Explicitly Perceiving and Preserving the Local Geometric Structures for 3D Point Cloud Attack

Daizong Liu and Wei Hu*

Wangxuan Institute of Computer Technology, Peking University
dzliu@stu.pku.edu.cn, forhuwei@pku.edu.cn

Abstract

Deep learning models for point clouds have shown to be vulnerable to adversarial attacks, which have received increasing attention in various safety-critical applications such as autonomous driving, robotics, and surveillance. Existing 3D attack methods generally employ global distance losses to implicitly constrain the point-wise perturbations for optimization. However, these simple losses are quite difficult to accurately measure and restrict the proper 3D geometry as point clouds are highly structured. Although few recent works try to exploit additional shape-aware surface knowledge to globally constrain the point position, they still fail to preserve the detailed point-to-point geometric dependency in different local regions. To this end, in this paper, we propose a novel Multi-grained Geometry-aware Attack (MGA), which *explicitly* captures the *local* topology characteristics in different 3D regions for adversarial constraint. Specifically, we first develop multi-scale local spectral filter banks adapting to different 3D object shapes to explore potential geometric structures in different local regions. Considering that objects may contain complex geometries, we then extend each filter bank into multi-layer ones to gradually capture different-granularity topology contexts of the same region in a coarse-to-fine manner. Hence, the focused local geometries will be highlighted in the coefficients calculated by the filtering process. At last, by restricting these coefficients between benign and adversarial samples, our MGA is able to properly measure and preserve the detailed geometry contexts in the whole 3D object with trivial perturbations. Experiments demonstrate that our attack can achieve superior performance on various 3D classification models, with satisfying adversarial imperceptibility and strong resistance to different defense methods.

Introduction

Deep neural networks have shown to be vulnerable to adversarial examples (Goodfellow, Shlens, and Szegedy 2014; Szegedy et al. 2013), which add visually indistinguishable perturbations to network inputs but lead to incorrect prediction results. Significant progress has been made in adversarial attacks on 2D images, where many methods (Dong et al. 2018; Madry et al. 2017; Kurakin, Goodfellow, and Bengio 2016; Tu et al. 2019) learn to add imperceptible pixel-wise noise. Nevertheless, adversarial attacks on 3D depth or point cloud data are still relatively underexplored. Different from images, point clouds are discrete representations of 3D objects or scenes (Huang, Liu, and Hu 2023; Hu, Liu, and Hu 2023), receiving increasing attention in various 3D applications such as autonomous driving (Chen et al. 2017b) and medical data analysis (Singh et al. 2020). Similarly to

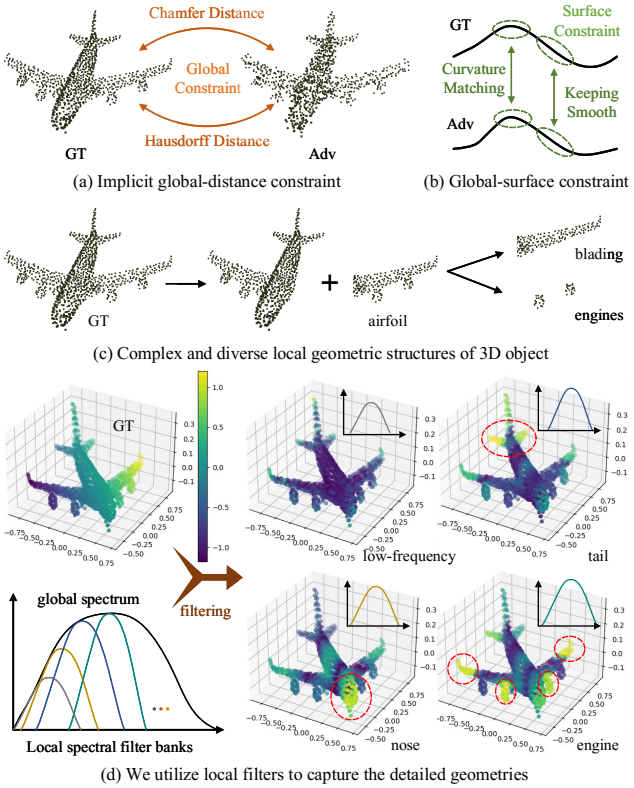

Figure 1: Existing 3D attackers simply utilize (a) global distance losses or (b) surface constraints to restrict the perturbations. Since 3D objects generally contain (c) diverse geometric characteristics in different local regions, they fail to comprehend and preserve corresponding detailed contexts. In this paper, we propose to exploit (d) local spectral filters to perceive the complex geometric structures in different regions for generating imperceptible adversarial samples.

*Corresponding Author.

their 2D counterparts, deep learning models trained for point clouds are often vulnerable to adversarial perturbations, increasing the risk in safety-critical 3D applications.

Most existing 3D point cloud attack methods (Xiang, Qi, and Li 2019; Wicker and Kwiatkowska 2019; Zhang et al. 2019a; Zheng et al. 2019; Tsai et al. 2020; Zhao et al. 2020; Zhou et al. 2020; Hamdi et al. 2020) generally adapt the existing 2D adversarial attacks into the 3D scenario. They either follow a point addition/dropping framework that identifies critical points from point clouds and modifies (add or delete) them to distort the most representative features, or follow the C&W framework (Goodfellow, Shlens, and Szegedy 2014) to learn to perturb the Euclidean coordinates of points by optimizing the gradient from end to end. However, these works directly utilize general global distance losses, *e.g.* Chamfer (Fan, Su, and Guibas 2017) and Hausdorff (Huttenlocher, Klanderman, and Rucklidge 1993) distances as shown in Figure 1 (a), to implicitly preserve the 3D original shape. It is still hard for them to accurately measure and generate the geometry-aware perturbations with these simple losses as point cloud is a highly structured data format, leading to noticeable noise. Although few works (Wen et al. 2020; Huang et al. 2022) try to exploit additional surface knowledge to restrict the point positions as shown in Figure 1 (b), they only consider the curve and smoothness characteristics and treat all points in a global manner for joint constraint. Since 3D object generally contains complex and diverse geometric characteristics in different regions as shown in Figure 1 (c), they fail to separately perceive corresponding different types of point-to-point dependencies for preserving the detailed local geometric structures.

Based on the above observations, in this paper, we make the attempt to *explicitly* capture the subtle geometric structures in different *local* regions to completely preserve the diverse 3D topology information for generating more imperceptible adversarial examples. As for exploring the 3D geometric characteristics, graph spectral tool (Hu, Liu, and Hu 2022; Hu et al. 2021) has shown great success in frequency-to-geometry analysis. However, it suffers from its global mapping that can only identify the geometric characteristics of the whole 3D object but fails to locate where they are in the spatial region. Therefore, to perceive different geometric structures corresponding to different regions in the point cloud, we introduce a local spectral tool with different filter banks to filter out and highlight the potential subtle geometries in local regions. As shown in Figure 1 (d), the local spectral filter banks can be decomposed from the global spectrum of the 3D object. By applying them to the original point cloud, we can separately obtain different highlighted regions (computed by the coefficients) to represent the components of different geometric structures. In this way, we are able to set individual geometry-aware constraints on each local region for jointly preserving the original object shape during the perturbation optimization.

To this end, we propose a novel Multi-grained Geometry-aware Attack (MGA), which explicitly captures the local topology characteristics in different 3D regions for adversarial constraint. Specifically, we first represent each point cloud in a graph format and then develop adaptive diffusion wavelets to serve as local spectral filters for capturing the subtle geometries. Considering that 3D objects generally contain complex and diverse geometric topologies, we further extend these local filters into a hierarchical one with multiple scales and multiple layers. In particular, the multi-scale filter banks are applied to the same object for perceiving different geometric characteristics in different local regions, while the multi-layer strategy is applied on each scale filter bank to gradually capture more grained-level geometric structures in the same region in a coarse-to-fine manner. At last, by restricting the different-level coefficients calculated by the hierarchical filter banks between benign and adversarial samples, our MGA attack is able to accurately measure and preserve the complete geometric shape of the 3D object. To sum up, our main contributions are three-fold:

- Instead of globally and implicitly perturbing the point clouds like previous 3D attackers, we propose a new attack method that explicitly captures the local geometric characteristics in different 3D regions. Compared to them, our attack is able to preserve more subtle and detailed geometric contexts for generating more imperceptible adversarial examples.
- To perceive different subtle geometric contexts of different local object components in the whole point cloud, we develop multi-layer multi-scale local spectral filter banks to capture potential geometric structures of all possible granularities. We restrict corresponding coefficients between point clouds for perturbation optimization without using any traditional global distance loss.
- We perform extensive experiments to validate the effectiveness of our MGA attack using several popular point cloud networks on both ModelNet40 and ShapNetPart datasets, which show superiority over the state-of-the-arts. We also demonstrate how our MGA attack outperforms others when targeted by currently available point-cloud defense methods.

Related Work

Adversarial attack on 3D point cloud. Many works (Xiang, Qi, and Li 2019; Wicker and Kwiatkowska 2019; Zhang et al. 2019a; Zheng et al. 2019; Tsai et al. 2020; Zhao et al. 2020; Liu, Hu, and Li 2022; Tao et al. 2023; Liu, Hu, and Li 2023) investigate the vulnerability of 3D point clouds from the perspective of adversarial attack. Early works (Xiang, Qi, and Li 2019; Zhang et al. 2019a; Wicker and Kwiatkowska 2019; Zheng et al. 2019) simply modify a few points in the point cloud to achieve attack. They either add limited synthesized points or drop the critical points based on the characteristics of 3D models. Although they can achieve 100% attack success rate, they easily lead to the outlier problem. To alleviate this limitation, recent works propose to add perturbations to the whole point cloud. These methods (Wen et al. 2020; Tsai et al. 2020; Hamdi et al. 2020; Liu, Yu, and Su 2019; Ma et al. 2020; Zhang et al. 2019b; Liu and Hu 2021) learn to perturb the Euclidean coordinates of each point by utilizing the C&W framework (Carlini and Wagner 2017) to shift the points to fool the 3D models. However, these methods implicitly utilize both

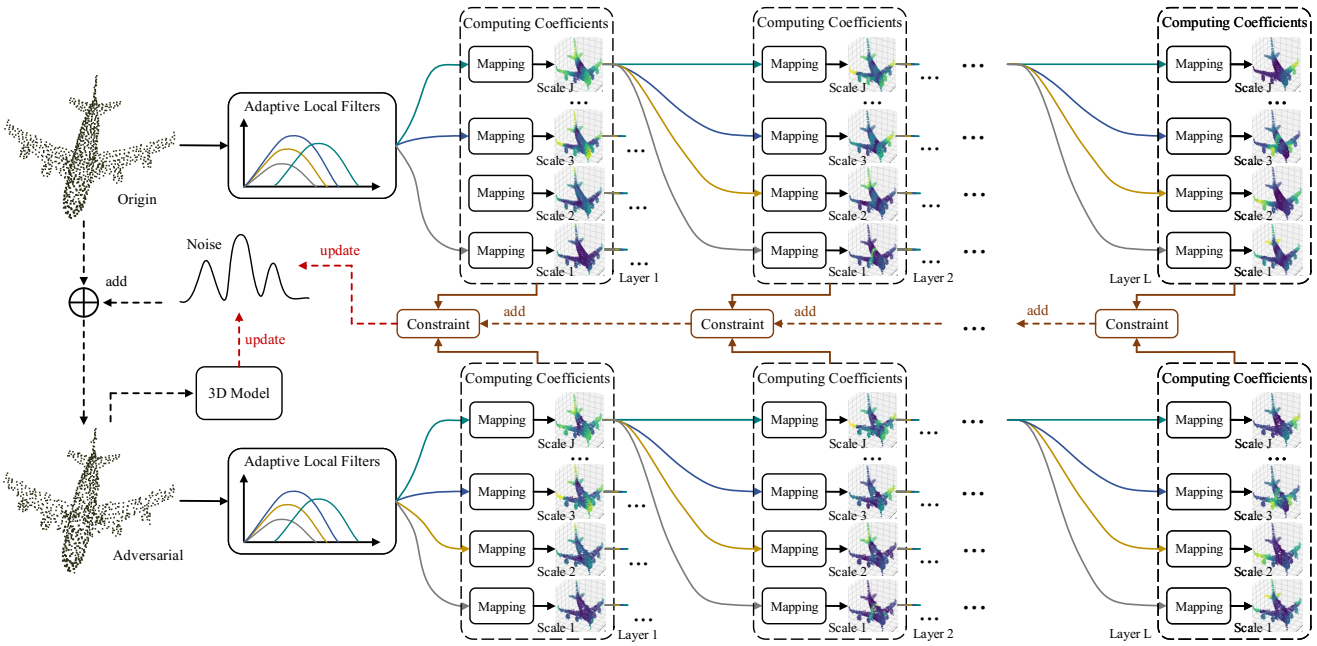

Figure 2: Overall pipeline of our proposed MGA attack. We develop novel hierarchical local spectral filters with multiple scales and layers, to perceive multi-grained local geometric components of the 3D object in a coarse-to-fine manner. By restricting the coefficients calculated by each grained filter between the original and adversarial samples, our MGA is able to preserve the local structures highlighted by different coefficients for completely measuring and constraining the geometry-aware perturbation.

Chamfer and Hausdorff distances as the constraint to keep the point cloud geometry. Some recent works (Wen et al. 2020; Huang et al. 2022) try to utilize additional surface knowledge to build shape prior to preserving the object contours. Since 3D object generally contains complex and diverse geometric characteristics in different regions, they fail to separately perceive corresponding different types of point-to-point dependencies for preserving the detailed local geometric structures. Different from them, we propose to explicitly capture the subtle geometric structures in different local regions to completely preserve the diverse 3D topology information for generating more imperceptible adversarial examples.

Spectral methods for 3D point cloud. There already exist methodologies that exploit spectral information to understand point clouds. Some 3D denoising methods (Rosman, Dubrovina, and Kimmel 2013; Zhang, Cui, and Ding 2020) transform the input point cloud into the graph spectral domain, where the rough shape of a point cloud is encoded into low-frequency components. The noisy shape can thus be reconstructed by the spectral filter. Other applications (Chen et al. 2017a; Ramasinghe et al. 2020) also represent the fine details of point clouds through transformed high-frequency components, and use them to detect contours or process redundant information. Recently, spectrum-based attack (Hu, Liu, and Hu 2022) is also introduced to selectively perturb the frequency bands for keeping the geometries in the data domain as invariant as possible. However, this spectral tool is globally based, which can only identify the geometric characteristics in the whole point cloud and cannot locate where they are in the local spatial region. Different from them, we develop a new local spectral strategy to perceive different geometries in different local regions.

The Proposed MGA Attack

Overview

Problem definition. Generally, a point cloud consists of an unordered set of points $\boldsymbol{P} = \{\boldsymbol{p}_i\}_{i=1}^{n} \in \mathbb{R}^{n\times 3}$ sampled from the surface of a 3D object or scene, where each point $\boldsymbol{p}_i \in \mathbb{R}^3$ is a vector that contains the coordinates (x, y, z) of point i, and n is the number of points. In this paper, we mainly focus on the basic point cloud classification task. Given a point cloud $\boldsymbol{P}$ as input, a learned classifier $f(\cdot)$ predicts a vector of confidence scores $f(\boldsymbol{P}) \in \mathbb{R}^C$. The final predicted label is $y = F(\boldsymbol{P}) = argmax_{i\in[C]} f(\boldsymbol{P})_i \in Y, Y = \{1, 2, 3, ..., C\}$ that represents the class of the original 3D object underlying the point cloud, where C is the number of classes. To attack such classification model, the general objective is to find a perturbation $\boldsymbol{\Delta} \in \mathbb{R}^{n\times 3}$ to generate an adversarial example $\boldsymbol{P'} = \boldsymbol{P} + \boldsymbol{\Delta}$ that $f(\boldsymbol{P'}) \neq y$ or $f(\boldsymbol{P'}) = y'$, where $y' \in Y$ but $y' \neq y$.

Our pipeline. To completely preserve the detailed geometries of each 3D object, we propose a novel Multi-grained Geometry-aware Attack (MGA) method, which develops multiple local spectral filters to separately perceive different local geometric characteristics of the whole 3D point cloud for constraint. The overall pipeline of our MGA is shown in Figure 2. Given the benign point cloud, we develop its adaptive local spectral filters with multiple scales and layers

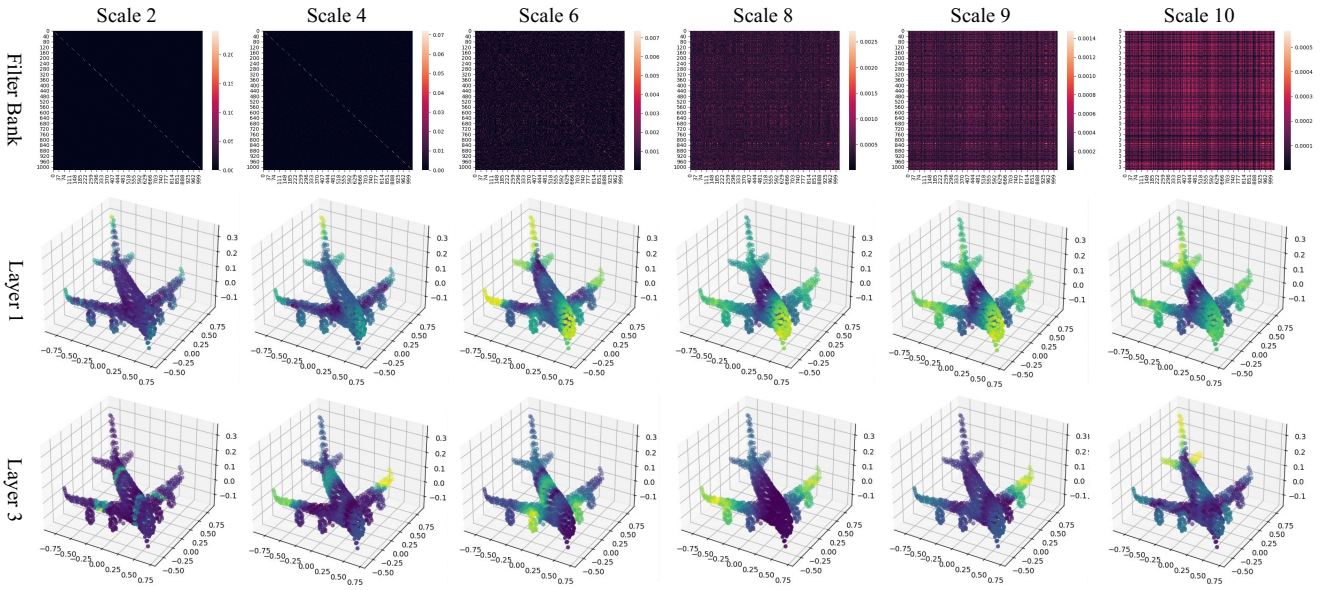

Figure 3: Visualization on the filter banks of different scales, and the coefficients computed by these scales of different layers.

to capture potential subtle 3D components. In particular, the multi-scale filter banks are applied to the object for perceiving different geometric characteristics in local regions, while the multi-layer strategy is applied on each scale filter bank to gradually capture multi-grained geometric structures in the same region in a coarse-to-fine manner. By restricting the highlighted coefficients of filters of the same scale and layer, MGA is able to preserve the corresponding local geometries for producing more imperceptible adversarial samples.

Transforming Point Cloud into Spectral Domain

Before devising the local spectral filters adapting to different 3D objects, we need to transform the point cloud data into the spectral domain. Following previous work (Hu et al. 2021; Hu, Liu, and Hu 2022), we utilize graph tool to transform the 3D data into the graph spectral domain. Specifically, we represent a point cloud $\boldsymbol{P} = \{\boldsymbol{p}_i\}_{i=1}^{n} \in \mathbb{R}^{n\times 3}$ consisting of n points over a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \boldsymbol{A}\}$, which is composed of a vertex set $\mathcal{V}$ of cardinality $|\mathcal{V}| = n$ representing points, an edge set $\mathcal{E}$ connecting vertices, and an adjacency matrix $\boldsymbol{A}$. Each entry $a_{i,j}$ in $\boldsymbol{A}$ represents the weight of the edge between vertices i and j, which often captures the similarity between adjacent vertices. Here, we construct an unweighted K-nearest-neighbor graph (K-NN graph), where each vertex is connected to its K nearest neighbors in terms of the Euclidean distance with weight 1. The coordinates of points in $\boldsymbol{P}$ are treated as graph signals. We define the combinatorial graph Laplacian operator (Shuman et al. 2013) as $\boldsymbol{L} := \boldsymbol{D} - \boldsymbol{A}$, where $\boldsymbol{D}$ is the *degree matrix*—a diagonal matrix where $d_{i,i} = \sum_{j=1}^{n} a_{i,j}$.

Developing Adaptive Local Spectral Filters

Graphs are flexible data structures that enable general metric structures and modeling non-Euclidean domains. Once we map the point cloud into the graph spectral domain, we are able to investigate the topology characteristics on the underlying graph domain, while capturing corresponding geometric regions on the data domain.

Since diffusion wavelets (Coifman and Maggioni 2006) provide a simple yet effective framework to define multi-resolution geometric analysis from powers of a diffusion operator defined on a graph, we follow it to first define a lazy diffusion operator on point cloud $\boldsymbol{P}$ as:

$$\boldsymbol{T} = \frac{1}{2}(\boldsymbol{I} + \widetilde{\boldsymbol{A}}), \tag{1}$$

where $\boldsymbol{I}$ is the identical diagonal matrix and $\widetilde{\boldsymbol{A}} = \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}$ is the normalized adjacency. Based on this, we construct a family of multi-scale local spectral filters by exploiting the powers of the diffusion operator $\boldsymbol{T}^{2m}$ as:

$$\begin{aligned} &\boldsymbol{\Psi}_0 = \boldsymbol{I} - \boldsymbol{T}, \\ &\boldsymbol{\Psi}_m = \boldsymbol{T}^{2^{m-1}}(\boldsymbol{I} - \boldsymbol{T}^{2^{m-1}}) = \boldsymbol{T}^{2^{m-1}} - \boldsymbol{T}^{2^m},\ m > 0 \end{aligned} \tag{2}$$

Here, representing point clouds in the diffusion wavelet domain is able to capture spectral differences corresponding to smooth regions as well as sharp transitions or oscillations of the 3D surface at increasingly spaced diffusion times. Specifically, each $\boldsymbol{\Psi}_m$ corresponds to a graph wavelet filter bank with special spatial localization on 3D object. The low-scale $\boldsymbol{\Psi}_0 = \frac{1}{2}(\boldsymbol{I} - \widetilde{\boldsymbol{A}})$ corresponds to one half of the normalized Laplacian operator, which is often used as a low-pass filter in graph signal processing and spectral graph theory. Therefore, $\boldsymbol{\Psi}_0$ can capture the low-frequency contexts of the 3D objects. By increasing spaced diffusion times, the higher-scale $\{\boldsymbol{\Psi}_m\}$ can explicitly focus on various higher bands than $\boldsymbol{\Psi}_0$. Each scale wavelet coefficient of the point cloud can be obtained by:

$$\boldsymbol{\Phi}_m(\boldsymbol{P}) = \boldsymbol{\Psi}_m\boldsymbol{P},\ m \in \{1, ..., M\}. \tag{3}$$

Since some 3D objects have complex and diverse geometric topologies, we further extend the multi-scale local filters into multi-layer ones to capture hierarchical multi-grained local structures. The l-th layer m-th scale wavelet coefficient of point clouds can be obtained by:

$$\mathbf{\Phi}_{m,l}(\boldsymbol{P}) = (\mathbf{\Psi}_m)^l \boldsymbol{P},\ l \in \{1, ..., L\}. \tag{4}$$

As shown in Figure 3, the frequency responses of different scale filter banks are quite different. Therefore, they are able to capture different kinds of geometric components in different local regions of the 3D object. Moreover, by applying multi-layer filtering, the coefficients gradually highlight more on the fine-grained subtle region in the higher layer. It demonstrates that such hierarchical filter banks are effective to perceive the local regions in a coarse-to-fine manner.

Hierarchical Coefficient Constraints

To make the adversarial sample imperceptible to humans, the attackers should keep its geometric shape same to the original 3D object as same as possible. Previous attack methods (Huang et al. 2022) either utilize global distance loss to implicitly generate geometry-aware perturbations, or utilize surface constraint to only keep the smoothness and ignore the complicated region. Different from them, we propose to explicitly preserve the complete geometric characteristics by separately perceiving and maintaining different geometries in different local regions. To achieve this goal, we utilize the obtained hierarchical local spectral filters to capture multi-grained local geometric components in both original and adversarial samples. We restrict the coefficients calculated by the same filter bank in the same layer between benign and adversarial samples, to individually and accurately measure and preserve the corresponding subtle geometric structure:

$$\mathcal{L}_{reg}^{m,l} = ||\mathbf{\Phi}_{m,l}(\boldsymbol{P}) - \mathbf{\Phi}_{m,l}(\boldsymbol{P'})||_2^2. \tag{5}$$

By restricting all coefficients of different local filters in the same way, we can perceive and preserve the complete geometric characteristics of the whole 3D point cloud as:

$$\mathcal{L}_{reg} = \frac{1}{\sum_{l=1}^{L} M^l} \sum_{l=1}^{L} \sum_{m=1}^{M^l} ||\mathbf{\Phi}_{m,l}(\boldsymbol{P}) - \mathbf{\Phi}_{m,l}(\boldsymbol{P'})||_2^2, \tag{6}$$

Here, each layer l contains M^l number of coefficients, which are calculated from the previous M^{l-1} coefficients of the former layer $l-1$ with M scale filter banks.

Discussion. The reason why this coefficient constraint is effective to restrict the geometry information in data domain is: The coefficients represent the strong point-to-point dependency in its highlighted region. Specifically, as for scale 0 filter $\mathbf{\Psi}_0$, it corresponds to the low-frequency component representing the basic object shape (Hu, Liu, and Hu 2022). By restricting its coefficients, the adversarial samples can preserve the object contour. The higher scale filter contributes more to the detailed structure. By restricting the special point-to-point dependency in each local region, the adversarial sample is able to keep the local geometry the same as the original one. By jointly restricting all the hierarchical coefficients, our strategy can explicitly preserve the complete geometric characteristics.

Generating Adversarial Examples

At last, we reformulate the task of generating adversarial sample $\boldsymbol{P'}$ as the following optimization problem:

$$\begin{aligned} &\min_{\boldsymbol{\Delta}} \mathcal{L}_{mis}(\boldsymbol{P'}, y) + \lambda \mathcal{L}_{reg}(\boldsymbol{P'}, \boldsymbol{P}), \\ &\text{s.t.} \quad \boldsymbol{P'} = \boldsymbol{P} + \boldsymbol{\Delta}, \end{aligned} \tag{7}$$

where $\mathcal{L}_{mis}$ is the cross-entropy loss to promote the misclassification of $\boldsymbol{P'}$, and λ is a parameter to strike a balance between the two terms in the objective. In particular, $\mathcal{L}_{mis}(\boldsymbol{P'}, y)$ is formulated as a cross-entropy loss:

$$\mathcal{L}_{mis}(\boldsymbol{P'}, y) = \begin{cases} -\log(p_{y'}(\boldsymbol{P'})), & \text{for targeted attack,} \\ \log(p_y(\boldsymbol{P'})), & \text{for untargeted attack,} \end{cases} \tag{8}$$

where $p(\cdot)$ is the softmax functioned on the output of the target model, *i.e.*, the probability with respect to adversarial class y' or clean class y.

Experiments

Experimental Setup

Datasets. We use ModelNet40 (Wu et al. 2015) and ShapeNetPart (Yi et al. 2016) to evaluate the performance of the adversarial point clouds generated by different attack methods. Specifically, ModelNet40 consists of 12,311 CAD models from 40 object categories, and ShapeNetPart consists of 16,881 pre-aligned shapes from 16 categories. Following previous works, we uniformly sample 1,024 points from the surface of each object.

3D victim models. Our approach is evaluated on three categories of the most popular 3D recognition models, *i.e.*, PointNet (Qi et al. 2017), DGCNN (Wang et al. 2019) and PointConv (Wu, Qi, and Fuxin 2019). We train them from scratch, and the test accuracy of each trained model is within 0.1% of the best accuracy reported in their original articles.

Implementation. We update the frequency perturbation $\boldsymbol{\Delta}$ with 500 iterations. We use Adam optimizer (Kingma and Ba 2015) to optimize the objective of our proposed attack in Eq. (7) with a fixed learning rate of 0.01, and the momentum is set as 0.9. We assign $K = 10$ to build a K-NN graph. The numbers of filter scale and layer are set to $M = 10, L = 4$, respectively. The parameter λ is set to 1.0. We focus on the targeted attack in the experiments. All experiments are implemented on a single NVIDIA RTX 2080Ti GPU.

Evaluation on Our MGA Attack

Performance comparison. We fairly compare our MGA attack with six competitive methods, including PGD (Madry et al. 2017), AdvPC (Hamdi et al. 2020), IFGM (Gu and Rigazio 2014), LG-GAN (Zhou et al. 2020), GSDA (Hu, Liu, and Hu 2022), and SI-Adv (Huang et al. 2022), and measure the perturbation on both ModelNet40 and ShapeNetPart datasets. The corresponding results are presented in Table 1 and Table 2. We can find that, our MGA attack generates adversarial point clouds with almost the lowest perturbation sizes in all evaluation metrics on three 3D victim models. This is because MGA explicitly perceives the

Attack Method	PointNet			DGCNN			PointConv		
	ASR↑	CD↓	HD↓	ASR↑	CD↓	HD↓	ASR↑	CD↓	HD↓
PGD (Madry et al. 2017)	100%	0.0018	0.0302	100%	0.0019	0.0202	100%	0.0019	0.0129
AdvPC (Hamdi et al. 2020)	100%	0.0013	0.0346	100%	0.0012	0.0186	100%	0.0010	0.0127
IFGM (Gu and Rigazio 2014)	100%	0.0009	0.0226	100%	0.0007	0.0137	100%	0.0005	0.0122
LG-GAN (Zhou et al. 2020)	99%	0.0011	0.0512	86%	0.0010	0.0725	78%	0.0008	0.0417
GSDA (Hu, Liu, and Hu 2022)	100%	0.0007	**0.0031**	100%	0.0014	0.0140	100%	0.0017	0.0218
SI-Adv (Huang et al. 2022)	100%	0.0002	0.0205	100%	0.0004	0.0054	100%	0.0003	0.0116
MGA (Ours)	**100%**	**0.0001**	0.0046	**100%**	**0.0002**	**0.0049**	**100%**	**0.0002**	**0.0085**

Table 1: Quantitative comparison on the perturbation size generated by different attack methods on ModelNet40 dataset.

Attack Method	PointNet			DGCNN			PointConv		
	ASR↑	CD↓	HD↓	ASR↑	CD↓	HD↓	ASR↑	CD↓	HD↓
PGD (Madry et al. 2017)	100%	0.0017	0.0434	100%	0.0019	0.0628	100%	0.0018	0.0442
AdvPC (Hamdi et al. 2020)	100%	0.0019	0.0543	100%	0.0029	0.0649	100%	0.0015	0.0404
IFGM (Gu and Rigazio 2014)	100%	0.0005	0.0411	100%	0.0006	0.0546	100%	**0.0002**	0.0385
LG-GAN (Zhou et al. 2020)	98%	0.0058	0.1122	75%	0.0083	0.1501	61%	0.0068	0.1156
GSDA (Hu, Liu, and Hu 2022)	95%	0.0023	0.0257	98%	0.0035	0.0388	94%	0.0026	0.0240
SI-Adv (Huang et al. 2022)	96%	0.0010	0.0433	95%	0.0009	0.0418	95%	0.0008	0.0125
MGA (Ours)	**100%**	**0.0003**	**0.0091**	**100%**	**0.0004**	**0.0118**	**100%**	0.0003	**0.0084**

Table 2: Quantitative comparison on the perturbation size generated by different attack methods on ShapeNetPart dataset.

Defense	PGD	GSDA	SI-Adv	MGA
SOR (Zhou et al. 2019)	52.3%	81.0%	**97.4%**	96.8%
DUP-Net (Zhou et al. 2019)	49.5%	68.9%	95.8%	**97.2%**
AT (Madry et al. 2017)	62.4%	86.3%	90.0%	**95.3%**
IF-Defence (Wu et al. 2020)	37.1%	50.1%	61.2%	**84.7%**

Table 3: Attack success rate (ASR) of different attacks on PointNet on ModelNet40 equipped with various defenses.

detailed and diverse local geometric structures of the 3D object for completely preserving the geometric characteristics, leading to more imperceptible adversarial examples.

Visualization results. We show the visualization results of the adversarial examples generated by different attack methods in Figure 4. We observe that our adversarial point clouds exhibit similar geometric structures to their corresponding benign point clouds, and are more imperceptible than other attacks. Since our MGA explicitly perceives detailed and diverse local geometric structures, we are able to completely and properly preserve the geometries of original 3D object.

Resistance to Defenses. To verify the attack performance when facing the well-defended 3D models, we conduct experiments on defense methods including standard statistical outlier removal (SOR) (Zhou et al. 2019), DUP-Net (Zhou et al. 2019), adversarial training (AT) (Madry et al. 2017), and IF-Defence (Wu et al. 2020). Through comparing to existing point cloud attack methods in Table 3, we find that our MGA attack outperforms the baselines a lot on almost all defense methods, which further demonstrates our multi-grained geometry-aware attack is effective to perceive the detailed and complete local geometric structures for better preserving the geometric characteristics.

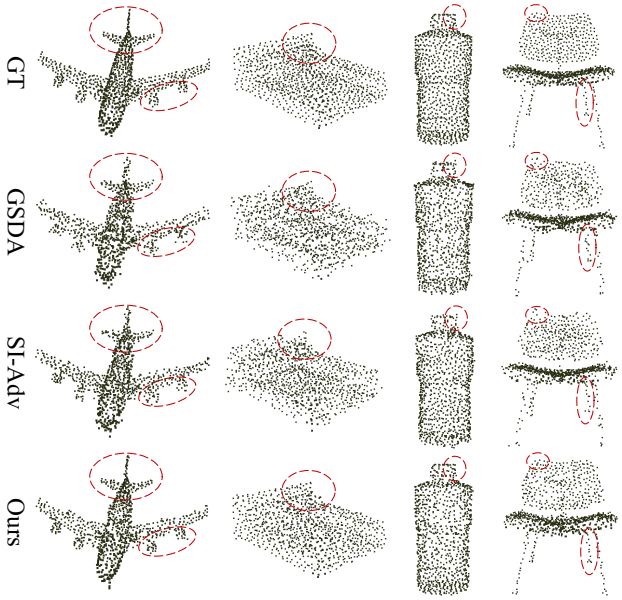

Figure 4: Visualization on the adversarial examples.

Ablation Study

Investigation on the choice of local spectral filter bank. To verify the effect of our local spectral filter bank, we conduct experiments by replacing our used Diffusion Wavelet with different strategies while maintaining the other modules the same. Specifically, two general local spectral tools are compared: Monic Cubic Wavelet (Hammond, Vandergheynst, and Gribonval 2011) and Tight Hann Wavelet (Shuman et al. 2015). As shown in Table 4, our utilized dif-

Local Filter	ASR↑	CD↓	HD↓
Monic Cubic Wavelet	100%	0.0008	0.0128
Tight Hann Wavelet	100%	0.0003	0.0079
Diffusion Wavelet	**100%**	**0.0001**	**0.0046**

Table 4: Investigation on different types of local spectral filter bank. Victim model: PointNet.

Module	Variant	ASR↑	CD↓	HD↓
Scale	$M = 2$	100%	0.0048	0.0630
	$M = 4$	100%	0.0019	0.0283
	$M = 6$	100%	0.0009	0.0135
	$M = 8$	100%	0.0002	0.0069
	$M = 10$	100%	**0.0001**	0.0046
	$M = 12$	100%	**0.0001**	**0.0043**
Layer	$L = 1$	100%	0.0021	0.0317
	$L = 2$	100%	0.0010	0.0141
	$L = 3$	100%	0.0004	0.0082
	$L = 4$	100%	**0.0001**	0.0046
	$L = 5$	100%	**0.0001**	**0.0044**

Table 5: Investigation on different numbers of scale and layer of the local filter banks. Victim model: PointNet.

Module	Variant	ASR↑	CD↓	HD↓
K-NN Graph	$K = 5$	100%	**0.0001**	0.0061
	$K = 10$	100%	**0.0001**	**0.0046**
	$K = 20$	100%	**0.0001**	**0.0046**
	$K = 40$	100%	**0.0001**	0.0052
Loss Function	$\lambda = 0.1$	100%	0.0006	0.0074
	$\lambda = 1.0$	100%	**0.0001**	**0.0046**
	$\lambda = 10.0$	100%	0.0002	0.0059

Table 6: Sensitive analysis on K, λ. Victim model: PointNet.

fusion wavelet achieves the smallest perturbations than other strategies in all metrics, this is because: diffusion wavelet is more suitable than Monic Cubic Wavelet and Tight Hann Wavelet for processing graph-structured data, allowing the extraction of valuable features from complex graphs.

Investigation on different numbers of scale and layer of the local filter banks. As shown in Table 5, we conduct the ablation studies on the numbers of both scale and layer of the local filter banks. Specifically, the scale number determines how many frequency-guided local geometries we can explore in the 3D object. By increasing its number M, our MGA can perceive more relevant local geometries and lead to more imperceptible results. Our model achieves the best performance when M is set to 12. However, the model with $M = 12$ is slightly better than the model with $M = 10$, but leads to much more time consumption. To balance both the performance and time cost, we choose $M = 10$ in our all experiments. Similarly, the layer number determines the highest granularity we can explore. Considering both the performance and time cost, we choose $L = 4$.

Sensitivity on the hyperparameters K, λ. As shown in Table 6, we investigate whether the adversarial effects vary with respect to different settings of the number K in the K-NN graph and the number λ in the loss function. We see that, the attack performance is insensitive to K since our attack with different K requires similar perturbation budgets CD and HD in the data domain. Therefore, we set $K = 10$ in all our experiments. As for the parameter λ, our MGA achieves the best performance when λ is set to 1.0.

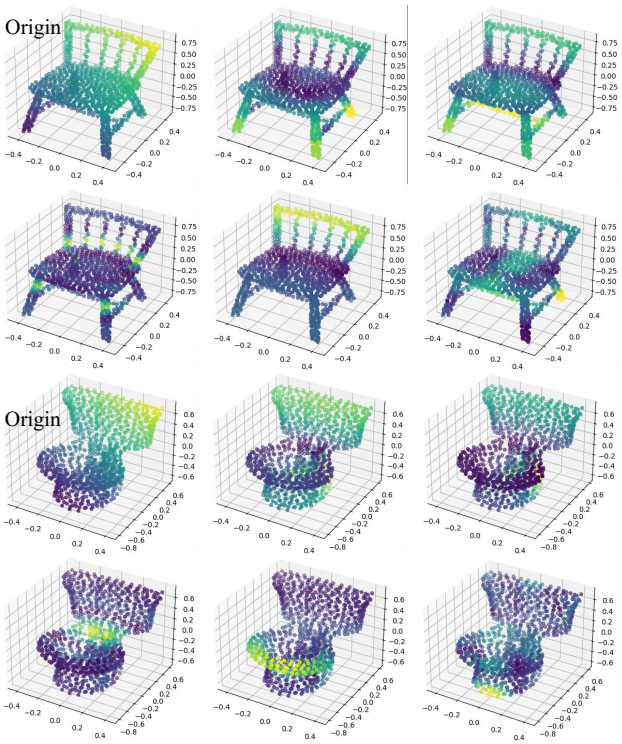

Figure 5: More visualization results on the coefficients.

More visualization on the coefficients. To investigate whether the local filters are able to properly perceive different local geometric structures in different regions, we provide more visualization results on the coefficients computed by local filters of different scales and layers in Figure 5.

Conclusion

In this paper, we propose a novel Multi-grained Geometry-aware Attack (MGA), which explicitly captures the local topology characteristics in different 3D regions for generating more imperceptible adversarial examples. To achieve this goal, we develop hierarchical local spectral filters to capture diverse individual local geometries in the whole 3D object. By restricting the coefficients computed by multi-scale and -layer filters between benign and adversarial point clouds, our MGA is able to properly measure and preserve the complete 3D geometric shape. Experiments validate both effectiveness and robustness of our MGA.

Acknowledgments

This work was supported by National Natural Science Foundation of China (61972009).

References

Carlini, N.; and Wagner, D. 2017. Towards evaluating the robustness of neural networks. In *2017 IEEE Symposium on Security and Privacy (SP)*, 39–57.

Chen, S.; Tian, D.; Feng, C.; Vetro, A.; and Kovačević, J. 2017a. Fast resampling of three-dimensional point clouds via graphs. *IEEE Transactions on Signal Processing*, 66(3): 666–681.

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

Coifman, R. R.; and Maggioni, M. 2006. Diffusion wavelets. *Applied and computational harmonic analysis*, 21(1): 53–94.

Dong, Y.; Liao, F.; Pang, T.; Su, H.; Zhu, J.; Hu, X.; and Li, J. 2018. Boosting adversarial attacks with momentum. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 9185–9193.

Fan, H.; Su, H.; and Guibas, L. J. 2017. A point set generation network for 3d object reconstruction from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 605–613.

Goodfellow, I. J.; Shlens, J.; and Szegedy, C. 2014. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*.

Gu, S.; and Rigazio, L. 2014. Towards deep neural network architectures robust to adversarial examples. *arXiv preprint arXiv:1412.5068*.

Hamdi, A.; Rojas, S.; Thabet, A.; and Ghanem, B. 2020. Advpc: Transferable adversarial perturbations on 3d point clouds. In *European Conference on Computer Vision (ECCV)*, 241–257.

Hammond, D. K.; Vandergheynst, P.; and Gribonval, R. 2011. Wavelets on graphs via spectral graph theory. *Appl. Comput. Harmonic Anal.*, 30(2): 129–150.

Hu, Q.; Liu, D.; and Hu, W. 2022. Exploring the Devil in Graph Spectral Domain for 3D Point Cloud Attacks. In *European Conference on Computer Vision (ECCV)*.

Hu, Q.; Liu, D.; and Hu, W. 2023. Density-Insensitive Unsupervised Domain Adaption on 3D Object Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17556–17566.

Hu, W.; Pang, J.; Liu, X.; Tian, D.; Lin, C.-W.; and Vetro, A. 2021. Graph Signal Processing for Geometric Data and Beyond: Theory and Applications. *IEEE Transactions on Multimedia*.

Huang, Q.; Dong, X.; Chen, D.; Zhou, H.; Zhang, W.; and Yu, N. 2022. Shape-invariant 3D Adversarial Point Clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15335–15344.

Huang, W.; Liu, D.; and Hu, W. 2023. Dense Object Grounding in 3D Scenes. In *Proceedings of the 31st ACM International Conference on Multimedia*, 5017–5026.

Huttenlocher, D. P.; Klanderman, G. A.; and Rucklidge, W. J. 1993. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9): 850–863.

Kingma, D. P.; and Ba, J. 2015. Adam: A Method for Stochastic Optimization. In *ICLR (Poster)*.

Kurakin, A.; Goodfellow, I.; and Bengio, S. 2016. Adversarial machine learning at scale. *arXiv preprint arXiv:1611.01236*.

Liu, D.; and Hu, W. 2021. Imperceptible Transfer Attack and Defense on 3D Point Cloud Classification. *arXiv preprint arXiv:2111.10990*.

Liu, D.; Hu, W.; and Li, X. 2022. Point cloud attacks in graph spectral domain: When 3d geometry meets graph signal processing. *arXiv preprint arXiv:2207.13326*.

Liu, D.; Hu, W.; and Li, X. 2023. Robust Geometry-Dependent Attack for 3D Point Clouds. *IEEE Transactions on Multimedia*.

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.

Ma, C.; Meng, W.; Wu, B.; Xu, S.; and Zhang, X. 2020. Efficient joint gradient based attack against sor defense for 3d point cloud classification. In *Proceedings of the 28th ACM International Conference on Multimedia*, 1819–1827.

Madry, A.; Makelov, A.; Schmidt, L.; Tsipras, D.; and Vladu, A. 2017. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*.

Qi, C. R.; Su, H.; Mo, K.; and Guibas, L. J. 2017. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 652–660.

Ramasinghe, S.; Khan, S.; Barnes, N.; and Gould, S. 2020. Spectral-gans for high-resolution 3d point-cloud generation. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 8169–8176. IEEE.

Rosman, G.; Dubrovina, A.; and Kimmel, R. 2013. Patch-Collaborative Spectral Point-Cloud Denoising. In *Computer Graphics Forum*, volume 32, 1–12. Wiley Online Library.

Shuman, D. I.; Narang, S. K.; Frossard, P.; Ortega, A.; and Vandergheynst, P. 2013. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Process. Mag.*, 30(3): 83–98.

Shuman, D. I.; Wiesmeyr, C.; Holighaus, N.; and Vandergheynst, P. 2015. Spectrum-adapted tight graph wavelet and vertex-frequency frames. *IEEE Transactions on Signal Processing*, 63(16): 4223–4235.

Singh, S. P.; Wang, L.; Gupta, S.; Goli, H.; Padmanabhan, P.; and Gulyás, B. 2020. 3D deep learning on medical images: a review. *Sensors*, 20(18): 5097.

Szegedy, C.; Zaremba, W.; Sutskever, I.; Bruna, J.; Erhan, D.; Goodfellow, I.; and Fergus, R. 2013. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*.

Tao, Y.; Liu, D.; Zhou, P.; Xie, Y.; Du, W.; and Hu, W. 2023. 3DHacker: Spectrum-based Decision Boundary Generation for Hard-label 3D Point Cloud Attack. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 14340–14350.

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.

Tu, C.-C.; Ting, P.; Chen, P.-Y.; Liu, S.; Zhang, H.; Yi, J.; Hsieh, C.-J.; and Cheng, S.-M. 2019. Autozoom: Autoencoder-based zeroth order optimization method for attacking black-box neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 742–749.

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 (TOG)*, 38(5): 1–12.

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 (TPAMI)*.

Wicker, M.; and Kwiatkowska, M. 2019. Robustness of 3d deep learning in an adversarial setting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 11767–11775.

Wu, W.; Qi, Z.; and Fuxin, L. 2019. Pointconv: Deep convolutional networks on 3d point clouds. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, 9621–9630.

Wu, Z.; Duan, Y.; Wang, H.; Fan, Q.; and Guibas, L. J. 2020. If-defense: 3d adversarial point cloud defense via implicit function based restoration. *arXiv preprint arXiv:2010.05272*.

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 (CVPR)*, 1912–1920.

Xiang, C.; Qi, C. R.; and Li, B. 2019. Generating 3d adversarial point clouds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 9136–9144.

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.

Zhang, Q.; Yang, J.; Fang, R.; Ni, B.; Liu, J.; and Tian, Q. 2019a. Adversarial attack and defense on point sets. *arXiv preprint arXiv:1902.10899*.

Zhang, S.; Cui, S.; and Ding, Z. 2020. Hypergraph spectral analysis and processing in 3D point cloud. *IEEE Transactions on Image Processing*, 30: 1193–1206.

Zhang, Y.; Liang, G.; Salem, T.; and Jacobs, N. 2019b. Defense-pointnet: Protecting pointnet against adversarial attacks. In *2019 IEEE International Conference on Big Data (Big Data)*, 5654–5660.

Zhao, Y.; Wu, Y.; Chen, C.; and Lim, A. 2020. On isometry robustness of deep 3d point cloud models under adversarial attacks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1201–1210.

Zheng, T.; Chen, C.; Yuan, J.; Li, B.; and Ren, K. 2019. Pointcloud saliency maps. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1598–1606.

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 Conference on Computer Vision and Pattern Recognition (CVPR)*, 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 *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1961–1970.