TPC: Transformation-Specific Smoothing for Point Cloud Models

Wenda Chu [1] **Linyi 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 $\pm 20°$) 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.

[1] Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, P. R. China (work done during remote internship at UIUC) [2] University of Illinois Urbana-Champaign (UIUC), Illinois, USA. Correspondence to: Wenda Chu <chuwd19@mails.tsinghua.edu.cn>, Linyi Li <linyi2@illinois.edu>, Bo Li <lbo@illinois.edu>.

Proceedings of the 39th International Conference on Machine Learning, Baltimore, Maryland, USA, PMLR 162, 2022.

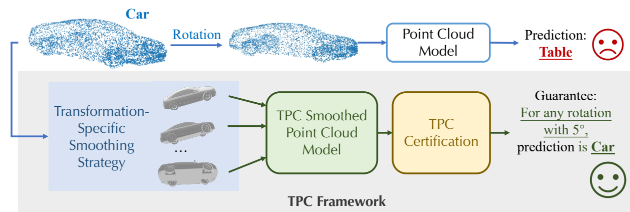


Figure 1. Overview of TPC framework. TPC includes smoothing and certification strategies to provide certified robustness for point cloud models against semantic transformations. Besides the rotation as shown in the figure, TPC provides strong robustness certification for a wide range of other semantic transformations.

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.

For example, regarding general rotation based attacks, we first prove that it is a type of *composable* transformations; we then propose a corresponding smoothing strategy and certify the robustness of the smoothed model with a novel sampling-based algorithm, which is shown to have sound and tight input-dependent sampling error bound. TPC achieves 69.2% certified robust accuracy for any rotation within $10°$. To the best of our knowledge, no prior work can provide robustness certification for rotations within such large angles.

In addition to our theoretical analysis for the certification against different types of semantic transformations, we conduct extensive experiments to evaluate TPC. Compared with existing baselines, our TPC achieves *substantially* higher certified robust accuracy. For example, for any twisting along z-axis within $\pm 20°$, we achieve the probabilistic certified accuracy of 83.8% with a high confidence level 99.9%, while the existing baseline (Lorenz et al., 2021) provides a deterministic certified accuracy of 20.3%. Furthermore, compared with prior works, we show that TPC can: (1) certify a more general class of semantic transformations; (2) certify large-size point clouds; and (3) certify under large perturbation magnitudes. We also show that TPC can certify the robustness for multiple tasks on 3D point clouds, including classification and part segmentation.

We illustrate our TPC framework in Figure 1, and we summarize the main technical **contributions** as follows.

- We propose a general robustness certification framework TPC for point cloud models. We categorize common semantic transformations of point clouds into three categories: additive, composable, and indirectly composable, and provide general smoothing and certification strategies for each.
- We concretize our framework TPC to provide transformation-specific smoothing and certification for various realistic common semantic transformations for point clouds, including rotation, shearing, twisting, and tapering as well as their compositions.
- We conduct extensive experiments and show that TPC (1) achieves significantly higher certified robust accuracy than baselines, (2) provides certification for large-size point clouds and large perturbation magnitudes, (3) provides efficient and effective certification for different tasks such as classification and part segmentation.

Related Work

Certified Robustness of Deep Neural Networks. To mitigate the threats of adversarial attacks on deep neural networks (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 ℓ_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 (Balunović 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 ℓ_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., $\|p_i\|_2 \leq 1$. We mainly consider classification tasks on the point clouds level. Such classification task is defined with a set of labels $\mathcal{Y} = \{1, \ldots, C\}$ and a classifier is defined by a deterministic function $h : \mathcal{X} \to \mathcal{Y}$. More extensions are in Section 5.2.4.

2.1. Semantic Transformations

Semantic transformations on point cloud models are defined as functions $\phi : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ 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 $\phi(x, z)$[1].

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 $\mathcal{Z}_{\text{robust}} \subseteq Z$ for a classifier $h : \mathcal{X} \to \mathcal{Y}$, such that

$$h(x) = h(\phi(x, z)), \forall z \in \mathcal{Z}_{\text{robust}}. \tag{1}$$

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

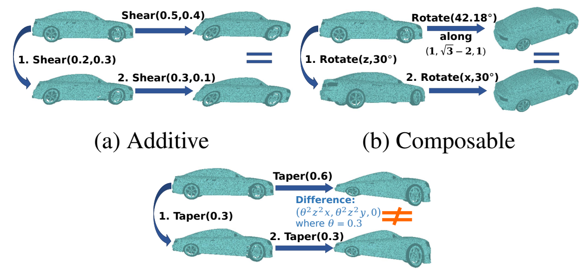
We apply transformation-specific smoothing to an arbitrary base classifier $h : \mathcal{X} \to \mathcal{Y}$ to construct a smoothed classifier. Specifically, the smoothed classifier g predicts the class with the highest conditional probability when the input x is perturbed by some random transformations.

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

$$g(x; \epsilon) = \arg\max_{y \in \mathcal{Y}} q(y|x, \epsilon) = \arg\max_{y \in \mathcal{Y}} \mathbb{E}_\epsilon(p(y|\phi(x, \epsilon))). \tag{2}$$

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.



(a) Additive (b) Composable

(c) Indirectly composable

Figure 2. Illustration of different types of transformations. (a) additive transformations (e.g., shearing), (b) composable transformations (e.g., rotation) and (c) indirectly composable transformations (e.g., tapering).

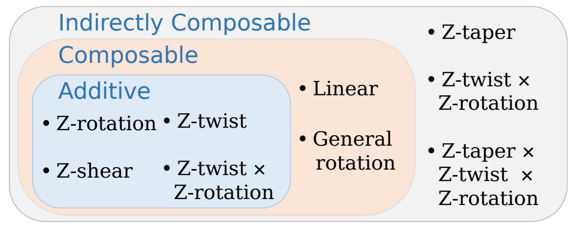


Figure 3. Taxonomy of common 3D semantic transformations for point clouds.

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 $\phi : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ is called **composable** if for any $\alpha \in \mathcal{Z}$ there exists an injective and continuously differentiable function $\gamma_\alpha : \mathcal{Z} \to \mathcal{Z}$ with non-vanishing Jacobian, such that

$$\phi(\phi(x, \alpha), \beta) = \phi(x, \gamma_\alpha(\beta)), \; \forall x \in \mathcal{X}, \beta \in \mathcal{Z}. \tag{3}$$

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

[1]Without loss of generality, we consider untargeted attacks here, and the targeted attack can be derived similarly.

to Euler's rotation theorem, we can always find another rotation $\gamma_\alpha(\beta) \in \mathcal{Z}$ for any two rotations $\alpha, \beta \in \mathcal{Z}$. 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 ϵ_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 ϵ_1, causing difficulties for the certification. Therefore, we focus on a subset of composable transformations, called *additive transformations*, for which it is straight-forward to certify by applying Theorem 6.

3.3. Additive Transformations

We are particularly interested in a subset of composable transformations that the function $\gamma_\alpha : \mathcal{Z} \to \mathcal{Z}$ 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 $\phi : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ is called **additive** if

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

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° along the x axis first and then 10° along the y axis does not equal rotating 20° 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 $\phi : \mathcal{X} \times \mathcal{Z}_\phi \to \mathcal{X}$ is **indirectly composable** if there is a set of composable transformations $\psi : \mathcal{X} \times \mathcal{Z}_\psi \to \mathcal{X}$, such that for any $x \in \mathcal{X}$, there exists a function $\delta_x : \mathcal{Z}_\phi \times \mathcal{Z}_\phi \to \mathcal{Z}_\psi$ with

$$\phi(x, \alpha) = \psi(\phi(x, \beta), \delta_x(\alpha, \beta)), \forall \alpha, \beta \in \mathcal{Z}_\phi. \quad (5)$$

This definition involves more kinds of transformations, since we can choose the transformation ψ as $\psi(x, \delta) = x + \delta$ and let $\delta_x(\alpha, \beta) = \phi(x, \alpha) - \phi(x, \beta)$. This specific assignment of ψ 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 $\phi_{Sz} : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ where $\mathcal{X} = \mathbb{R}^{N\times 3}$ is the space of the point clouds with N points and $\mathcal{Z} = \mathbb{R}^2$ is the parameter space. For any $z = (\theta_1, \theta_2)$, z-shear acting on a point cloud $x \in \mathcal{X}$ with $x = \{p_i\}_{i=1}^N$ yields ($p_i = (x_i, y_i, z_i)^T$)

$$\phi_{Sz}(p_i, z) = (x_i + \theta_1 z_i, y_i + \theta_2 z_i, z_i). \quad (6)$$

Z-twist transformation $\phi_{Tz} : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ is defined similarly but with parameter space $\mathcal{Z} = \mathbb{R}$. 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 ℓ_2 norm for the transformation parameters.

Corollary 1. *Suppose a classifier $g : \mathcal{X} \to \mathcal{Y}$ is smoothed by a transformation $\phi : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ with $\epsilon \sim \mathcal{N}(0, \sigma^2 \mathbb{1}_d)$. Assume its class probability satisfies Equation (24). If the transformation is z-rotation, z-shear, or z-twist ($\phi = \phi_{Sz}, \phi_{Tz}$ or ϕ_{Rot-z}), then it is guaranteed that $g(\phi(x, \alpha); \epsilon) = g(x; \epsilon)$, if the following condition holds:*

$$\|\alpha\|_2 \leq \frac{\sigma}{2}\Big(\Phi^{-1}(p_A) - \Phi^{-1}(p_B)\Big), \alpha \in \mathcal{Z}. \tag{8}$$

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 **z-taper** transformation $\phi_{TP} : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ as tapering along the z-axis, with its parameter space defined by $\mathcal{Z} = \mathbb{R}$. For any point cloud $x = \{p_i\}_{i=1}^N \in \mathcal{X}$ ($p_i = (x_i, y_i, z_i)$) and for any $\theta \in \mathcal{Z}$,

$$\phi_{TP}(p_i, \theta) = \big(x_i(1 + \theta z_i), y_i(1 + \theta z_i), z_i\big). \tag{9}$$

However, z-taper is not a composable transformation, since the composition of two z-taper transformations contains terms with z_i^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 $\mathcal{Z}$ 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 a ϵ-smoothed classifier with random noises $\epsilon \sim \mathcal{N}(0, \sigma^2 \mathbb{1}_{3\times N})$, which predicts $g(x; \epsilon) = \arg\max_y q(y|x; \epsilon) = \arg\max \mathbb{E}_\epsilon p(y|x + \epsilon)$. Let $\{\theta_j\}_{j=0}^M$ be a set of transformation parameters and $\theta_j = (\frac{2j}{M} - 1)R$. Suppose for any i,*

$$q(y_A|\phi_{TP}(x, \theta_j); \epsilon) \geq p_A^{(j)} > p_B^{(j)} \geq \max_{y \neq y_A} q(y|\phi_{TP}(x, \theta_j); \epsilon) \tag{10}$$

Then it is guaranteed that $\forall \theta \in [-R, R]$: $y_A = \arg\max_y q(y|\phi_{TP}(x, \theta); \epsilon)$ if for all $j = 1, \ldots, M$,

$$\frac{\sigma}{2}\left(\Phi^{-1}\left(p_A^{(j)}\right) - \Phi^{-1}\left(p_B^{(j)}\right)\right) \geq \frac{R\sqrt{N}}{2M}. \tag{11}$$

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; and then provide a concrete certification protocol for smoothing and certifying the robustness against this type of transformation.

We define **general rotation** transformations as $\phi_R : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ where $\mathcal{Z} = S^2 \times \mathbb{R}^+$ is the parameter space of rotations. For a rotation $z \in \mathcal{Z}$, 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 $p_i \in \mathbb{R}^3$,

$$\phi_R(p_i, z) = Rot(k, \theta)p_i, \ z = (k, \theta). \tag{12}$$

where $Rot(k, \theta)$ is the rotation matrix that rotates by θ 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 ϵ_0, a semantic attack with parameter $\alpha \in \mathcal{Z}$ results in $\phi_R(\phi_R(x, \alpha), \epsilon_0) = \gamma_\alpha(\epsilon_0)$, 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 $\mathcal{Z} = S^2 \times \mathbb{R}^+$ and certify robustness in the neighborhood of each sample, we are able to certify a large and continuous subspace $\mathcal{Z}_{\text{robust}} \subseteq \mathcal{Z}$. 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 $\phi_R : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ be a general rotation transformation. Let $g : \mathcal{X} \to \mathcal{Y}$ be a classifier smoothed by random noises $\epsilon \sim \mathcal{N}(0, \sigma^2 \mathbb{1}_{3\times N})$, which predicts $g(x; \epsilon) = \arg\max_y q(y|x; \epsilon) = \arg\max_y \mathbb{E}(p(y|x + \epsilon))$. Let $\{z_j\}_{j=1}^M$ 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*

$$\forall k \in S^2, \theta \in [0, R], \exists k_j, \theta_j \text{ s.t. } \langle k, k_j \rangle \leq \epsilon, |\theta - \theta_j| \leq \delta \tag{13}$$

Suppose for any j, the smoothed classifier g has class probabilities that satisfy

$$q(y_A|\phi_R(x, z_j); \epsilon) \geq p_A^{(j)} > p_B^{(j)} \geq \max_{y \neq y_A} q(y|\phi_R(x, z_j); \epsilon). \tag{14}$$

Then it is guaranteed that for any z with rotation angle $\theta < R$: $y_A = \arg\max_y q(y|\phi_R(x, z); \epsilon)$ if $\forall j$,

$$\frac{\sigma}{2}\left(\Phi^{-1}\left(p_A^{(j)}\right) - \Phi^{-1}\left(p_B^{(j)}\right)\right) \geq \pi\sqrt{\frac{\delta^2}{4} + \frac{\epsilon^2 R^2}{8}}\|x\|_2. \tag{15}$$

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 $\|\phi(x, z_j) - \phi(x, z)\|_2 = \|\phi(x, z') - x\|_2 \leq \theta'(\sum_i^N \|p_i\|_2^2)^{1/2}$, 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 θ' 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 $\phi_L : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ has a parameter space of $\mathcal{Z} = \mathbb{R}^{3\times 3}$. For any point cloud $x = \{p_i\}_{i=1}^N \in \mathcal{X}$ and for any $\mathbf{A} \in \mathcal{Z}$,

$$\phi_L(p_i, \mathbf{A}) = (\mathbf{I} + \mathbf{A})p_i. \tag{16}$$

Equation (16) describes any linear transformation with a bounded perturbation $\mathbf{A}$ from the identity transformation $\mathbf{I}$. A natural threat model is considered by (Reisizadeh et al., 2020) that the perturbation matrix $\mathbf{A}$ has a bounded Frobenius norm $\|\mathbf{A}\|_F \leq \epsilon$, 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 $\phi_L : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ where $\mathcal{Z} = \mathbb{R}^{3\times 3}$, with a Gaussian random variable $\epsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_9)$. 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*

$$R = \frac{\sigma\Big(\Phi^{-1}(\tilde{p}_A) - \Phi^{-1}(1-\tilde{p}_A)\Big)}{2 + \sigma\Big(\Phi^{-1}(\tilde{p}_A) - \Phi^{-1}(1-\tilde{p}_A)\Big)}. \tag{17}$$

$\tilde{p}_A$ 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-twist ∘ z-rotation, z-taper ∘ z-rotation and z-twist ∘ z-taper ∘ z-rotation.

Notice that z-twist ∘ z-rotation is an additive function:

$$\begin{aligned}&\phi_{Tz}(\phi_{Rot-z}(\phi_{Tz}(\phi_{Rot-z}(x,\theta_1),\alpha_1),\theta_2),\alpha_2)\\ &= \phi_{Tz}(\phi_{Rot-z}(x,\theta_1+\theta_2),\alpha_1+\alpha_2).\end{aligned} \tag{18}$$

Therefore, we directly apply Corollary 3 in Appendix A.1 to certify z-twist ∘ z-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-twist ∘ z-rotation $\phi : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ where the parameter space $\mathcal{Z} = \mathcal{Z}_{Twist} \times \mathcal{Z}_{Rot-z} = \mathbb{R}^2$. The random variable for smoothing is $\epsilon \sim \mathcal{N}(0, \mathrm{diag}(\sigma_1^2, \sigma_2^2))$. If the class probability of g satisfies Equation (24), then it is guaranteed that $g(\phi(x,\alpha);\epsilon) = g(x;\epsilon)$ for all $(\alpha_1,\alpha_2) \in \mathcal{Z}$, if the following condition holds:*

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

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

Theorem 4. *We denote z-taper ∘ z-rotation by $\phi : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}, \phi = \phi_{TP} \circ \phi_{Rot-z}$ with a parameter space of $\mathcal{Z} = \mathcal{Z}_{TP} \times \mathcal{Z}_{Rot-z} = \mathbb{R}^2$. Let $g : \mathcal{X} \to \mathcal{Y}$ be a classifier smoothed by random noises $\epsilon \sim \mathcal{N}(0, \sigma^2 \mathbb{1}_{3\times N})$.*

For a subspace in the parameter space, $S = [-\varphi, \varphi] \times [-\theta, \theta] \subseteq \mathcal{Z}$, we uniformly sample $\varphi\theta M^2$ 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

$$q(y_A|\phi(x,z_{jk});\epsilon) \geq p_A^{(jk)} > p_B^{(jk)} \geq \max_{y\neq y_A} q(y|\phi(x,z_{jk});\epsilon), \tag{20}$$

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

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

We also consider the composition of three transformations: z-twist ∘ z-taper ∘ z-rotation.

Theorem 5. *We define the composite transformation z-twist ∘ z-taper ∘ z-rotation by $\phi : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$, with input space $\mathcal{X} = \mathbb{R}^{3\times N}$ and parameter space $\mathcal{Z} = \mathcal{Z}_{Twist} \times \mathcal{Z}_{Taper} \times \mathcal{Z}_{Rot-z} = \mathbb{R}^3$. Let $g : \mathcal{X} \to \mathcal{Y}$ be a classifier smoothed by random noises $\epsilon \sim \mathcal{N}(0, \sigma^2 \mathbb{1}_{3\times N})$, which predicts $g(x;\epsilon) = \arg\max_y q(y|x;\epsilon) = \arg\max_y \mathbb{E}(p(y|x+\epsilon))$.*

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

$$\begin{aligned}q(y_A|\phi(x,z_{jkl});\epsilon) \geq p_A^{(jkl)} &> p_B^{(jkl)}\\ &\geq \max_{y\neq y_A} q(y|\phi(x,z_{jkl});\epsilon),\end{aligned} \tag{22}$$

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

$$\frac{\sigma}{2}\Big(\Phi^{-1}\Big(p_A^{(jkl)}\Big) - \Phi^{-1}\Big(p_B^{(jkl)}\Big)\Big) \geq \frac{\sqrt{N(1 + \frac{27}{4}(1+\alpha)^2)}}{2M} \tag{23}$$

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 pre-processing 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

Table 1. Comparison of certified accuracy 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.

Transformation	Attack radius	Certified Accuracy (%) TPC	Certified Accuracy (%) DeepG3D
ZYX-rotation	2°	**81.4**	61.6
	5°	**69.2**	49.6
General rotation	5°	**78.5**	-
	10°	**69.2**	-
	15°	**55.5**	-
Z-rotation	20°	**84.2**	81.8
	60°	**83.8**	81.0
	180°	**81.3**	-
Z-shear	0.03	**83.4**	59.8
	0.1	**82.2**	-
	0.2	**77.7**	-
Z-twist	20°	**83.8**	20.3
	60°	**80.1**	-
	180°	**64.3**	-
Z-taper	0.1	**78.1**	69.0
	0.2	**76.5**	23.9
	0.5	**66.0**	-
Linear	0.1	**74.0**	-
	0.2	**59.9**	-
Z-twist ∘ Z-rotation	20°, 1°	**78.9**	13.8
	20°, 5°	**78.5**	-
	50°, 5°	**76.9**	-
Z-taper ∘ Z-rotation	0.1, 1°	**76.1**	58.2
	0.2, 1°	**72.9**	17.5
Z-twist ∘ Z-taper ∘ Z-rotation	10°, 0.1, 1°	**68.8**	17.5
	20°, 0.2, 1°	**63.1**	4.6

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 ℓ_2 or ℓ_∞ norm bounded 3D perturbations for different point clouds 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 2. Comparison of certified ratio as well as certified accuracy for z-rotation transformations. "-" denotes the settings which the baselines cannot scale up to.

Radius	Certified Ratio (%)		Certified Accuracy (%)	
	TPC	DeepG3D	TPC	DeepG3D
20°	**99.0**	96.7	**84.2**	81.8
60°	**98.1**	95.7	**83.8**	81.0
180°	**95.2**	-	**81.3**	-

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 $\pm\theta$. For z-shear with a parameter space of $\mathbb{R}^2$, we report the certified accuracy against attacks in a certain ℓ_2 parameter radius. For the linear transformation with a parameter space of $\mathbb{R}^{3\times3}$, 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 $\pm20^\circ$.

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 $\pm180^\circ$, 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 $\pm\theta$ (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 $S^2 \times [0, 2\theta]$ of general rotation strictly contains the space defined by $\pm\theta$ 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) $\theta = \pm3^\circ$ compared with DeepG3D (Lorenz et al., 2021)

Points	16	32	64	128	256	512	1024
TPC	**83.2**	**83.8**	**86.6**	**87.4**	**89.4**	**89.8**	**90.5**
DeepG3D	75.4	78.4	79.1	69.4	57.5	42.8	32.3

(b) Certified accuracy of TPC under $\theta = \pm180^\circ$

Points	16	32	64	128	256	512	1024
TPC	73.6	79.3	81.3	81.8	83.0	84.6	83.8

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 $\pm3^\circ$ (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 $\theta = \pm3^\circ$, which cannot fully illustrate the capability of our method. Therefore, we also report our experimental results for z-rotations in $\theta = \pm180^\circ$ 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 ℓ_p norm Bounded 3D-Perturbations

In addition to semantic transformations, we also provide robustness certification for point cloud models against ℓ_p perturbations. Defenses have been proposed for point cloud models against ℓ_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 ℓ_∞ norm. However, a certification bound similar to the baseline (Lorenz et al., 2021) can still be derived,

Table 4. Certified accuracy of TPC for point cloud models under ℓ_2 attacks. The certified accuracy increases as the size of point cloud models increases.

Attack	Radius	Certified Accuracy (%)		
		16	64	256
ℓ_2	0.05	74.1	82.2	84.2
ℓ_2	0.1	61.9	70.8	77.3

Table 5. Comparison of point-wise certified accuracy for the *part segmentation* task. "-" denotes the settings that the baseline does not consider or cannot scale up to.

Transformation	Radius	Certified Accuracy (%)	
		TPC	DeepG3D
Z-rotation	5°	**87.8**	85.7
Z-rotation	10°	**86.1**	84.8
Z-rotation	180°	**70.8**	-
Z-shear	0.2	**86.1**	-
Z-twist	180°	**74.5**	-

using the loose inequality $\|\theta\|_\infty \leq \sqrt{3N}\|\theta\|_2$. Here, we exhibit the certified accuracy for bounded ℓ_2 norm in Table 4 as a benchmark; and omit more details for ℓ_∞ norm to Appendix C. We show that under the ℓ_2 norm bounded 3D-perturbations, TPC certifies even better as the point clouds sizes increase and achieve a high certified accuracy of 77.3% for perturbations with ℓ_2 norm bounded by 0.1.

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 $\pm 5°$ and $\pm 10°$, while we present robustness guarantees for any z-rotation ($\pm 180°$) 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

Aoki, Y., Goforth, H., Srivatsan, R. A., and Lucey, S. Pointnetlk: Robust & efficient point cloud registration using pointnet. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7163–7172, 2019.

Balunović, M., Baader, M., Singh, G., Gehr, T., and Vechev, M. Certifying geometric robustness of neural networks. *Advances in Neural Information Processing Systems 32*, 2019.

Cao, Y., Xiao, C., Cyr, B., Zhou, Y., Park, W., Rampazzi, S., Chen, Q. A., Fu, K., and Mao, Z. M. Adversarial sensor attack on lidar-based perception in autonomous driving. In *Proceedings of the 2019 ACM SIGSAC conference on computer and communications security*, pp. 2267–2281, 2019.

Cao, Y., Wang, N., Xiao, C., Yang, D., Fang, J., Yang, R., Chen, Q. A., Liu, M., and Li, B. Invisible for both camera and lidar: Security of multi-sensor fusion based perception in autonomous driving under physical-world attacks. In *2021 IEEE Symposium on Security and Privacy (SP)*, pp. 176–194. IEEE, 2021.

Chang, A. X., Funkhouser, T. A., Guibas, L. J., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L., and Yu, F. Shapenet: An information-rich 3d model repository. *CoRR*, abs/1512.03012, 2015. URL http://arxiv.org/abs/1512.03012.

Chen, S., Liu, B., Feng, C., Vallespi-Gonzalez, C., and Wellington, C. 3d point cloud processing and learning for autonomous driving: Impacting map creation, localization, and perception. *IEEE Signal Processing Magazine*, 38(1):68–86, 2020.

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

Clopper, C. J. and Pearson, E. S. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4):404–413, 1934.

Cohen, J., Rosenfeld, E., and Kolter, Z. Certified adversarial robustness via randomized smoothing. In Chaudhuri, K. and Salakhutdinov, R. (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 1310–1320. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/cohen19c.html.

Eykholt, K., Evtimov, I., Fernandes, E., Li, B., Rahmati, A., Xiao, C., Prakash, A., Kohno, T., and Song, D. Robust physical-world attacks on deep learning visual classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1625–1634, 2018.

Fang, J., Yang, R., Chen, Q. A., Liu, M., Li, B., et al. Invisible for both camera and lidar: Security of multi-sensor fusion based perception in autonomous driving under physical-world attacks. *arXiv preprint arXiv:2106.09249*, 2021.

Fischer, M., Baader, M., and Vechev, M. Certified defense to image transformations via randomized smoothing. *Advances in Neural Information Processing Systems 33 proceedings (NeurIPS 2020)*, 2020.

Fischer, M., Baader, M., and Vechev, M. T. Scalable certified segmentation via randomized smoothing. *CoRR*, abs/2107.00228, 2021. URL https://arxiv.org/abs/2107.00228.

Hendrycks, D. and Dietterich, T. Benchmarking neural network robustness to common corruptions and perturbations. In *International Conference on Learning Representations*, 2018.

Li, B. 3d fully convolutional network for vehicle detection in point cloud. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1513–1518. IEEE, 2017.

Li, H., Xu, X., Zhang, X., Yang, S., and Li, B. Qeba: Query-efficient boundary-based blackbox attack. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1221–1230, 2020a.

Li, H., Li, L., Xu, X., Zhang, X., Yang, S., and Li, B. Nonlinear gradient estimation for query efficient blackbox attack. In *International Conference on Artificial Intelligence and Statistics (AISTATS 2021)*, Proceedings of Machine Learning Research. PMLR, 13–15 Apr 2021a.

Li, L., Zhong, Z., Li, B., and Xie, T. Robustra: Training provable robust neural networks over reference adversarial space. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI 2019)*, pp. 4711–4717. International Joint Conferences on Artificial Intelligence Organization, 7 2019. doi: 10.24963/ijcai.2019/654. URL https://doi.org/10.24963/ijcai.2019/654.

Li, L., Xie, T., and Li, B. Sok: Certified robustness for deep neural networks. *arXiv preprint arXiv:2009.04131*, 2020b.

Li, L., Weber, M., Xu, X., Rimanic, L., Kailkhura, B., Xie, T., Zhang, C., and Li, B. TSS: transformation-specific smoothing for robustness certification. In Kim, Y., Kim, J., Vigna, G., and Shi, E. (eds.), *CCS '21: 2021 ACM SIGSAC Conference on Computer and Communications Security, Virtual Event, Republic of Korea, November 15 - 19, 2021*, pp. 535–557. ACM, 2021b. doi: 10.1145/3460120.3485258. URL https://doi.org/10.1145/3460120.3485258.

Liu, H., Jia, J., and Gong, N. Z. Pointguard: Provably robust 3d point cloud classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6186–6195, 2021.

Lorenz, T., Ruoss, A., Balunovic, M., Singh, G., and Vechev, M. T. Robustness certification for point cloud models. *CoRR*, abs/2103.16652, 2021. URL https://arxiv.org/abs/2103.16652.

Mirman, M., Gehr, T., and Vechev, M. Differentiable abstract interpretation for provably robust neural networks. In *International Conference on Machine Learning*, pp. 3578–3586. PMLR, 2018.

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.

Qi, C. R., Su, H., Mo, K., and Guibas, L. J. 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, 2017.

Qiu, H., Xiao, C., Yang, L., Yan, X., Lee, H., and Li, B. Semanticadv: Generating adversarial examples via attribute-conditioned image editing. In *European Conference on Computer Vision*, pp. 19–37. Springer, 2020.

Reisizadeh, A., Farnia, F., Pedarsani, R., and Jadbabaie, A. Robust federated learning: The case of affine distribution shifts. *CoRR*, abs/2006.08907, 2020. URL https://arxiv.org/abs/2006.08907.

Salman, H., Yang, G., Li, J., Zhang, P., Zhang, H., Razenshteyn, I., and Bubeck, S. Provably robust deep learning via adversarially trained smoothed classifiers. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, pp. 11292–11303, 2019.

Singh, G., Gehr, T., Püschel, M., and Vechev, M. An abstract domain for certifying neural networks. *Proc. ACM Program. Lang.*, 3(POPL), jan 2019. doi: 10.1145/3290354. URL https://doi.org/10.1145/3290354.

Sun, J., Cao, Y., Chen, Q. A., and Mao, Z. M. Towards robust lidar-based perception in autonomous driving: General black-box adversarial sensor attack and countermeasures. In *29th USENIX Security Symposium (USENIX Security 20)*, pp. 877–894, 2020.

Sun, J., Cao, Y., Choy, C., Yu, Z., Anandkumar, A., Mao, Z., and Xiao, C. Adversarially robust 3d point cloud recognition using self-supervisions. In Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, 2021a. URL https://openreview.net/forum?id=6_sF7BuscXe.

Sun, J., Koenig, K., Cao, Y., Chen, Q. A., and Mao, Z. M. On adversarial robustness of 3d point cloud classification under adaptive attacks. In *The 32nd British Machine Vision Conference, BMVC 2021*, 2021b.

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

Tramer, F., Carlini, N., Brendel, W., and Madry, A. On adaptive attacks to adversarial example defenses. *Advances in Neural Information Processing Systems*, 33, 2020.

Tsuzuku, Y., Sato, I., and Sugiyama, M. Lipschitz-margin training: scalable certification of perturbation invariance for deep neural networks. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pp. 6542–6551, 2018.

Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., and Solomon, J. M. Dynamic graph cnn for learning on point clouds. *Acm Transactions On Graphics (tog)*, 38 (5):1–12, 2019.

Weng, L., Zhang, H., Chen, H., Song, Z., Hsieh, C.-J., Daniel, L., Boning, D., and Dhillon, I. Towards fast computation of certified robustness for relu networks. In *International Conference on Machine Learning*, pp. 5276–5285. PMLR, 2018.

Wong, E. and Kolter, Z. Provable defenses against adversarial examples via the convex outer adversarial polytope. In *International Conference on Machine Learning*, pp. 5286–5295. PMLR, 2018.

Wong, E., Schmidt, F. R., Metzen, J. H., and Kolter, J. Z. Scaling provable adversarial defenses. In *NeurIPS*, 2018.

Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., and Xiao, J. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1912–1920, 2015.

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

Xiang, T., Zhang, C., Song, Y., Yu, J., and Cai, W. Walk in the cloud: Learning curves for point clouds shape analysis, 2021. URL https://arxiv.org/abs/2105.01288.

Xiao, C., Deng, R., Li, B., Yu, F., Liu, M., and Song, D. Characterizing adversarial examples based on spatial consistency information for semantic segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 217–234, 2018.

Xiao, C., Yang, D., Li, B., Deng, J., and Liu, M. Meshadv: Adversarial meshes for visual recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6898–6907, 2019.

Yang, G., Duan, T., Hu, J. E., Salman, H., Razenshteyn, I., and Li, J. Randomized smoothing of all shapes and sizes. In *International Conference on Machine Learning*, pp. 10693–10705. PMLR, 2020.

Zhang, B., Cai, T., Lu, Z., He, D., and Wang, L. Towards certifying l-infinity robustness using neural networks with l-inf-dist neurons. In *International Conference on Machine Learning*, pp. 12368–12379. PMLR, 2021.

Zhang, H., Weng, T.-W., Chen, P.-Y., Hsieh, C.-J., and Daniel, L. Efficient neural network robustness certification with general activation functions. *Advances in Neural Information Processing Systems*, 31:4939–4948, 2018.

Zhang, J., Li, L., Li, H., Zhang, X., Yang, S., and Li, B. Progressive-scale boundary blackbox attack via projective gradient estimation. *ICML*, 2022.

Zhou, Y. and Tuzel, O. Voxelnet: End-to-end learning for point cloud based 3d object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4490–4499, 2018.

Zhu, Q., Li, Y., Hu, H., and Wu, B. Robust point cloud classification based on multi-level semantic relationships for urban scenes. *ISPRS journal of photogrammetry and remote sensing*, 129:86–102, 2017.

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 $\mathcal{Z}$-valued random variables with probability density function f_0 and f_1. Let $\phi : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ be a semantic transformation. Suppose a classifier smoothed by the transformation ϕ predicts $y_A = g(x;\epsilon)$, and that*

$$q(y_A|x,\epsilon) \geq p_A > p_B \geq \max_{y \neq y_A} q(y|x,\epsilon). \tag{24}$$

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

$$\xi(p) := \sup\{\mathbb{P}_1(S) : \underline{S}_{\tau_p} \subseteq S \subseteq \overline{S}_{\tau_p}\} \tag{25}$$

$$\text{where } \tau_p := \inf\{t \geq 0 : \mathbb{P}_0(\overline{S}_t) \geq p\}. \tag{26}$$

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

$$\xi(p_A) + \xi(1-p_B) > 1. \tag{27}$$

Intuitively, $\xi(p_A)$ computes a lower bound for the class probability of y_A when the smoothing distribution changes from ϵ_0 to ϵ_1. Suppose we want to certify an ϵ_0-smoothed classifier against an composable transformation ϕ and $\phi(\phi(x,\alpha),\beta) = \phi(x,\gamma_\alpha(\beta))$. For any attack $\alpha \in \mathcal{Z}$, we assign $\epsilon_1 = \gamma_\alpha(\epsilon_0)$ and check for the condition of Equation (27). Moreover, if ϕ is additive and ϵ_0 is a Gaussian random variable, we have the following corollary:

Corollary 3 (Corollary 7 (Li et al., 2021b)). *Let $\phi : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ be an additive transformation and $\mathcal{Z} = \mathbb{R}^m$. Suppose classifier g is smoothed by a random variable $\epsilon_0 \sim \mathcal{N}(0, \mathrm{diag}(\sigma_1^2, \ldots, \sigma_m^2))$. Assume that the class probability satisfies:*

$$q(y_A|x,\epsilon_0) \geq p_A > p_B \geq \max_{y \neq y_A}(y|x,\epsilon_0). \tag{28}$$

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

$$\sqrt{\sum_{i=1}^{m}\left(\frac{\alpha_i}{\sigma_i}\right)^2} < \frac{1}{2}\Big(\Phi^{-1}(p_A) - \Phi^{-1}(p_B)\Big). \tag{29}$$

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

$$q(y_A|\phi(x,\alpha_i),\epsilon) \geq p_A^{(i)} \geq p_B^{(i)} \geq \max_{y \neq y_A} q(y|\phi(x,\alpha_i),\epsilon). \tag{30}$$

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

$$\mathcal{M}_{\mathcal{S}} := \max_{\alpha \in \mathcal{S}} \min_{1 \leq i \leq N} \mathcal{M}(\alpha, \alpha_i) \tag{31}$$

$$= \max_{\alpha \in \mathcal{S}} \min_{1 \leq i \leq N} \|\phi(x,\alpha) - \phi(x,\alpha_i)\|_2 \tag{32}$$

$$\text{satisfies } \mathcal{M}_{\mathcal{S}} < R := \frac{\sigma}{2} \min_{1 \leq i \leq N} \left(\Phi^{-1}(p_A^{(i)}) - \Phi^{-1}(p_B^{(i)})\right) \tag{33}$$

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 $\phi_{TP} : \mathcal{X} \times R \to \mathcal{X}$ 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

$$\phi_{TP}(p_i, \theta) = \begin{pmatrix} x_i(1+\theta z_i) \\ y_i(1+\theta z_i) \\ z_i \end{pmatrix}, \text{ if } p_i = (x_i, y_i, z_i)^T. \tag{34}$$

We calculate the interpolation error between two parameters θ and θ_j by

$$\begin{aligned} \mathcal{M}(\theta, \theta_j) &= \|\phi_{TP}(x,\theta) - \phi_{TP}(x,\theta_j)\|_2 & (35)\\ &= \left(\sum_{i=1}^N \|\phi_{TP}(p_i,\theta) - \phi_{TP}(p_i,\theta_j)\|_2^2\right)^{1/2} & (36)\\ &= \left(\sum_{i=1}^N (x_i^2+y_i^2)z_i^2(\theta-\theta_j)^2\right)^{1/2} & (37)\\ &\le \frac{\sqrt{N}|\theta-\theta_j|}{2}. & (38)\end{aligned}$$

The last inequality holds because we assume point clouds are normalized into a unit ball, so $(x_i^2+y_i^2)z_i^2 \le \frac{1}{4}$. Recall that we choose $\theta_j = (\frac{2j}{M}-1)R$ 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

$$\begin{aligned} \mathcal{M}_S &= \max_{\theta\in[-R,R]} \min_j \mathcal{M}(\theta,\theta_j) & (39)\\ &\le \frac{R\sqrt{N}}{2M}. & (40)\end{aligned}$$

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 $\forall j$,

$$\frac{\sigma}{2}\left(\Phi^{-1}\left(p_A^{(j)}\right) - \Phi^{-1}\left(p_B^{(j)}\right)\right) \ge \frac{R\sqrt{N}}{2M}. \tag{41}$$

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

Proof. We first recall the definition of general rotation: $\phi_R : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$. The parameter space $\mathcal{Z} = S^2 \times \mathbb{R}^+$ 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:

$$\phi_R(\phi_R(x, z_1), z_2) = \phi_R(x, z_3) \tag{42}$$

$$\text{where } \begin{cases} k_3 = \text{normalize}(\sin\frac{\theta_1}{2}\cos\frac{\theta_2}{2}k_1 + \cos\frac{\theta_1}{2}\sin\frac{\theta_2}{2}k_2 + \sin\frac{\theta_1}{2}\sin\frac{\theta_2}{2}k_2 \times k_1) \\ \theta_3 = 2\arccos(\cos\frac{\theta_1}{2}\cos\frac{\theta_2}{2} - \sin\frac{\theta_1}{2}\sin\frac{\theta_2}{2}k_1\cdot k_2) \end{cases} \tag{43}$$

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:

$$\begin{aligned}
\mathcal{M}(z, z_j) &= \|\phi_R(x, z) - \phi_R(x, z_j)\|_2 && (44)\\
&= \|\phi_R(\phi_R(x, z), z_j^{-1}) - x\|_2, \quad (\text{where } z_j^{-1} = (-k_j, \theta_j)) && (45)\\
&= \left(\sum_{i=1}^{N} \|\phi_R(\phi_R(p_i, z), z_j^{-1}) - p_i\|_2^2\right)^{1/2} && (46)\\
&= \left(\sum_{i=1}^{N} \|\phi_R(p_i, z') - p_i\|_2^2\right)^{1/2} \quad (\text{Let } z' = (k', \theta') \text{ be the composition of } z, z_j^{-1}) && (47)\\
&\le \left(\sum_{i=1}^{N} (\theta' \|p_i\|_2)^2\right)^{1/2} = \theta' \|x\|_2. && (48)
\end{aligned}$$

Assuming $\langle k, k_i\rangle \le \epsilon$, $|\theta - \theta_i| \le \delta$, we derive that for $\theta, \theta_i \in [0, R]$,

$$\begin{aligned}
\cos\frac{\theta'}{2} &= \cos\frac{\theta_j}{2}\cos\frac{\theta}{2} + \cos\langle k, k_i\rangle \sin\frac{\theta_j}{2}\sin\frac{\theta}{2}, && (\langle k, k_i\rangle \le \epsilon) && (49)\\
&\ge \cos\frac{\theta_j - \theta}{2} - \frac{\epsilon^2}{2}\sin\frac{\theta_j}{2}\sin\frac{\theta}{2} && (\text{since } \cos\epsilon \ge 1 - \frac{\epsilon^2}{2}) && (50)\\
&\ge 1 - \left(\frac{\theta_j - \theta}{2}\right)^2 - \frac{\epsilon^2\theta\theta_j}{8}. && (\text{since } \sin x \le x) && (51)\\
&\ge 1 - \frac{\delta^2}{4} - \frac{\epsilon^2 R^2}{8}. && && (52)
\end{aligned}$$

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

$$\begin{aligned}
\theta' &\le 2\arccos\left(1 - \frac{\delta^2}{4} - \frac{\epsilon^2 R^2}{8}\right) && (53)\\
&\le \pi\sqrt{\frac{\delta^2}{4} + \frac{\epsilon^2 R^2}{8}}. && (54)
\end{aligned}$$

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

$$\begin{aligned}
\mathcal{M}_S &= \max_z \min_j \mathcal{M}(z, z_j) && (55)\\
&\le \pi\sqrt{\frac{\delta^2}{4} + \frac{\epsilon^2 R^2}{8}}\|x\|_2. && (56)
\end{aligned}$$

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 πM number of $a_r \in [0, \pi]$.
- For each a_r, uniformly sample $2\pi M \sin\theta_s$ points $b_{rs} \in [0, 2\pi]$.
- Uniformly sample M number of $\theta_t \in [0, R]$.
- Draw $O(M^3)$ samples in total: $z_j = (k_j, \theta_t)$ 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 $\epsilon = \frac{\sqrt{2}}{2M}$ and $\delta = \frac{R}{2M}$ for the conditions in Equation (13).

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

$$\frac{\sigma}{2}\left(\Phi^{-1}\left(p_A^{(j)}\right) - \Phi^{-1}\left(p_B^{(j)}\right)\right) \ge \frac{\sqrt{2}\pi R\|x\|_2}{4M}. \tag{57}$$

Remark. In practise, we implement a tighter bound that $\arccos(1 - x) \le \sqrt{2x} + (\frac{\pi}{2} - \sqrt{2})x^{\frac{3}{2}}$ 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: $\phi_{ZYX-rot} : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ with parameter space $\mathcal{Z} = \mathbb{R}^3$. Specifically, for $z = (\alpha, \beta, \gamma) \in \mathcal{Z}$ and $x = \{p_i\}_{i=1}^N \in \mathcal{X}$,

$$\phi_{ZYX-rot}(p_i, z) = R_z(\gamma)R_y(\beta)R_x(\alpha)p_i, \text{ where } R_z, R_y, R_x \text{ are the rotation matrix along x,y,z axes.} \tag{58}$$

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

$$|\theta| = \arccos\left(\frac{tr(R) - 1}{2}\right) \tag{59}$$

The trace of the rotation matrix for ZYX-rotation is

$$f(\alpha, \beta, \gamma) = tr(R_z(\gamma)R_y(\beta)R_x(\alpha)) = \cos\alpha\cos\beta + \cos\alpha\cos\gamma + \cos\beta\cos\gamma - \sin\alpha\sin\beta\sin\gamma. \tag{60}$$

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

$$\min_{\alpha,\beta,\gamma\in[-\varphi,\varphi]} f(\alpha, \beta, \gamma) = 3\cos^2\varphi - \sin^3\varphi \tag{61}$$

Thus,

$$\begin{aligned}
\cos\theta &= \frac{tr(R) - 1}{2} && (62)\\
&\geq \frac{3\cos^2\varphi - \sin^3\varphi - 1}{2} && (63)\\
&= \frac{(2\cos^2\varphi - 2\sin^2\varphi) + (\sin^2\varphi - \sin^3\varphi)}{2} && (64)\\
&\geq \cos 2\varphi. && (65)
\end{aligned}$$

The rotation angle θ is thus bounded by $\theta \leq 2\varphi$. Hence, any transformation $\phi_{ZYX-rot}$ with $z \in [-\theta, \theta]^3$ and $\theta \in [0, \pi/2]$ belongs to the set of general rotations ϕ_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 $\phi_L : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ where the parameter space is $\mathcal{Z} = \mathbb{R}^{3\times3}$ and $\phi_L(p_i, A) = (I + A)p_i$. The composition of two linear transformations can be expressed by

$$\begin{aligned}
\phi(B)\phi(A) &= I + A + B + BA && (66)\\
&= I + A + (b_{ij} + \sum_k b_{ik}a_{kj})_{ij} && (67)
\end{aligned}$$

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 $\tilde{E}$.

$$g(\phi(A, x); E) = g(x; A + \tilde{E}) \tag{68}$$

Suppose $E \sim \mathcal{N}(0, \sigma^2\mathcal{I}_9)$, then

$$\tilde{E} = \begin{pmatrix} I + A^T & O & O \\ O & I + A^T & O \\ O & O & I + A^T \end{pmatrix} E = SE. \tag{69}$$

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

$$\sqrt{\Sigma_{i=1}^{9}(k_i - 1)^2} \leq \|A\|_F \tag{70}$$

Therefore, the problem is reduced to the diagonal case such that the covariance matrix is diagonal. Assume $\epsilon_0 \sim (0, \sigma^2 I_9)$ and $\epsilon_1 \sim (0, \sigma^2 \cdot \mathrm{diag}(k_1^2, k_2^2, \ldots, k_9^2))$. Let $A = \mathrm{diag}(k_1, k_2, \ldots, k_9)$ 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

$$\sqrt{\sum_{i=1}^{9} \left(\frac{a_i}{k_i}\right)^2} \leq \frac{\sigma}{2}\Big(\Phi^{-1}(\tilde{p}_A) - \Phi^{-1}(\tilde{p}_B)\Big) \tag{71}$$

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

$$\sqrt{\sum_{i=1}^{9} a_i^2} = \|A\|_F \leq \frac{\sigma(1 - \|A\|_F)}{2}\Big(\Phi^{-1}(\tilde{p}_A) - \Phi^{-1}(\tilde{p}_B)\Big) \tag{72}$$

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

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

□

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 $\epsilon_0 \sim \mathcal{N}(0, \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2))$, while another classifier g' is smoothed by a Gaussian random variable $\epsilon_1 \sim \mathcal{N}(0, \mathrm{diag}(k^2\sigma_1, \ldots, k^2\sigma_m^2, \sigma_{m+1}^2, \ldots, \sigma_n^2))$. If the smoothed classifier g predicts $p_A = q(y_A|x; \epsilon_0)$, then*

$$p'_A = q(y|x; \epsilon_1) \geq \begin{cases} F_{\chi^2_m}(\frac{1}{k^2} F^{-1}_{\chi^2_m}(p_A)) & (k \geq 1) \\ 1 - F_{\chi^2_m}(\frac{1}{k^2} F^{-1}_{\chi^2_m}(1 - p_A)) & (k < 1). \end{cases} \tag{74}$$

Proof. We denote by f_0 and f_1 the probability density function of ϵ_0 and ϵ_1. We define a level function as $\Lambda(z) := \frac{f_1(z)}{f_0(z)}$. Suppose $\epsilon_0 \sim \mathcal{N}(0, \Sigma)$ and $\epsilon_1 \sim \mathcal{N}(0, A^2\Sigma)$. Therefore, A is a diagonal matrix with $A_{ii} = k$ if $i \leq m$ and $A_{ii} = 1$ otherwise.

$$\Lambda(z) = \frac{f_1(z)}{f_0(z)} \tag{75}$$

$$= \frac{\big((2\pi)^n |A^2\Sigma|\big)^{-1/2} \exp(-\frac{1}{2}(z^T(A^2\Sigma)^{-1}z))}{((2\pi)^n|\Sigma|)^{-1/2}\exp(-\frac{1}{2}z^T(\Sigma)^{-1}z)} \tag{76}$$

$$= \frac{\big((2\pi)^n \sigma^{2n} k^{2m}\big)^{-1/2} \exp(-\frac{1}{2}\sum_{i=1}^m \frac{z_i^2}{k^2\sigma_i^2})}{((2\pi)^n\sigma^{2n})^{-1/2}\exp(-\frac{1}{2}\sum_{i=1}^m \frac{z_i^2}{\sigma_i^2})} \tag{77}$$

$$= \frac{1}{k^{2m}} \exp\left(\sum_{i=1}^{m} \frac{z_i^2}{2\sigma_i^2}\left(1 - \frac{1}{k^2}\right)\right) \tag{78}$$

We define lower level sets as $S_t := \{z \in \mathcal{Z} : \Lambda(z) \le t\}$. When $k > 1$, we can write the series of lowe level sets as

$$S_t = \left\{ z \;\Big|\; \sum_{i=1}^{m} \frac{z_i^2}{\sigma^2} \le t \right\}, t \ge 0, \tag{79}$$

while for $k < 1$,

$$S_t = \left\{ z \;\Big|\; \sum_{i=1}^{m} \frac{z_i^2}{\sigma^2} \ge t \right\}, t \ge 0. \tag{80}$$

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

$$q(y|x;\epsilon_1) \ge \mathbb{P}_1(S_{t^*}), \text{ where } \mathbb{P}_0(S_{t^*}) = q(y|x;\epsilon_0). \tag{81}$$

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

$$\mathbb{P}_0\Big(\sum_{i=1}^{m} \frac{z_i^2}{\sigma_i^2} \le t' \Big) = F_{\chi^2_m}(t') \tag{82}$$

And

$$\mathbb{P}_1\Big(\sum_{i=1}^{m} \frac{z_i^2}{\sigma_i^2} \le t' \Big) = F_{\chi^2_m}(\frac{t'}{k^2}) \tag{83}$$

Thus, for $k > 1$,

$$q(y|x;\epsilon_1) \ge \mathbb{P}_1(S_{t'}) \tag{84}$$

$$= F_{\chi^2_m}(\frac{1}{k^2} F^{-1}_{\chi^2_m}(p_A)). \tag{85}$$

For $k < 1$, however, we have $\mathbb{P}_0(S_t) = 1 - F_{\chi^2_m}(t')$, so

$$q(y|x;\epsilon_1) \ge \mathbb{P}_1(S_{t'}) \tag{86}$$

$$= 1 - F_{\chi^2_m}(\frac{1}{k^2} F^{-1}_{\chi^2_m}(1 - p_A)). \tag{87}$$

□

Remember that our goal is to estimate $\tilde{p}_A = q(y|x;\tilde{E})$. Since its covariance matrix Σ 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 $E \sim \mathcal{N}(0, \sigma^2 I^9)$. If $p_A = q(y_A|x, E)$ and $\tilde{p}_A = q(y_A|x;\tilde{E})$ with $\tilde{E} \sim \mathcal{N}(0, \mathrm{diag}(k_1^2 I_3, k_2^2 I_3, k_3^2 I_3)\sigma^2)$. If $(k_1-1)^2+(k_2-1)^2+(k_3-1)^2 \le R$, then*

$$\tilde{p}_A \ge \min\{p_1, p_2, p_3, p_4\}. \tag{88}$$

where

$$p_1 = F_{\chi^2_3}\left(\frac{1}{(1+\frac{R}{\sqrt{3}})^6} F^{-1}_{\chi^2_3}(p_A) \right). \tag{89}$$

$$p_2 = 1 - \frac{1}{(1-\frac{R}{\sqrt{3}})^6} F^{-1}_{\chi^2_3}(1 - p_A). \tag{90}$$

$$p_3 = \inf_{r_1^2+r_2^2=R^2, r_1, r_2 \ge 0} F_{\chi^2_3}\left(\frac{1}{(1+\frac{r_1}{\sqrt{2}})^4} F^{-1}_{\chi^2_3}\left(1 - F_{\chi^2_3}\left(\frac{1}{(1-r_2)^2} F^{-1}_{\chi^2_3}(1-p_A) \right)\right)\right) \tag{91}$$

$$p_4 = \inf_{r_1^2+r_2^2=R^2, r_1, r_2 \ge 0} 1 - F_{\chi^2_3}\left(\frac{1}{(1-\frac{r_1}{\sqrt{2}})^4} F^{-1}_{\chi^2_3}\left(1 - F_{\chi^2_3}\left(\frac{1}{(1+r_2)^2} F^{-1}_{\chi^2_3}(p_A) \right)\right)\right) \tag{92}$$

Proof. The four probabilities p_1, p_2, p_3, p_4 corresponds to four different cases for k_1, k_2, k_3.

- $k_1, k_2, k_3 \geq 1$. In this case

$$F^{-1}_{\chi^2_3}(\tilde{p}_A) = \frac{1}{k_1^2 k_2^2 k_3^2} F^{-1}_{\chi^2_3}(p_A). \tag{93}$$

Thus,

$$\inf_{k_1,k_2,k_3 \geq 1} \tilde{p}_A = F_{\chi^2_3}\left(\frac{1}{(1+\frac{R}{\sqrt{3}})^6} F^{-1}_{\chi^2_3}(p_A)\right). \tag{94}$$

- $k_1, k_2, k_3 < 1$. In this case

$$F^{-1}_{\chi^2_3}(1-\tilde{p}_A) = \frac{1}{k_1^2 k_2^2 k_3^2} F^{-1}_{\chi^2_3}(1-p_A) \leq \frac{1}{(1-\frac{R}{\sqrt{3}})^6} F^{-1}_{\chi^2_3}(1-p_A). \tag{95}$$

Thus,

$$\inf_{k_1,k_2,k_3 < 1} \tilde{p}_A = 1 - F_{\chi^2_3}\left(\frac{1}{(1-\frac{R}{\sqrt{3}})^6} F^{-1}_{\chi^2_3}(1-p_A)\right) \tag{96}$$

- $k_1, k_2 \geq 1, k_3 < 1$. Let $r_1 = \sqrt{(k_1-1)^2 + (k_2-2)^2}$ and $r_2 = 1 - k_3$.

$$\tilde{p}_A = F_{\chi^2_3}\Big(\frac{1}{k_1^2 k_2^2} F^{-1}_{\chi^2_3}(1-p'_A)\Big) \geq F_{\chi^2_3}\Big(\frac{1}{(1+\frac{r_1}{\sqrt{2}})^4} F^{-1}_{\chi^2_3}(1-p'_A)\Big). \tag{97}$$

where $p'_A = 1 - F_{\chi^2_3}(\frac{1}{k_3^2} F^{-1}_{\chi^2_3}(1-p_A))$. Hence,

$$\inf_{k_1,k_2 \geq 1, k_3 < 1} \tilde{p}_A = \inf_{r_1^2 + r_2^2 = R^2, r_1, r_2 \geq 0} F_{\chi^2_3}\left(\frac{1}{(1+\frac{r_1}{\sqrt{2}})^4} F^{-1}_{\chi^2_3}\left(1 - F_{\chi^2_3}\left(\frac{1}{(1-r_2)^2} F^{-1}_{\chi^2_3}(1-p_A)\right)\right)\right). \tag{98}$$

- $k_1, k_2 < 1$ and $k_3 \geq 1$. Let $r_1 = \sqrt{(1-k_1)^2 + (1-k_2)^2}$ and $r_2 = k_3 - 1$, so

$$\tilde{p}_A = 1 - F_{\chi^2_3}\Big(\frac{1}{k_1^2 k_2^2} F^{-1}_{\chi^2_3}(1-p'_A)\Big) \geq 1 - F_{\chi^2_3}\Big(\frac{1}{(1-\frac{r_1}{\sqrt{2}})^4} F^{-1}_{\chi^2_3}(1-p'_A)\Big). \tag{99}$$

where $p'_A = F_{\chi^2_3}\Big(\frac{1}{k_3^2} F^{-1}_{\chi^2_3}(p_A)\Big)$.

$$\inf_{k_1,k_2 < 1, k_3 \geq 1} \tilde{p}_A = \inf_{r_1^2 + r_2^2 = R^2, r_1, r_2 \geq 0} 1 - F_{\chi^2_3}\left(\frac{1}{(1-\frac{r_1}{\sqrt{2}})^4} F^{-1}_{\chi^2_3}\left(1 - F_{\chi^2_3}\left(\frac{1}{(1+r_2)^2} F^{-1}_{\chi^2_3}(p_A)\right)\right)\right) \tag{100}$$

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

Lemma B.2 bridge $\tilde{p}_A$ 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 θ, respectively. If we sample $\theta_i = (i/\epsilon) \cdot \frac{\pi}{2}$, then $p_3 \geq \min_i \mathcal{F}_3(\theta_i, p_A) - \frac{\epsilon L_3}{2}$ and $p_4 \geq \min_i \mathcal{F}_4(\theta_i, p_A) - \frac{\epsilon L_4}{2}$. By taking partial derivative on $\mathcal{F}_3$ and $\mathcal{F}_4$, the Lipschitz constants are bounded by

$$L_3 \leq \frac{2Ru}{\sqrt{\pi e}} + \frac{\sqrt{2} R F^{-1}_{\chi^2_3}(1-p_A)}{\sqrt{\pi}(1-R)^3 \sqrt{u} e^{-\frac{u}{2}+1}}. \tag{101}$$

$$L_4 \leq \frac{\sqrt{2}Ru}{\sqrt{\pi e}} + \frac{2R F^{-1}_{\chi^2_3}(1-p_A)}{\sqrt{\pi}(1-\frac{R}{\sqrt{2}})^5 \sqrt{u} e^{-\frac{u}{2}+1}.}, \text{ where } u = F^{-1}_{\chi^2_3}(p_A). \tag{102}$$

B.5. Proof of Theorem 4: Certifying Z-taper ∘ Z-rotation

Proof. Consider the composite transformation $\phi_{TP} \circ \phi_{Rot-z}$ with parameter space $\mathcal{Z} = \mathcal{Z}_{TP} \times \mathcal{Z}_{Rot-z} = \mathbb{R}^2$. As stated in Theorem 4, we sample $\varphi\theta M^2$ parameters $z_{jk} = (\varphi_j, \theta_k)$ in the subspace $S = [-\varphi, \varphi] \times [-\theta, \theta] \subseteq \mathcal{Z}$, 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:

$$\mathcal{M}(z_{jk}, z') = \left(\sum_{i=1}^N \|\phi_{TP}(\phi_{Rot-z}(p_i, \theta_k), \varphi_j) - \phi_{TP}(\phi_{Rot-z}(p_i, \theta'), \varphi')\|_2^2\right)^{1/2} \tag{103}$$

$$= \left(\sum_{i=1}^N \|\phi_{TP}(p_i', \varphi_j) - \phi_{TP}(\phi_{Rot-z}(p_i', \theta' - \theta_k), \varphi')\|_2^2\right)^{1/2} \quad \text{where } p_i' = \phi_{Rot-z}(p_i, \theta_k) \tag{104}$$

$$= \left(\sum_{i=1}^N \left[(1 + \varphi_j z_i')^2 r_i'^2 + (1 + \varphi' z_i')^2 r_i'^2 - 2(1 + \varphi_j z_i')(1 + \varphi' z_i') r_i'^2 \cos(\theta' - \theta_k)\right]\right)^{1/2} \quad (r_i'^2 = x_i'^2 + y_i'^2) \tag{105}$$

$$\leq \left(\sum_{i=1}^N \left[(\varphi' - \varphi_j)^2 z_i'^2 r_i'^2 + (\theta' - \theta_k)^2 r_i'^2 (1 + \varphi_j z_i')(1 + \varphi' z_i')\right]\right)^{1/2} \tag{106}$$

Equation (105) uses the law of cosine to compute the ℓ_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, $z_i'^2 r_i'^2 \leq \frac{1}{4}$ for $z_i'^2 + r_i'^2 \leq 1$. Therefore, the interpolation error

$$\mathcal{M}_S = \max_{z=(\varphi',\theta') \in S} \min_{j,k} \mathcal{M}(z_{jk}, z') \tag{107}$$

$$\leq \left(\sum_{i=1}^N \left(\frac{1}{4M^2} + \frac{(1+\varphi)^2}{M^2}\right)\right)^{1/2} \tag{108}$$

$$= \frac{\sqrt{N(4\varphi^2 + 8\varphi + 5)}}{2M} \tag{109}$$

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

$$\frac{\sigma}{2}\left(\Phi^{-1}\left(p_A^{(jk)}\right) - \Phi^{-1}\left(p_B^{(jk)}\right)\right) \geq \frac{\sqrt{N(4\varphi^2 + 8\varphi + 5)}}{2M}. \tag{110}$$

B.6. Proof of Theorem 5: Certifying Z-twist ∘ Z-taper ∘ Z-rotation

Proof. The composite transformation $\phi_{Tz} \circ \phi_{TP} \circ \phi_{Rot-z}$ has a parameter space of $\mathcal{Z} = \mathcal{Z}_{Twist} \times \mathcal{Z}_{Taper} \times \mathcal{Z}_{Rot-z} = \mathbb{R}^3$. 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.)

$$\mathcal{M}(z', z_{jkl}) = \|\phi(x, z_{jkl}) - \phi(x, z')\|_2 \tag{111}$$
$$= \left(\sum_{i=1}^{N} \|\phi(p_i, z_{jkl}) - \phi(p_i, z')\|_2^2\right)^{1/2} \tag{112}$$
$$= \left(\sum_{i=1}^{N} \|\phi_{TP}(p_i', \alpha_k) - \phi_{TP}(\phi_{Tz}(\phi_{Rot-z}(p_i', \theta' - \theta_l), \varphi' - \varphi_j), \alpha')\|_2^2\right)^{1/2} \tag{113}$$
$$= \left(\sum_{i=1}^{N} \left[(1+\alpha_k z_i')^2 r_i'^2 + (1+\alpha' z_i')^2 r_i'^2 - 2(1+\alpha_k z_i')(1+\alpha' z_i') r_i'^2 \cos((\varphi' - \varphi_j) z_i' + \theta' - \theta_l)\right]\right)^{1/2} \tag{114}$$
$$\leq \left(\sum_{i=1}^{N} \left[(\alpha_k - \alpha')^2 z_i'^2 r_i'^2 + (1+\alpha_k z_i')(1+\alpha' z_i')((\varphi' - \varphi_j) z_i' + \theta' - \theta_l)^2 r_i'^2\right]\right)^{1/2} \tag{115}$$

Equation (114) uses the law of cosine for computing the ℓ_2 distance. Following the sampling strategy, for $z' = (\varphi', \alpha', \theta') \in S = [-\varphi, \varphi, -\alpha, \alpha, -\theta, \theta]$, we have $\max_{\varphi'} \min_j |\varphi' - \varphi_j| = \frac{1}{M}$, $\max_{\alpha'} \min_k |\alpha' - \alpha_k| = \frac{1}{M}$ and $\max_{\theta'} \min_l |\theta' - \theta_l| = \frac{1}{M}$. Hence, the maximum interpolation error for the subspace S is

$$\mathcal{M}_S = \max_{z' \in S} \min_{j,k,l} \mathcal{M}(z', z_{jkl}) \tag{116}$$
$$\leq \left(\sum_{i=1}^{N} \left[\frac{z_i'^2 r_i'^2}{M^2} + \frac{(1+\alpha z_i')^2 (1+z_i')^2 r_i'^2}{M^2}\right]\right)^{1/2} \tag{117}$$
$$\leq \left(\sum_{i=1}^{N} \left(\frac{1}{4M^2} + \frac{(1+\alpha)^2 \times \frac{27}{16}}{M^2}\right)\right) = \frac{\sqrt{N(1 + \frac{27}{4}(1+\alpha^2))}}{2M} \tag{118}$$

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

$$\frac{\sigma}{2}\left(\Phi^{-1}\left(p_A^{(jkl)}\right) - \Phi^{-1}\left(p_B^{(jkl)}\right)\right) \geq \frac{\sqrt{N(1 + \frac{27}{4}(1+\alpha)^2)}}{2M} \tag{119}$$

C. Discussion on ℓ_p Norm Bounded Perturbations

We exhibit the certified robust accuracy under attacks with restricted ℓ_p norm in Table 6. Our `TPC` framework is directly applicable for certifying ℓ_2 norm bounded attacks. We smooth a base classifier by additive noise $\epsilon \sim \mathcal{N}(0, \mathbb{1}_{3\times N})$ so its class probability is $q(y|x;\epsilon) = \mathbb{E}_\epsilon p(y|x+\epsilon)$. Since additive noise is a additive transformation, the smoothed classifier must be robust for any attacks $\alpha \in \mathbb{R}^{3\times N}$ with

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

Though our `TPC` framework cannot directly be applied to certify against ℓ_∞ 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 ℓ_∞ 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 ℓ_∞ 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 ℓ_p attacks. We achieve similar certification bound for ℓ_∞ norm bounded attack as DeepG3D (Lorenz et al., 2021).

Attack	Radius	TPC			DeepG3D	
		16	64	256	64	256
ℓ_2	0.05	74.1	82.2	**84.2**	-	-
ℓ_2	0.1	61.9	70.8	**77.3**	-	-
ℓ_∞	0.01	**70.9**	64.4	47.0	**70.9**	67.0

Table 7. Inference time of TPC for different transformations

Transformation	Attack radius	Inference time (ms)
General rotation	15°	9.56
Z-rotation	180°	3.12
Z-shear	0.2	3.46
Z-twist	180°	3.76
Z-taper	0.5	10.61
Linear	0.2	3.94
Z-twist ∘ Z-rotation	50°, 5°	5.34
Z-taper ∘ Z-rotation	0.2, 1°	8.72
Z-twist ∘ Z-taper ∘ Z-rotation	20°, 0.2, 1°	9.40

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 $\underline{p_A}$) 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.

Transformation	Attack radius	Certified Ratio (%)	
		TPC	DeepG3D
ZYX-rotation	2°	**92.6**	72.8
	5°	**79.5**	58.7
General rotation	5°	**89.4**	-
	10°	**79.5**	-
	15°	**63.1**	-
Z-rotation	20°	**99.0**	96.7
	60°	**98.1**	95.7
	180°	**95.2**	-
Z-shear	0.03	**98.6**	70.7
	0.1	**97.1**	-
	0.2	**91.8**	-
Z-twist	20°	**100.0**	23.9
	60°	**95.6**	-
	180°	**77.5**	-
Z-taper	0.1	**95.2**	81.5
	0.2	**93.3**	28.3
	0.5	**91.2**	-
Z-twist ∘ Z-rotation	20°, 1°	**96.5**	16.3
	20°, 5°	**96.0**	-
	50°, 5°	**95.0**	-
Z-taper ∘ Z-rotation	0.1, 1°	**89.5**	68.5
	0.2, 1°	**86.1**	20.7
Z-twist ∘ Z-taper ∘ Z-rotation	10°, 0.1, 1°	**74.9**	20.7
	20°, 0.2, 1°	**68.7**	5.4

Table 9. Certified robust accuracy of TPC on CurveNet

Transformation	Attack radius	Certified Accuracy (%)	
		PointNet	CurveNet
Z-rotation	180°	81.3	**85.4**
Z-shear	0.2	77.7	**87.8**
Z-twist	180°	64.3	**86.2**
Z-taper	0.2	76.5	**88.6**
Linear	0.2	59.9	**77.7**