TPC: Transformation-Specific Smoothing for Point Cloud Models

Wenda Chu 1  Linxi Li 2  Bo Li 2

Abstract

Point cloud models with neural network architectures have achieved great success and have been widely used in safety-critical applications, such as Lidar-based recognition systems in autonomous vehicles. However, such models are shown to be vulnerable to adversarial attacks that aim to apply stealthy semantic transformations such as rotation and tapering to mislead model predictions. In this paper, we propose a transformation-specific smoothing framework TPC, which provides tight and scalable robustness guarantees for point cloud models against semantic transformation attacks. We first categorize common 3D transformations into three categories: additive (e.g., shearing), composable (e.g., rotation), and indirectly composable (e.g., tapering), and we present generic robustness certification strategies for all categories respectively. We then specify unique certification protocols for a range of specific semantic transformations and their compositions. Extensive experiments on several common 3D transformations show that TPC significantly outperforms state of the art. For example, our framework boosts the certified accuracy against twisting transformation along the z-axis (within $\pm 20^\circ$) from 20.3% to 83.8%. Codes and models are available at https://github.com/Qianhewu/Point-Cloud-Smoothing.

1. Introduction

Deep neural networks that take point clouds data as inputs (point cloud models) are widely used in computer vision (Qi et al., 2017; Wang et al., 2019; Zhou & Tuzel, 2018) and autonomous driving (Li, 2017; Chen et al., 2017; 2020). For instance, modern autonomous driving systems are equipped with LiDAR sensors that generate point cloud inputs to feed into point cloud models (Cao et al., 2019). Despite their successes, point cloud models are shown to be vulnerable to adversarial attacks that mislead the model’s prediction by adding stealthy perturbations to point coordinates or applying semantic transformations (e.g., rotation, shearing, tapering) (Cao et al., 2019; Xiang et al., 2019; Xiao et al., 2019; Fang et al., 2021). Specifically, semantic transformation based attacks can be easily operated on point cloud models by simply manipulating sensor positions or orientations (Cao et al., 2019; 2021). These attacks may lead to severe consequences such as forcing an autonomous driving vehicle to steer toward the cliff (Pei et al., 2017). A wide range of empirical defenses against these attacks has been studied (Zhu et al., 2017; Aoki et al., 2019; Sun et al., 2020; 2021b; a), while defenses with robustness guarantees are less explored (Lorenz et al., 2021; Fischer et al., 2021) and provides loose and less scalable certification.

In this paper, we propose a transformation-specific smoothing framework TPC that provides tight and scalable probabilistic robustness guarantees for point cloud models against a wide range of semantic transformation attacks. We first categorize common semantic transformations into three categories: additive (e.g., shearing), composable (e.g., rotation), and indirectly composable (e.g., tapering). For each category, our framework proposes novel smoothing and robustness certification strategies. With TPC, for each common semantic transformation or composition, we prove the corresponding robustness conditions that yield efficient and tight robustness certification.
We illustrate our TPC works (Szegedy et al., 2013; Tramer et al., 2020; Eykholt et al., 2018; Qiu et al., 2020; Li et al., 2020a; Zhang et al., 2022; Li et al., 2021a; Xiao et al., 2018), efforts have been made toward certifying and improving the certified robustness of DNNs (Cohen et al., 2019; Li et al., 2020b; 2019). Existing works mainly focus on image classification models against \( \ell_p \) bounded perturbations. For such threat models, the robustness certification can be roughly divided into two types: deterministic and probabilistic, where deterministic methods are mainly based on feasible region relaxation (Wong & Kolter, 2018; Weng et al., 2018; Zhang et al., 2018), abstract interpretation (Mirman et al., 2018; Singh et al., 2019), or Lipschitz bounds (Tsuzuku et al., 2018; Zhang et al., 2021); and probabilistic methods provide certification that holds with high probability, and they are mainly based on randomized smoothing (Cohen et al., 2019; Yang et al., 2020). Along with the certification methods, there are several robust training methods that aim to train DNNs to be more certifiably robust (Wong et al., 2018; Li et al., 2019; Salman et al., 2019).

Semantic Transformation Attacks and Certified Robustness on Point Cloud Models. Our TPC aims to generalize the model robustness certification to point cloud models against a more generic family of practical attacks – semantic transformation attacks. The semantic transformation attacks have been shown feasible for both image classification models and point cloud models (Hendrycks & Dietterich, 2018; Cao et al., 2019; Xiang et al., 2019), and certified robustness against such attacks is mainly studied for 2D image classification models (Balanuvić et al., 2019; Fischer et al., 2020; Li et al., 2021b). For point cloud models, some work considers point addition and removal attacks (Xiang et al., 2019) and provides robustness certification against such attacks (Liu et al., 2021). The randomized smoothing technique is applied to certify point cloud models on segmentation tasks by (Fischer et al., 2021). However, their certification only covers points edition with bounded \( \ell_2 \) norm and rotations along a fixed axis. For general semantic transformation attacks, to the best of our knowledge, the only work that can provide robustness certification against them is DeepG3D (Lorenz et al., 2021), which is based on linear bound relaxations. In this work, we derive novel randomized smoothing techniques on point clouds models to provide probabilistic robustness certification against semantic transformations. In Section 5, we conduct extensive experiments to show that our framework is more general and provides significantly higher certified robust accuracy than DeepG3D under different settings.

2. Semantic Transformation Attacks on Point Cloud Models

We denote the space of point cloud inputs as \( \mathcal{X} = \mathbb{R}^{N \times 3} \) where \( N \) is the number of points the point cloud has. A point cloud with \( N \) points is denoted by \( x = \{p_i\}_{i=1}^{N} \) with \( p_i \in \mathbb{R}^3 \). Unless otherwise noted, we assume all point...
cloud inputs are normalized to be within a unit ball, i.e., \( \|p_x\|_2 \leq 1 \). We mainly consider classification tasks on the point clouds level. Such classification task is defined with a set of labels \( Y = \{1, \ldots, C\} \) and a classifier is defined by a deterministic function \( h : \mathcal{X} \rightarrow Y \). More extensions are in Section 5.2.4.

### 2.1. Semantic Transformations

Semantic transformations on point cloud models are defined as functions \( \phi : \mathcal{X} \times \mathcal{Z} \rightarrow \mathcal{X} \) where \( \mathcal{Z} \) is the parameter space for transformations. The semantic transformations discussed in this paper may change the three-dimensional coordinate of each point (usually in a point-wise manner) but do not increase or decrease the number of points. In Section 3, we will further categorize different semantic transformations based on their intrinsic properties.

### 2.2. Threat Model and Certification Goal

We consider semantic transformation attacks that an adversary can apply arbitrary semantic transformations to the point cloud data according to a parameter \( z \in \mathcal{Z} \). The adversary then performs evasion attacks to a classifier \( h \) with the transformed point cloud \( \phi(x, z) \). The attack is successful if \( h \) predicts different labels on \( x \) and \( \phi(x, z) \).

The main goal of this paper is to certify the robustness of point cloud classifiers against all semantic attacks within a certain transformation parameter space. Formally, our certification goal is to find a subset \( Z_{\text{robust}} \subseteq \mathcal{Z} \) for a classifier \( h : \mathcal{X} \rightarrow Y \), such that

\[
\forall x \in \mathcal{X} \quad \exists z \in Z_{\text{robust}} \quad h(x) = h(\phi(x, z)).
\]

### 3. Transformation Specific Smoothing for Point Cloud Models

In this section, we first introduce the proposed randomized smoothing techniques for general semantic transformations. Next, we categorize the semantic transformations into three types: composable, additive, and indirectly composable transformations. We then derive the smoothing-based certification strategies for each type.

#### 3.1. Transformation Specific Smoothed Classifier

We apply transformation-specific smoothing to an arbitrary base classifier \( h : \mathcal{X} \rightarrow Y \) to construct a smoothed classifier. Specifically, the smoothed classifier \( g \) predicts the class with the highest conditional probability when the input \( x \) is perturbed by some random transformations.

**Definition 1** (Transformation Specific Smoothed Classifier). Let \( \phi : \mathcal{X} \times \mathcal{Z} \rightarrow \mathcal{X} \) be a semantic transformation. Let \( \epsilon \) be a random variable in the parameter space \( \mathcal{Z} \). Suppose we have a base classifier that learns a conditional probability distribution, \( h(x) = \arg \max_{y \in Y} p(y|x) \). Applying transformation specific smoothing to the base classifier \( h \) yields a smoothed classifier \( g : \mathcal{X} \rightarrow \mathcal{Z} \), which predicts

\[
g(x; \epsilon) = \arg \max_{y \in Y} q(y|x, \epsilon) = \arg \max_{y \in Y} E_{\epsilon} (p(y|\phi(x, \epsilon))).
\]

We recall the theorem proved (Li et al., 2021b) in Appendix A.1, which provides a generic certification bound for the transformation-specific smoothed classifier based on the Neyman-Pearson lemma (Neyman & Pearson, 1933).

Next, we will categorize the semantic transformations into different categories based on their intrinsic properties as shown in Figure 3, and we will then discuss the certification principles for each specific category.

#### 3.2. Composable Transformations

A set of semantic transformations is called composable if it is closed under composition.

**Definition 2.** A set of semantic transformations defined by \( \phi : \mathcal{X} \times \mathcal{Z} \rightarrow \mathcal{X} \) is called composable if for any \( \alpha \in \mathcal{Z} \) there exists an injective and continuously differentiable function \( \gamma_{\alpha} : \mathcal{Z} \rightarrow \mathcal{Z} \) with non-vanishing Jacobian, such that

\[
\phi(\phi(x, \alpha), \beta) = \phi(x, \gamma_{\alpha}(\beta)), \quad \forall x \in \mathcal{X}, \beta \in \mathcal{Z}.
\]

Common semantic transformations for point cloud data that are composable include rotation, shearing along a fixed axis, and twisting along a fixed axis. For example, according to...
to Euler’s rotation theorem, we can always find another rotation $\gamma_\alpha(\beta) \in Z$ for any two rotations $\alpha, \beta \in Z$. Therefore, rotations belong to the composable transformations as shown in Figure 2 (b).

In general, composable transformations can be certified against by Theorem 6 stated in Appendix A.1. For a classifier $g(x; \epsilon_0)$ smoothed by the composable transformation, we can simply replace the random variable $\epsilon_1$ by $\gamma_\alpha(\epsilon)$ in Theorem 6 to derive a robustness certification condition. However, some composable transformations with complicated $\gamma_\alpha(\beta)$ function result in intractable distribution for $\epsilon_1$, causing difficulties for the certification. Therefore, we focus on a subset of composable transformations, called additive transformations, for which it is straightforward to certify by applying Theorem 6.

3.3. Additive Transformations

We are particularly interested in a subset of composable transformations that the function $\gamma_\alpha : Z \rightarrow Z$ defined in Definition 2 satisfies $\gamma_\alpha(\beta) = \alpha + \beta$ as shown in Figure 2 (a) where the one step rotation above is equivalent to the two step transformations below.

Definition 3. A set of semantic transformations $\phi : X \times Z \rightarrow X$ is called additive if

$$\phi(\phi(x, \alpha), \beta) = \phi(x, \alpha + \beta), \forall x \in X, \alpha, \beta \in Z. \quad (4)$$

An additive transformation must be composable, but the reverse direction does not hold. For instance, the set of general rotations from the SO(3) group is composable, but not additive. Rotating $10^\circ$ along the $x$ axis first and then $10^\circ$ along the $y$ axis does not equal rotating $20^\circ$ along the $xy$ axis. Thus, general rotations cannot be categorized as an additive transformation. However, rotating along any fixed axis is additive. Based on this observation, we discuss $z$-rotation (i.e., rotation along $z$ axis) and general rotations separately in Section 4.

All additive transformations can be certified following the same protocol derived from Theorem 6. We omit Corollary 3 for certified robustness against additive transformation in Appendix A.1.

3.4. Indirectly Composable Transformations

As shown in Section 3.2, composable transformations can be certified following Theorem 6. However, some semantic transformations of point clouds do not have such closure property under composition and thus do not fall in this category as shown in Figure 2 (c). For example, the tapering transformation which we will discuss in Section 4 is not composable and cannot be certified directly using Theorem 6. This kind of transformation is therefore categorized as a more general class called indirectly composable transformations.

Definition 4. A set of transformations $\phi : X \times Z \rightarrow X$ is indirectly composable if there is a set of composable transformations $\psi : X \times Z \rightarrow X$, such that for any $x \in X$, there exists a function $\delta_\alpha : Z \times Z \rightarrow Z$ with

$$\phi(x, \alpha) = \psi(\phi(x, \beta), \delta_\alpha(\alpha, \beta)), \forall \alpha, \beta \in Z. \quad (5)$$

This definition involves more kinds of transformations, since we can choose the transformation $\psi$ as $\psi(x, \delta) = x + \delta$ and let $\delta_\alpha(\alpha, \beta) = \phi(x, \alpha) - \phi(x, \beta)$. This specific assignment of $\psi$ leads to a useful theorem (Li et al., 2021b) in Appendix A.2, which we use to certify against some more complicated transformations, such as tapering in Section 4.

The theorem states that the overall robustness can be guaranteed if we draw multiple samples within the parameter space and certify the neighboring distribution of each sampled parameter separately.

4. Certifying Point Cloud Models against Specific Semantic Transformations

In this section, we certify the point cloud models against several specific semantic transformations that are commonly seen for point cloud data, including rotation, shearing, twisting, and tapering. We do not analyze scaling and translation, since the point cloud models are usually inherently invariant to them due to the standard pre-processing pipeline (Qi et al., 2017). For each transformation, we specify a corresponding certification protocol based on the categorization they belong to introduced in Section 3.

4.1. Rotation, Shearing, and Twisting along a Fixed Axis

Rotation, shearing, and twisting are all common 3D transformations that are performed pointwise on point clouds. Without loss of generality, we consider performing these transformations along the $z$-axis.

Specifically, we define $z$-shear transformation as $\phi_{Sz} : X \times Z \rightarrow X$ where $X = \mathbb{R}^{N \times 3}$ is the space of the point clouds with $N$ points and $Z = \mathbb{R}^2$ is the parameter space. For any $z = (\theta_1, \theta_2)$, $z$-shear acting on a point cloud $x \in X$ with $x = \{p_i\}_{i=1}^N$ yields $(p_i(z)) = (x_i + \theta_1 z_i, y_i, + \theta_2 z_i, z_i). \quad (6)$

$z$-twist transformation $\phi_{Tz} : X \times Z \rightarrow X$ is defined similarly but with parameter space $Z = \mathbb{R}$. For any $\theta \in Z$ and $p_i = (x_i, y_i, z_i)^T$,

$$\phi_{Tz}(p_i, \theta) = \begin{pmatrix} x_i \cos(\theta z_i) - y_i \sin(\theta z_i) \\ x_i \sin(\theta z_i) + y_i \cos(\theta z_i) \\ z_i \end{pmatrix}. \quad (7)$$

Note that $z$-rotation, $z$-shear and $z$-twist are all additive transformations. Hence, we present the following corollary based on Corollary 3, which certifies the robustness of point clouds models with bounded $\ell_2$ norm for the transformation parameters.
Corollary 1. Suppose a classifier \( g : \mathcal{X} \to \mathcal{Y} \) is smoothed by a transformation \( \phi : \mathcal{X} \times Z \to \mathcal{X} \) with \( \epsilon \sim N(0, \sigma^2 I_d) \). Assume its class probability satisfies Equation (24). If the transformation is z-rotation, z-shear, or z-twist (\( \phi = \phi_{xz}, \phi_{xz} \), or \( \phi_{rot-z} \)), then it is guaranteed that \( g(\phi(x, \alpha); \epsilon) = g(x; \epsilon) \), if the following condition holds:
\[
\|\alpha\|_2 \leq \frac{\sigma}{2} \left( \Phi^{-1}(p_A) - \Phi^{-1}(p_B) \right), \alpha \in \mathbb{Z}. \tag{8}
\]

4.2. Tapering along a Fixed Axis

Tapering a point keeps the coordinate of a specific axis \( k \), but scales the coordinates of other axes proportional to \( k \)'s coordinate. For clarity, we define a z-taper transformation \( \phi_{TP} : \mathcal{X} \times Z \to \mathcal{X} \) as tapering along the z-axis, with its parameter space defined as \( Z = \mathbb{R} \). For any point cloud \( x = \{p_i\}^N_{i=1} \in \mathcal{X} \) and any \( \theta \in Z \),
\[
\phi_{TP}(p, \theta) = (x_i(1 + \theta z_i), y_i(1 + \theta z_i), z_i). \tag{9}
\]

However, z-taper is not a composable transformation, since the composition of two z-taper transformations contains terms with \( z^2 \) component. Therefore, we propose a specific certification protocol for z-taper based on Theorem 7. To achieve this goal, we specify a sampling strategy in the parameter space \( Z \) and bound the interpolation error (Equation (31)) of the sampled z-taper transformations.

Theorem 1. Let \( \phi_{TP} : \mathcal{X} \times \mathbb{R} \to \mathcal{X} \) be a z-taper transformation. Let \( g : \mathcal{X} \to \mathcal{Y} \) be an \( \epsilon \)-smoothed classifier with random noises \( \epsilon \sim N(0, \sigma^2 I_{3xN}) \), which predicts \( g(x; \epsilon) = \arg \max \epsilon_y g(y|x; \epsilon) = \arg \max \epsilon_y \mathbb{E} g(y|x + \epsilon). \)

Let \( \{\theta_j\}_{j=0}^M \) be a set of transformation parameters and \( \theta_j = (2^j - 1)R. \) Suppose for any \( i \),
\[
q(y_A|\phi_{TP}(x, \theta_j); \epsilon) \geq p_A^{(j)} > p_B^{(j)} \geq \max_{\neq y_A} q(y|\phi_{TP}(x, \theta_j); \epsilon) \tag{10}
\]

Then it is guaranteed that \( \forall \theta \in [-R, R] : y_A = \arg \max \epsilon \epsilon_y q(y|\phi_{TP}(x, \theta_j); \epsilon) \) if for all \( j = 1, \ldots, M, \)
\[
\frac{\sigma}{2} \left( \Phi^{-1}(p_A^{(j)}) - \Phi^{-1}(p_B^{(j)}) \right) \geq \frac{R\sqrt{N}}{2M}. \tag{11}
\]

Detailed proof for Theorem 1 can be found in Appendix B.1.

4.3. General Rotation

Rotation is one of the most common transformations for point cloud data. Therefore, we hope the classifier is robust not only against rotation attacks along a fixed axis, but also those along arbitrary axes. In this section, we first define general rotation and show its universality for rotations as well as their composition; then provide a concrete certification protocol for smoothing and certifying the robustness against this type of transformation.

We define general rotation transformations as \( \phi_R : \mathcal{X} \times Z \to \mathcal{X} \) where \( Z = S^2 \times \mathbb{R}^+ \) is the parameter space of rotations. For a rotation \( z \in Z \), its rotation axis is defined by a unit vector \( k \in S^2 \) and its rotation angle is \( \theta \in \mathbb{R}^+ \). For any 3D point \( p_i \in \mathbb{R}^3 \),
\[
\phi_R(p_i, z) = \text{Rot}(k, \theta)p_i, z = (k, \theta). \tag{12}
\]

where Rot\((k, \theta)\) is the rotation matrix that rotates by \( \theta \) along axis \( k \). General rotations are composable transformations since the composition of any two 3D rotations can be expressed by another 3D rotation.

However, certifying against the general rotation is more challenging, since the general rotation is not additive and the expression of their composition is extremely complicated. In particular, if we smooth a base classifier with a random variable \( \epsilon_0 \), a semantic attack with parameter \( \alpha \in Z \) results in \( \phi_R(\phi_R(x, \alpha), \epsilon_0) = \gamma_0(\epsilon_0) \), which is a bizarre distribution in the parameter space. Therefore, we cannot directly apply Theorem 6 to certify general rotation.

On the other hand, as Theorem 7 shows, if we uniformly sample many parameters in a subspace of \( Z = S^2 \times \mathbb{R}^+ \) and certify robustness in the neighborhood of each sample, we are able to certify a large and continuous subspace \( Z_{robust} \subseteq Z \). As a result, we propose a sampling-based certification strategy, together with a tight bound for the interpolation error of general rotation transformations, which we summarize in the following theorem.

Theorem 2. Let \( \phi_R : \mathcal{X} \times Z \to \mathcal{X} \) be a general rotation transformation. Let \( g : \mathcal{X} \to \mathcal{Y} \) be a classifier smoothed by random noises \( \epsilon \sim N(0, \sigma^2 I_{3xN}) \), which predicts \( g(x; \epsilon) = \arg \max \epsilon_y g(y|x; \epsilon) = \arg \max \epsilon_y \mathbb{E} g(y|x + \epsilon). \)

Let \( \{z_j\}_{j=1}^M \) be a set of transformation parameters with \( z_j = (k_j, \theta_j), k_j \in S^2, \theta_j \in \mathbb{R}^+ \) such that
\[
\forall k \in S^2, \theta \in [0, R], \exists k_j, \theta_j : (k, \theta) \leq \epsilon, |\theta - \theta_j| \leq \delta \tag{13}
\]

Suppose for any \( j \), the smoothed classifier \( g \) has class probabilities that satisfy
\[
q(y_A|\phi_R(x, z_j); \epsilon) \geq p_A^{(j)} > p_B^{(j)} \geq \max_{\neq y_A} q(y|\phi_R(x, z_j); \epsilon). \tag{14}
\]

Then it is guaranteed that for any \( z \) with rotation angle \( \theta < R; y_A = \arg \max \epsilon \epsilon_y q(y|\phi_R(x, z); \epsilon) \) if for all \( \forall j \),
\[
\frac{\sigma}{2} \left( \Phi^{-1}(p_A^{(j)}) - \Phi^{-1}(p_B^{(j)}) \right) \geq \pi \sqrt{\frac{\delta^2}{4} + \frac{\epsilon^2 R^2}{8} ||x||_2}. \tag{15}
\]

We present a proof sketch here and leave the details in Appendix B.2. Notice that the interpolation error between two transformations on a point cloud \( x = \{p_i\}_{i=1}^N \) can be calculated by \( ||\phi(x, z_j) - \phi(x, z)||_2 = ||\phi(x, z') - x||_2 \leq \theta' (\sum_i ||p_i||^2_2)^{1/2} \), where \( z' = (k', \theta') \) is the composition of the rotation with parameter \( z \) and the reverse rotation \( z_j^{-1} \). Combined with the generic theorem for indirectly composable transformations (Theorem 7), bounding \( \theta' \) using Equation (13) yields Theorem 2.
4.4. Linear Transformations

Here, we consider a broader class of semantic transformations that contains all linear transformations applied to a 3D point. Formally, a linear transformation \( \phi_L : \mathcal{X} \times Z \to \mathcal{X} \) has a parameter space of \( Z = \mathbb{R}^{3 \times 3} \). For any point cloud \( x = \{ p_i \}_{i=1}^N \in \mathcal{X} \) and for any \( A \in Z \),

\[
\phi_L(p_i, A) = (I + A)p_i. \tag{16}
\]

Equation (16) describes any linear transformation with a bounded perturbation \( A \) from the identity transformation \( I \). A natural threat model is considered by (Reisizadeh et al., 2020) that the perturbation matrix \( A \) has a bounded Frobenius norm \( \| A \|_F \leq \epsilon \), for which we present a certification protocol in this paper. Linear transformations are composable because their compositions are also linear, but the fact that they are not additive prohibits a direct usage of Corollary 3. Nevertheless, these transformations can still be certified with a more complicated protocol, if Gaussian smoothing is applied.

**Theorem 3.** Suppose a classifier \( g \) is smoothed by random linear transformations \( \phi_L : \mathcal{X} \times Z \to \mathcal{X} \) where \( Z = \mathbb{R}^{3 \times 3} \), with a Gaussian random variable \( \epsilon \sim \mathcal{N}(0, \sigma^2 I_3) \). If the class probability satisfies Equation (24), then it is guaranteed that \( g(\phi(x, \alpha); \epsilon) = g(x; \epsilon) \) for all \( \| \alpha \|_F \leq R \), where

\[
R = \frac{\sigma}{2 + \sigma} \left( \Phi^{-1}(\tilde{p}_A) - \Phi^{-1}(1 - \tilde{p}_A) \right). \tag{17}
\]

\( \tilde{p}_A \) is a function of \( p_A \) as explained in Lemma B.2.

4.5. Compositions of Different Transformations

In addition to certifying against a single transformation, we also provide certification protocols for composite transformations, including \( z\text{-twist} \circ z\text{-rotation}, z\text{-taper} \circ z\text{-rotation} \) and \( z\text{-twist} \circ z\text{-taper} \circ z\text{-rotation} \).

Notice that \( z\text{-twist} \circ z\text{-rotation} \) is an additive function:

\[
\phi_{Tz}(\phi_{Rot}(\phi_{Rot}(x, \theta_1), \theta_1), \theta_2) = \phi_{Tz}(\phi_{Rot}(x, \theta_1 + \theta_2), \theta_1 + \theta_2). \tag{18}
\]

Therefore, we directly apply Corollary 3 in Appendix A.1 to certify \( z\text{-twist} \circ z\text{-rotation} \) transformation. The concrete corollary is stated as below.

**Corollary 2.** Suppose a classifier \( g : \mathcal{X} \to \mathcal{Y} \) is smoothed by random transformations \( z\text{-twist} \circ z\text{-rotation} \) \( \phi : \mathcal{X} \times Z \to \mathcal{X} \) where the parameter space \( Z = \mathbb{R}^{z_{Twist} \times z_{Rot}} = \mathbb{R}^3 \). The random variable for smoothing is \( \epsilon \sim \mathcal{N}(0, \text{diag}(\sigma_1^2, \sigma_2^2)) \). If the class probability of \( g \) satisfies Equation (24), then it is guaranteed that \( g(\phi(x, \alpha); \epsilon) = g(x; \epsilon) \) for all \( (\alpha_1, \alpha_2) \in Z \), if the following condition holds:

\[
\sqrt{\frac{\alpha_1}{\sigma_1}^2 + \frac{\alpha_2}{\sigma_2}^2} \leq \frac{\sigma}{2} \left( \Phi^{-1}(p_A) - \Phi^{-1}(p_B) \right). \tag{19}
\]

Another composite transformation \( z\text{-taper} \circ z\text{-rotation} \) first rotates the point cloud along \( z\)-axis, and then taper along \( z\)-axis. As \( z\text{-taper} \) is not composable with itself, this composite transformation is also not composable. Similar to \( z\text{-taper} \), we certify the composite transformation \( z\text{-taper} \circ z\text{-rotation} \) by upper-bounding the interpolation error in Equation (31).

**Theorem 4.** We denote \( z\text{-taper} \circ z\text{-rotation} by \( \phi : \mathcal{X} \times Z \to \mathcal{X}, \phi = \phi_{Tz} \circ \phi_{Rot} \) with a parameter space of \( Z = \mathbb{R}^{Tz} \times Z_{Rot} = \mathbb{R}^2 \). Let \( g : \mathcal{X} \to \mathcal{Y} \) be a classifier smoothed by random noises \( \epsilon \sim \mathcal{N}(0, \sigma^2 I_{3 \times N}) \).

For a subspace in the parameter space, \( S = [-\varphi, \varphi] \times [-\theta, \theta] \subseteq Z \), we uniformly sample \( \varphi \theta N^2 \) parameters \( \{ z_{jk} \} \) in \( S \). That is, \( z_{jk} = (\varphi_j, \theta_k) \) where \( \varphi_j = \frac{2j}{M} - \varphi \) and \( \theta_k = \frac{2k}{M} - \theta \). Suppose for any \( j, k \) the smoothed classifier \( g \) has class probability that satisfy

\[
q(y_A(g(x, z_{jk})); \epsilon) \leq p_{A}^{jk} > p_{B}^{jk} \geq \max_{y \neq y_A} q(y|\phi(x, z_{jk}); \epsilon), \tag{20}
\]

then it is guaranteed that \( y_A = \arg \max_y q(y|\phi(x, z); \epsilon) \) if \( y_j, k \) and \( \forall z \in S \).

\[
\frac{\sigma}{2} \left( \Phi^{-1}(p_A^{jk}) - \Phi^{-1}(p_B^{jk}) \right) \geq \frac{\sqrt{N(4\sigma^2 + 8\varphi + 5)}}{2M}. \tag{21}
\]

We also consider the composition of three transformations: \( z\text{-twist} \circ z\text{-taper} \circ z\text{-rotation} \).

**Theorem 5.** We define the composite transformation \( z\text{-taper} \circ z\text{-rotation} \circ z\text{-rotation} \ by \( \phi : \mathcal{X} \times Z \to \mathcal{X} \), with input space \( X = \mathbb{R}^{3 \times N} \) and parameter space \( Z = \mathbb{R}^{Tz} \times \mathbb{R}^{z_{Rot}} \times \mathbb{R}^3 \). Let \( g : \mathcal{X} \to \mathcal{Y} \) be a classifier smoothed by random noises \( \epsilon \sim \mathcal{N}(0, \sigma^2 I_{3 \times N}) \), which predicts \( g(x; \epsilon) = \arg \max_y q(y|x; \epsilon) = \arg \max_y \mathbb{E}(p(y|x + \epsilon)) \).

Let \( \{ z_{jkl} \in Z : z = (\varphi_j, \alpha_k, \theta_l) \} \) be a set of parameters with \( \varphi_j = \frac{2j}{M} - \alpha \), \( \alpha_k = \frac{2k}{M} - \alpha \) and \( \theta_l = \frac{2l}{M} - \theta \). Therefore \( (\varphi_j, \alpha_k, \theta_l) \) distribute uniformly in the subspace \( Z_{\text{robust}} = [-\varphi, \varphi] \times [-\alpha, \alpha] \times [-\theta, \theta] \subseteq Z \). Suppose for any \( j, k, l \), the smoothed classifier \( g \) has class probability that satisfy:

\[
q(y_A(g(x, z_{jkl})); \epsilon) \leq p_{A}^{jkl} > p_{B}^{jkl} \geq \max_{y \neq y_A} q(y|\phi(x, z_{jkl}); \epsilon), \tag{22}
\]

then it is guaranteed that for any \( z \in Z_{\text{robust}} : y_A = \arg \max_y q(y|\phi(x, z); \epsilon) \), for any \( i, j, k \),

\[
\frac{\sigma}{2} \left( \Phi^{-1}(p_A^{jkl}) - \Phi^{-1}(p_B^{jkl}) \right) \geq \sqrt{N(1 + \frac{2\varphi}{M}(1 + \alpha))}. \tag{23}
\]

Both Theorem 4 and Theorem 5 are based on our proposed approach of sampling parameters in the parameter space and certifying the neighboring distributions of the samples separately by bounding the interpolation error (Equation (31)). These two theorems are rigorously proved in Appendix B.5 and Appendix B.6.
5. Experiments

We conduct extensive experiments on different 3D semantic transformations and models to evaluate the certified robustness derived from our TPC framework. We show that TPC significantly outperforms the state-of-the-art in terms of the certified robustness against a range of semantic transformations, and the results also lead to some interesting findings.

5.1. Experimental setup

Dataset. We perform experiments on the ModelNet40 dataset (Wu et al., 2015), which includes different 3D objects of 40 categories. We follow the standard preprocessing pipeline that places the point clouds in the center and scales them into a unit sphere.

We also conduct experiments for part segmentation tasks, for which the ShapeNet dataset (Chang et al., 2015) is used for evaluation. It contains 16681 meshes from 16 categories and also 50 predefined part labels. The experiment results are presented in Section 5.2.4.

Models. We run our experiments for point cloud classification on PointNet models (Qi et al., 2017) with different point cloud sizes. We apply data augmentation training for each transformation combined with consistency regularization to train base classifiers. We then employ our TPC framework to smooth these models and derive robustness certification bounds against various transformations. TPC does not depend on specific model selection and can be directly applied to certify other point cloud model architectures. We present certification results for other architectures (e.g., CurveNet (Xiang et al., 2021)) in Appendix E.3.

Evaluation Metrics. To evaluate the robustness of point clouds classification, we pick a fixed random subset of the ModelNet40 test dataset. We report the certified accuracy defined by the fraction of point clouds that are classified both correctly and consistently within certain transformation space. The baseline we compare with (Lorenz et al., 2021) only presents certified ratio, which is the fraction of test samples classified consistently. We believe that the certified accuracy is a more rigorous metric for evaluation based on existing standard certification protocols in the image domain (Cohen et al., 2019). We thus calculate the certified accuracy for baselines based on the results reported in the paper (Lorenz et al., 2021) for comparison. Besides, we also report the certified ratio comparison in Table 2 and Appendix E.2. We remark that TPC provides a probabilistic certification for point cloud models and we use a high confidence level of 99.9% in all experiments; while the baseline DeepG3D (Lorenz et al., 2021) yields a deterministic robustness certification.

For the part segmentation task, we evaluate our method using a fixed random subset of the ShapeNet test dataset. As the part segmentation task requires assigning a part category to each point in a point cloud, we report the point-wise certified accuracy defined as the fraction of points that are classified correctly and consistently. Note that other common metrics such as IoU can be easily derived based on our bound as well. We will focus on the point-wise certified accuracy for the convenience of comparison with baseline (Lorenz et al., 2021).

5.2. Main Results

In this section, we present our main experimental results. Concretely, we show that: (1) the certified accuracy of TPC under a range of semantic transformations is significantly higher than the baseline, and TPC is able to certify under some transformation space where the baseline cannot be applied; (2) the certified accuracy of TPC always outperforms the baseline for different point cloud sizes, and more interestingly, the certified accuracy of TPC increases with the increasing of point cloud size while that of the baseline decreases due to relaxation; (3) TPC is also capable of certifying against $\ell_2$ or $\ell_\infty$ norm bounded 3D perturbations for different point cloud sizes; (4) on the part segmentation task, TPC still outperforms the baseline against different semantic transformations and is able to certify some transformation parameter space that the baseline is not applicable.

<table>
<thead>
<tr>
<th>Transformation</th>
<th>Attack radius</th>
<th>Certified Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\ell_\infty$</td>
<td>TPC</td>
</tr>
<tr>
<td>ZYX-rotation</td>
<td>$2^\circ$</td>
<td>81.4</td>
</tr>
<tr>
<td></td>
<td>$5^\circ$</td>
<td>69.2</td>
</tr>
<tr>
<td>General rotation</td>
<td>$10^\circ$</td>
<td>69.2</td>
</tr>
<tr>
<td></td>
<td>$15^\circ$</td>
<td>55.5</td>
</tr>
<tr>
<td>Z-rotation</td>
<td>$20^\circ$</td>
<td>84.2</td>
</tr>
<tr>
<td></td>
<td>$60^\circ$</td>
<td>83.8</td>
</tr>
<tr>
<td></td>
<td>$180^\circ$</td>
<td>81.3</td>
</tr>
<tr>
<td>Z-shear</td>
<td>0.03</td>
<td>83.4</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>82.2</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>77.7</td>
</tr>
<tr>
<td>Z-twist</td>
<td>$20^\circ$</td>
<td>83.8</td>
</tr>
<tr>
<td></td>
<td>$60^\circ$</td>
<td>80.1</td>
</tr>
<tr>
<td></td>
<td>$180^\circ$</td>
<td>64.3</td>
</tr>
<tr>
<td>Z-taper</td>
<td>0.1</td>
<td>78.1</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>76.5</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>66.0</td>
</tr>
<tr>
<td>Linear</td>
<td>0.1</td>
<td>74.0</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>59.9</td>
</tr>
<tr>
<td>Z-twist $\circ$</td>
<td>$20^\circ$, $1^\circ$</td>
<td>78.9</td>
</tr>
<tr>
<td>Z-rotation</td>
<td>$20^\circ$, $5^\circ$</td>
<td>78.5</td>
</tr>
<tr>
<td></td>
<td>$50^\circ$, $5^\circ$</td>
<td>76.9</td>
</tr>
<tr>
<td>Z-taper $\circ$</td>
<td>$0.1$, $1^\circ$</td>
<td>76.1</td>
</tr>
<tr>
<td>Z-rotation</td>
<td>$0.2$, $1^\circ$</td>
<td>72.9</td>
</tr>
<tr>
<td>Z-twist $\circ$ Z-taper</td>
<td>$10^\circ$, $0.1$, $1^\circ$</td>
<td>68.8</td>
</tr>
<tr>
<td></td>
<td>$20^\circ$, $0.2$, $1^\circ$</td>
<td>63.1</td>
</tr>
</tbody>
</table>
Table 2. Comparison of certified ratio as well as certified accuracy for z-rotation transformations. “-” denotes the settings which the baselines cannot scale up to.

<table>
<thead>
<tr>
<th>Radius</th>
<th>Certified Ratio (%)</th>
<th>Certified Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TPC</td>
<td>DeepG3D</td>
</tr>
<tr>
<td>20°</td>
<td>99.0</td>
<td>96.7</td>
</tr>
<tr>
<td>60°</td>
<td>98.1</td>
<td>95.7</td>
</tr>
<tr>
<td>180°</td>
<td>95.2</td>
<td>-</td>
</tr>
</tbody>
</table>

5.2.1. COMPARISON OF CERTIFIED ACCURACY

Table 1 shows the certified accuracy we achieved for different transformations compared with prior works. We train a PointNet model with 64 points, which is consistent with the baseline. For transformations characterized by one parameter, such as z-rotation, z-twist, and z-taper, we report the certified accuracy against attacks in ±θ. For z-shear with a parameter space of \( \mathbb{R}^2 \), we report the certified accuracy against attacks in a certain \( \ell_2 \) parameter radius. For the linear transformation with a parameter space of \( \mathbb{R}^{3 \times 3} \), we report the certified accuracy against attacks in a certain Frobenius norm radius.

The highlighted results in Table 1 demonstrate that our framework TPC significantly outperforms the state of the art in every known semantic transformation. For example, we improve the certified accuracy from 59.8% to 83.4% for z-shear in ±0.03 and from 20.3% to 83.8% for z-twist in ±20°.

Besides, we also report the certified accuracy for larger attack radius for which the baseline cannot certify (cells with “-”). For instance, we achieve 81.3% certified accuracy on z-rotation within ±180°, which is essentially every possible z-rotation transformation.

The general rotation transformations we define in Section 4.3 includes rotations along any axis with bounded angles. Lorenz et al. (2021) consider ZYX-rotation, the composition of three rotations within ±θ (Euler angles) along \( x, y, z \) axes instead (2021), which results in a different geometric shape for the certified parameter space. However, the parameter space restricted by \( S^2 \times [0, 2\pi] \) of general rotation strictly contains the space defined by ±θ for three Euler angles. (See Appendix B.3 for proof.) The derived results for ZYX-rotation are also shown in Table 1 for comparison.

Comparison of Certified Ratio. Aside from the certified accuracy, we also consider the certified ratio as another metric according to the baseline. This metric measures the tightness of certification bounds but fails to take the classification accuracy into account which is important. Therefore, we mainly present the comparison based on the certified ratio for z-rotations in Table 2 only for comparison and leave the full comparison in Appendix E.2.

Table 3. Certification of z-rotation for different point cloud sizes. The certified accuracy achieved by our TPC increases as the size of the point cloud model increases.

(a) \( \theta = \pm 3° \) compared with DeepG3D (Lorenz et al., 2021)

<table>
<thead>
<tr>
<th>Points</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>128</th>
<th>256</th>
<th>512</th>
<th>1024</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPC</td>
<td>83.2</td>
<td>83.8</td>
<td>86.6</td>
<td>87.4</td>
<td>89.4</td>
<td>89.8</td>
<td>90.5</td>
</tr>
<tr>
<td>DeepG3D</td>
<td>75.4</td>
<td>78.4</td>
<td>79.1</td>
<td>69.4</td>
<td>57.5</td>
<td>42.8</td>
<td>32.3</td>
</tr>
</tbody>
</table>

(b) Certified accuracy of TPC under \( \theta = \pm 180° \)

<table>
<thead>
<tr>
<th>Points</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>128</th>
<th>256</th>
<th>512</th>
<th>1024</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPC</td>
<td>73.6</td>
<td>79.3</td>
<td>81.3</td>
<td>81.8</td>
<td>83.0</td>
<td>84.6</td>
<td>83.8</td>
</tr>
</tbody>
</table>

5.2.2. CERTIFICATION ON POINT CLOUDS WITH DIFFERENT SIZES

Here we show that our certification framework naturally scales up to larger point cloud models. A basic principle of our TPC framework is that deriving the certification bound for a smoothed classifier only depends on the predicted class probability. In other words, it does not rely on specific model architectures.

The relaxation-based verifiers (Lorenz et al., 2021; Singh et al., 2019) have worse certification guarantees for larger point clouds due to the precision loss during relaxation, especially for pooling layers that are heavily used in point cloud model architectures. For example, the DeepG3D verifier guarantees 79.1% certified accuracy for a 64-point model on z-rotation with ±3° (without splitting); but the certification drops to 32.3% for a 1024-point model (Lorenz et al., 2021). In contrast, using our TPC framework, the certified accuracy tends to increase with a larger number of points in point clouds. This is because larger PointNet models predict more accurately and yield higher class probability after smoothing. We compare our TPC framework with the baseline in terms of certified accuracy for different point cloud sizes in Table 3a. The baseline DeepG3D only presents results for z-rotations in \( \theta = \pm 3° \), which cannot fully illustrate the capability of our method. Therefore, we also report our experimental results for z-rotations in \( \theta = \pm 180° \) in Table 3b. It shows that our method can scale up to larger point cloud models to accommodate real-world scenarios.

5.2.3. CERTIFICATION AGAINST \( \ell_p \) NORM BOUNDED 3D-PERTURBATIONS

In addition to semantic transformations, we also provide robustness certification for point cloud models against \( \ell_p \) perturbations. Defenses have been proposed for point cloud models against \( \ell_2 \) norm bounded perturbations (Fischer et al., 2021), which is a special case of our TPC framework regarding an additive transformation \( \phi(x, z) = x + z \).

We cannot directly certify against perturbations with bounded \( \ell_\infty \) norm. However, a certification bound similar to the baseline (Lorenz et al., 2021) can still be derived,
We evaluate our method using the ShapeNet part dataset (Chang et al., 2015). We train a segmentation version Point-TPC for point cloud models against a diverse range of semantic transformations. Our theoretical and empirical analysis show that Point-TPC is more scalable and able to provide much tighter certification under different settings and tasks.

### 5.2.4. Certification for Part Segmentation

Part segmentation is a common 3D recognition task in which a model is in charge of assigning each point or face of a 3D mesh to one of the predefined categories. As our TPC framework is independent of concrete model architectures, it can be naturally extended to handle this task.

We evaluate our method using the ShapeNet part dataset (Chang et al., 2015). We train a segmentation version PointNet (Qi et al., 2017) with 64 points, which predicts a part category for each point in the point cloud. The certified accuracy reported in Table 5 denotes the percentage of points guaranteed to be labeled correctly. We can see that for the part segmentation task, TPC consistently outperforms the baseline against different semantic transformations. The baseline only reports the result for z-rotations in ±5° and ±10°, while we present robustness guarantees for any z-rotation (±180°) as well as other transformations including shearing and twisting.

### 6. Conclusion

In this work, we propose a unified certification framework TPC for point cloud models against a diverse range of semantic transformations. Our theoretical and empirical analysis show that TPC is more scalable and able to provide much tighter certification under different settings and tasks.

### Acknowledgements

The authors thank the anonymous reviewers for their valuable feedback. This work is partially supported by NSF grant No.1910100, NSF CNS No.2046726, C3 AI, and the Alfred P. Sloan Foundation. WC would like to thank the support from Institute for Interdisciplinary Information Sciences, Tsinghua University.

### References


Neyman, J. and Pearson, E. S. Ix. on the problem of the most efficient tests of statistical hypotheses. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 231(694-706):289–337, 1933.

Pei, K., Cao, Y., Yang, J., and Jana, S. Deepxplore: Automated whitebox testing of deep learning systems. In proceedings of the 26th Symposium on Operating Systems Principles, pp. 1–18, 2017.


Transformation-Specific Smoothing for Point Cloud Models


A. Generic Theorems for Composable and Indirectly Composable Transformations

A.1. Main Theorem for Transformation Specific Smoothing

**Theorem 6** (Theorem 1 (Li et al., 2021b)). Let $\epsilon_0 \sim \mathbb{P}_0$ and $\epsilon_1 \sim \mathbb{P}_1$ be $\mathbb{Z}$-valued random variables with probability density function $f_0$ and $f_1$. Let $\phi : \mathbb{X} \times \mathbb{Z} \to \mathbb{X}$ be a semantic transformation. Suppose a classifier smoothed by the transformation $\phi$ predicts $y_A = g(x; \epsilon)$, and that

$$q(y_A | x, \epsilon) \geq p_A > p_B \geq \max_{y \neq y_A} q(y | x, \epsilon).$$

(24)

For $t \geq 0$, we define sets $S_t, \overline{S}_t \subseteq \mathbb{Z}$ by $S_t := \{ f_1 / f_0 < t \}$ and $\overline{S}_t := \{ f_1 / f_0 \leq t \}$. Also define a function $\xi : [0, 1] \to [0, 1]$ by

$$\xi(p) := \sup \{ \mathbb{P}_1(S) : \mathbb{S}_t \subseteq S \subseteq \mathbb{S}_{\tau} \}$$

(25)

where $\tau_p := \inf \{ t \geq 0 : \mathbb{P}_0(\mathbb{S}_t) \geq p \}$.

Then it is guaranteed that $g(x; \epsilon_1) = g(x; \epsilon_0)$ if the following condition holds:

$$\xi(p_A) + \xi(1 - p_B) > 1.$$  

(27)

Intuitively, $\xi(p_A)$ computes a lower bound for the class probability of $y_A$ when the smoothing distribution changes from $\epsilon_0$ to $\epsilon_1$. Suppose we want to certify an $\epsilon_0$-smoothed classifier against an composable transformation $\phi$ and $\phi(\phi(x, \alpha), \beta) = \phi(x, \gamma_\alpha(\beta))$. For any attack $\alpha \in \mathbb{S}$, we assign $\epsilon_1 = \gamma_\alpha(\epsilon_0)$ and check for the condition of Equation (27). Moreover, if $\phi$ is additive and $\epsilon_0$ is a Gaussian random variable, we have the following corollary:

**Corollary 3** (Corollary 7 (Li et al., 2021b)). Let $\phi : \mathbb{X} \times \mathbb{Z} \to \mathbb{X}$ be an additive transformation and $\mathbb{Z} = \mathbb{R}^m$. Suppose classifier $g$ is smoothed by a random variable $\epsilon_0 \sim \mathcal{N}(0, \sigma_{1}^2, \ldots, \sigma_{m}^2)$. Assume that the class probability satisfies:

$$q(y_A | x, \epsilon_0) \geq p_A > p_B \geq \max_{y \neq y_A} q(y | x, \epsilon_0).$$

(28)

Then it holds that $g(x; \epsilon_0) = g(\phi(x, \alpha); \epsilon_0)$ if the attack parameter $\alpha$ satisfies:

$$\sqrt{\sum_{i=1}^{m} \left( \frac{\alpha_i}{\sigma_i} \right)^2} < \frac{1}{2} \left( \Phi^{-1}(p_A) - \Phi^{-1}(p_B) \right).$$

(29)

We direct readers to (Li et al., 2021b) for the rigorous proof on Theorem 6 and Corollary 3.

A.2. Theorem for Certifying Indirectly Composable Transformations

**Theorem 7** (Corollary 2 (Li et al., 2021b)). Let $\psi(x, \delta) = x + \delta$ and $\epsilon \sim \mathcal{N}(0, \sigma^2 \mathbb{1}_d)$. $\phi : \mathbb{X} \times \mathbb{Z} \to \mathbb{X}$ is a indirectly composable transformation. Construct a smoothed classifier with additive noise $\psi(x, \delta)$ and suppose it predicts $y = \arg \max_{y \in \mathbb{Y}} q(y | x; \epsilon)$. Draw $\mathcal{N}$ samples $\{ \alpha_i \}_{i=1}^{\mathcal{N}}$ from a set $\mathbb{S} \subseteq Z$. Assume

$$q(y_A | \phi(x, \alpha_i), \epsilon) \geq p_A^{(i)} > p_B^{(i)} \geq \max_{y \neq y_A} q(y | \phi(x, \alpha_i), \epsilon).$$

(30)

Then it is guaranteed that $\forall \alpha \in \mathbb{S} : y_A = \arg \max_{y \in \mathbb{Y}} q(y | \phi(x, \alpha); \epsilon)$ if the maximum interpolation error

$$\mathcal{M}_{\mathcal{S}} := \max_{\alpha \in \mathbb{S}} \min_{1 \leq i \leq \mathcal{N}} \mathcal{M}(\alpha, \alpha_i)$$

(31)

$$= \max_{\alpha \in \mathbb{S}} \min_{1 \leq i \leq \mathcal{N}} \| \phi(x, \alpha) - \phi(x, \alpha_i) \|_2$$

(32)

satisfies

$$\mathcal{M}_{\mathcal{S}} < \mathcal{R} := \frac{\sigma}{2} \min_{1 \leq i \leq \mathcal{N}} \left( \Phi^{-1}(p_A^{(i)}) - \Phi^{-1}(p_B^{(i)}) \right).$$

(33)

B. Proofs for Certifying Specific Transformations

Here, we present proofs for the theorems and corollaries proposed in Section 4, including concrete protocols for common 3D transformations, such as z-rotation, z-twist, z-taper, etc.
B.1. Proof of Theorem 1: Certifying Z-taper

Proof. The z-taper transformation is defined as $\phi_{TP} : \mathcal{X} \times \mathbb{R} \rightarrow \mathcal{X}$ where $\mathcal{X} = \mathbb{R}^{3 \times N}$ is the space for input point clouds. A z-taper transformation acting on a point cloud $x = \{p_i\}_{i=1}^N \in \mathcal{X}$ in fact performs point-wise transformation to each $p_i$, where

$$
\phi_{TP}(p_i, \theta) = \begin{pmatrix}
  x_i(1 + \theta z_i) \\
  y_i(1 + \theta z_i) \\
  z_i 
\end{pmatrix}, \text{ if } p_i = (x_i, y_i, z_i)^T. 
$$

(34)

We calculate the interpolation error between two parameters $\theta$ and $\theta_j$ by

$$
M(\theta, \theta_j) = ||\phi_{TP}(x, \theta) - \phi_{TP}(x, \theta_j)||_2
$$

(35)

$$
= \left( \sum_{i=1}^N \|\phi_{TP}(p_i, \theta) - \phi_{TP}(p_i, \theta_j)\|_2^2 \right)^{1/2}
$$

(36)

$$
= \left( \sum_{i=1}^N (x_i^2 + y_i^2)z_i^2(\theta - \theta_j)^2 \right)^{1/2}
$$

(37)

$$
\leq \sqrt{N} |\theta - \theta_j|.
$$

(38)

The last inequality holds because we assume point clouds are normalized into a unit ball, so $(x_i^2 + y_i^2)z_i^2 \leq \frac{1}{4}$. Recall that we choose $\theta_j = \left( \frac{2j}{M} - 1 \right)R$ and $j = 0, 1, \ldots, M$. Hence, $\max_{\theta \in [-R, R]} \min_j |\theta - \theta_j| < \frac{R}{M}$. The maximal interpolation error is thus bounded by

$$
\mathcal{M}_S = \max_{\theta \in [-R, R]} \min_j M(\theta, \theta_j)
$$

(39)

$$
\leq \frac{R\sqrt{N}}{2M}.
$$

(40)

According to Theorem 7, it is guaranteed for all $\theta \in [-R, R]$ that $y_A = \arg \max_y q(y|\phi_{TP}(x, \theta); \epsilon)$, if $\forall j$,

$$
\frac{\sigma}{2} \left( \Phi^{-1}\left(p_A^{(j)}\right) - \Phi^{-1}\left(p_B^{(j)}\right) \right) \geq \frac{R\sqrt{N}}{2M}.
$$

(41)

B.2. Proof of Theorem 2: Certifying General Rotation

Proof. We first recall the definition of general rotation: $\phi_R : \mathcal{X} \times \mathbb{Z} \rightarrow \mathcal{X}$. The parameter space $\mathbb{Z} = S^2 \times \mathbb{R}^+$ where $S^2$ characterizes the rotation axis and $\mathbb{R}^+$ stands for the rotation angle. By Euler’s theorem, general rotations are composable transformations; and the composition of two rotations can be expressed by:

$$
\phi_R(\phi_R(x, z_1), z_2) = \phi_R(x, z_3)
$$

(42)

where

$$
k_3 = \text{normalize}(\sin \frac{\theta_3}{2} \cos \frac{\theta_2}{2} k_1 + \cos \frac{\theta_3}{2} \sin \frac{\theta_2}{2} k_2 + \sin \frac{\theta_3}{2} \sin \frac{\theta_2}{2} k_1 \times k_2)
$$

$$
\theta_3 = 2 \arccos \left( \cos \frac{\theta_3}{2} \cos \frac{\theta_2}{2} - \sin \frac{\theta_3}{2} \sin \frac{\theta_2}{2} k_1 \cdot k_2 \right).
$$

(43)
The interpolation error of a point cloud \( x = \{ p_i \}_{i=1}^N \) between two transformations with parameters \( z = (k, \theta) \) and \( z_j = (k_j, \theta_j) \) is bounded by:

\[
\mathcal{M}(z, z_j) = \| \phi_R(x, z) - \phi_R(x, z_j) \|_2
\]

\[
= \| \phi_R(\phi_R(x, z), z_j^{-1}) - x \|_2, \quad \text{(where } z_j^{-1} = (-k_j, \theta_j) \text{)}
\]

\[
= \left( \sum_{i=1}^N \| \phi_R(\phi_R(p_i, z), z_j^{-1}) - p_i \|_2^2 \right)^{1/2}
\]

\[
= \left( \sum_{i=1}^N \| \phi_R(p_i, z') - p_i \|_2^2 \right)^{1/2} \quad \text{(Let } z' = (k', \theta') \text{ be the composition of } z, z_j^{-1} \text{)}
\]

\[
\leq \left( \sum_{i=1}^N (\theta' \| p_i \|_2^2) \right)^{1/2} = \theta' \| x \|_2.
\]

Assuming \( (k, k_i) \leq \epsilon, |\theta - \theta_i| \leq \delta \), we derive that for \( \theta, \theta_i \in [0, R] \),

\[
\cos \frac{\theta'}{2} = \cos \frac{\theta_j}{2} \cos \frac{\theta_i}{2} + \cos(k, k_i) \sin \frac{\theta_j}{2} \sin \frac{\theta_i}{2}
\]

\[
\geq \cos \frac{\theta_j}{2} - \frac{\epsilon^2}{2} \sin \frac{\theta_j}{2} \sin \frac{\theta_i}{2} \quad \text{(since } \cos \epsilon \geq 1 - \frac{\epsilon^2}{2} \text{)}
\]

\[
\geq 1 - \left( \frac{\theta_j - \theta_i}{2} \right)^2 - \frac{\epsilon^2 \theta_j}{8}. \quad \text{(since } \sin x \leq x \text{)}
\]

\[
\geq 1 - \frac{\delta^2}{4} - \frac{\epsilon^2 R^2}{8}.
\]

Note that \( \arccos(1 - x) \leq \frac{x}{2} \sqrt{x} \) when \( x \in [0, 1] \), we have

\[
\theta' \leq 2 \arccos \left( 1 - \frac{\delta^2}{4} - \frac{\epsilon^2 R^2}{8} \right)
\]

\[
\leq \pi \sqrt{\frac{\delta^2}{4} + \frac{\epsilon^2 R^2}{8}}.
\]

Combining Equation (48) and Equation (54), the maximal interpolation error for \( z \in S^2 \times [0, R] \) satisfies

\[
\mathcal{M}_S = \max_{z} \min_{j} \mathcal{M}(z, z_j)
\]

\[
\leq \pi \sqrt{\frac{\delta^2}{4} + \frac{\epsilon^2 R^2}{8} \| x \|_2^2}.
\]

Theorem 2 thus holds combining Equation (55) with Theorem 6.

Moreover, we specify a sampling strategy to satisfy the condition of Equation (13).

- Uniformly sample \( \pi M \) number of \( a_r \in [0, \pi] \).
- For each \( a_r \), uniformly sample \( 2 \pi M \sin \theta_a \) points \( b_{rs} \in [0, 2\pi] \).
- Uniformly sample \( M \) number of \( \theta_i \in [0, R] \).
- Draw \( O(M^3) \) samples in total: \( z_j = (k_j, \theta_i) \) with \( k_j = (\cos b_{rs} \sin a_r, \sin b_{rs} \sin a_r, \cos a_r) \).

Following this strategy, the sampled parameters distribute evenly in the subspace of \( S^2 \times [0, R] \), which guarantees \( \epsilon = \frac{\sqrt{\pi}}{2M} \) and \( \delta = \frac{R}{2M} \) for the conditions in Equation (13).

To sum up, it is guaranteed that for all \( z \in S^2 \times [0, R], x \in X : y_A = \arg \max_y q(y \| \phi_R(x, z); \epsilon) \), if \( \forall j \),

\[
\frac{\sigma}{2} \left( \Phi^{-1}(p_A^{(j)}) - \Phi^{-1}(p_B^{(j)}) \right) \geq \frac{\sqrt{2\pi R} \| x \|_2}{4M}.
\]

Remark. In practice, we implement a tighter bound that \( \arccos(1 - x) \leq \sqrt{2x + (\frac{x}{2} - \sqrt{2})x^2} \) for \( x \in [0, 1] \).
B.3. From General Rotation to ZYX-rotation

ZYX-rotation, the composition of three rotations along x, y and z axes, is defined by: \( \phi_{\text{ZYX-rot}} : \mathcal{X} \times \mathbb{R}^3 \rightarrow \mathcal{X} \) with parameter space \( \mathcal{Z} = \mathbb{R}^3 \). Specifically, for \( z = (\alpha, \beta, \gamma) \in \mathcal{Z} \) and \( x = \{p_i\}_{i=1}^N \in \mathcal{X} \),

\[
\phi_{\text{ZYX-rot}}(p_i, z) = R_z(\gamma)R_y(\beta)R_x(\alpha)p_i, \quad \text{where } R_z, R_y, R_x \text{ are the rotation matrix along x,y,z axes.}
\]  

(58)

Note that the rotation angle for any rotation matrix \( R \) can be calculated by:

\[
|\theta| = \arccos \left( \frac{tr(R) - 1}{2} \right)
\]  

(59)

The trace of the rotation matrix for ZYX-rotation is

\[
f(\alpha, \beta, \gamma) = tr(R_z(\gamma)R_y(\beta)R_x(\alpha)) = \cos \alpha \cos \beta + \cos \alpha \cos \gamma + \cos \beta \cos \gamma - \sin \alpha \sin \beta \sin \gamma.
\]  

(60)

We assume \( \alpha, \beta, \gamma \in [-\frac{\pi}{2}, \frac{\pi}{2}] \). \( \frac{\partial f}{\partial \alpha} = \frac{\partial f}{\partial \beta} = \frac{\partial f}{\partial \gamma} = 0 \) yields \( \alpha = \beta = \gamma = 0, \pm \frac{\pi}{2} \). Therefore, for \( \alpha, \beta, \gamma \in [-\varphi, \varphi] \), the minimum of \( f(\alpha, \beta, \gamma) \) can only be on \( \alpha, \beta = \pm \varphi \) or \( \alpha = \beta = \gamma = 0 \). Since \( \alpha = \beta = \gamma = 0 \) yields the maximum \( f(\alpha, \beta, \gamma) \), we have

\[
\min_{\alpha, \beta, \gamma \in [-\varphi, \varphi]} f(\alpha, \beta, \gamma) = 3 \cos^2 \varphi - \sin^3 \varphi
\]  

(61)

Thus,

\[
\cos \theta = \frac{tr(R) - 1}{2}
\]  

(62)

\[
\geq \frac{3 \cos^2 \varphi - \sin^3 \varphi - 1}{2}
\]  

(63)

\[
= \frac{(2 \cos^2 \varphi - 2 \sin^2 \varphi) + (\sin^2 \varphi - \sin^3 \varphi)}{2}
\]  

(64)

\[
\geq \cos 2\varphi.
\]  

(65)

The rotation angle \( \theta \) is thus bounded by \( \theta \leq 2\varphi \). Hence, any transformation \( \phi_{\text{ZYX-rot}} \) with \( z \in [-\theta, \theta]^3 \) and \( \theta \in [0, \pi/2] \) belongs to the set of general rotations \( \phi_R \) with parameter space \( \mathcal{Z}_R = S^2 \times [0, 2\theta] \).

B.4. Proof of Theorem 3: Certifying Linear Transformations

Proof. The set of linear transformations is defined by \( \phi_L : \mathcal{X} \times \mathbb{R}^3 \rightarrow \mathcal{X} \) where the parameter space is \( \mathcal{Z} = \mathbb{R}^{3 \times 3} \) and \( \phi_L(p_i, A) = (I + A)p_i \). The composition of two linear transformations can be expressed by

\[
\phi(B)\phi(A) = I + A + B + BA
\]  

(66)

\[
= I + A + (b_{ij} + \sum_k b_{ik}a_{kj})_{ij}
\]  

(67)

To certify these transformations, we smooth the classifier with \( \phi(E) \), where \( e_{ij} \sim \mathcal{N}(0, \sigma^2) \). The smoothed classifier is denoted as \( g(x; E) \). Then the smoothed classifier can be viewed as smoothed additively by an equivalent random variable \( \tilde{E} \).

\[
g(\phi(A, x); E) = g(x; A + \tilde{E})
\]  

(68)

Suppose \( E \sim \mathcal{N}(0, \sigma^2I_n) \), then

\[
\tilde{E} = \begin{pmatrix}
I + A^T & O & O \\
O & I + A^T & O \\
O & O & I + A^T
\end{pmatrix} E = SE.
\]  

(69)
The covariance matrix $\Sigma$ of $\tilde{E}$ is thus $\Sigma = SS^T$. This symmetric covariance matrix can be decomposed by $\Sigma = QDQ^T$. Suppose the singular values of $\Sigma$ are $k_i$. Then by the Mirsky Theorem for matrix perturbation, we have

$$\sqrt{\sum_{i=1}^{n} (k_i - 1)^2} \leq \|A\|_F$$  \hfill (70)

Therefore, the problem is reduced to the diagonal case such that the covariance matrix is diagonal. Assume $\epsilon_0 \sim (0, \sigma^2 I_0)$ and $\epsilon_1 \sim (0, \sigma^2 \cdot \text{diag}(k_1^2, k_2^2, \ldots, k_9^2))$. Let $A = \text{diag}(k_1, k_2, \ldots, k_9)$ and $\Sigma = \sigma^2 I_9$. Also, we assume without loss of generality by symmetry that $k_1 = k_2 = k_3, k_4 = k_5 = k_6$ and $k_7 = k_8 = k_9$.

Therefore, according to Corollary 3, the smoothed classifier is guaranteed to be robust under attack $A$, if

$$\sqrt{\sum_{i=1}^{n} \left(\frac{a_i}{k_i}\right)^2} \leq \frac{\sigma}{2} \left(\Phi^{-1}(\tilde{p}_A) - \Phi^{-1}(\tilde{p}_B)\right)$$  \hfill (71)

where $\tilde{p}_A = q(y|x; \tilde{E})$ given $p_A = q(y|x; E)$. Since $k_i \geq 1 - \|A\|_F, \forall i$, the condition can be simplified to

$$\sqrt{\sum_{i=1}^{n} \sigma_i^2} = \|A\|_F \leq \frac{\sigma \left(\Phi^{-1}(\tilde{p}_A) - \Phi^{-1}(\tilde{p}_B)\right)}{2 + \sigma \left(\Phi^{-1}(\tilde{p}_A) - \Phi^{-1}(\tilde{p}_B)\right)}.$$  \hfill (72)

which is equivalent to the following condition when $\|A\|_F \leq 1$.

$$\|A\|_F \leq \frac{\sigma \left(\Phi^{-1}(\tilde{p}_A) - \Phi^{-1}(\tilde{p}_B)\right)}{2 + \sigma \left(\Phi^{-1}(\tilde{p}_A) - \Phi^{-1}(\tilde{p}_B)\right)}.$$  \hfill (73)

Next we find a functional relation between $\tilde{p}_A$ and $p_A$. To do so, we first state and prove a lemma.

**Lemma B.1.** Suppose a classifier $g$ is smoothed by a Gaussian random variable $\epsilon_0 \sim N(0, \text{diag}(\sigma_1^2, \ldots, \sigma_n^2))$, while another classifier $g'$ is smoothed by a Gaussian random variable $\epsilon_1 \sim N(0, \text{diag}(k_1^2 \sigma_1^2, \ldots, k_n^2 \sigma_n^2))$. If the smoothed classifier $g$ predicts $p_A = q(y_A|x; \epsilon_0)$, then

$$p_A' = q(y|x; \epsilon_1) \geq \begin{cases} F_{X_n} \left( \frac{1}{2} \Phi^{-1}(p_A) \right) & (k \geq 1) \\ 1 - F_{X_n} \left( \frac{1}{2} \Phi^{-1}(1 - p_A) \right) & (k < 1). \end{cases}$$  \hfill (74)

**Proof.** We denote by $f_0$ and $f_1$ the probability density function of $\epsilon_0$ and $\epsilon_1$. We define a level function as $\Lambda(z) := \frac{f_1(z)}{f_0(z)}$. Suppose $\epsilon_0 \sim N(0, \Sigma)$ and $\epsilon_1 \sim N(0, A^2 \Sigma)$. Therefore, $A$ is a diagonal matrix with $A_{ii} = k$ if $i \leq m$ and $A_{ii} = 1$ otherwise.

$$\Lambda(z) = \frac{f_1(z)}{f_0(z)} = \frac{((2\pi)^n |A^2 \Sigma|)^{-1/2} \exp(-\frac{1}{2} (z^T (A^2 \Sigma)^{-1} z))}{((2\pi)^n |\Sigma|)^{-1/2} \exp(-\frac{1}{2} z^T \Sigma^{-1} z)}$$  \hfill (75)

$$= \frac{((2\pi)^n \sigma_2^2 k^2 m)^{-1/2} \exp(-\frac{1}{2} \sum_{i=1}^{m} z_i^2 \frac{1}{k^2 \sigma_i^2})}{((2\pi)^n \sigma_2^2)^{-1/2} \exp(-\frac{1}{2} \sum_{i=1}^{m} z_i^2 \frac{1}{\sigma_i^2})}$$  \hfill (76)

$$= \frac{1}{k^2 m} \exp \left( \sum_{i=1}^{m} \frac{z_i^2}{2 k^2 \sigma_i^2} \left(1 - \frac{1}{k^2} \right) \right).$$  \hfill (77)
We define lower level sets as $S_t := \{ z \in \mathcal{Z} : \Lambda(z) \leq t \}$. When $k > 1$, we can write the series of low level sets as

$$S_t = \left\{ z \mid \sum_{i=1}^{m} \frac{z_i^2}{\sigma_i^2} \leq t \right\}, \quad t \geq 0,$$

while for $k < 1$,

$$S_t = \left\{ z \mid \sum_{i=1}^{m} \frac{z_i^2}{\sigma_i^2} \geq t \right\}, \quad t \geq 0.$$

By Neyman Pearson lemma, we can lower bound $q(y|x; \epsilon_1)$ by

$$q(y|x; \epsilon_1) \geq P_1(S_{t'}) \text{, where } P_0(S_{t'}) = q(y|x; \epsilon_0).$$

We denote by $F_{\chi^2_m}(\cdot)$ the cumulative density function of a chi-square distribution with $m$ degree of freedom. Then,

$$P_0 \left( \sum_{i=1}^{m} \frac{z_i^2}{\sigma_i^2} \leq t' \right) = F_{\chi^2_m}(t')$$

And

$$P_1 \left( \sum_{i=1}^{m} \frac{z_i^2}{\sigma_i^2} \leq t' \right) = F_{\chi^2_m}(\frac{t'}{k^2})$$

Thus, for $k > 1$,

$$q(y|x; \epsilon_1) \geq P_1(S_{t'}) = F_{\chi^2_m}(\frac{1}{k^2} F_{\chi^2_m}^{-1}(p_A)).$$

For $k < 1$, however, we have $P_0(S_t) = 1 - F_{\chi^2_m}(t')$, so

$$q(y|x; \epsilon_1) \geq P_1(S_{t'}) = 1 - F_{\chi^2_m}(\frac{1}{k^2} F_{\chi^2_m}^{-1}(1 - p_A)).$$

Remember that our goal is to estimate $\tilde{p}_A = q(y|\hat{E})$. Since its covariance matrix $\Sigma$ only has three unique eigenvalues, leveraging Lemma B.1 three times helps address the class probability $\tilde{p}_A$.

**Lemma B.2.** Suppose a classifier $g$ is smoothed by random linear transformations with random variable $E \sim \mathcal{N}(0, \sigma^2 I^3)$. If $p_A = q(y_A|x,E)$ and $\tilde{p}_A = q(y_A|x,\hat{E})$ with $\hat{E} \sim \mathcal{N}(0, \text{diag}(k_1 I_3, k_2 I_3, k_3 I_3) \sigma^2)$. If $(k_1-1)^2 + (k_2-1)^2 + (k_3-1)^2 \leq R$, then

$$\tilde{p}_A \geq \min\{p_1, p_2, p_3, p_4\}.$$  

where

$$p_1 = F_{\chi^3} \left( \frac{1}{(1 + \frac{R}{\sqrt{3}})^6} F_{\chi^3}^{-1}(p_A) \right).$$

$$p_2 = 1 - \frac{1}{(1 - \frac{R}{\sqrt{3}})^6} F_{\chi^3}^{-1}(1 - p_A).$$

$$p_3 = \inf_{r_1^2 + r_2^2 = R^2, r_1, r_2 \geq 0} F_{\chi^3} \left( \frac{1}{(1 + \frac{r_1}{\sqrt{2}})^4} F_{\chi^3}^{-1} \left( 1 - F_{\chi^3} \left( \frac{1}{(1 + \frac{r_2}{\sqrt{2}})^2} F_{\chi^3}^{-1}(1 - p_A) \right) \right) \right)$$

$$p_4 = \inf_{r_1^2 + r_2^2 = R^2, r_1, r_2 \geq 0} 1 - F_{\chi^3} \left( \frac{1}{(1 - \frac{r_1}{\sqrt{2}})^4} F_{\chi^3}^{-1} \left( 1 - F_{\chi^3} \left( \frac{1}{(1 + \frac{r_2}{\sqrt{2}})^2} F_{\chi^3}^{-1}(p_A) \right) \right) \right)$$
Proof. The four probabilities $p_1, p_2, p_3, p_4$ correspond to four different cases for $k_1, k_2, k_3$.

- $k_1, k_2, k_3 \geq 1$. In this case
  
  \begin{equation}
  F_{x_3}^{-1}(\hat{p}_A) = \frac{1}{k_1^2 k_2^2 k_3^2} F_{x_3}^{-1}(p_A).
  \end{equation}

  Thus,
  
  \begin{equation}
  \inf_{k_1, k_2, k_3 \geq 1} \hat{p}_A = F_{x_3} \left( \frac{1}{(1 + \frac{R}{\sqrt{3}})^6} F_{x_3}^{-1}(p_A) \right).\end{equation}

- $k_1, k_2, k_3 < 1$. In this case
  
  \begin{equation}
  F_{x_3}^{-1}(1 - \hat{p}_A) = \frac{1}{k_1^2 k_2^2 k_3^2} F_{x_3}^{-1}(1 - p_A) \leq \frac{1}{(1 + \frac{R}{\sqrt{3}})^6} F_{x_3}^{-1}(1 - p_A).
  \end{equation}

  Thus,
  
  \begin{equation}
  \inf_{k_1, k_2, k_3 < 1} \hat{p}_A = 1 - F_{x_3} \left( \frac{1}{(1 + \frac{R}{\sqrt{3}})^6} F_{x_3}^{-1}(1 - p_A) \right).\end{equation}

- $k_1, k_2 \geq 1$, $k_3 < 1$. Let $r_1 = \sqrt{(k_1 - 1)^2 + (k_2 - 2)^2}$ and $r_2 = 1 - k_3$.
  
  \begin{equation}
  \hat{p}_A = F_{x_3} \left( \frac{1}{k_1 k_2^2} F_{x_3}^{-1}(1 - p_A') \right) \geq F_{x_3} \left( \frac{1}{(1 + \frac{R}{\sqrt{2}})^4} F_{x_3}^{-1}(1 - p_A) \right).
  \end{equation}

  where $p_A' = 1 - F_{x_3} \left( \frac{1}{k_1 k_2^2} F_{x_3}^{-1}(1 - p_A) \right)$. Hence,
  
  \begin{equation}
  \inf_{k_1, k_2 \geq 1, k_3 < 1} \hat{p}_A = \inf_{r_1^2 + r_2^2 = R^2, r_1, r_2 \geq 0} F_{x_3} \left( \frac{1}{(1 + \frac{R}{\sqrt{2}})^4} F_{x_3}^{-1} \left( 1 - F_{x_3} \left( \frac{1}{(1 - r_2)^2} F_{x_3}^{-1}(1 - p_A) \right) \right) \right).
  \end{equation}

- $k_1, k_2 < 1$ and $k_3 \geq 1$. Let $r_1 = \sqrt{(1 - k_1)^2 + (1 - k_2)^2}$ and $r_2 = k_3 - 1$, so
  
  \begin{equation}
  \hat{p}_A = 1 - F_{x_3} \left( \frac{1}{k_1 k_2^2} F_{x_3}^{-1}(1 - p_A') \right) \geq 1 - F_{x_3} \left( \frac{1}{(1 + \frac{R}{\sqrt{2}})^4} F_{x_3}^{-1}(1 - p_A) \right).
  \end{equation}

  where $p_A' = F_{x_3} \left( \frac{1}{k_1 k_2^2} F_{x_3}^{-1}(p_A) \right)$.

  \begin{equation}
  \inf_{k_1, k_2 < 1, k_3 \geq 1} \hat{p}_A = \inf_{r_1^2 + r_2^2 = R^2, r_1, r_2 \geq 0} 1 - F_{x_3} \left( \frac{1}{(1 + \frac{R}{\sqrt{2}})^4} F_{x_3}^{-1} \left( 1 - F_{x_3} \left( \frac{1}{(1 + r_2)^2} F_{x_3}^{-1}(p_A) \right) \right) \right).
  \end{equation}

As a result, $\hat{p}_A \geq \inf_{k_1, k_2, k_3} \hat{p}_A = \min\{p_1, p_2, p_3, p_4\}$.

Lemma B.2 bridge $\hat{p}_A$ with $p_A$ with a functional relation. However, the infimum used in the definition of $p_3$ and $p_4$ makes them hard to compute. Fortunately, we can draw samples in $\{(r_1, r_2) | r_1^2 + r_2^2 = R^2, r_1, r_2 \geq 0\}$ and calculate lower bounds for $p_3$ and $p_4$. Let $r_1 = R \cos \theta$ and $r_2 = R \sin \theta$ where $\theta \in [0, \frac{\pi}{2}]$. Suppose minimized functions are $L_3$ and $L_4$, Lipschitz in terms of $\theta$, respectively. If we sample $\theta_i = (i/\epsilon) \cdot \frac{\pi}{2}$, then $p_3 \geq \min_i F_3(\theta_i, p_A) - \frac{\epsilon}{2}\epsilon$ and $p_4 \geq \min_i F_4(\theta_i, p_A) - \frac{\epsilon}{2}\epsilon$. By taking partial derivative on $F_3$ and $F_4$, the Lipschitz constants are bounded by

\begin{equation}
L_3 \leq \frac{2Ru}{\sqrt{3}e} + \frac{\sqrt{2}RF_{x_3}^{-1}(1 - p_A)}{\sqrt{3}(1 - R)^{3/2} \sqrt{u}e^{-\frac{u}{2} + 1}},
\end{equation}

\begin{equation}
L_4 \leq \frac{\sqrt{2}Ru}{\sqrt{3}e} + \frac{2RF_{x_3}^{-1}(1 - p_A)}{\sqrt{3}(1 - R)^{3/2} \sqrt{u}e^{-\frac{u}{2} + 1}}, \text{ where } u = F_{x_3}^{-1}(p_A).
\end{equation}
B.5. Proof of Theorem 4: Certifying Z-taper $\circ$ Z-rotation

Proof. Consider the composite transformation $\phi_{TP} \circ \phi_{Rot-z}$ with parameter space $Z = Z_{TP} \times Z_{Rot-z} = \mathbb{R}^2$. As stated in Theorem 4, we sample $\varphi \theta M^2$ parameters $z_{jk} = (\varphi_j, \theta_k)$ in the subspace $S = [-\varphi, \varphi] \times [-\theta, \theta] \subseteq Z$, with $\varphi_j = \frac{2j}{M} - \varphi$ and $\theta_k = \frac{2k}{M} - \theta$. The interpolation error of a point cloud $x = \{p_i\}_{i=1}^N (p_i = x_i, y_i, z_i)$ between two transformations $z_{jk} = (\varphi_j, \theta_k)$ and $z' = (\varphi', \theta')$ is:

$$\mathcal{M}(z_{jk}, z') = \left( \sum_{i=1}^N \| \phi_{TP}(\phi_{Rot-z}(p_i, \theta_k), \varphi_j) - \phi_{TP}(\phi_{Rot-z}(p_i, \theta'), \varphi') \|_2^2 \right)^{1/2}$$

(103)

$$= \left( \sum_{i=1}^N \| \phi_{TP}(p'_i, \varphi_j) - \phi_{TP}(\phi_{Rot-z}(p'_i, \theta - \theta_k), \varphi') \|_2^2 \right)^{1/2} \text{ where } p'_i = \phi_{Rot-z}(p_i, \theta_k)$$

(104)

$$= \left( \sum_{i=1}^N \left[ (1 + \varphi_j z'_i)^2 r_i^2 + (1 + \varphi' z'_i)^2 r_i^2 - 2(1 + \varphi_j z'_i)(1 + \varphi' z'_i)r_i^2 \cos(\theta' - \theta_k) \right] \right)^{1/2} \text{ (105)}$$

$$\leq \left( \sum_{i=1}^N \left[ (\varphi' - \varphi_j)^2 z'_i r_i^2 + (\theta' - \theta_k)^2 r_i^2 (1 + \varphi_j z'_i)(1 + \varphi' z'_i) \right] \right)^{1/2}$$

(106)

Equation (105) uses the law of cosine to compute the $\ell_2$ distance. Note that $\max_{\theta'} \min_k |\theta' - \theta_k| = \frac{1}{M}$ and $\max_{\varphi'} \min_j |\varphi' - \varphi_j| = \frac{1}{M}$. Also, $z_i^2 r_i^2 \leq 1$ for $z_i^2 + r_i^2 \leq 1$. Therefore, the interpolation error

$$\mathcal{M}_S = \max_{z=(\varphi', \theta') \in S} \min_{j,k} \mathcal{M}(z_{jk}, z')$$

(107)

$$\leq \left( \sum_{i=1}^N \left( \frac{1}{4M^2} + \frac{(1 + \varphi)^2}{M^2} \right) \right)^{1/2}$$

(108)

$$= \frac{\sqrt{N(4\varphi^2 + 8\varphi + 5)}}{2M}$$

(109)

It thus follows from Theorem 7 that for any $z \in S$, $y_A = \arg \max_y q(y|\phi(x,z); \epsilon)$; if $\forall j,k$ :

$$\frac{\sigma}{2} \left( \Phi^{-1}(p_A^{(jk)}) - \Phi^{-1}(p_B^{(jk)}) \right) \geq \frac{\sqrt{N(4\varphi^2 + 8\varphi + 5)}}{2M}.$$ 

(110)

B.6. Proof of Theorem 5: Certifying Z-twist $\circ$ Z-taper $\circ$ Z-rotation

Proof. The composite transformation $\phi_{Tz} \circ \phi_{TP} \circ \phi_{Rot-z}$ has a parameter space of $Z = Z_{Twist} \times Z_{Taper} \times Z_{Rot-z} = \mathbb{R}^3$. We calculate the interpolation error of a point cloud $x = \{p_i\}_{i=1}^N$ between two transformations $z_{jkl} = (\varphi_j, \alpha_k, \theta_l)$ and
\[ z' = (\varphi', \alpha', \theta'). \] (Note that z-twist, z-taper and z-rotation are pairwise commutative.)

\[
\mathcal{M}(z', z_{jkl}) = \| \phi(x, z_{jkl}) - \phi(x, z') \|_2^2 \\
= \left( \sum_{i=1}^{N} \| \phi(p_i, z_{jkl}) - \phi(p_i, z') \|_2^2 \right)^{1/2} \\
= \left( \sum_{i=1}^{N} \| \phi_{TP}(p_i', \alpha_k) - \phi_{TP}(\phi_{Rot-z}(p_i', \theta' - \theta_i), \varphi' - \varphi_j, \alpha') \|_2^2 \right)^{1/2} \\
= \left( \sum_{i=1}^{N} \left[ (1 + \alpha_k z_i')^2 r_i^2 + (1 + \alpha' z_i')^2 r_i^2 - 2(1 + \alpha_k z_i')(1 + \alpha' z_i')r_i^2 \cos((\varphi' - \varphi_j)z_i' + \theta' - \theta_i) \right] \right)^{1/2} \\
\leq \left( \sum_{i=1}^{N} \left[ (\alpha_k - \alpha')^2 z_i^2 r_i^2 + (1 + \alpha_k z_i')(1 + \alpha' z_i')((\varphi' - \varphi_j)z_i' + \theta' - \theta_i)^2 r_i^2 \right] \right)^{1/2} \\
\tag{115}
\]

Equation (114) uses the law of cosine for computing the \( \ell_2 \) distance. Following the sampling strategy, for \( z' = (\varphi', \alpha', \theta') \in S = [-\varphi, \varphi, -\alpha, \alpha, -\theta, \theta] \), we have \( \max_{\varphi'} \min_{\varphi} |\varphi' - \varphi| = \frac{1}{4\varphi}, \max_{\alpha'} \min_{\alpha} |\alpha' - \alpha| = \frac{1}{4\alpha}, \max_{\theta'} \min_{\theta} |\theta' - \theta| = \frac{1}{4\theta} \). Hence, the maximum interpolation error for the subspace \( S \) is

\[
\mathcal{M}_S = \max_{\varphi, \varphi', \alpha, \alpha', \theta, \theta'} \min_{\varphi, \alpha, \theta} \mathcal{M}(z', z_{jkl}) \\
\leq \left( \sum_{i=1}^{N} \left[ \frac{z_i^2 r_i^2}{M^2} + \frac{(1 + \alpha z_i')(1 + \alpha' z_i')r_i^2}{M^2} \right] \right)^{1/2} \\
\leq \left( \sum_{i=1}^{N} \left( \frac{1}{4M^2} + \frac{(1 + \alpha)^2 \times \frac{27}{16}}{M^2} \right) \right) = \sqrt{\frac{N(1 + \frac{27}{16}(1 + \alpha^2))}{2M}} \\
\tag{118}
\]

Applying Theorem 7, it is guaranteed that for any \( z \in S, y_A = \arg \max_y(q|\phi(x, z); \epsilon) \), if \( \forall j, k, l, \)

\[
\sigma^2 \left( \Phi^{-1}(p_A^{(jkl)}) - \Phi^{-1}(p_B^{(jkl)}) \right) \geq \sqrt{\frac{N(1 + \frac{27}{16}(1 + \alpha^2))}{2M}} \\
\tag{119}
\]

C. Discussion on \( \ell_p \) Norm Bounded Perturbations

We exhibit the certified robust accuracy under attacks with restricted \( \ell_p \) norm in Table 6. Our TPC framework is directly applicable for certifying \( \ell_2 \) norm bounded attacks. We smooth a base classifier by additive noise \( \epsilon \sim \mathcal{N}(0, I_{3 \times N}) \) so its class probability is \( q(y|x; \epsilon) = \mathbb{E}_\epsilon p(y|x + \epsilon) \). Since additive noise is an additive transformation, the smoothed classifier must be robust for any attacks \( \alpha \in \mathbb{R}^{3 \times N} \) with

\[
\| \alpha \|_2 \leq \frac{\sigma}{2} \left( \Phi^{-1}(p_A) - \Phi^{-1}(p_B) \right). \\
\tag{120}
\]

Though our TPC framework cannot directly be applied to certify against \( \ell_\infty \) norm bounded perturbations, we can still derive a certification bound for point clouds \( x \in \mathbb{R}^{3 \times N} \) by a loose relaxation \( \| \theta \|_\infty \leq \sqrt{3N} \| \theta \|_2 \). In fact, this certification bound for \( \ell_\infty \) is the best we can get in terms of dimension dependence, as (Yang et al., 2020) pointed out that no smoothing techniques can certify nontrivial accuracy within a radius of \( \Omega(N^{-1/2}) \). However, this relaxation is too imprecise when applying it in a high-dimensional space. As a result, the certified accuracy for \( \ell_\infty \) norm drops as the point cloud size increases.

D. Implementation Details

In the implementation of TPC, we need to figure out \( p_A, p_B \) for (additive transformations), or \( p_A^{(i)}, p_B^{(i)} \) for other transformations. Following convention (Cohen et al., 2019), we set \( p_B = 1 - p_A \) which gives an upper bound of true \( p_B \) and
Table 6. Certified robustness for point cloud models under different $\ell_p$ attacks. We achieve similar certification bound for $\ell_\infty$ norm bounded attack as DeepG3D (Lorenz et al., 2021).

<table>
<thead>
<tr>
<th>Attack</th>
<th>Radius</th>
<th>TPC</th>
<th>DeepG3D</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ell_2$</td>
<td>0.05</td>
<td>74.1</td>
<td>82.2</td>
</tr>
<tr>
<td>$\ell_2$</td>
<td>0.1</td>
<td>61.9</td>
<td>70.8</td>
</tr>
<tr>
<td>$\ell_\infty$</td>
<td>0.01</td>
<td><strong>70.9</strong></td>
<td>64.4</td>
</tr>
</tbody>
</table>

Table 7. Inference time of TPC for different transformations

<table>
<thead>
<tr>
<th>Transformation</th>
<th>Attack radius</th>
<th>Inference time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>General rotation</td>
<td>15°</td>
<td>9.56</td>
</tr>
<tr>
<td>Z-rotation</td>
<td>180°</td>
<td>3.12</td>
</tr>
<tr>
<td>Z-shear</td>
<td>0.2</td>
<td>3.46</td>
</tr>
<tr>
<td>Z-twist</td>
<td>180°</td>
<td>3.76</td>
</tr>
<tr>
<td>Z-taper</td>
<td>0.5</td>
<td>10.61</td>
</tr>
<tr>
<td>Linear</td>
<td>0.2</td>
<td>3.94</td>
</tr>
<tr>
<td>Z-twist $\circ$ Z-rotation</td>
<td>50°, 5°</td>
<td>5.34</td>
</tr>
<tr>
<td>Z-taper $\circ$ Z-rotation</td>
<td>0.2, 1°</td>
<td>8.72</td>
</tr>
<tr>
<td>Z-twist $\circ$ Z-taper $\circ$ Z-rotation</td>
<td>20°, 0.2, 1°</td>
<td>9.40</td>
</tr>
</tbody>
</table>

thus resulting in a sound relaxation. To figure out $p_A$, we use Monte-Carlo sampling and Clopper-Pearson confidence interval (Clopper & Pearson, 1934) to obtain a high-confidence lower bound (denoted as $p_A$) of $p_A$, which implies a high-confidence robustness certification. Specifically, we sample $N = 10^3$ times for additive transformations and set confidence level $1 - \alpha = 99.9\%$. Regarding other transformations, to guarantee that overall certification holds with confidence level 99.9%, suppose there are $M$ samples of $z_j$, then for each sampled parameter $z_j$, the $p_A^{(j)}$ estimation uses $N = 10^4$ samples with confidence level $(1 - \alpha/M)$.

E. More Experimental Details

E.1. Runtime

One drawback of randomized smoothing techniques is their extra computation overheads for certification. In our implementation, all transformations are processed in batch to accelerate computation. We report the average inference time of TPC for each point cloud input in Table 7.

E.2. Certified Ratio

The certified ratio is defined as the fraction of test point clouds classified *consistently*, but not necessarily *correctly* under a set of attacks. We compare the certified ratio achieved by our TPC method with the baseline, DeepG3D (Lorenz et al., 2021) in Table 8.

E.3. Evaluation of other architectures

In Section 5 of the main text, we only discuss one particular architecture for the point cloud model for clarity of comparing with the baseline (Lorenz et al., 2021). However, TPC is model-agnostic and can be directly applied to other architectures. We also conduct experiments on a more complicated architecture, CurveNet (Xiang et al., 2021) to show the flexibility and scalability of TPC. The certified robust accuracy for CurveNet is shown in Table 9.
Table 8. Comparison of certified ratio achieved by our transformation-specific smoothing framework TPC and the baseline, DeepG3D (Lorenz et al., 2021). “-” denotes the settings where the baselines cannot scale up to.

<table>
<thead>
<tr>
<th>Transformation</th>
<th>Attack radius</th>
<th>Certified Ratio (%)</th>
<th>TPC</th>
<th>DeepG3D</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZYX-rotation</td>
<td>2°</td>
<td>92.6</td>
<td>72.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5°</td>
<td>79.5</td>
<td>58.7</td>
<td></td>
</tr>
<tr>
<td>General rotation</td>
<td>5°</td>
<td>89.4</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10°</td>
<td>79.5</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15°</td>
<td>63.1</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Z-rotation</td>
<td>20°</td>
<td>99.0</td>
<td>96.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>60°</td>
<td>98.1</td>
<td>95.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>180°</td>
<td>95.2</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Z-shear</td>
<td>0.03</td>
<td>98.6</td>
<td>70.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>97.1</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>91.8</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Z-twist</td>
<td>20°</td>
<td>100.0</td>
<td>23.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>60°</td>
<td>95.6</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>180°</td>
<td>77.5</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Z-taper</td>
<td>0.1</td>
<td>95.2</td>
<td>81.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>93.3</td>
<td>28.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>91.2</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Z-twist ⊕ Z-rotation</td>
<td>20°, 1°</td>
<td>96.5</td>
<td>16.3</td>
<td></td>
</tr>
<tr>
<td>Z-rotation</td>
<td>20°, 5°</td>
<td>96.0</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50°, 5°</td>
<td>95.0</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Z-taper ⊕ Z-rotation</td>
<td>0.1, 1°</td>
<td>89.5</td>
<td>68.5</td>
<td></td>
</tr>
<tr>
<td>Z-rotation</td>
<td>0.2, 1°</td>
<td>86.1</td>
<td>20.7</td>
<td></td>
</tr>
</tbody>
</table>

Table 9. Certified robust accuracy of TPC on CurveNet

<table>
<thead>
<tr>
<th>Transformation</th>
<th>Attack radius</th>
<th>Certified Accuracy (%)</th>
<th>PointNet</th>
<th>CurveNet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z-rotation</td>
<td>180°</td>
<td>81.3</td>
<td>85.4</td>
<td></td>
</tr>
<tr>
<td>Z-shear</td>
<td>0.2</td>
<td>77.7</td>
<td>87.8</td>
<td></td>
</tr>
<tr>
<td>Z-twist</td>
<td>180°</td>
<td>64.3</td>
<td>86.2</td>
<td></td>
</tr>
<tr>
<td>Z-taper</td>
<td>0.2</td>
<td>76.5</td>
<td>88.6</td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>0.2</td>
<td>59.9</td>
<td>77.7</td>
<td></td>
</tr>
</tbody>
</table>