GEOMETRY-AWARE GENERATION OF ADVERSARIAL AND COOPERATIVE POINT CLOUDS

Anonymous authors
Paper under double-blind review

ABSTRACT

Recent studies show that machine learning models are vulnerable to adversarial examples. In 2D image domain, these examples are obtained by adding imperceptible noises to natural images. This paper studies adversarial generation of point clouds by learning to deform those approximating object surfaces of certain categories. As 2D manifolds embedded in the 3D Euclidean space, object surfaces enjoy the general properties of smoothness and fairness. We thus argue that in order to achieve imperceptible surface shape deformations, adversarial point clouds should have the same properties with similar degrees of smoothness/fairness to the benign ones, while being close to the benign ones as well when measured under certain distance metrics of point clouds. To this end, we propose a novel loss function to account for imperceptible, geometry-aware deformations of point clouds, and use the proposed loss in an adversarial objective to attack representative models of point set classifiers. Experiments show that our proposed method achieves stronger attacks than existing methods, without introduction of noticeable outliers and surface irregularities. In this work, we also investigate an opposite direction that learns to deform point clouds of object surfaces in the same geometry-aware, but cooperative manner. Cooperatively generated point clouds are more favored by machine learning models in terms of improved classification confidence or accuracy. We present experiments verifying that our proposed objective succeeds in learning cooperative shape deformations.

1 INTRODUCTION

The existence of adversarial examples shows the vulnerability of machine learning models, and triggers a great amount of research attention paid to either attacking and defense studies for safety-critical issues, or robustness analysis on machine learning models themselves. In existing literature, adversarial examples generally satisfy two key properties: visually imperceptible by humans and capable to mislead machine learning models. For 2D adversarial images (Szegedy et al., 2013; Carlini & Wagner, 2017; Madry et al., 2017), their imperceptible nature indicates less sensitivity by humans to the texture changes of the original images, which can be viewed as noises with small magnitudes anchored on RGB values of pixels.

Different from texture changes in adversarial images, adversarial point clouds as approximate shape representations of object surface demand imperceptible shape deformations. Heuristic manners of existing several works are taken to allow shape deformations in arbitrary directions for adversarial attack effects. Consequently, they produce adversarial examples that contain obvious point outliers (Xiang et al., 2019; Liu et al., 2019a), which can thus be easily defended by simple outlier removal method such as SOR (Zhou et al., 2018).

In this work, we aim to adversarially perturbing benign point clouds with awareness of the underlying geometric properties, which can be generally described as some sort of approximate geometric smoothness, *e.g.,* point-wise normals (Wang et al., 2018), curvatures (Bae, 2006), or geodesics for pairs of points (He et al., 2019), in view of the nature of object surface as 2-manifold embedded in the 3D space. To this end, we propose a novel geometry-aware loss in combination with the attack one. Our proposed geometry-aware loss is composed of three terms: the Chamfer Distance (CD) and Hausdorff Distance (HD) terms prevent geometric/topological changes of global shape via constraining point-wise perturbations, while the term encouraging consistency of local curvatures

between the adversarial and benign point clouds achieves smooth local surface deformations. Extensive experiments on the ModelNet40 dataset (Wu et al., 2015) can verify the imperceptibility and effectiveness of geometry-aware adversaries.

Our success in generating less noticeable point clouds inspires us to think of an opposite direction: is it possible to perturb raw point clouds in a cooperative, geometry-aware manner such that the resulting imperceptible perturbations of point clouds can improve classification performance with existing machine learning models? To the best of our knowledge, very few existing works are pursuing this direction, at least in the domain of 3D shape analysis. Technically, the objective of cooperative generation on point clouds is achieved by learning point perturbations in favor of either improved classification confidence or corrected class predictions. In view of their geometry-aware nature, our cooperative point clouds could be practically meaningful in terms of guiding design of certain objects such that they can be better perceived by in-built machine learning systems in physical world. Experiments on the ModelNet40 again verify our motivation on cooperative generation of point clouds.

2 Related Works

A number of adversarial attack algorithms have been proposed on 3D semantic analysis (Xiao et al., 2019; Liu et al., 2019a; Xiang et al., 2019; Yang et al., 2019; Zheng et al., 2018; Wicker & Kwiatkowska, 2019; Liu et al., 2019b). Beyond (Xiao et al., 2019) manipulating both texture and shape on the mesh level to attack image classifiers and detectors, most of adversarial algorithms on point clouds attack 3D classifiers via points attachment (Xiang et al., 2019; Yang et al., 2019), point detachment (Yang et al., 2019; Zheng et al., 2018; Wicker & Kwiatkowska, 2019) or point-wise perturbation (Xiang et al., 2019; Liu et al., 2019a; Yang et al., 2019; Liu et al., 2019b). Existing methods fail to exploit geometric properties during adversarial example generation, and thus have evident outliers and can be easily defended by SOR defense (Zhou et al., 2018), which encourage our adversaries (see Figure 2 and Table 1).

A number of deep models have been proposed for point-based surface reconstruction such as Point Set Generation (PSG) (Fan et al., 2017), the AtlasNet (Groueix et al., 2018b), and Deep Cascade Generation (DCG) (Wang et al., 2019a) from 2D images. These methods exploiting local mesh structure into point clouds generation have inspired our geometry-aware concept, but they are directly regressing from the latent vector encoded from images rather than point-wise deformations in the proposed method.

Data argumentation on point clouds such as jittering, rotation and random scale firstly introduced in (Qi et al., 2017a) shares similar concept as our cooperative deformations, both of which generate new data to boost classification. However, the key differences between our cooperative learning and data argumentation lie in two folds: 1) optimization based vs. simple physical processing; and 2) deformations on the underlying shape vs. point density changes. Simply put, the proposed cooperative deformation can be a promising learning paradigm, while data argumentation is an effective pre-processing step to mitigate sparse sample distributions.

3 Methodology

Our problem setting assumes the availability of a collection of point clouds in the input space $\mathcal{X}$, which are approximate shape representations of object surfaces of certain categories. Each point cloud $\mathcal{P} \in \mathcal{X}$ contains an orderless set of n points $\{\boldsymbol{p}_i\}_{i=1}^n$, with the corresponding label $y \in \mathcal{Y}$ of object category, where any $\boldsymbol{p}$ denotes the coordinates (x, y, z) in the 3D Euclidean space. In this work, we focus on machine learning models of 3D point set classification (Qi et al., 2017a;b; Wang et al., 2019b), which learn a classifier $f : \mathcal{X} \to \mathcal{Y}$ and expect $f(\mathcal{P}) = y$ for any input $\mathcal{P}$ with the true label y.

$\mathcal{P}$ is a discrete approximation of an object surface that satisfies the general properties of *smoothness* and *fairness* (Botsch et al., 2010), which concern with the continuity and variation of (partial) derivatives of a parametric surface function; or in a more intuitive way, the properties concern with the curvatures of local surface patches. Our objective in this work is to obtain from $\mathcal{P}$ a deformed point cloud $\mathcal{P}'$ via learning to perturb individual points $\{\boldsymbol{p}_i\}_{i=1}^n$ of $\mathcal{P}$. Since the deformation is

expected to be imperceptible by humans, we argue that $\mathcal{P}'$ should have the aforementioned surface properties with similar degrees of smoothness/fairness to local surface patches of $\mathcal{P}$, while being globally close to $\mathcal{P}$ as well, measured by certain distance metrics of point sets; otherwise, humans would notice either the global, possibly topological changes of part configuration of the object surface, or those of local surface details. We present shortly in section 3.1 our technical solutions to the above objective of *geometry-aware point perturbations*, and discuss how such solutions can be used to generate either *adversarial* or *cooperative* point clouds in sections 3.2 and 3.3 respectively.

3.1 Geometry-aware Point Perturbations

In 2D image domain, pixel-wise l_p-norms are usually used to constrain the noise magnitudes of adversarial examples (Carlini & Wagner, 2017; Madry et al., 2017). Due to the sharply different data nature of point clouds, we consider the following learning criteria to perturb points $\{\boldsymbol{p}_i\}_{i=1}^n$ of $\mathcal{P}$ in a geometry-aware manner. These criteria are to constrain the deformation magnitudes of the resulting $\mathcal{P}'$, while taking into account prevention of point outliers and smooth regularization of local point neighborhoods. Combined use of these criteria leads to point cloud deformations that are less noticeable by humans, as verified in the comparative experiments in section 4.

Chamfer Distance Given two point sets $\mathcal{P}$ and $\mathcal{P}'$, the Chamfer distance computes

$$C_{\texttt{Chamfer}}(\mathcal{P}', \mathcal{P}) = \frac{1}{n} \sum_{\boldsymbol{p}' \in \mathcal{P}'} \min_{\boldsymbol{p} \in \mathcal{P}} \|\boldsymbol{p}' - \boldsymbol{p}\|_2^2 + \frac{1}{n} \sum_{\boldsymbol{p} \in \mathcal{P}} \min_{\boldsymbol{p}' \in \mathcal{P}'} \|\boldsymbol{p} - \boldsymbol{p}'\|_2^2, \tag{1}$$

which shows that Chamfer distance is symmetric w.r.t. $\mathcal{P}$ and $\mathcal{P}'$. Although the Chamfer distance (1) is not a strict distance function, since the triangle inequality does not hold, it is popularly used in the recent literature of learning based 3D shape generation (Yang et al., 2018; Groueix et al., 2018a; Fan et al., 2017). It measures the distance between the two point sets by averaging over the individual deviation of any $\boldsymbol{p} \in \mathcal{P}$ from $\mathcal{P}'$ and that of any $\boldsymbol{p}' \in \mathcal{P}'$ from $\mathcal{P}$. However, Chamfer distance is less effective in prevention of outlier points in $\mathcal{P}'$, since a small portion of outliers in $\mathcal{P}'$ increases the distance (1) negligibly — one can intuitively think of outliers of a point cloud as those away from the object surface, represented by the point cloud, with relatively large distances. This shortcoming of Chamfer distance motivates us to additionally use the Hausdorff distance as introduced below.

Hausdorff Distance In this work, we use a non-symmetric Hausdorff distance between the two point sets $\mathcal{P}$ and $\mathcal{P}'$ that computes

$$C_{\texttt{Hausdorff}}(\mathcal{P}', \mathcal{P}) = \max_{\boldsymbol{p}' \in \mathcal{P}'} \min_{\boldsymbol{p} \in \mathcal{P}} \|\boldsymbol{p}' - \boldsymbol{p}\|_2^2, \tag{2}$$

since only the deformation of $\mathcal{P}'$ is concerned. As (2) indicates, the Hausdorff distance finds the largest one among the smallest distances of individual $\boldsymbol{p}' \in \mathcal{P}'$ from $\mathcal{P}$. It is thus sensitive to generation of outliers in the resulting $\mathcal{P}'$.

Distances computed by the functions (1) and (2) rely on those between the individual points $\{\boldsymbol{p}_i'\}_{i=1}^n$ and $\{\boldsymbol{p}_i\}_{i=1}^n$, which does not involve geometries of local surface patches associated with these individual points; consequently, the resulting $\mathcal{P}'$ could be close to $\mathcal{P}$ when measured by (1) and/or (2), but changes of geometric details at certain local patches could be clearly visible, causing failure to achieve the objective of imperceptible deformation.

Consistency of Local Curvatures Our way of achieving geometry-aware imperceptibility is to constrain the point cloud deformation such that local patches of the surface approximated by the resulting $\mathcal{P}'$ have curvatures whose magnitudes are similar to those of the corresponding patches of the surface approximated by $\mathcal{P}$. Since computations in this work are conducted on the discrete surface approximation of point clouds, we propose discrete notions that approximately characterize curvatures of local surface patches.

More specifically, for any point $\boldsymbol{p}' \in \mathcal{P}'$, we find its closest point $\boldsymbol{p} \in \mathcal{P}$ by $\boldsymbol{p} = \arg\min_{\boldsymbol{p} \in \mathcal{P}} \|\boldsymbol{p}' - \boldsymbol{p}\|_2$. There exist local point neighborhoods $\mathcal{N}_{\boldsymbol{p}'}' \subset \mathcal{P}'$ and $\mathcal{N}_{\boldsymbol{p}} \subset \mathcal{P}$ respectively associated with $\boldsymbol{p}'$ and $\boldsymbol{p}$, which are obtained in this work by searching k nearest neighbors, suggesting $|\mathcal{N}_{\boldsymbol{p}'}'| = |\mathcal{N}_{\boldsymbol{p}}| = k$. To capture the local geometry of $\mathcal{N}_{\boldsymbol{p}}$, we rely on the following discrete notion

$$\kappa(\boldsymbol{p}; \mathcal{P}) = \frac{1}{k} \sum_{\boldsymbol{q} \in \mathcal{N}_{\boldsymbol{p}}} \left| \langle (\boldsymbol{q} - \boldsymbol{p}) / \|\boldsymbol{q} - \boldsymbol{p}\|_2, \boldsymbol{n}_{\boldsymbol{p}} \rangle \right|, \tag{3}$$

where $\boldsymbol{n_p}$ denotes the unit normal vector of the surface at $\boldsymbol{p}$. The term (3) intuitively measures the averaged angles between the normal vector and the vector defined by pointing $\boldsymbol{p}$ towards each $\boldsymbol{q}$ of its neighboring points. Indeed, since the normal vector $\boldsymbol{n_p}$ is orthogonal to the tangent plane of the surface at $\boldsymbol{p}$, each inner product in (3) characterizes how the normals vary directionally in the local neighborhood $\mathcal{N}_{\boldsymbol{p}}$, thus approximately measuring the local, directional curvature, and an average of $|\mathcal{N}_{\boldsymbol{p}}|$ inner products in (3) approximately measures the local, mean curvature. Note that unit normal vector $\boldsymbol{n_p}$ in (3) can be computed from $\mathcal{N}_{\boldsymbol{p}}$ via eigen-decomposition of the set $\mathcal{N}(\boldsymbol{p})$ (Hoppe et al., 1992). We compute $\kappa'(\boldsymbol{p}';\mathcal{P}')$ in the same way as (3), with a subtle difference that instead of computing $\boldsymbol{n}'_{\boldsymbol{p}'}$ from $\mathcal{N}'_{\boldsymbol{p}'}$, we directly use $\boldsymbol{n_p}$, i.e., the unit normal vector of the point in $\mathcal{P}$ that is closest to $\boldsymbol{p}'$, as a surrogate of $\boldsymbol{n}'_{\boldsymbol{p}'}$, since normal vectors of $\mathcal{P}$ can be pre-computed and efficiently retrieved during the deformation learning.

Given $\kappa'(\boldsymbol{p}';\mathcal{P}',\mathcal{P})$ and $\kappa(\boldsymbol{p};\mathcal{P})$, we use the following criterion to encourage the consistency of local geometries between any $\boldsymbol{p}' \in \mathcal{P}'$ and its closest point $\boldsymbol{p} \in \mathcal{P}$

$$C_{\texttt{Curvature}}(\mathcal{P}',\mathcal{P}) = \frac{1}{n}\sum_{\boldsymbol{p}'\in\mathcal{P}'} \|\kappa'(\boldsymbol{p}';\mathcal{P}',\mathcal{P}) - \kappa(\boldsymbol{p};\mathcal{P})\|_2^2 \ \text{ s.t. } \ \boldsymbol{p} = \arg\min_{\boldsymbol{p}\in\mathcal{P}} \|\boldsymbol{p}' - \boldsymbol{p}\|_2. \tag{4}$$

where we write $\kappa'(\boldsymbol{p}';\mathcal{P}',\mathcal{P})$ since the normal vector involved in its computation is from the corresponding one of $\mathcal{P}$. Note that terms similar to (3) are also used in (Wang et al., 2018; Tang et al., 2019) for single-view surface reconstruction. Our use of the term (3) in (4) is to encourage the consistency of local surface geometries between $\mathcal{P}'$ and $\mathcal{P}$, rather than to directly minimize (3) as in (Wang et al., 2018; Tang et al., 2019).

The Combined Geometry-aware Objective We use the following combined objective to learn the deformed $\mathcal{P}'$ by either adversarial or cooperative perturbations of individual points of $\mathcal{P}$

$$C_{Geometry}(\mathcal{P}',\mathcal{P}) = C_{\texttt{Hausdorff}}(\mathcal{P}',\mathcal{P}) + \alpha \cdot C_{\texttt{Chamfer}}(\mathcal{P}',\mathcal{P}) + \beta \cdot C_{\texttt{Curvature}}(\mathcal{P}',\mathcal{P}), \tag{5}$$

where α and β are the weighting parameters.

3.2 Generation of Adversarial Point Clouds

Assume a point cloud classification model $f : \mathcal{X} \to \mathcal{Y}$. An adversarial example of a point cloud $\mathcal{P}$ is a crafted malicious input $\mathcal{P}'$ to the model $f(\cdot)$, with imperceptible deformation, such that $\mathcal{P}'$ is misclassified by the model. Let the true label of $\mathcal{P}$ be $y \in \mathcal{Y}$, it means that $f(\mathcal{P}') \neq y$. Adversarial examples can be generated either by untargeted or targeted attacks. Similar to (Xiang et al., 2019), we focus in this work on the more difficult task of targeted attack for point cloud data, which generates $\mathcal{P}'$ such that it is classified as a specified class $y' \neq y$, i.e., $f(\mathcal{P}') = y'$.

Among various attack models proposed for 2D image classification (Madry et al., 2017; Carlini & Wagner, 2017), we adopt the state-of-the-art framework of C&W attack (Carlini & Wagner, 2017). Its objective can be generally written as

$$\min_{\boldsymbol{x}'} C_{Adv}(\boldsymbol{x}') + \lambda \cdot \|\boldsymbol{x}' - \boldsymbol{x}\|_p, \tag{6}$$

where $\boldsymbol{x}$ is the benign signal (e.g., an image) and $\boldsymbol{x}'$ is the adversarial example to be optimized. The misclassification loss $C_{Adv}(\boldsymbol{x}')$ is regularized by a λ weighted, l_p-norm based term that constrains the noise magnitude of adversarial $\boldsymbol{x}'$. To specify $C_{Adv}(\boldsymbol{x}')$, we assume a classification model $f(\cdot)$ be implemented as a deep network, and denote as $g(\cdot)$ the function that outputs the network logits, i.e., $g(\cdot)$ includes all layers of the network except the final softmax. Let the targeted label attacking $\boldsymbol{x}$ as y', C&W attack commonly uses a margin based loss function as $C_{Adv}(\boldsymbol{x}') = \max\{\max_{i\neq y'} g_i(\boldsymbol{x}') - g_{y'}(\boldsymbol{x}'), 0\}$.

In this work, we adopt the C&W attack framework, and propose to replace the term of l_p-norm in (6) with our geometry-aware objective (5), in order to generate adversarial point clouds with imperceptible shape deformations, giving

$$\min_{\mathcal{P}'} C_{Adv}(\mathcal{P}') + \lambda \cdot C_{Geo}(\mathcal{P}',\mathcal{P}). \tag{7}$$

The original choice of margin based $C_{Adv}(\mathcal{P}')$ in C&W attack ceases pursuing more adversarial examples once $\max_{i\neq y'} g_i(\mathcal{P}') - g_{y'}(\mathcal{P}') \leq 0$, assuming that further optimization would reduce

the imperceptibility of the resulting $\mathcal{P}'$. Our proposed geometry-aware (5) allows us to take a more aggressive strategy, and we propose to use the following term as our misclassification loss

$$C_{Adv}(\mathcal{P}') = -\log\left(\exp(g_{y'}(\mathcal{P}'))/\sum_i \exp(g_i(\mathcal{P}'))\right). \tag{8}$$

Our proposed objective (7) with term (8) continues to pursue more adversarial, arguably less defendable, point clouds without introducing noticeable shape deformations, as empirically verified by our experiments in section 4.

3.3 Generation of Cooperative Point Clouds

The adversarial objective (7) motivates us to think of an *opposite* direction of geometry-aware point cloud deformations that are less noticable by humans. Specifically, is it possible to perturb points of any $\mathcal{P}$ in a *cooperative, geometry-aware manner* such that the resulting $\mathcal{P}'$ is more favored by machine learning models in terms of improved classification confidence/accuracy? Technically, for a given classifier $f(\cdot)$, this seems to be achieved trivially by optimizing

$$\min_{\mathcal{P}'} C_{Coop}(\mathcal{P}') + \lambda \cdot C_{Geo}(\mathcal{P}', \mathcal{P}), \tag{9}$$

with the cooperative loss term as

$$C_{Coop}(\mathcal{P}') = -\log\left(\exp(g_{y}(\mathcal{P}'))/\sum_i \exp(g_i(\mathcal{P}'))\right), \tag{10}$$

where we simply use the true label y of $\mathcal{P}$ to replace the targeted attack label $y' \neq y$ in (8).

Given knowledge of the true label y of $\mathcal{P}$, the objective (9) seems only learn a deformed $\mathcal{P}'$ that overfits the classifier $f(\cdot)$, which could be practically less meaningful. In this work, we present the following interesting, but less explored investigation based on (9), which suggests that subtle deformations of shape instances of common object categories (e.g., those in ShapeNet (Chang et al., 2015)) could lead to practical meanings in terms of being better perceived by existing 3D point set classification models (Qi et al., 2017a;b; Wang et al., 2019b).

Assume we are given a set of m point clouds $\{\mathcal{P}_i\}_{i=1}^m$ of different object categories. We take the following procedure on $\{\mathcal{P}_i\}_{i=1}^m$.

1. We divide $\{\mathcal{P}_i\}_{i=1}^m$ evenly into s subsets, with consideration of class balance.
2. We use the first $s-1$ subsets as training data to train a classifier $f^s(\cdot)$, and use the objective (9) to cooperatively deform point clouds in the sth subset (the validation set) w.r.t. $f^s(\cdot)$.
3. We perform s times of step 2 by using each of the s subsets as the validation set, obtaining $f^i(\cdot)$, $i = 1, \ldots, s$, and the aggregated collection of deformed $\{\mathcal{P}'_i\}_{i=1}^m$.

The above procedure can optionally be conducted for multiple times to further increase the degree of deformations, with awareness of geometric imperceptibility. We use the obtained $\{\mathcal{P}'_i\}_{i=1}^m$ in a standard training-and-test setting of point set classification. That is, we split $\{\mathcal{P}'_i\}_{i=1}^m$ as training and test data, whose indices are the same as those of the original $\{\mathcal{P}_i\}_{i=1}^m$, and train a new classifier $\hat{f}(\cdot)$ to evaluate $\hat{f}(\cdot)$ on the split test data.

The obtained $\{\mathcal{P}'_i\}_{i=1}^m$ are indeed cooperative when performance of $\hat{f}(\cdot)$ is improved over that of the original classifier $f(\cdot)$. Experiments in Table 2 show that it is just the case. In fact, we have further conducted cross-model experiments by learning $\{\mathcal{P}'_i\}_{i=1}^m$ via the above procedure with PointNet (Qi et al., 2017a), and train and evaluate $\hat{f}(\cdot)$ of other representative point set classifiers (e.g., PointNet++ (Qi et al., 2017b) and DGCNN (Wang et al., 2019b)). We have also conducted experiments by converting the obtained $\{\mathcal{P}'_i\}_{i=1}^m$ to their mesh representations (Edelsbrunner et al., 1983), and uniformly re-sample points from the meshes to form $\{\widetilde{\mathcal{P}'}_i\}_{i=1}^m$; classification performance again improves by training and testing on $\{\widetilde{\mathcal{P}'}_i\}_{i=1}^m$, confirming that we are indeed cooperatively deforming the underlying object surfaces. These experiments are also presented in Table 2.

While our above investigations remain in the digital space, cooperative deformations of point clouds could be physically achieved either by changes of shape design for common rigid objects, or by designing special devices that actively control the reflections of laser pulses from LiDAR (Cao et al., 2019). Physical-world adversarial attacks (Athalye et al., 2017; Eykholt et al., 2018) are actively pursued recently in the 2D image domain. Deformations of geometry-aware imperceptibility make sense here since we are not to change the geometric and/or topological surface structures of the objects; otherwise the deformations would have influence on human perception or functional attributes of the objects.

4 EXPERIMENTS

Dataset We use point clouds of object instances from ModelNet40 (Wu et al., 2015) to evaluate our proposed algorithms. The dataset consists of $12,311$ CAD models belonging to 40 semantic categories. For each CAD model, $1,024$ points are uniformly sampled from its surface as the working point clouds, which are re-scaled into a unit ball following (Qi et al., 2017b).

Models and Protocols We evaluate the adversarial and cooperative generation of point clouds based on three representative classifiers, namely PointNet (Qi et al., 2017a), PointNet++ (Qi et al., 2017b), and DGCNN (Wang et al., 2019b). In our adversarial setting, we follow the official data split for 3D classification (Qi et al., 2017a;b) and use $9,843$ samples of point clouds for training classifiers that are to be attacked, and for testing, we follow (Xiang et al., 2019) and randomly select 25 samples from each testing set of the top-10 object classes (ordered by sample sizes of these classes, giving rise to *airplane, bed, bookshelf, bottle, chair, monitor, sofa, table, toilet, vase*); all adversarial attack algorithms are evaluated with a white-box, targeted attack protocol. In our cooperative setting, we set the subset number as $s = 5$.

Evaluation Metrics We respectively use the `attack success rate` (i.e., misclassification rate) and `classification accuracy` to evaluate the algorithmic effectiveness in our adversarial and cooperative settings. For both metrics, the higher, the better.

Implementation Details We set $k = 16$ to define local point neighborhoods for computation of normals and approximate curvatures in (3). We fix $\alpha = 0.1$ and $\beta = 1.0$ for our proposed geometry-aware loss (5), and the trade-off parameter λ in (7) and (9) is optimized via 10-step binary search. We use Adam optimizer (Kingma & Ba, 2014) to train networks, and set its learning rate as 0.01.

4.1 EVALUATION ON GENERATION OF ADVERSARIAL POINT CLOUDS

In this section, we evaluate the efficacy of our proposed geometry-aware objective (7) for generation of adversarial point clouds. We dub our method as GeoAwareAdv and compare with the method (Xiang et al., 2019), which is among the only few existing works (Liu et al., 2019a;b) addressing adversarial point clouds and takes the same white-box, targeted attack as our method does. Objective of (Xiang et al., 2019) strictly follows the C&W attack (6) (Carlini & Wagner, 2017), with a margin-based misclassification loss regularized by l_2-norm constraining magnitude of point perturbations. We also compare with a degenerate version of our method, dubbed GeoDegenerateAdv, which replaces the geometry-aware term (5) with a similar l_2-norm as in (Xiang et al., 2019). These experiments are conducted using PointNet (Qi et al., 2017a) as the model of classifier.

Ablation Studies To investigate how different terms in our proposed geometry-aware loss (5), we conduct ablation studies by removing each of them from (5), and use the remaining ones into the adversarial objective (7) for point cloud generation. Figure 1) shows that these terms all together contribute to smooth and outlier-free results. Without using CD, the shape deformations go to unexpected twists. The use of HD and curvature terms largely removes generation of outliers, with the later one further improving the surface smoothness.

Comparative Results We report comparative results of our GeoAwareAdv, GeoDegenerateAdv, and the method (Xiang et al., 2019) in Table 1 and Figure 2. Table 1 compares the attack success rates under two defense methods of SOR (Zhou et al., 2018) and Random Romoval (RR). SOR is the current state-of-the-art method to defend attacking of point clouds, which works by dropping certain points from a point cloud based on statistical analysis. RR simply drops points from a point cloud randomly. Based on our stronger adversarial term (8), both of our methods achieve better success

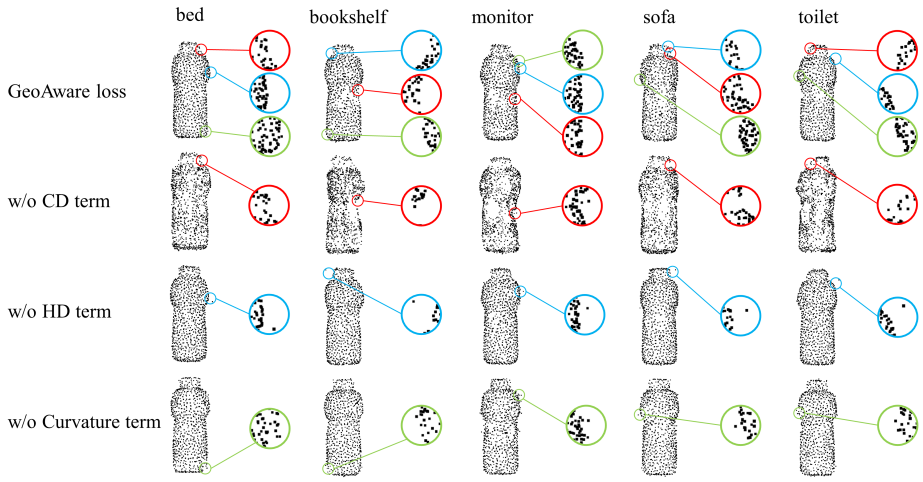

Figure 1: Ablation studies on the effectiveness of different terms in our geometry-aware loss (5) for generation of adversarial point clouds. See the main text for experiment settings of these results.

rates than those of (Xiang et al., 2019) under different dropping settings of the defenses SOR and RR. More significantly, the stronger attacking effects of our GeoAwareAdv are achieved without introduction of noticeable outliers and surface irregularities, as shown in Figure 2; in contrast, the use of simple l_2 norm in (Xiang et al., 2019) produces its adversarial results with clear outliers. Advantages of our GeoAwareAdv are essentially due to the use of HD and curvature terms in our geometry-aware loss (5).

Table 1: Success rates (%) of different methods attacking the classifier of PointNet (Qi et al., 2017a) under defenses of SOR (Zhou et al., 2018) and Random Romoval (RR). We report results with different numbers of dropping points for the two defense methods. For all methods, attack success rates are 100% without defense.

Method	Defense method	Attack success rate (%) defense by dropping different numbers of points								
		1	2	4	8	16	32	64	128	256
The method (Xiang et al., 2019)	RR	91.78	87.02	80.71	70.44	58.31	47.07	34.89	23.82	16.00
GeoDegenerateAdv (our vairant)	RR	98.17	96.63	94.61	90.06	84.89	76.65	61.32	38.64	17.38
GeoAwareAdv (our method)	RR	**99.87**	**99.78**	**99.60**	**99.38**	**98.57**	**97.77**	**94.65**	**79.97**	**47.48**
The method (Xiang et al., 2019)	SOR	12.71	7.29	3.69	1.96	1.51	1.24	0.71	0.67	1.56
GeoDegenerateAdv (our vairant)	SOR	29.37	13.73	6.02	3.25	1.83	0.94	0.89	0.80	1.25
GeoAwareAdv (our method)	SOR	**98.84**	**97.86**	**96.08**	**92.74**	**87.62**	**77.06**	**61.29**	**41.38**	**28.20**

4.2 Evaluation on Generating Geometry-aware Cooperative Point Clouds

Table 2: Classification accuracy (%) of different networks trained on different source of data, including vanilla point clouds, cooperative point clouds and its variant generated based on PointNet. The state-of-the-art Geo-CNN (Lan et al., 2019) achieves the classification accuracy of 93.90%.

Data	Classification accuracy (%)		
	PointNet	PointNet++	DGCNN
Vanilla point clouds	87.96	89.50	90.64
Cooperative point clouds (ours)	96.88	94.32	96.27
Re-sampling variant (ours)	94.81	92.78	94.12

Comparative Evaluation Cooperative point clouds are generated based on the PointNet via optimizing objective function (9), and a variant can be re-sampled point clouds from its mesh surface. A number of recent classifiers are trained and evaluated on vanilla and cooperative point clouds respectively, whose results are reported in Figure 3 and Table 2. It can be observed from Table 2, both cooperative point clouds significantly outperform the vanilla ones by a large margin, *i.e.,* increase at

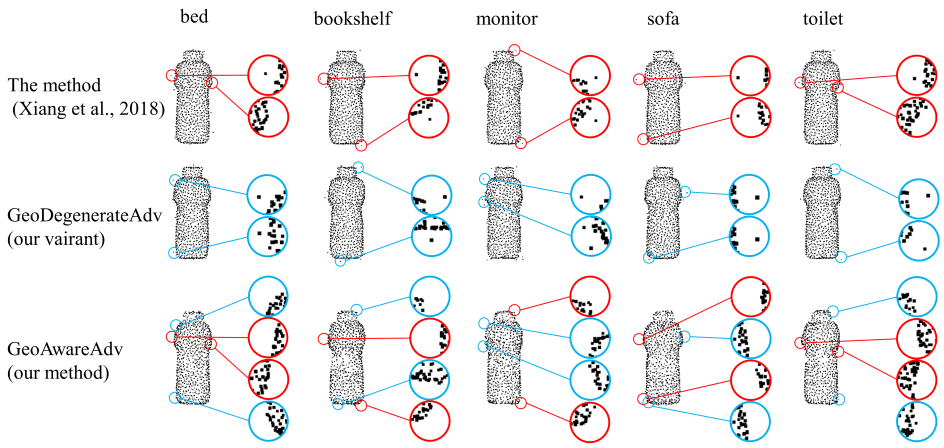

Figure 2: Comparative quality results of different methods successfully attacking the classifier of PointNet (Qi et al., 2017a).

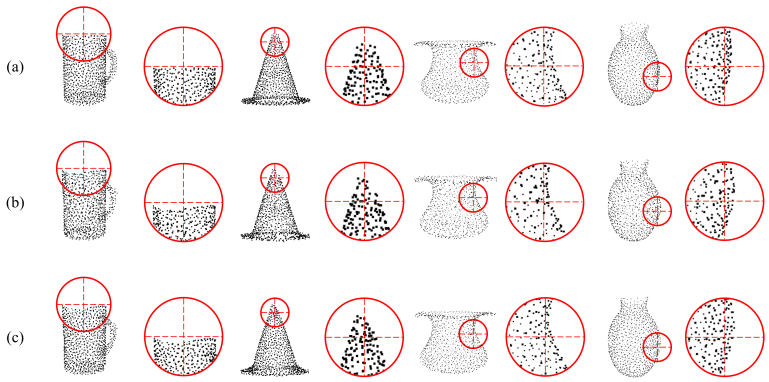

Figure 3: Visualization of cooperative point clouds generated based on PointNet. (a) Vanilla point clouds, (b) Cooperative point clouds, and (c) Uniformly re-sampling on the mesh surface of (b).

least 6% on classification accuracy. Moreover, performance gap between two types of cooperative point clouds can be explained by both approximation errors of mesh reconstruction and the changes on point distribution, which encourage us to consider adversarial shape deformations on the mesh level.

Cross-model experiments We conduct one more cross-model experiment to test transferable characteristics of cooperative point clouds across classifiers, *i.e.,* generating cooperative examples on PointNet, while train and evaluate other classifiers. Results in Table 2 show consistent improvement on classification accuracy achieved for different neural classifiers, which reveal that cooperative shape deformations on point clouds can capture discriminative geometric patterns favor for semantic object classification, which are independent on neural models.

5 Conclusions

In this paper, we propose a compact geometry-aware loss to constrain point-wise imperceptible perturbations to achieve similar geometric smoothness and fairness of local patches and their global topological configuration in deformed point clouds as the original ones. Experiment results reveal the rationale of three components in the proposed geometry-aware loss for generating adversaries without evident outliers and shape irregularities, which can thus achieve adversarial effects even confronting the SOR defense. Moreover, cooperative generation on point clouds also demonstrates its positive effects on improving classification with aware of geometric properties.

REFERENCES

Anish Athalye, Logan Engstrom, Andrew Ilyas, and Kevin Kwok. Synthesizing robust adversarial examples. *arXiv preprint arXiv:1707.07397*, 2017.

Kwang-Ho Bae. *Automated registration of unorganised point clouds from terrestrial laser scanners*. PhD thesis, Curtin University, 2006.

Mario Botsch, Leif Kobbelt, Mark Pauly, Pierre Alliez, and Bruno Lévy. *Polygon mesh processing*. AK Peters/CRC Press, 2010.

Yulong Cao, Chaowei Xiao, Benjamin Cyr, Yimeng Zhou, Won Park, Sara Rampazzi, Qi Alfred Chen, Kevin Fu, and Z Morley Mao. Adversarial sensor attack on lidar-based perception in autonomous driving. *arXiv preprint arXiv:1907.06826*, 2019.

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

Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. ShapeNet: An Information-Rich 3D Model Repository. Technical Report arXiv:1512.03012 [cs.GR], Stanford University — Princeton University — Toyota Technological Institute at Chicago, 2015.

Herbert Edelsbrunner, David Kirkpatrick, and Raimund Seidel. On the shape of a set of points in the plane. *IEEE Transactions on information theory*, 29(4):551–559, 1983.

Kevin Eykholt, Ivan Evtimov, Earlence Fernandes, Bo Li, Amir Rahmati, Chaowei Xiao, Atul Prakash, Tadayoshi Kohno, and Dawn Song. Robust physical-world attacks on deep learning visual classification. June 2018.

Haoqiang Fan, Hao Su, and Leonidas J Guibas. 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*, pp. 605–613, 2017.

Thibault Groueix, Matthew Fisher, Vladimir G. Kim, Bryan Russell, and Mathieu Aubry. AtlasNet: A Papier-Mâché Approach to Learning 3D Surface Generation. 2018a.

Thibault Groueix, Matthew Fisher, Vladimir G. Kim, Bryan C. Russell, and Mathieu Aubry. A papier-mache approach to learning 3d surface generation. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun 2018b. doi: 10.1109/cvpr.2018.00030. URL http://dx.doi.org/10.1109/cvpr.2018.00030.

Tong He, Haibin Huang, Li Yi, Yuqian Zhou, Chihao Wu, Jue Wang, and Stefano Soatto. Geonet: Deep geodesic networks for point cloud analysis. pp. 6888–6897, 2019.

Hugues Hoppe, Tony DeRose, Tom Duchamp, John McDonald, and Werner Stuetzle. *Surface reconstruction from unorganized points*. ACM, 1992.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Shiyi Lan, Ruichi Yu, Gang Yu, and Larry S Davis. Modeling local geometric structure of 3d point clouds using geo-cnn. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 998–1008, 2019.

Daniel Liu, Ronald Yu, and Hao Su. Extending adversarial attacks and defenses to deep 3d point cloud classifiers. pp. 2279–2283, 09 2019a. doi: 10.1109/ICIP.2019.8803770.

Daniel Liu, Ronald Yu, and Hao Su. Adversarial point perturbations on 3d objects, 2019b.

Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*, 2017.

Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 652–660, 2017a.

Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in Neural Information Processing Systems*, pp. 5099–5108, 2017b.

Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*, 2013.

Jiapeng Tang, Xiaoguang Han, Junyi Pan, Kui Jia, and Xin Tong. A skeleton-bridged deep learning approach for generating meshes of complex topologies from single rgb images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4541–4550, 2019.

Kaiqi Wang, Ke Chen, and Kui Jia. Deep cascade generation on point sets. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pp. 3726–3732. International Joint Conferences on Artificial Intelligence Organization, 7 2019a. doi: 10.24963/ijcai.2019/517. URL https://doi.org/10.24963/ijcai.2019/517.

Nanyang Wang, Yinda Zhang, Zhuwen Li, Yanwei Fu, Wei Liu, and Yu-Gang Jiang. Pixel2mesh: Generating 3d mesh models from single rgb images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 52–67, 2018.

Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (TOG)*, 2019b.

Matthew Wicker and Marta Kwiatkowska. Robustness of 3d deep learning in an adversarial setting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 11767–11775, 2019.

Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *CVPR*, 2015.

Chong Xiang, Charles R Qi, and Bo Li. Generating 3d adversarial point clouds. pp. 9136–9144, 2019.

Chaowei Xiao, Dawei Yang, Bo Li, Jia Deng, and Mingyan Liu. Meshadv: Adversarial meshes for visual recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6898–6907, 2019.

Jiancheng Yang, Qiang Zhang, Rongyao Fang, Bingbing Ni, Jinxian Liu, and Qi Tian. Adversarial attack and defense on point sets. *arXiv preprint arXiv:1902.10899*, 2019.

Yaoqing Yang, Chen Feng, Yiru Shen, and Dong Tian. Foldingnet: Point cloud auto-encoder via deep grid deformation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 206–215, 2018.

Tianhang Zheng, Changyou Chen, Junsong Yuan, Bo Li, and Kui Ren. Pointcloud saliency maps. 2018.

Hang Zhou, Kejiang Chen, Weiming Zhang, Han Fang, Wenbo Zhou, and Nenghai Yu. Dup-net: Denoiser and upsampler network for 3d adversarial point clouds defense. 2018.

A APPENDIX

A.1 VISUALIZATION OF MORE GEOMETRY-AWARE ADVERSARIAL EXAMPLES.

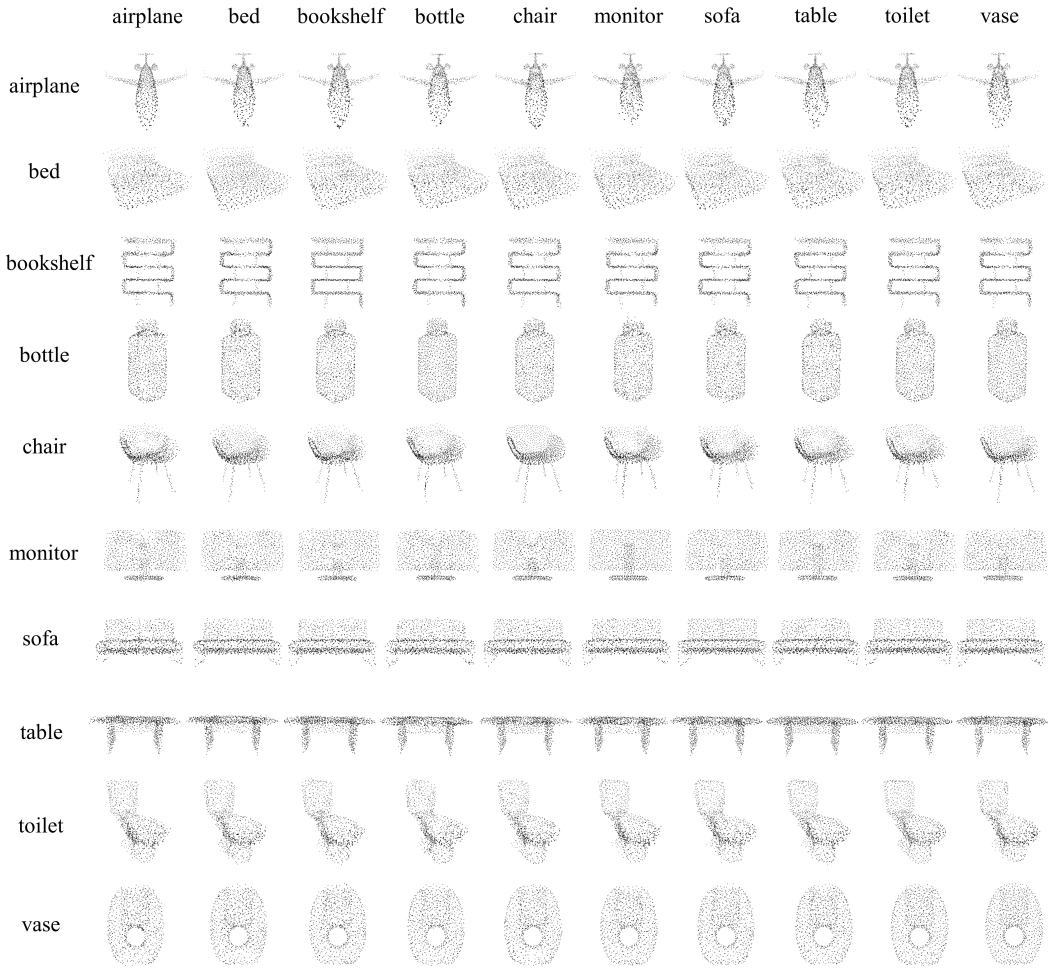

Figure 4: Visualization of our proposed geometry-aware adversarial examples based on PointNet. We show all kinds of our selected categories here targeted on all the other classes. The diagonal examples are corresponding benign ones.

A.2 VISUALIZATION OF MORE GEOMETRY-AWARE COOPERATIVE EXAMPLES.

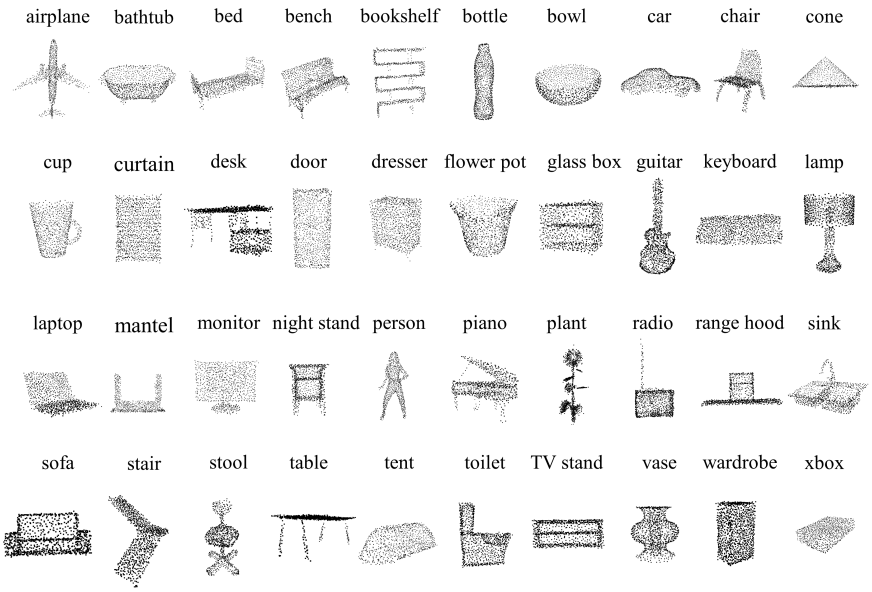

Figure 5: Visualization of our proposed geometry-aware cooperative examples based on PointNet. We show all kinds of categories in the dataset.

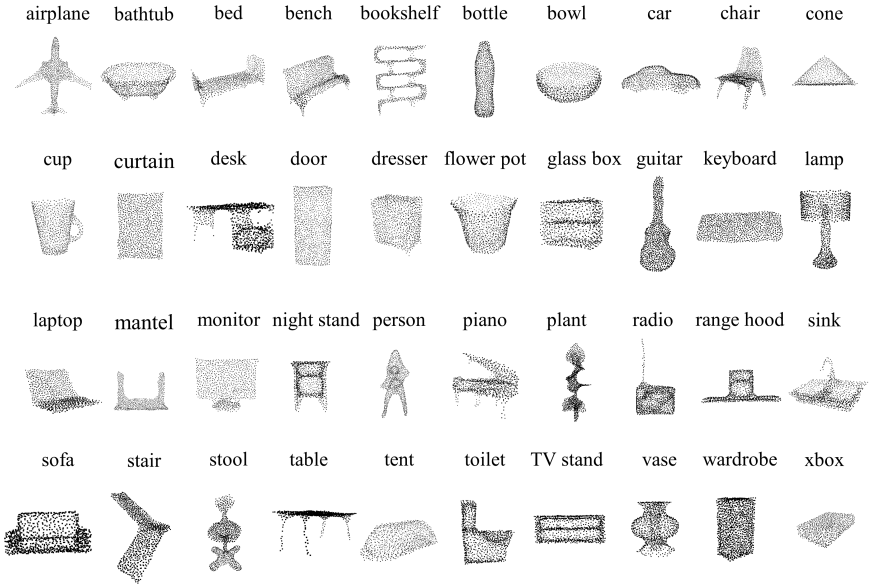

Figure 6: Visualization of our proposed resampling-after-reconstruction on geometry-aware cooperative examples based on PointNet. We show all kinds of categories in the dataset.