

# Certified L2-Norm Robustness of 3D Point Cloud Recognition in the Frequency Domain

Liang Zhou<sup>1</sup>, Qiming Wang<sup>1\*</sup>, Tianze Chen<sup>1</sup>

<sup>1</sup>School of Mathematics, Hohai University, China  
zeallll1127@gmail.com, wqm@hhu.edu.cn, 241319020003@hhu.edu.cn

## Abstract

3D point cloud classification is a fundamental task in safety-critical applications such as autonomous driving, robotics, and augmented reality. However, recent studies reveal that point cloud classifiers are vulnerable to structured adversarial perturbations and geometric corruptions, posing risks to their deployment in safety-critical scenarios. Existing certified defenses limit point-wise perturbations but overlook subtle geometric distortions that preserve individual points yet alter the overall structure, potentially leading to misclassification. In this work, we propose FreqCert, a novel certification framework that departs from conventional spatial domain defenses by shifting robustness analysis to the frequency domain, enabling structured certification against global  $\ell_2$ -bounded perturbations. FreqCert first transforms the input point cloud via the graph Fourier transform (GFT), then applies structured frequency-aware subsampling to generate multiple sub-point clouds. Each sub-cloud is independently classified by a standard model, and the final prediction is obtained through majority voting, where sub-clouds are constructed based on spectral similarity rather than spatial proximity, making the partitioning more stable under  $\ell_2$  perturbations and better aligned with the object’s intrinsic structure. We derive a closed-form lower bound on the certified  $\ell_2$  robustness radius and prove its tightness under minimal and interpretable assumptions, establishing a theoretical foundation for frequency domain certification. Extensive experiments on the ModelNet40 and ScanObjectNN datasets demonstrate that FreqCert consistently achieves higher certified accuracy and empirical accuracy under strong perturbations. Our results suggest that spectral representations provide an effective pathway toward certifiable robustness in 3D point cloud recognition.

## Introduction

3D point clouds, which represent unordered sets of spatial coordinates, are widely adopted in safety-critical applications such as autonomous driving, robotics, and industrial inspection (Poulose, Baek, and Han 2022; Wang et al. 2022; Bin Zaydie, Low, and Lin 2023; Sepulveda-Valdez et al. 2023). Due to their sparse, irregular, and non-Euclidean structure, point clouds pose unique challenges for perception

tasks, prompting the development of dedicated deep learning architectures (Charles et al. 2017; Qi et al. 2017; Li et al. 2018; Wang et al. 2019a; Thomas et al. 2019; Zhao et al. 2021).

While these models achieve high classification accuracy under clean conditions, recent studies have revealed their alarming vulnerability to adversarial attacks—small, imperceptible perturbations in the point cloud can easily cause models to make erroneous predictions (Xiang, Qi, and Li 2019; Liu, Yu, and Su 2019). These adversarial manipulations, in addition to natural corruptions such as occlusions and noise, pose serious security risks due to their subtlety and targeted attack nature (Tsai et al. 2020; Wen et al. 2020).

To address this, a variety of empirical defense strategies have been proposed. Typical methods involve input transformation or purification strategies, such as DUP-Net which combines outlier removal and upsampling to defend against adversarial perturbations (Zhou et al. 2019), IF-Defense which utilizes implicit surfaces for input projection (Wu et al. 2021), and diffusion-based purification such as PointDP (Sun et al. 2023). Despite improving empirical robustness, these methods lack certified guarantees, leaving them susceptible to adaptive attacks (Pérez et al. 2022; Lorenz et al. 2021; Li et al. 2022). Recent test-time adaptation methods, such as Purified Self-Training (PST) (Lin et al. 2024), attempt to improve robustness during inference, but still fail to provide provable robustness certificates in most practical scenarios.

To bridge this gap, certified defenses have been proposed to provide formal robustness guarantees. Certified defenses aim to provide provable guarantees that a model’s prediction remains stable under bounded perturbations. In the 3D point cloud domain, voting-based methods such as PointGuard (Liu, Jia, and Gong 2021) and PointCert (Zhang et al. 2023) construct multiple sub-clouds via random sampling or hashing, and aggregate their predictions through majority vote to tolerate a limited number of corrupted points. While effective against point additions or deletions, these methods cannot certify robustness under  $\ell_2$ -bounded perturbations, which preserve the number of points but subtly shift their positions. For example, a typical  $\ell_2$  attack perturbs point coordinates along adversarial gradients, altering the geometry without changing the point count (Xiang, Qi, and Li 2019). In contrast, randomized smooth-

\*Corresponding author.

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

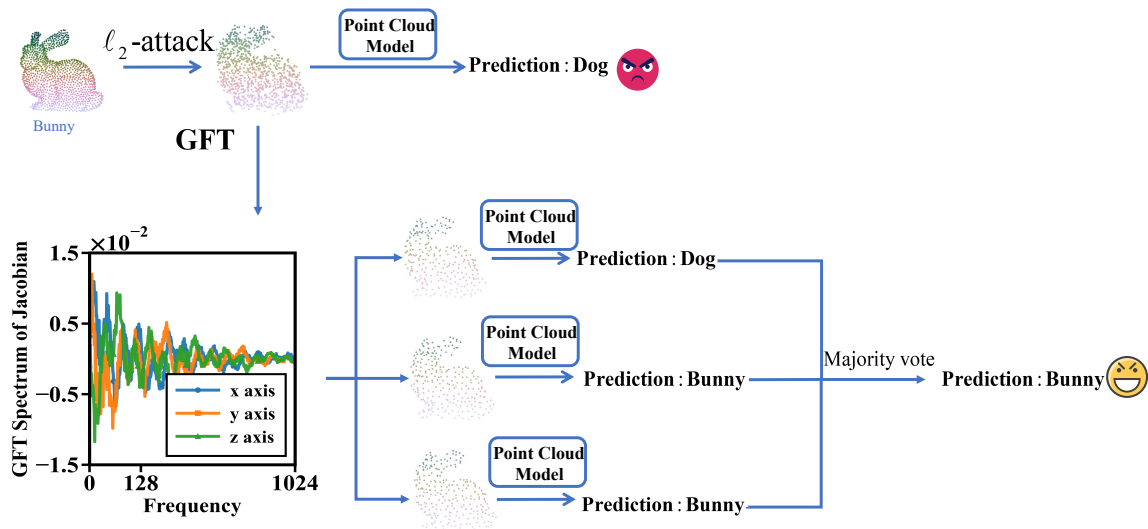


Figure 1: Illustration of our framework. The input point cloud is processed into spectral slices via graph Fourier transform (GFT), each of which is independently classified. The final prediction is obtained by majority voting.

ing(Cohen, Rosenfeld, and Kolter 2019) widely effective in 2D image classification—relies on fixed input dimensionality (pixel count), which does not apply to point clouds: even slight coordinate perturbations can disrupt neighborhood relations or trigger preprocessing changes (e.g. sampling or outlier removal), invalidating the theoretical guarantees. Consequently, smoothing-based approaches fail to reliably certify robustness in point cloud settings.

Spatial-domain point clouds are highly redundant, allowing  $\ell_2$  perturbations to distort geometry imperceptibly while preserving point count—challenging existing spatial defenses. Therefore, in this paper, we analyze point cloud robustness in the frequency domain. Prior study(Miao et al. 2024) shows that low-frequency components capture semantic structure, while high-frequency components are more vulnerable to adversarial perturbations. We leverage this by applying graph Fourier transform (GFT)(Shuman et al. 2013a) to project point clouds onto a spectral basis, then we design a frequency-guided sampling strategy called *dense-overlapping spectral window* (d-OSW). Specifically, we first compute the GFT coefficients of each point to identify its dominant spectral response. Based on these responses, we assign points to overlapping frequency bands, ensuring that each slice captures a different portion of the spectrum while retaining sufficient spatial coverage. Each frequency-aware slice is then independently classified using a shared point cloud model. Finally, the overall prediction is determined by majority voting across slices, which improves robustness by reducing the impact of any single corrupted subset.

Theoretically, we establish certified  $\ell_2$  robustness guarantees for the final prediction based on the stability of individual slices. Specifically, we derive closed-form expressions for two robustness radii against additive,  $\ell_2$ -bounded perturbations. The first, a conservative bound, ensures that the prediction remains unchanged as long as all slices are stable. The second, tighter bound leverages the majority vot-

ing mechanism: it certifies robustness even when a subset of slices is affected, as long as more than half retain the correct prediction. This analysis connects the spectral structure of the input to provable robustness at the classification level, and represents, to the best of our knowledge, the first certification framework for point clouds under continuous  $\ell_2$  perturbations.

We conduct extensive experiments on the ModelNet40 and ScanObjectNN datasets to validate the effectiveness of our method under  $\ell_2$  perturbations. Results show that **Fre-qCert** significantly improves certified robustness across various backbone networks, especially under strong attacks, demonstrating the effectiveness of our method for robust point cloud recognition.

In summary, our contributions are threefold: (1) We develop a frequency-guided sub-sampling strategy for point clouds based on graph Fourier transform and dense-overlapping spectral windows, and integrate it into **Fre-qCert**, a novel certification framework that classifies each spectral slice independently and aggregates predictions via majority voting; (2) We derive closed-form certified  $\ell_2$  robustness bounds that explicitly account for the voting mechanism; (3) We conduct extensive experiments on ModelNet40 and ScanObjectNN, demonstrating significant gains in certified accuracy and empirical accuracy under  $\ell_2$  perturbations.

## Related Work

**Deep Learning on 3D Point Clouds.** Early works on 3D point cloud learning adapted 2D CNNs via voxelization (Maturana and Scherer 2015; Su et al. 2015; Tchapmi et al. 2017), but suffered from sparsity and high computational cost. OctNet (Riegler, Ulusoy, and Geiger 2017) and sparse convolutions (Choy, Gwak, and Savarese 2019) alleviated these issues through efficient 3D data structures. A major breakthrough came with PointNet (Charles et al.

2017), which directly processed raw points using symmetric functions. Its extension PointNet++ (Qi et al. 2017) captured local structures hierarchically. Subsequent architectures such as PointCNN (Li et al. 2018), KPConv (Thomas et al. 2019), and DGCNN (Wang et al. 2019a) improved geometric modeling through learned kernels and graph-based features. GDANet (Xu et al. 2021) further introduced deformation-aware graph convolution to enhance feature learning under geometric variations. CurveNet (Xiang et al. 2021) enhances local geometric modeling by leveraging curve-based neighborhoods, leading to improved robustness against fine-grained surface deformations.

Recently, transformer-based models have emerged as state-of-the-art, including Point Transformer (Zhao et al. 2021) and LFT-Net (Gao et al. 2022), which integrate self-attention for local-global reasoning. Others, such as PVT (Zhang et al. 2022), combine voxel and point-based transformers for enhanced feature extraction.

### Adversarial Attacks & Defense on 3D Point Clouds.

Adversarial attacks on 3D point clouds aim to mislead classifiers by manipulating point coordinates. Gradient-based methods such as C&W and PGD have been adapted to the 3D setting through geometry-aware loss functions (Xiang, Qi, and Li 2019; Wen et al. 2020; Tsai et al. 2020), while black-box approaches leverage generative models for improved transferability (Zhou et al. 2020). To counter these threats, various defenses have been developed. Filtering-based methods (Zhou et al. 2019) remove outlier points, while input transformations and adversarial detectors have also been explored (Dong et al. 2020; Liu, Yu, and Su 2019). Purification-based strategies (Wu et al. 2021; Li et al. 2022; Sun et al. 2023) aim to restore clean geometry before classification. Certified defenses provide formal robustness guarantees. Existing efforts include robustness certification under pose variations and rigid transformations (Lorenz et al. 2021; Pérez et al. 2022), as well as under point-wise and additive perturbations (Liu, Jia, and Gong 2021; Zhang et al. 2023). However, most current methods remain ineffective against subtle yet structured  $\ell_2$ -bounded attacks.

## Method

The proposed **FreqCert** is a certified defense framework that leverages frequency-aware subsampling and majority voting to certify the robustness of point cloud classifiers against  $\ell_2$ -bounded perturbations. In this section, we first introduce the graph Fourier transform (GFT) as the spectral foundation of our method. Then, we present a frequency-guided subsampling strategy that extracts multiple low-frequency sub-point clouds. Finally, we derive a closed-form robustness radius under the  $\ell_2$  threat model and majority voting, offering a tight certification guarantee.

### Graph Fourier transform

Unlike images that lie on regular grids and can be naturally processed with the discrete Fourier transform (DFT) (Rasheed et al. 2020), 3D point clouds are unordered and reside in irregular metric spaces, which limits the use of standard frequency tools. To address this, we leverage

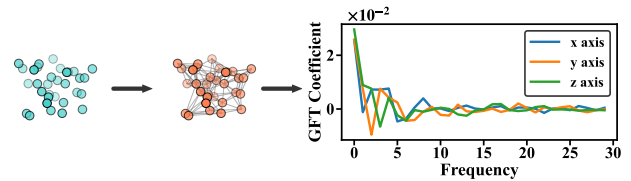


Figure 2: An intuitive visualization of our spectral transformation pipeline. A raw point cloud (left) is first transformed into a  $k$ -nearest neighbor graph (middle), followed by a graph Fourier transform (right) to derive frequency domain representations.

the graph Fourier transform (GFT) (Shuman et al. 2013a), which generalizes the concept of Fourier analysis to non-Euclidean domains by treating point clouds as graphs.

Let  $\mathcal{P} = \{p_i\}_{i=1}^n \subset \mathbb{R}^3$  be a point cloud of  $n$  points in Euclidean space. We construct an undirected  $k$ -nearest neighbor graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , where each node  $v_i \in \mathcal{V}$  corresponds to a point  $p_i \in \mathcal{P}$ , and edges  $(v_i, v_j) \in \mathcal{E}$  are formed by connecting each point to its  $k$  nearest neighbors based on Euclidean distance.

To capture local geometric similarity, edge weights are defined as  $w_{ij} = \exp(-\|p_i - p_j\|_2^2)$ , forming a weighted adjacency matrix  $W$ . The degree matrix  $D$  is diagonal with entries  $D_{ii} = \sum_j W_{ij}$ , and the combinatorial graph Laplacian is defined as  $L = D - W$ .

The Laplacian matrix  $L$  is symmetric and positive semi-definite. It can be diagonalized via eigendecomposition as  $L = U\Lambda U^T$ , where  $U \in \mathbb{R}^{n \times n}$  is an orthonormal matrix of eigenvectors and  $\Lambda$  is a diagonal matrix of non-negative eigenvalues. The columns of  $U$  form the graph Fourier basis. The entries in  $\Lambda$  represent the corresponding graph frequencies, arranged in ascending order from low to high.

Given a graph signal  $X \in \mathbb{R}^{n \times c}$ , such as the 3D coordinates of the point cloud, the graph Fourier transform is defined as  $\hat{X} = U^T X$ . Each row in  $\hat{X}$  represents the projection of  $X$  onto a different frequency component. The inverse transform is given by  $X = U\hat{X}$ , allowing full reconstruction from the frequency domain. Figure 2 visually demonstrates our spectral transformation pipeline, which converts raw point clouds into graph structures and further into frequency representations via GFT.

This frequency decomposition enables a compact and structured analysis of point cloud geometry, which is particularly useful for guiding robust subsampling and perturbation-aware reasoning in downstream tasks.

### Spectral Subsampling

To certify robustness against  $\ell_2$  perturbations, we must ensure that small changes to the input cannot simultaneously affect all parts of the model. One effective approach is to divide the point cloud into multiple sub-point clouds and base the final decision on a majority vote. If most sub-clouds remain stable under perturbation, the overall prediction remains unchanged.

A key question is how to perform this division. Sampling in the spatial domain, such as by random grouping or clus-

tering in Euclidean space offers little control over how individual point movements affect the slices. A single perturbed point may be assigned to very different slices depending on its shifted location, making the behavior under perturbation difficult to analyze or bound. In addition, spatial closeness alone can be misleading: points that are close in Euclidean space may belong to very different parts of the shape. In contrast, points with similar structural functions may be far apart spatially but share similar spectral responses.

We propose a frequency-guided subsampling algorithm, termed **dense-overlapping spectral windows (d-OSW)**, to construct robust and analyzable sub-point clouds for certification. Each point in the cloud typically exhibits a dominant response at a particular graph frequency, reflecting its geometric role within the global structure. By identifying these dominant frequencies, we group points based on spectral similarity rather than spatial proximity. To this end, we define a set of overlapping frequency bands and assign each point a soft weight for each band based on the alignment between its dominant frequency and the band center. Sampling is then performed proportionally to these weights, resulting in overlapping sub-point clouds focused on distinct frequency ranges.

We begin by computing the graph Fourier basis of the input point cloud and retaining the first  $K$  frequency components. These components are indexed by  $0, 1, \dots, K-1$ . Following Miao et al. (2024), lower frequencies primarily capture global shape information while higher frequencies encode finer geometric details and potential noise. To balance computational efficiency and structural fidelity, we retain only the first  $K = 128$  components.

The spectral range is evenly divided into  $m$  bands. The center of the  $b$ -th band is given by

$$\mu_b = \left(b + \frac{1}{2}\right) \cdot \frac{K}{m}, \quad b = 0, 1, \dots, m-1. \quad (1)$$

Each point is assigned a dominant frequency index  $\nu_i^*$ , defined as the index at which its squared GFT coefficient is maximized. We measure how well this dominant frequency aligns with band  $b$  using a Gaussian weight:

$$\gamma_b(i) = \exp\left(-\frac{(\nu_i^* - \mu_b)^2}{2\sigma^2}\right), \quad (2)$$

where  $\sigma$  is a bandwidth parameter controlling the overlap between adjacent bands. A smaller  $\sigma$  yields sharper, more selective assignments, while a larger  $\sigma$  produces smoother, more redundant coverage. Inspired by spectral kernel designs (Hammond, Vandergheynst, and Gribonval 2011; Shuman et al. 2013b), we set  $\sigma = 0.6 \cdot \frac{K}{m}$  to ensure effective but not excessive overlap between neighboring bands. This empirically balances slice diversity and stability, and avoids degenerate hard assignments at small  $m$ .

After computing spectral weights, we normalize them across all points. For each band  $b$ , we sample exactly  $n$  points without replacement from the full point cloud using the normalized weights as sampling probabilities. This yields one sub-point cloud per band.

Repeating the above sampling across all  $m$  bands yields  $m$  sub-point clouds in total. Since a point may have non-zero weights for multiple bands, it can be included in mul-

---

Algorithm 1: Dense-Overlapping Spectral Windows (d-OSW)

---

**Input:** Point cloud  $P = \{p_i\}_{i=1}^N$ , GFT basis  $\Phi \in \mathbb{R}^{N \times K}$ , number of bands  $m$ , slice size  $n$ , bandwidth  $\sigma$

**Output:** Set of sub-point clouds  $\{S_b\}_{b=0}^{m-1}$

---

```

1: for  $i = 1$  to  $N$  do
2:    $\nu_i^* \leftarrow \arg \max_{\nu} |\Phi_{\nu}(i)|^2$  // dominant frequency of point  $p_i$ 
3: end for
4: for  $b = 0$  to  $m - 1$  do
5:    $\mu_b \leftarrow \left(b + \frac{1}{2}\right) \cdot \frac{K}{m}$ 
6:   for  $i = 1$  to  $N$  do
7:      $\gamma_b(i) \leftarrow \exp\left(-\frac{(\nu_i^* - \mu_b)^2}{2\sigma^2}\right)$ 
8:   end for
9:   Normalize  $\{\gamma_b(i)\}_{i=1}^N$  to form a probability distribution
10:   $S_b \leftarrow$  sample  $n$  points from  $P$  without replacement according to weights  $\gamma_b$ 
11: end for
12: return  $\{S_b\}_{b=0}^{m-1}$ 

```

---

iple slices. This overlap introduces controlled redundancy and improves robustness under perturbations.

We refer to this overall procedure and the resulting collection of slices as **dense-overlapping spectral windows (d-OSW)**. The construction ensures complete spectral coverage while maintaining structural coherence. It also enables precise frequency domain robustness analysis, which we leverage to derive certified  $\ell_2$  perturbation bounds.

To ensure consistency between training and certification, we apply the same spectral subsampling strategy during training. Specifically, for each training example, we randomly extract  $m$  frequency-guided sub-clouds using d-OSW. The model is trained to correctly classify all sub-clouds. This encourages the network to learn features that are invariant across different spectral slices—an essential property for accurate majority voting at test time. Without such training-time alignment, the model may overfit to global features and fail to generalize when presented with subsampled views during certification.

## Certified Robustness

Frequency-aware subsampling offers more than just a way to construct diverse inputs—it enables precise reasoning about how perturbations affect the classifier. Each point is assigned to slices based on its spectral response, and small  $\ell_2$  perturbations lead to gradual, predictable changes in this assignment. This continuity allows us to control how input noise propagates through the slicing and voting pipeline. Such analysis is difficult in the spatial domain. When slices are formed by Euclidean grouping or random selection, even minor displacements may abruptly reassign points, making the system sensitive and analytically intractable.

By contrast, the structure imposed by frequency-guided sampling ensures that slice memberships evolve smoothly under perturbation, and redundancy across overlapping

slices provides natural resilience. These properties open the door to formal certification.

In the FreqCert framework, robustness is achieved by aggregating the predictions of multiple frequency-aware sub-point clouds. To move beyond empirical robustness and provide formal guarantees, we seek to certify that the final decision remains unchanged under any bounded perturbation of the input.

Let  $P = \{p_i\}_{i=1}^N$  be the input point cloud. Using Algorithm 1, we construct  $m$  overlapping frequency slices  $\{S_j\}_{j=1}^m$ , where each  $S_j$  is a subset of points sampled based on spectral similarity. Each slice is passed through a shared base classifier  $f$ , and the final output is determined by majority voting:

$$h(P) = \text{MajorityVote}(\{f(S_j)\}_{j=1}^m). \quad (3)$$

We define certified robustness with respect to the prediction function  $h(P)$ , which aggregates the results from  $m$  frequency-based slices. A prediction is certifiably robust at  $P \in \mathbb{R}^{N \times 3}$  if it remains unchanged under any  $\ell_2$ -bounded perturbation to the input. Specifically, we consider perturbed point clouds  $P' = P + \Delta$ , where each row  $\Delta_i$  satisfies  $\|\Delta_i\|_2 \leq \varepsilon$ . The goal is to find the largest  $\varepsilon > 0$  such that

$$h(P') = h(P), \quad \forall P' \text{ with } \|\Delta_i\|_2 \leq \varepsilon \text{ for all } i. \quad (4)$$

This definition corresponds to a per-point  $\ell_2$  ball of radius  $\varepsilon$  and reflects the total robustness of the system under structured perturbations.

To characterize certified robustness, we analyze how input perturbations influence slice assignments. Each point  $p_i$  is grouped based on its dominant graph frequency—the frequency at which its graph Fourier coefficient is maximized.

The sensitivity of this assignment can be quantified by the **spectral margin**  $g_i$ , which measures how close point  $p_i$  is to changing its slice under perturbation. Let  $\nu_i^*$  be the index of the frequency at which  $p_i$  exhibits the largest spectral energy, i.e., the dominant frequency. Then the spectral margin  $g_i$  is defined as the minimum distance from  $\nu_i^*$  to the center of any other band:

$$g_i = \min_{b \in \{0, \dots, m-1\}} |\nu_i^* - \mu_b|. \quad (5)$$

A larger  $g_i$  indicates that  $p_i$  is farther from band boundaries, making it more robust to perturbations that could alter its slice assignment.

The effect of perturbation on spectral responses is further governed by the graph Laplacian’s eigengap  $\Delta\lambda$ , which controls how sharply frequencies vary with changes in geometry.

During slice construction, each point may softly contribute to multiple bands. Let  $\kappa$  be the maximum number of slices a point can appear in; if  $t$  points shift their dominant frequencies under perturbation, at most  $\kappa t$  slices are affected.

We next provide two formal guarantees that characterize the certified  $\ell_2$  robustness of our method.

**Theorem 1** (Certified Perturbation Size for Slice Stability). *Let  $P$  be an input point cloud, and suppose the Frobenius*

*norm of the perturbation satisfies*

$$\|\Delta P\|_F < R_{\text{slice}}, \quad R_{\text{slice}} := \frac{g_{\min} \Delta\lambda}{4\sqrt{K}}, \quad (6)$$

*where  $g_{\min}$  is the minimum spectral margin across all points,  $\Delta\lambda$  is the Laplacian eigengap at index  $K$ , and  $K$  is the number of retained frequencies.*

*Then no slice changes its prediction, and the overall classification remains unaffected.*

*Moreover, the bound is tight: any perturbation with norm greater than  $R_{\text{slice}}$  may change the predicted label of at least one slice.*

The result above guarantees robustness by freezing all slice assignments, which can be overly conservative in practice. A more flexible guarantee can be obtained by allowing a limited number of slice changes—as long as the majority vote remains unaffected. The following result formalizes this relaxed but stronger form of certified robustness.

**Theorem 2** (Certified Perturbation Size for  $\ell_2$  Robustness). *Let  $g_{(1)} \leq \dots \leq g_{(N)}$  be the sorted spectral margins of the points. Define*

$$\alpha = \left\lfloor \frac{m-1}{2} \right\rfloor, \quad p = \left\lfloor \frac{\alpha}{\kappa} \right\rfloor.$$

*Then for*

$$R^* := \frac{\Delta\lambda \sqrt{K}}{8} \left( \sum_{i=1}^{p+1} g_{(i)}^2 \right)^{1/2}, \quad (7)$$

*any perturbation  $\Delta P$  with  $\|\Delta P\|_F < R^*$  affects at most  $p$  points, and thus no more than  $\kappa p \leq \alpha$  slices. The majority vote, and hence the prediction  $h(P)$ , remains unchanged.*

*This radius is tight: any uniform improvement of  $R^*$  would fail in the worst case without additional assumptions.*

*Proof.* The detailed proofs of Theorem 1 and Theorem 2 are provided in the extended version of this paper (Zhou, Wang, and Chen 2025).  $\square$

Eqs.(7) reveal that the certified radius depends on three key factors: the spectral margins of individual points, the Laplacian eigengap  $\Delta\lambda$ , and the overlap multiplicity  $\kappa$ . Larger margins and a wider eigengap strengthen robustness, while a lower  $\kappa$  reduces the number of slices affected by each perturbation. These insights suggest that designing d-OSW to amplify spectral separation and limit redundancy—without sacrificing coverage—can directly improve the certified  $\ell_2$  radius.

## Experiments

### Experimental Setup

**Datasets and moudles** We conduct experiments on two widely used datasets: **ModelNet40** (Wu et al. 2015) and **ScanObjectNN** (Uy et al. 2019). ModelNet40 consists of 12,311 synthetic 3D CAD models from 40 object categories, with 9,843 samples for training and 2,468 for testing. Each object is uniformly sampled into 1,024 surface points. To assess robustness under more realistic conditions,

we also evaluate on ScanObjectNN, which contains 2,902 real-world scanned objects from 15 categories. We adopt the PB\_T50\_RS variant, which includes background clutter and partial occlusions. Each object is represented by 2,048 points, and the dataset is split into 2,048 training and 881 test samples. We use PointNet (Charles et al. 2017) and DGCNN (Wang et al. 2019b) as representative classification backbones due to their widespread use in point cloud recognition tasks.

**Compared methods** We compare FreqCert with two baselines. The undefended classifier refers to the standard point cloud model without any robustness enhancement, trained on clean data. We evaluate it to assess the impact of our certification strategy. In addition, we include Randomized Smoothing (RS) (Cohen, Rosenfeld, and Kolter 2019), a widely used certified defense for  $\ell_2$  perturbations. RS adds isotropic Gaussian noise to the input and performs majority voting over multiple predictions. It provides a certified radius within which the classifier’s prediction remains provably unchanged.

**Evaluation Metrics** We evaluate robustness from both certified and empirical perspectives. *Certified accuracy* is the proportion of test samples that are correctly classified and provably robust against all  $\ell_2$  perturbations within a radius  $\epsilon$ , providing a worst-case guarantee. In contrast, *empirical accuracy* reflects the proportion of samples that remain correctly classified under specific adversarial attacks of strength  $\epsilon$ . As it accounts for all possible perturbations, certified accuracy is always a conservative lower bound on empirical accuracy at the same  $\epsilon$ .

In our experiments, empirical robustness is evaluated using adversarial examples generated by projected gradient descent (PGD) (Xiang, Qi, and Li 2019), a widely adopted attack in the point cloud literature. Given a clean input point cloud  $\mathbf{X}_o \in \mathbb{R}^{N \times 3}$ , consisting of  $N$  points in 3D space, we generate a perturbed version  $\mathbf{X}_o^{\text{adv}}$  such that the total displacement is constrained by  $\|\mathbf{X}_o^{\text{adv}} - \mathbf{X}_o\|_2 \leq \epsilon$ , where  $\epsilon$  controls the attack strength. We evaluate robustness across a range of  $\epsilon$  values and report both certified and empirical accuracy as functions of  $\epsilon$  in the final robustness plots.

**Parameter setting** Main experiments use DGCNN (Wang et al. 2019b) as the backbone, with a 20-nearest neighbor graph, trained for 250 epochs using the Adam optimizer with a batch size of 32 and an initial learning rate of 0.001. For FreqCert, each point cloud is decomposed into  $m = 32$  frequency-guided sub-clouds, and each sub-cloud contains  $n = 128$  points. These values are chosen to balance spectral coverage and certification tightness. In particular,  $n$  must not be too small under  $\ell_2$ -norm perturbations, since low-resolution sub-clouds are more sensitive to global geometric distortions, which degrades prediction stability and reduces certified accuracy. The same subsampling strategy is applied during training to ensure consistency between learning and certification. For the Randomized Smoothing baseline, we adopt  $\sigma = 0.5$ , which provides a reasonable trade-off between clean accuracy and certified robustness.

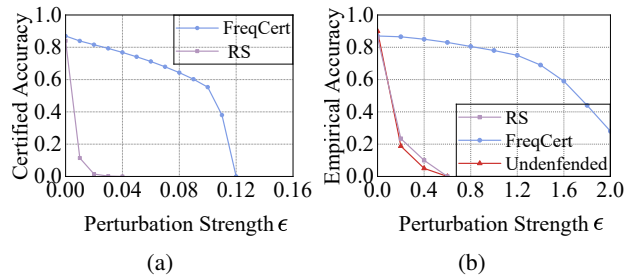


Figure 3: Certified accuracy (a) and empirical accuracy (b) on the ModelNet40 dataset under increasing  $\ell_2$  perturbation strength  $\epsilon$ . FreqCert consistently achieves higher robustness than randomized smoothing across all tested perturbation levels. Results on the ScanObjectNN dataset are provided in the extended version of this paper (Zhou, Wang, and Chen 2025).

## Experimental Results

**Main results:** As shown in Fig. 3, we compare the certified and empirical accuracy of FreqCert and randomized smoothing under  $\ell_2$ -norm bounded point perturbations on ModelNet40. At  $\epsilon = 0$ , the undefended model achieves the highest empirical accuracy, but its performance quickly drops as  $\epsilon$  increases, revealing its vulnerability to structured input distortions.

In terms of certified robustness, FreqCert demonstrates a clear advantage. It maintains non-trivial certified accuracy up to  $\epsilon = 0.12$ , whereas randomized smoothing fails to certify any samples beyond  $\epsilon = 0.02$ . The certified accuracy curve of FreqCert decays smoothly, with a noticeable drop-off near  $\epsilon = 0.10$ , indicating that its robustness degrades in a controlled manner as perturbations increase. In contrast, the certification curve of randomized smoothing remains nearly flat and close to zero throughout, reflecting its poor ability to provide formal guarantees on 3D data.

The sharp contrast between the two curves highlights a key limitation of additive-noise-based methods when applied to point clouds: even weak Gaussian perturbations can disrupt the underlying geometry, breaking the conditions required for certification. In contrast, FreqCert leverages frequency-domain decomposition and structured voting to certify a larger fraction of test samples across all tested radii. The smooth decay of its certification curve suggests that the method is inherently more resilient to localized distortions and better captures global structural properties.

On the empirical side, FreqCert maintains high robustness against PGD attacks, preserving over 70% accuracy at  $\epsilon = 1.0$  and retaining non-trivial performance even at  $\epsilon = 2.0$ . Compared to randomized smoothing, which collapses rapidly and becomes ineffective beyond  $\epsilon = 0.4$ , FreqCert degrades more gradually. This behavior reflects the advantage of frequency-based subsampling: by partitioning the input into multiple overlapping spectral slices, localized perturbations are prevented from dominating the prediction, and the majority vote mechanism mitigates adversarial impact across slices.

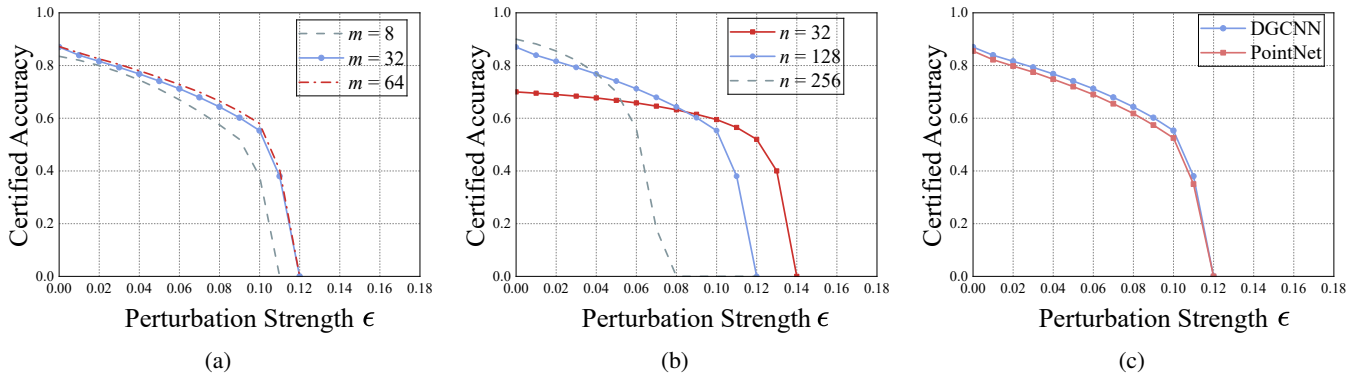


Figure 4: Impact of key factors on certified accuracy. (a) Impact of  $m$  on FreqCert. (b) Impact of  $n$  on FreqCert. (c) Comparing certified accuracy of FreqCert across different backbone architectures

		PointNet	DGCNN	CurveNet
FreqCert	PGD	<b>78.9</b>	<b>74.8</b>	<b>70.5</b>
	AdvPC	<b>69.4</b>	<b>71.7</b>	<b>73.3</b>
IF-Defense	PGD	74.1	70.6	69.6
	AdvPC	69.3	64.7	57.9
DUP-Net	PGD	74.8	68.6	69.2
	AdvPC	68.9	67.0	68.4

Table 1: Empirical accuracy (%) under PGD and AdvPC attacks with  $\ell_2$  perturbation of strength  $\epsilon = 1.25$ .

**Comparison with Empirical Defenses:** To further evaluate the empirical robustness of FreqCert, we compare it with several representative empirical defenses under strong adversarial attacks, including DUP-Net (Zhou et al. 2019) and IF-Defense (Wu et al. 2021). DUP-Net enhances robustness through outlier removal and upsampling, while IF-Defense projects the input onto an implicit surface to mitigate adversarial distortions.

In addition to the standard PGD attack, we also include AdvPC (Hamdi et al. 2020), a point-cloud-specific method that generates semantically meaningful perturbations by approximating latent geometric deformations.

All defenses are evaluated under a fixed perturbation strength of  $\epsilon = 1.25$ , defined as the  $\ell_2$  norm of the total displacement applied to the point cloud. For fair comparison, all methods use the same set of backbones: PointNet (Charles et al. 2017), DGCNN (Wang et al. 2019b), and CurveNet (Xiang et al. 2021).

As shown in Table 1, FreqCert consistently achieves the highest empirical accuracy across all three backbones under both PGD and AdvPC attacks. The performance gap is especially pronounced under AdvPC, where IF-Defense suffers notable drops (e.g., 57.9% on CurveNet), while FreqCert maintains robust performance (e.g., 73.3%). These results indicate that FreqCert generalizes better to strong, shape-aware attacks compared to existing empirical defenses.

**Impact of  $m$ :** As shown in Fig. 4(a), the  $m=32$  curve consistently outperforms  $m=8$  in the mid-range region  $\epsilon \in$

$[0.06, 0.10]$ , where the separation is most pronounced. For  $\epsilon < 0.04$ , all three curves are nearly identical. The  $m=8$  configuration drops sharply near  $\epsilon \approx 0.10$ , while  $m=32$  and  $m=64$  remain valid up to  $\epsilon \approx 0.12$ . Between them,  $m=64$  shows a slight but marginal advantage. This reflects a trade-off: increasing  $m$  improves voting robustness by raising  $\alpha = \lfloor (m-1)/2 \rfloor$  and  $q = \lfloor \alpha/\kappa \rfloor + 1$ , but also increases the overlap factor  $\kappa$ , limiting further gains.

**Impact of  $n$ :** As shown in Fig. 4(b), larger slices ( $n=256$ ) achieve the highest certified accuracy at small perturbations but degrade rapidly, with the certificate vanishing around  $\epsilon \approx 0.08$ . In contrast, smaller slices ( $n=32$ ) maintain non-zero certified accuracy up to  $\epsilon \approx 0.14$ , albeit with significantly lower performance in the low-perturbation regime. This trade-off arises because large slices capture more global structure but are more vulnerable to global geometric distortions, while small slices are more robust to such perturbations but suffer from higher prediction variance due to limited resolution.

**Backbone comparison:** Fig. 4(c) shows that DGCNN stays consistently above PointNet across the whole perturbation range ( $\epsilon \leq 0.10$ ), but the gap is small (typically below 2%). Both curves drop abruptly and reach zero around  $\epsilon \approx 0.12$ . This indicates that FreqCert is largely architecture-agnostic—the certificate is dominated by the spectral subsampling and voting scheme—while the stronger local feature extraction of DGCNN provides a slight advantage.

## Conclusion

In this work, we propose FreqCert, a certification framework based on frequency-domain subsampling for 3D point clouds. A novel spectral slicing algorithm enables closed-form  $\ell_2$  robustness certificates against structured perturbations without adversarial training or randomized smoothing. Theoretical analysis proves tightness under minimal assumptions, and experiments on ModelNet40 and ScanObjectNN demonstrate consistent improvements in certified and empirical robustness.

## Acknowledgements

This work was supported by the National Key R&D Program of China (No. 2022YFB3207400). We also thank the anonymous program committee for their constructive feedback and valuable suggestions.

## References

- Bin Zaydie, A. H.; Low, M. Y. H.; and Lin, W. 2023. Perception Pipeline of Autonomous Mobile Robots Using 3D Point Cloud Classification. In *2023 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*, 1–7.
- Charles, R. Q.; Su, H.; Kaichun, M.; and Guibas, L. J. 2017. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 77–85.
- Choy, C.; Gwak, J.; and Savarese, S. 2019. 4D Spatio-Temporal ConvNets: Minkowski Convolutional Neural Networks. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3070–3079.
- Cohen, J.; Rosenfeld, E.; and Kolter, Z. 2019. Certified Adversarial Robustness via Randomized Smoothing. In Chaudhuri, K.; and Salakhutdinov, R., eds., *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, 1310–1320. PMLR.
- Dong, X.; Chen, D.; Zhou, H.; Hua, G.; Zhang, W.; and Yu, N. 2020. Self-robust 3d point recognition via gather-vector guidance. In *2020 IEEE/CVF conference on computer vision and pattern recognition (cvpr)*, 11513–11521. IEEE.
- Gao, Y.; Liu, X.; Li, J.; Fang, Z.; Jiang, X.; and Huq, K. M. S. 2022. LFT-Net: Local feature transformer network for point clouds analysis. *IEEE transactions on intelligent transportation systems*, 24(2): 2158–2168.
- Hamdi, A.; Rojas, S.; Thabet, A.; and Ghanem, B. 2020. AdvPC: Transferable Adversarial Perturbations on 3D Point Clouds. In Bischof, H.; Brox, T.; Frahm, J.; and Vedaldi, A., eds., *COMPUTER VISION - ECCV 2020, PT XII*, volume 12357 of *Image Processing Computer Vision Pattern Recognition and Graphics*, 241–257. ISBN 978-3-030-58609-6; 978-3-030-58610-2. 16th European Conference on Computer Vision-ECCV-Biennial, ELECTR NETWORK, AUG 23-28, 2020.
- Hammond, D. K.; Vandergheynst, P.; and Gribonval, R. 2011. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2): 129–150.
- Li, K.; Zhang, Z.; Zhong, C.; and Wang, G. 2022. Robust structured declarative classifiers for 3d point clouds: Defending adversarial attacks with implicit gradients. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15294–15304.
- Li, Y.; Bu, R.; Sun, M.; Wu, W.; Di, X.; and Chen, B. 2018. PointCNN: Convolution On X-Transformed Points. In Bengio, S.; Wallach, H.; Larochelle, H.; Grauman, K.; Cesa-Bianchi, N.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc.
- Lin, J.; Yang, X.; Li, T.; and Xu, X. 2024. Improving Adversarial Robustness for 3D Point Cloud Recognition at Test-Time through Purified Self-Training. ArXiv:2409.14940 [cs].
- Liu, D.; Yu, R.; and Su, H. 2019. Extending adversarial attacks and defenses to deep 3d point cloud classifiers. In *2019 IEEE International Conference on Image Processing (ICIP)*, 2279–2283. IEEE.
- Liu, H.; Jia, J.; and Gong, N. Z. 2021. PointGuard: Provably Robust 3D Point Cloud Classification. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 6182–6191.
- Lorenz, T.; Ruoss, A.; Balunović, M.; Singh, G.; and Vechev, M. 2021. Robustness certification for point cloud models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 7608–7618.
- Maturana, D.; and Scherer, S. 2015. VoxNet: A 3D Convolutional Neural Network for real-time object recognition. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 922–928.
- Miao, Y.; Dong, Y.; Zhang, J.; Yu, L.; Yang, X.; and Gao, X.-S. 2024. Improving Robustness of 3D Point Cloud Recognition from a Fourier Perspective. In Globerson, A.; Mackey, L.; Belgrave, D.; Fan, A.; Paquet, U.; Tomczak, J.; and Zhang, C., eds., *Advances in Neural Information Processing Systems*, volume 37, 68183–68210. Curran Associates, Inc.
- Pérez, J. C.; Alfara, M.; Giancola, S.; Ghanem, B.; et al. 2022. 3deformrs: Certifying spatial deformations on point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15169–15179.
- Poulose, A.; Baek, M.; and Han, D. S. 2022. Point Cloud Map Generation and Localization for Autonomous Vehicles Using 3D Lidar Scans. In *2022 27th Asia Pacific Conference on Communications (APCC)*, 336–341.
- Qi, C. R.; Yi, L.; Su, H.; and Guibas, L. J. 2017. PointNet++: deep hierarchical feature learning on point sets in a metric space. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17*, 5105–5114. Red Hook, NY, USA: Curran Associates Inc. ISBN 9781510860964.
- Rasheed, M. H.; Salih, O. M.; Siddeq, M. M.; and Rodrigues, M. A. 2020. Image compression based on 2D Discrete Fourier Transform and matrix minimization algorithm. *Array*, 6: 100024.
- Riegler, G.; Ulusoy, A. O.; and Geiger, A. 2017. OctNet: Learning Deep 3D Representations at High Resolutions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6620–6629.
- Sepulveda-Valdez, C.; Sergiyenko, O.; Alaniz-Plata, R.; Núñez-López, J. A.; Tyrsa, V.; Flores-Fuentes, W.; Rodríguez-Quíñonez, J. C.; Mercorelli, P.; Kolendovska, M.; Kartashov, V.; Miranda-Vega, J. E.; and Murrieta-Rico, F. N. 2023. Laser Scanning Point Cloud Improvement by Implementation of RANSAC for Pipeline Inspection Application. In *IECON 2023- 49th Annual Conference of the IEEE Industrial Electronics Society*, 1–6.

- Shuman, D. I.; Narang, S. K.; Frossard, P.; Ortega, A.; and Vandergheynst, P. 2013a. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE signal processing magazine*, 30(3): 83–98.
- Shuman, D. I.; Narang, S. K.; Frossard, P.; Ortega, A.; and Vandergheynst, P. 2013b. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE signal processing magazine*, 30(3): 83–98.
- Su, H.; Maji, S.; Kalogerakis, E.; and Learned-Miller, E. 2015. Multi-view Convolutional Neural Networks for 3D Shape Recognition. In *2015 IEEE International Conference on Computer Vision (ICCV)*, 945–953.
- Sun, J.; Wang, J.; Nie, W.; Yu, Z.; Mao, Z.; and Xiao, C. 2023. A Critical Revisit of Adversarial Robustness in 3D Point Cloud Recognition with Diffusion-Driven Purification. In *Proceedings of the 40th International Conference on Machine Learning*, 33100–33114. PMLR. ISSN: 2640-3498.
- Tchapmi, L.; Choy, C.; Armeni, I.; Gwak, J.; and Savarese, S. 2017. SEGCloud: Semantic Segmentation of 3D Point Clouds. In *2017 International Conference on 3D Vision (3DV)*, 537–547.
- Thomas, H.; Qi, C. R.; Deschaud, J.-E.; Marcotegui, B.; Goulette, F.; and Guibas, L. 2019. KPConv: Flexible and Deformable Convolution for Point Clouds. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 6410–6419.
- Tsai, T.; Yang, K.; Ho, T.-Y.; and Jin, Y. 2020. Robust adversarial objects against deep learning models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 954–962.
- Uy, M. A.; Pham, Q.-H.; Hua, B.-S.; Nguyen, D. T.; and Yeung, S.-K. 2019. Revisiting Point Cloud Classification: A New Benchmark Dataset and Classification Model on Real-World Data. In *International Conference on Computer Vision (ICCV)*.
- Wang, Y.; Sun, Y.; Liu, Z.; Sarma, S. E.; Bronstein, M. M.; and Solomon, J. M. 2019a. Dynamic Graph CNN for Learning on Point Clouds. *ACM Trans. Graph.*, 38(5): 146:1–146:12.
- Wang, Y.; Sun, Y.; Liu, Z.; Sarma, S. E.; Bronstein, M. M.; and Solomon, J. M. 2019b. Dynamic Graph CNN for Learning on Point Clouds. *ACM Trans. Graph.*, 38(5).
- Wang, Y.; Xu, S.; Zuo, Z.; and Li, Z. 2022. Monocular 3D Object Detection Based on Pseudo-LiDAR Point Cloud for Autonomous Vehicles. In *2022 41st Chinese Control Conference (CCC)*, 5469–5474.
- Wen, Y.; Lin, J.; Chen, K.; Chen, C. P.; and Jia, K. 2020. Geometry-aware generation of adversarial point clouds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(6): 2984–2999.
- Wu, Z.; Duan, Y.; Wang, H.; Fan, Q.; and Guibas, L. J. 2021. IF-Defense: 3D Adversarial Point Cloud Defense via Implicit Function based Restoration. ArXiv:2010.05272 [cs].
- 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.
- Xiang, C.; Qi, C. R.; and Li, B. 2019. Generating 3D Adversarial Point Clouds. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9128–9136. Long Beach, CA, USA: IEEE.
- Xiang, T.; Zhang, C.; Song, Y.; Yu, J.; and Cai, W. 2021. Walk in the cloud: Learning curves for point clouds shape analysis. In *Proceedings of the IEEE/CVF international conference on computer vision*, 915–924.
- Xu, M.; Zhang, J.; Zhou, Z.; Xu, M.; Qi, X.; and Qiao, Y. 2021. Learning geometry-disentangled representation for complementary understanding of 3d object point cloud. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 3056–3064.
- Zhang, C.; Wan, H.; Shen, X.; and Wu, Z. 2022. PVT: Point-voxel transformer for point cloud learning. *International Journal of Intelligent Systems*, 37(12): 11985–12008.
- Zhang, J.; Jia, J.; Liu, H.; and Gong, N. Z. 2023. PointCert: Point Cloud Classification with Deterministic Certified Robustness Guarantees. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9496–9505. Vancouver, BC, Canada: IEEE. ISBN 9798350301298.
- 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.
- Zhou, H.; Chen, D.; Liao, J.; Chen, K.; Dong, X.; Liu, K.; Zhang, W.; Hua, G.; and Yu, N. 2020. Lg-gan: Label guided adversarial network for flexible targeted attack of point cloud based deep networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10356–10365.
- Zhou, H.; Chen, K.; Zhang, W.; Fang, H.; Zhou, W.; and Yu, N. 2019. DUP-Net: Denoiser and Upsampler Network for 3D Adversarial Point Clouds Defense. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 1961–1970. Seoul, Korea (South): IEEE. ISBN 978-1-72814-803-8.
- Zhou, L.; Wang, Q.; and Chen, T. 2025. Certified L2-Norm Robustness of 3D Point Cloud Recognition in the Frequency Domain. ArXiv:2511.07029 [cs].