i.e.*, for any measurable subset $U \subset \mathcal{M}$, $\mathrm{Vol}_{\mathcal{M}}(U) = \mathrm{Vol}_{\mathcal{N}}(f(U))$.

This theorem shows that a Volume-Preserving diffeomorphism preserve not only the topological structure but also volume-related properties. This is critical for point cloud augmentation, as it ensures smoothness, continuity, and density of local regions, enabling the generation of diverse yet semantically faithful samples aligned with the underlying manifold. Thus, **Theorem 1** provides a solid theoretical basis for designing geometrically-aware deformations that produce structurally reliable augmented data.

Mathematical Framework of ManiPoint

The point cloud is usually located on the low-dimensional manifold $\mathcal{M} \subset \mathbb{R}^3$, and its geometry is described by the local geometry and the global topology. By treating point cloud as manifolds, we can explicitly construct diffeomorphism to design manifold-preserving deformation augmentation methods. We are given a general framework:

Given that the original data is located on the manifold $\mathcal{M} \subset \mathbb{R}^n$, the mapping $f : \mathbb{R}^n \to \mathbb{R}^n$ as follows:

$$f^i(x^i) = x^i + \alpha_i g^i(x^1, x^2, ..., x^{i-1}, x^{i+1}, ..., x^n), \quad (1)$$

where $\sum_{i=1}^{n} \alpha_i = 1$ and $\alpha_i \in \{0, 1\}$. $g^i \in C^\infty(\mathbb{R}^{n-1})$. The augmented data is located in $\mathcal{N} = f(\mathcal{M})$, and the Jacobian determinant $det(J_f) = 1$.

Theorem 2. If $g \in C^\infty(\mathbb{R}^{n-1})$ is a continuously differentiable function, then f is a diffeomorphism. Moreover, the augmented point cloud $f(\mathcal{M})$ still lies in the smooth manifold $\mathcal{N} \subset \mathbb{R}^{n-1}$.

Theorem 2 establishes a foundational framework for our diffeomorphism-based augmentation paradigm. We construct point cloud deformations through parameterized function spaces f that satisfy the conditions of diffeomorphisms, ensuring C^∞-smooth transformations with invertible Jacobian determinants. As shown in Figure 3, this mathematical guarantee enables us to obtain continuous residual representations through the composition of diffeomorphisms, which are then used to generate augmented coordinates via coordinate-wise addition.

In 3D point cloud tasks, we only consider three-dimensional space, and we assume only one dimension z is deformed to study point cloud deformation. Specifically, for a given point cloud $P = [x, y, z]$, we only need to compute the residual coordinates $O(P)$ through the diffeomorphism f. Thus, the augmentation process becomes:

$$P' = f(P), \quad (2)$$
$$P' = [x, y, z + g(x, y)], \quad (3)$$
$$P' = [x, y, z] + [0, 0, g(x, y)], \quad (4)$$
$$P' = P + O(P). \quad (5)$$

We just need to design different deformation functions g. g can be a polynomial function, a multimodal function, a combinatorial function, etc. As shown in Figure 4, with different functions, we can generate a continuous and realistic variety of shapes. However, as can be seen at the bottom of Figure 1, too small or too large a deformation will reduce the performance gain of the augmented sample to the model. Therefore, we propose the deformation rate to further constrain the feature space neighborhood of the point clouds.

Controllable Deformation Rate

Deformation Rate (DR) is introduced to quantify the local geometric changes of the point cloud during the deformation process. As shown in Figure 5, this ensures the loss of geometric meaning due to excessive deformation.

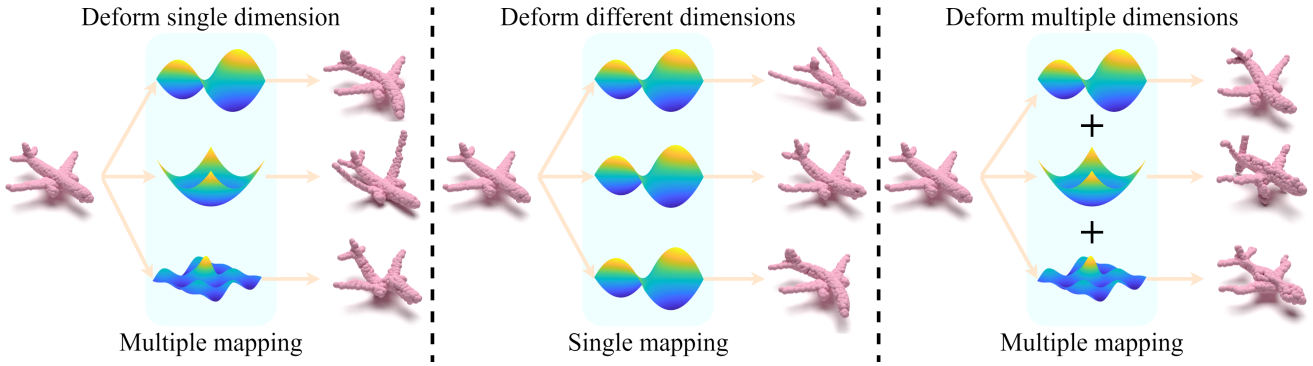

Figure 4: Our method can directly use different functions to locally deform point cloud at a fine-grained level. By using different continuously differentiable functions and applying different combination functions, different point clouds can be generated. This explicit deformation is controllable and reduces the parameter space of deformation. The specific function f is in the supplementary material. For instance, f can be defined as $x^2 - y^2$, $x^2 + y^2$, or $0.2\sin(4x)\cos(4y)$.

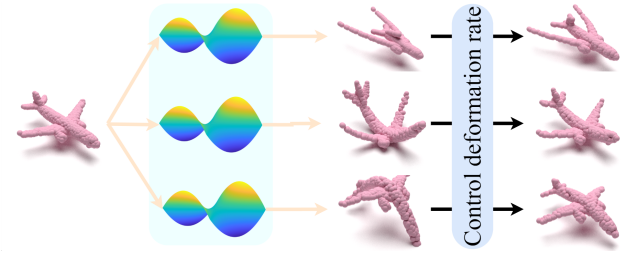

Figure 5: We use a controllable deformation rate to further constrain the deformation space.

Definition 2. Deformation Rate is measured by the ratio of the amount of displacement of each point to the distance from the original point to the D. Given a point cloud $P \in \mathbb{R}^{N\times 3}$ and a displacement $D \in \mathbb{R}^{N\times 3}$, then the point cloud deformation rate $R_i = \frac{\|D_i\|}{\|P_i\|}, i = 1, 2, ..., N$.

Definition 3. Average Deformation Rate is the average of the rate of deformation at all points. The deformation rate of the point cloud is $R_i = \frac{\|D_i\|}{\|P_i\|}, i = 1, 2, ..., N$. Therefore, the average deformation rate is $R_{avg} = \frac{1}{N}\sum_{i=1}^{N} R_i$.

Scaled Deformation Rate. We can control the degree of deformation by constraining the average deformation rate, thereby preventing the loss of geometric information due to excessive deformation.

To ensure the average deformation rate $R_{avg} \leq rate$, we need to scale the displacement D. Set a scaling factor $\lambda > 0$, such that the scaled displacement $D' = \lambda D$ meets:

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\parallel \lambda D_i \parallel}{\parallel P_i \parallel} \leq rate. \tag{6}$$

where $\|\cdot\|$ indicates the $l^2 - norm$. The deformation after scaling is λD_i, then $\parallel \lambda D_i \parallel = \lambda \parallel D_i \parallel$. After plugging in, we get the scaling factor λ:

$$\lambda \leq \frac{rate}{\frac{1}{N}\sum_{i=1}^{N}\frac{\|D_i\|}{\|P_i\|}}. \tag{7}$$

Finally, the augmented point cloud P' is:

$$P' = P + \lambda D. \tag{8}$$

From Empirical Risk Minimization to ManiPoint. In machine learning, the goal of the model is to minimize expected risk (Vapnik 1999):

$$R(f) = \int \mathcal{L}(f(x), y)\mathrm{d}P_{data}(x, y), \tag{9}$$

where P_{data} is the real data distribution, $\mathcal{L}$ is the loss function, and f is the model. Since the true distribution P_{data} is unknown, empirical risk is often used to approximate:

$$\hat{R}(f) = \frac{1}{N}\sum_{i=1}^{N}\mathcal{L}(f(x_i), y_i), \tag{10}$$

where $\{(x_i, y_i)\}_{i=1}^{N}$ is the training dataset.

$$\lim_{N\to\infty}\hat{R}(f) = R(f). \tag{11}$$

Data augmentation expands the coverage of empirical risks by generating diverse training samples. The increased data enables the model to access more transformation forms during training, thus reducing the overfitting of the model to the original training set. Especially when the original data set is small or noisy (Zhong et al. 2020), additional deformation samples help to improve the robustness of the model, so that the model does not just remember the details of the training data, but can extract more general features.

Because the diffeomorphism preserves the manifold structure of the data, the augmented sample and the original sample have similar structures in the local region, which allows the model to learn transformation-invariant features rather than memorizing rigid local appearances (Szegedy et al. 2014). The diffeomorphism T preserves the manifold structure, ensuring that the augmented data x' remains on the original manifold M. Therefore, the augmented empirical risk $\hat{R}_{aug}(f)$ is still a valid estimate of the expected risk $R(f)$. Finally, we summarize **Corollary 1** as follows:

Corollary 1. (Topologically preserving reduces empirical risk) For a diffeomorphism f, $det(J_f) = 1$, the empirical

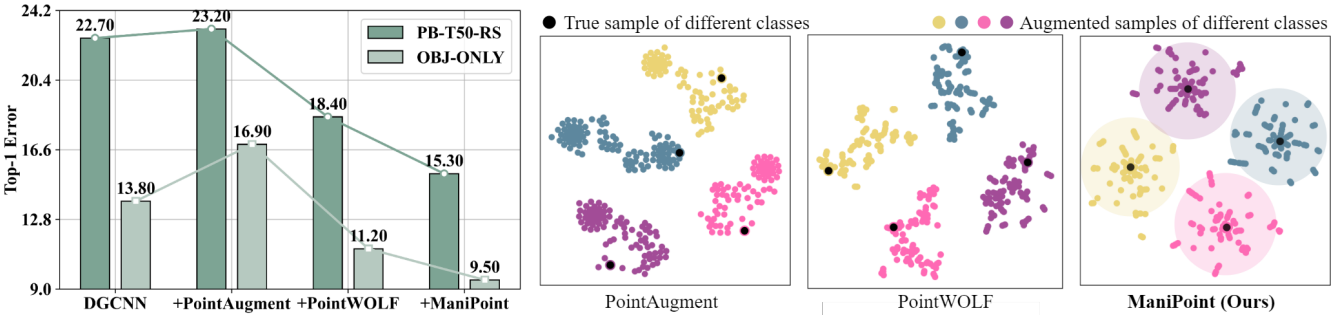

Figure 6: T-SNE visualization. Samples generated by our method cluster around the real samples, while those from PointAugment and PointWOLF exhibit significant deviation, often lying far from the true distribution and causing semantic distortion.

Method	Diff.	Non-Diff.
Geometric smooth	✓	✗
Topological consistency	✓	✗
Data distribution	debiased	biased
Sample diversity	high	redundant, low
Theoretical support	✓	✗

Table 1: The advantages of our diffeomorphism-based method. Diff. is diffeomorphism.

risk R_{emp} of the model trained with the generated data is lower than that of the model trained with transformation g that does not guarantee topological consistency, *i.e.*,

$$\hat{R}_{aug}(f) < \hat{R}_{aug}(g). \tag{12}$$

From Vicinal Risk Minimization to ManiPoint. Vicinal risk (Chapelle et al. 2000) is a powerful approach for improving generalization by focusing not only on the individual data points but also on their neighborhoods. The neighborhood $\mathbb{N}(x, \epsilon)$ can be defined as the set of data points less than ϵ from x at metric d, that is, $\mathbb{N}(x, \epsilon) = \{x' : d(x, x') < \epsilon\}$. The formula for vicinal risk is:

$$\hat{R}_v(f, \epsilon) = \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(f(x_i'), y_i), x_i' \sim P(\mathbb{N}(x_i, \epsilon)), \tag{13}$$

where $P(\mathbb{N}(x_i, \epsilon))$ represents the probability distribution over the neighborhood.

Our contribution is to propose a neighborhood distribution based on diffeomorphism and constrain the range of the neighborhood by deformation rate. The diffeomorphic deformation T can be viewed as a kind of proximity distribution $V(x_i, y_i)$, generating a virtual sample (x_i', y_i), where $x_i' = T(x_i)$. Therefore, the new vicinal risk is as follows:

$$\hat{R}_v(f, \epsilon) = \frac{1}{N} \sum_{i=1}^{n} \mathcal{L}(f(T(x_i)), y_i). \tag{14}$$

We provide an comparison of the diffeomorphism-based method with other methods in Table 1. Diffeomorphism generates augmented samples through a topological preservation, which increases diversity while preserving the geometric features of the original data in local areas. As shown in Figure 6, we use t-SNE to visualize the distribution of these augmented samples. Our method can be better distributed around the real samples and effectively approximate the real data distribution. However, PointAugment (Li et al. 2020) and PointWOLF (Kim et al. 2021) failed to maintain topological consistency, resulting in a deviation of the sample distribution. Thus, our approach can improve the model's adaptability to data changes by minimizing vicinal risk.

Experiments

In this section, we demonstrate the effectiveness of our ManiPoint, with various datasets and baselines. First, we evaluate the generalization performance and robustness in shape classification. Next, we extend ManiPoint to object-level and scene-level segmentation tasks to verify scalability. **More ablation studies and implementation details are provided in the supplementary material.**

Point Cloud Classification

Comparisons with State-Of-The-Art Point Cloud Augmentation Methods. We compared ManiPoint against existing point cloud augmentation methods (Li et al. 2020; Chen et al. 2020; Sheshappanavar, Singh, and Kambhamettu 2021; Kim et al. 2021; Lee et al. 2021; Ren, Pan, and Liu 2022; Lee et al. 2022; Zhang et al. 2022a; Hong, Zhang, and Ma 2023; Wang et al. 2024). We report the overall accuracy (OA) of each model. As shown in Table 2, ManiPoint achieves SOTA performance on both PointNet++ (Qi et al. 2017b) and DGCNN (Wang et al. 2019). Notably, it achieves a maximum OA improvement of 2.8% on the synthetic dataset ModelNet40 (Wu et al. 2015) and 7.4% on the real-world dataset PB-T50-RS (Uy et al. 2019). These significant improvements further validate that manifold-preserving shape perturbations can deliver gains for baselines.

Shape Classification on ScanObjectNN. We verify the performance of our method under supervised learning (Qi et al. 2017a,b; Xu et al. 2018; Li et al. 2018; Qiu, Anwar, and Barnes 2021; Cheng et al. 2021; Zhao et al. 2021) and self-supervised learning (Yu et al. 2022) frameworks, respectively. We use multiple robust classification models as our baseline, including DGCNN (Wang et al. 2019), PointMLP (Ma et al. 2022), DeLA (Chen et al. 2023), PointMAE (Pang et al. 2022), and PointM2AE (Zhang et al.

Method	Synthetic ModelNet40	Real-world OBJ-ONLY	Real-world PB-T50-RS
PointNet++	90.7	84.3	79.4
+PointMixup	92.3	88.5	80.6
+RSMix	91.6	-	-
+SageMix	93.3	88.7	83.7
+PointCutMix	93.4	-	-
+PointAugment	92.9	85.4	77.9
+PatchAugment	92.4	87.1	81.0
+PointWOLF	93.2	89.7	84.1
+WOLFMix	93.1	-	-
+PCSalMix	93.1	-	-
+PointPatchMix	92.9	-	-
+ManiPoint(Ours)	**93.5 (↑ 2.8)**	**90.3 (↑ 6.0)**	**84.6 (↑ 5.2)**
DGCNN	92.2	86.2	77.3
+PointMixup	92.9	-	-
+RSMix	93.5	-	-
+SageMix	93.6	88.0	83.6
+PointCutMix	93.2	-	-
+PointAugment	93.4	83.1	76.8
+PatchAugment	93.1	86.9	79.1
+PointWOLF	93.2	88.8	81.6
+WOLFMix	93.2	-	-
+PCSalMix	93.2	-	-
+PointPatchMix	93.3	-	-
+ManiPoint(Ours)	**94.0 (↑ 1.8)**	**90.5 (↑ 4.3)**	**84.7 (↑ 7.4)**

Table 2: Comparison with other point cloud augmentation methods. We report the classification accuracy. ↑ indicates the performance improvement relative to the baseline.

Methods	OBJ-BG	OBJ-ONLY	PB-T50-RS	#Params
	Top-1 Error			
Supervised methods				
PointNet	26.7	20.8	32	3.4M
SpiderCNN	22.9	20.5	26.3	-
PointNet++	17.7	15.7	22.1	1.5M
PointCNN	13.9	14.5	21.5	-
GBNet	-	-	19.5	8.4M
PRANet	-	-	19	-
PointTrans.	20.2	19.5	11.8	-
DGCNN	17.2	13.8	21.9	1.82M
+ManiPoint	**14.1**	**9.5**	**15.3**	**1.82M**
PointMLP	10.7	12.1	14.8	12.6M
+ManiPoint	**9.0**	**9.2**	**12.5**	**12.6M**
DeLA	-	-	9.6	5.3M
+ManiPoint	-	-	**9.3**	**5.3M**
Self-supervised methods				
Transformer	20.1	19.4	22.8	-
+OcCo	15.2	14.5	21.2	-
Point-BERT	12.6	11.9	17.0	22.1M
Point-MAE	10.0	11.8	14.9	22.1M
+ManiPoint	**9.3**	**9.7**	**13.5**	**22.1M**
Point-M2AE	8.8	11.2	13.6	12.9M
+ManiPoint	**8.6**	**8.9**	**13.1**	**12.9M**

Table 3: Comparison with state-of-the-art methods, including supervised and self-supervised methods. We report the Top-1 Error (%) on three splits of ScanObjectNN.

2022b). As demonstrated in Table 3, our experimental results on ScanObjectNN reveal consistent performance improvements and outperform all compared advanced models. Our method only enhances the diversity of training samples without introducing additional parameters and achieves significant performance improvements across both supervised and self-supervised learning frameworks.

Few-shot Classification. We conducted few-shot classification (Wang et al. 2019, 2021; Afham et al. 2022; Yu et al. 2022) under the transformer framework on ModelNet40 (Wu et al. 2015) to evaluate the performance of ManiPoint with limited fine-tuning data. Transformers are sensitive to local structures, and they can study the impact of local diversity on the model. Table 4 shows that ManiPoint outperforms all baselines (Pang et al. 2022; Zhang et al. 2022b) in every few-shot setting, demonstrating its effectiveness in enhancing local geometric diversity through limited training data and providing valuable discriminative features.

Method	5-way 10-shot	5-way 20-shot	10-way 10-shot	10-way 20-shot
DGCNN	91.8 ±3.7	93.4 ±3.2	86.3 ±6.2	90.9 ±5.1
+OcCo	91.9 ±3.3	93.9 ±3.1	86.4 ±5.4	91.3 ±4.6
+CrossPoint	92.5 ±3.0	94.9 ±2.1	83.6 ±5.3	87.9 ±4.2
Transformer	87.8 ±5.2	93.3 ±4.3	84.6 ±5.5	89.4 ±6.3
+OcCo	94.0 ±3.6	95.9 ±2.3	89.4 ±5.1	92.4 ±4.6
Point-BERT	94.6 ±3.1	96.3 ±2.7	91.0 ±5.4	92.7 ±5.1
Point-MAE	96.3 ±2.5	97.8 ±1.8	92.6 ±4.1	95.0 ±3.0
+ManiPoint	**96.8 ±2.3**	**97.9 ±1.7**	**92.8 ±3.5**	**95.2 ±2.7**
Point-M2AE	96.8 ±1.8	98.3 ±1.4	92.3 ±4.5	95.0 ±3.0
+ManiPoint	**96.9 ±2.0**	**98.5 ±1.5**	**92.5 ±3.8**	**95.2 ±2.5**

Table 4: Few-shot classification on ModelNet40 (Wu et al. 2015). We report the average accuracy (%) and standard deviation (%) of 10 independent experiments.

Scalability in Point Cloud Segmentation

Object-level Part Segmentation. As shown in Table 5, we evaluate ManiPoint on the canonical 3D shape part segmentation task on ShapeNetPart (Yi et al. 2016). Our results reveal consistent superiority over baseline architectures (Wang et al. 2019; Ma et al. 2022; Chen et al. 2023), demonstrating that local geometric perturbations remain critical for low-level feature learning in 3D segmentation. In addition, we give the results of a comparison with the most advanced methods in the supplemental materials, where our ManiPoint consistently outperforms other models.

Scene-level Scene Segmentation. To explore the generalizability of our method beyond specialized domains, we conduct experiments on scene segmentation. We decompose training scenes into constituent objects, apply spatially localized diffeomorphic transformations to individual entities, and reconstruct augmented composite inputs. As shown in Table 5, our approach achieved a significant performance improvement over the baseline (Choy, Gwak, and Savarese 2019; Tang et al. 2020) on SemanticKITTI (Behley et al.

	Method	mIoU
ShapeNetPart	DGCNN	84.8
	+ManiPoint(Ours)	**85.6**
	PointMLP	85.8
	+ManiPoint(Ours)	**86.1**
	DeLA	87.0
	+ManiPoint(Ours)	**87.4**
SemanticKITTI	MinkNet	55.9
	+ManiPoint(Ours)	**68.1**
	SPVCNN	58.0
	+ManiPoint(Ours)	**68.8**
nuScenes	MinkNet	67.1
	+ManiPoint(Ours)	**73.5**
	SPVCNN	68.4
	+ManiPoint(Ours)	**73.8**

Table 5: Reported results on enhanced model performance by our ManiPoint on 3D Point cloud segmentation.

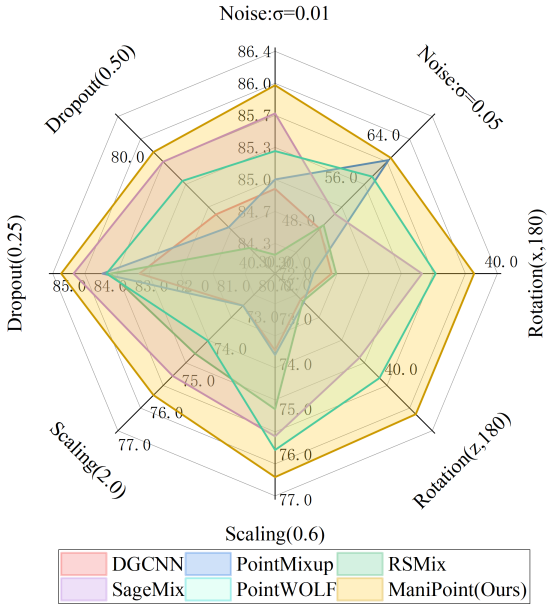

Figure 7: Our method is more stable and robust.

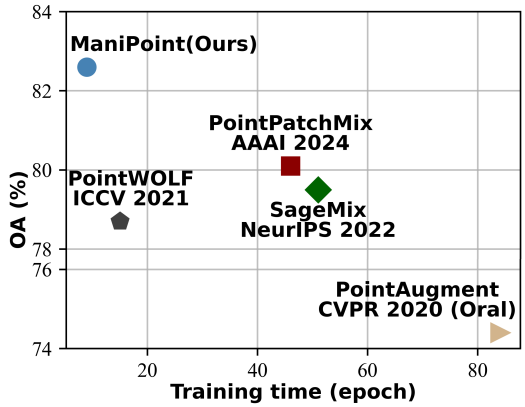

Figure 8: Since our method is a direct deform of the point clouds, our time complexity is only O(1). And we use the least amount of training time to get the best performance.

DGCNN	Diff.	ADR	Function		OA
			Single	Multiple	
Baseline					85.829
+PointWOLF	✗	✗			87.653
+PointWOLF	✗	✓			88.796
+ManiPoint	✗	✓	✓	✗	88.812
+ManiPoint	✓	✗	✓	✗	88.643
+ManiPoint	✓	✓	✓	✗	90.478
+ManiPoint	✓	✓	✗	✓	89.673

Table 6: The advantages of the diffeomorphism-based method. Diff. is diffeomorphism.

2019) and nuScenes (Caesar et al. 2020). More experimental results are in the supplemental materials.

Ablation Studies

Robustness. We evaluate the robustness of ManiPoint using DGCNN on OBJ-ONLY. As shown in Figure 7, we conducted robustness tests under eight scenarios across four types of corruption. Compared with state-of-the-art methods, DGCNN augmented with ManiPoint demonstrates superior robustness. We believe that the topologically preserved diverse samples generated by ManiPoint enable models to learn more robust features against corruption.

Training Efficiency. As shown in Figure 8, we compare other methods by convergence accuracy and single training time. Notably, ManiPoint achieved SOTA performance while significantly reducing training time.

Ablation of Component. As shown in Table 6, PointWOLF improves topological consistency by controlling the deformation rate, leading to better model performance. Moreover, performance drops significantly when non-differentiable functions are used, highlighting the importance of maintaining topological consistency. Finally, using a single map yields the best results because excessive variations can cause ambiguity in local semantics.

Conclusion

We propose ManiPoint, a novel point cloud augmentation method based on manifold structure preservation that perturbs point cloud shapes through arbitrary continuous and differentiable functions. Simultaneously, we introduce a deformation rate metric to construct an effective feature neighborhood, thereby minimizing vicinal risk for models. Extensive experiments confirm that ManiPoint enhances the performance of various models across multiple datasets. Theoretical analysis and empirical validation establish that ManiPoint achieves superior generalization performance while maintaining visual coherence. This corroborate the critical role of diffeomorphism in point cloud augmentation.

Acknowledgments

This work was supported by the National Science Fund of China under Grant Nos. 62361166670, U24A20330 and 62276135.

References

Afham, M.; Dissanayake, I.; Dissanayake, D.; Dharmasiri, A.; Thilakarathna, K.; and Rodrigo, R. 2022. Crosspoint: Self-supervised cross-modal contrastive learning for 3d point cloud understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9902–9912.

An, X.; Zhao, L.; Gong, C.; Li, J.; and Yang, J. 2025. Pre-Training a Density-Aware Pose Transformer for Robust LiDAR-based 3D Human Pose Estimation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(2): 1755–1763.

Behley, J.; Garbade, M.; Milioto, A.; Quenzel, J.; Behnke, S.; Stachniss, C.; and Gall, J. 2019. Semantickitti: A dataset for semantic scene understanding of lidar sequences. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9297–9307.

Bi, J.; Wu, Q.; Li, X.; Chen, S.; Qian, J.; lei luo; and Yang, J. 2025a. Rethinking Point Cloud Data Augmentation: Topologically Consistent Deformation. In *Forty-second International Conference on Machine Learning*.

Bi, J.; Wu, Q.; Qian, J.; Luo, L.; and Yang, J. 2025b. Dual Manifold Regularization Steered Robust Representation Learning for Point Cloud Analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2, 1844–1852.

Bi, J.; Wu, Q.; Qian, J.; Luo, L.; and Yang, J. 2025c. Structure-Aware Spherical Density Steered Cross-Domain Learning for Effective Point Cloud Understanding. *Pattern Recognition*, 112527.

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.

Chapelle, O.; Weston, J.; Bottou, L.; and Vapnik, V. 2000. Vicinal risk minimization. *Advances in Neural Information Processing Systems*, 13.

Chen, B.; Xia, Y.; Zang, Y.; Wang, C.; and Li, J. 2023. Decoupled local aggregation for point cloud learning. *arXiv preprint arXiv:2308.16532*.

Chen, X.; Ma, H.; Wan, J.; Li, B.; and Xia, T. 2017. Multi-view 3d object detection network for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1907–1915.

Chen, Y.; Hu, V. T.; Gavves, E.; Mensink, T.; Mettes, P.; Yang, P.; and Snoek, C. G. 2020. Pointmixup: Augmentation for point clouds. In *European Conference on Computer Vision*, 330–345. Springer.

Cheng, S.; Chen, X.; He, X.; Liu, Z.; and Bai, X. 2021. Pranet: Point relation-aware network for 3d point cloud analysis. *IEEE Transactions on Image Processing*, 30: 4436–4448.

Choy, C.; Gwak, J.; and Savarese, S. 2019. 4d spatio-temporal convnets: Minkowski convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3075–3084.

Hong, T.; Zhang, Z.; and Ma, J. 2023. PCSalmix: Gradient Saliency-Based Mix Augmentation for Point Cloud Classification. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1–5. IEEE.

Hu, X.; Tai, Y.; Zhao, X.; Zhao, C.; Zhang, Z.; Li, J.; Zhong, B.; and Yang, J. 2025. Exploiting multimodal spatial-temporal patterns for video object tracking. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 4, 3581–3589.

Kim, S.; Lee, S.; Hwang, D.; Lee, J.; Hwang, S. J.; and Kim, H. J. 2021. Point cloud augmentation with weighted local transformations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 548–557.

Lee, D.; Lee, J.; Lee, J.; Lee, H.; Lee, M.; Woo, S.; and Lee, S. 2021. Regularization strategy for point cloud via rigidly mixed sample. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15900–15909.

Lee, S.; Jeon, M.; Kim, I.; Xiong, Y.; and Kim, H. J. 2022. Sagemix: Saliency-guided mixup for point clouds. *Advances in Neural Information Processing Systems*, 35: 23580–23592.

Li, R.; Li, X.; Heng, P.-A.; and Fu, C.-W. 2020. Pointaugment: an auto-augmentation framework for point cloud classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6378–6387.

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.

Lian, J.; Du, X.; Liu, J.; Hui, L.; and Yang, J. 2025. Cross-Modal Driven Object Restoration for 3D Point Cloud Backdoor Defense. *IEEE Transactions on Information Forensics and Security*, 20: 11006–11018.

Lian, J.; Wang, D.-H.; Wu, Y.; and Zhu, S. 2024. Multi-Branch Enhanced Discriminative Network for Vehicle Re-Identification. *IEEE Transactions on Intelligent Transportation Systems*, 25(2): 1263–1274.

Ma, X.; Qin, C.; You, H.; Ran, H.; and Fu, Y. 2022. Rethinking network design and local geometry in point cloud: A simple residual MLP framework. *arXiv preprint arXiv:2202.07123*.

Navarro-Serment, L. E.; Mertz, C.; and Hebert, M. 2010. Pedestrian detection and tracking using three-dimensional ladar data. *The International Journal of Robotics Research*, 29(12): 1516–1528.

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. Springer.

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/CVF 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.

Qiu, S.; Anwar, S.; and Barnes, N. 2021. Geometric back-projection network for point cloud classification. *IEEE Transactions on Multimedia*, 24: 1943–1955.

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

Sheshappanavar, S. V.; Singh, V. V.; and Kambhamettu, C. 2021. Patchaugment: Local neighborhood augmentation in point cloud classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2118–2127.

Szegedy, C.; Zaremba, W.; Sutskever, I.; Bruna, J.; Erhan, D.; Goodfellow, I.; and Fergus, R. 2014. Intriguing properties of neural networks. In *2nd International Conference on Learning Representations, ICLR 2014*.

Tang, H.; Liu, Z.; Zhao, S.; Lin, Y.; Lin, J.; Wang, H.; and Han, S. 2020. Searching efficient 3d architectures with sparse point-voxel convolution. In *European Conference on Computer Vision*, 685–702. Springer.

Uy, M. A.; Pham, Q.-H.; Hua, B.-S.; Nguyen, T.; and Yeung, S.-K. 2019. 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*, 1588–1597.

Vapnik, V. N. 1999. An overview of statistical learning theory. *IEEE transactions on neural networks*, 10(5): 988–999.

Wang, H.; Liu, Q.; Yue, X.; Lasenby, J.; and Kusner, M. J. 2021. Unsupervised point cloud pre-training via occlusion completion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9782–9792.

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.

Wang, Y.; Wang, J.; Li, J.; Zhao, Z.; Chen, G.; Liu, A.; and Heng, P. A. 2024. Pointpatchmix: Point cloud mixing with patch scoring. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 5686–5694.

Wang, Z.; Yan, Z.; Pan, J.; Gao, G.; Zhang, K.; and Yang, J. 2025. DORNet: A Degradation Oriented and Regularized Network for Blind Depth Super-Resolution. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 15813–15822.

Wang, Z.; Yan, Z.; and Yang, J. 2024. Sgnet: Structure guided network via gradient-frequency awareness for depth map super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 6, 5823–5831.

Wu, Q.; Jiang, H.; Ding, Y.; Luo, L.; Xie, J.; and Yang, J. 2023. Diff-pcr: Diffusion-based correspondence searching in doubly stochastic matrix space for point cloud registration. *arXiv preprint arXiv:2401.00436*.

Wu, Q.; Jiang, H.; Luo, L.; Li, J.; Ding, Y.; Xie, J.; and Yang, J. 2024. Diff-reg: Diffusion model in doubly stochastic matrix space for registration problem. In *European Conference on Computer Vision*, 160–178. Springer.

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/CVF Conference on Computer Vision and Pattern Recognition*, 1912–1920.

Xu, Y.; Fan, T.; Xu, M.; Zeng, L.; and Qiao, Y. 2018. Spidercnn: Deep learning on point sets with parameterized convolutional filters. In *European Conference on Computer Vision*, 87–102.

Yang, Z.; Jiang, Z.; Li, X.; Zhou, H.; Dong, J.; Zhang, H.; and Du, Y. 2024. D^4-VTON: Dynamic Semantics Disentangling for Differential Diffusion Based Virtual Try-On. In *ECCV*, 36–52. Springer.

Yang, Z.; Li, Y.; He, S.; Li, X.; Xu, Y.; Dong, J.; and Du, Y. 2025. OmniVTON: Training-Free Universal Virtual Try-On. *ICCV*.

Yi, L.; Kim, V. G.; Ceylan, D.; Shen, I.-C.; Yan, M.; Su, H.; Lu, C.; Huang, Q.; Sheffer, A.; and Guibas, L. 2016. A scalable active framework for region annotation in 3d shape collections. *ACM Transactions on Graphics (ToG)*, 35(6): 1–12.

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, J.; Chen, L.; Ouyang, B.; Liu, B.; Zhu, J.; Chen, Y.; Meng, Y.; and Wu, D. 2022a. Pointcutmix: Regularization strategy for point cloud classification. *Neurocomputing*, 505: 58–67.

Zhang, R.; Guo, Z.; Gao, P.; Fang, R.; Zhao, B.; Wang, D.; Qiao, Y.; and Li, H. 2022b. Point-m2ae: multi-scale masked autoencoders for hierarchical point cloud pre-training. *Advances in Neural Information Processing Systems*, 35: 27061–27074.

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.

Zheng, Y.; Zhan, J.; He, S.; Dong, J.; and Du, Y. 2023. Curricular contrastive regularization for physics-aware single image dehazing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5785–5794.

Zhong, Z.; Zheng, L.; Kang, G.; Li, S.; and Yang, Y. 2020. Random erasing data augmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 07, 13001–13008.