Pixel-wise Smoothing for Certified Robustness against Camera Motion Perturbations

Hanjiang Hu
Machine Learning Department,
Carnegie Mellon University

Zuxin Liu
Mechanical Engineering,
Carnegie Mellon University

Linyi Li
Computer Science, University of
Illinois at Urbana-Champaign

Jiacheng Zhu
Computer Science and Artificial Intelligence Laboratory,
Massachusetts Institute of Technology

Ding Zhao
Mechanical Engineering,
Carnegie Mellon University

Abstract

Deep learning-based visual perception models lack robustness when faced with camera motion perturbations in practice. The current certification process for assessing robustness is costly and time-consuming due to the extensive number of image projections required for Monte Carlo sampling in the 3D camera motion space. To address these challenges, we present a novel, efficient, and practical framework for certifying the robustness of 3D-2D projective transformations against camera motion perturbations. Our approach leverages a smoothing distribution over the 2D pixel space instead of in the 3D physical space, eliminating the need for costly camera motion sampling and significantly enhancing the efficiency of robustness certifications. With the pixel-wise smoothed classifier, we are able to fully upper bound the projection errors using a technique of uniform partitioning in camera motion space. Additionally, we extend our certification framework to a more general scenario where only a single-frame point cloud is required in the projection oracle. Through extensive experimentation, we validate the trade-off between effectiveness and efficiency enabled by our proposed method. Remarkably, our approach achieves approximately 80% certified accuracy while utilizing only 30% of the projected image frames.

Proceedings of the 27[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2024, Valencia, Spain. PMLR: Volume 238.

1 INTRODUCTION

Visual perception has been boosted in recent years by leveraging the power of neural networks, with broad applications in robotics and autonomous driving. Despite the success of deep learning based perception, robust perception suffers from external sensing uncertainty in real-world settings, e.g. glaring illumination [Sun et al., 2022, Hu et al., 2023b], sensor placement perturbation [Hu et al., 2022a, Li et al., 2023], pose attack [Xu et al., 2022], motion blurring and corruptions [Sayed and Brostow, 2021, Mintun et al., 2021, Kong et al., 2023], etc. Besides, the internal vulnerability of deep neural networks has been well studied for years, and rich literature reveals that deep neural networks can be easily fooled by adversarial examples. The victim perception models predict incorrect results under stealthy perturbation on pixels [Goodfellow et al., 2014, Szegedy et al., 2013, Xiao et al., 2018] or 2D semantic transformation, e.g. image rotation, scaling, translation, and other pixel-wise deformations of vector fields [Pei et al., 2017, Dreossi et al., 2018, Hosseini and Poovendran, 2018, Engstrom et al., 2019, Hendrycks and Dietterich, 2018, Kanbak et al., 2018, Liu et al., 2018].

In response to such internal model vulnerability from ℓ_p-bounded pixel-wise perturbations, in parallel to many empirical defense methods [Madry et al., 2018, Tramèr et al., 2018, Ma et al., 2018, Tramer et al., 2020], lots of recent work provide provable robustness guarantees and certifications for any bounded perturbations within certain ℓ_p norm threshold [Cohen et al., 2019, Tjeng et al., 2018, Zhang et al., 2018, Dathathri et al., 2020]. As for the challenging 2D semantic transformations including 2D geometric transformation, deterministic verification [Balunović et al., 2019,

Mohapatra et al., 2020, Ruoss et al., 2021, Yang et al., 2022, Hu et al., 2023a] and probabilistic certification methods [Fischer et al., 2020, Li et al., 2021, Alfarra et al., 2021, Hao et al., 2022] are recently proposed to guarantee such robustness, which is more relevant and important to real-world applications.

However, rare literature focuses on the more commonly-seen 3D-2D projective transformation caused by external sensing uncertainty of camera motion perturbations, which commonly exist in real-world applications such as autonomous driving and robotics. The external perturbations may cause severe consequences in safety-critical scenarios if the perception models are fooled by certain translation or rotation changes of the onboard cameras. Recent work CMS [Hu et al., 2022b] studies how the camera motions influence the perception models and presents a resolvable robustness certification framework for such motion perturbations through randomized smoothing in the parameterized camera motion space. However, although CMS gives the tight certification as an upper bound due to the formulation of the resolvable projective transformation [Li et al., 2021, Hao et al., 2022], the high computational cost of "camera shaking" induced by the Monte Carlo sampling in the camera motion space and the requirement of the entire dense point cloud of the object model restricts its practical applications.

To this end, as shown in Figure 1, we introduce a new efficient robustness certification framework with pixel-wise smoothing against camera motion perturbations in a non-resolvable manner. We first construct smoothed classifiers over the pixel level, reducing the cost of redundant image projections in Monte Carlo sampling. Then we propose to use the uniform partitions technique with the consistent camera motion interval to fully cover the projection error based on the pixel-wise smoothed classifier. To further avoid the projection oracle where the whole dense point cloud must be required, we leverage the Lipschitz property of the projection function to approximate the upper bound of the partitioning intervals. This results in the successful certification given the projection oracle with knowing only one-frame point cloud, which is more convenient to obtain in many real-world applications. In addition to the theoretical contributions, we conduct extensive experiments to show the significant trade-off of required projection frames and certified accuracy compared to the resolvable baseline as an upper bound, validating the efficiency and effectiveness of our proposed method. Our contributions are summarized below.

- We propose a new efficient robustness certification framework against camera motion perturbations using pixel-wise smoothing based on the uniform partitioning of camera motion, avoiding the substantial expenses associated with projection sampling in camera motion space.
- We further derive a Lipschitz-based approximation for the upper bound of the partitioning interval and extend our theoretical framework to the general case with only the one-frame point cloud known in the projection oracle.
- Comprehensive comparison experiments show that our method is much more efficient on the MetaRoom dataset: it requires only 30% projected image frames to achieve 80% certified accuracy compared to the upper bound of the resolvable baseline.

2 RELATED WORK

Provable Defenses against ℓ_p-bounded Attacks. Compared to empirical defense approaches [Madry et al., 2018, Tramèr et al., 2018, Ma et al., 2018, Samangouei et al., 2018, Tramer et al., 2020, Pang et al., 2022] to train robust models against specific adversarial perturbations, defense with provable guarantees aims to guarantee the accuracy for all perturbations within some ℓ_p-bounded attack radius [Li et al., 2020, Liu et al., 2019]. The complete certifications [Katz et al., 2017, Ehlers, 2017] are NP-complete for deep neural networks [Li et al., 2020, Zhang et al., 2022b], though they guarantee finding existing attacks. Incomplete certifications are more practical as they provide relaxed guarantees to find non-certifiable examples, which can be categorized into deterministic and probabilistic ones [Tjeng et al., 2018, Wong and Kolter, 2018, Singh et al., 2019, Dathathri et al., 2020, Müller et al., 2022, Zhang et al., 2022a]. Deterministic certifications adopt linear programming [Salman et al., 2019b, Zhang et al., 2018] or semi-definite programming [Raghunathan et al., 2018a, Raghunathan et al., 2018b], but they cannot scale up well to large datasets. Probabilistic certifications [Cohen et al., 2019] based on randomized smoothing show impressive scalability and promising advantages through adversarial training [Salman et al., 2019a] and consistency regularization [Jeong and Shin, 2020]. Lots of work show that the robustness certification can be boosted by improving robustness against Gaussian noise with some denoisers [Salman et al., 2020, Carlini et al., 2022].

Semantic Transformation Robustness Certification. The robustness of deep neural networks against real-world semantic transformations, e.g. translation, rotation, and scaling [Pei et al., 2017, Dreossi et al., 2018, Hosseini and Poovendran, 2018,

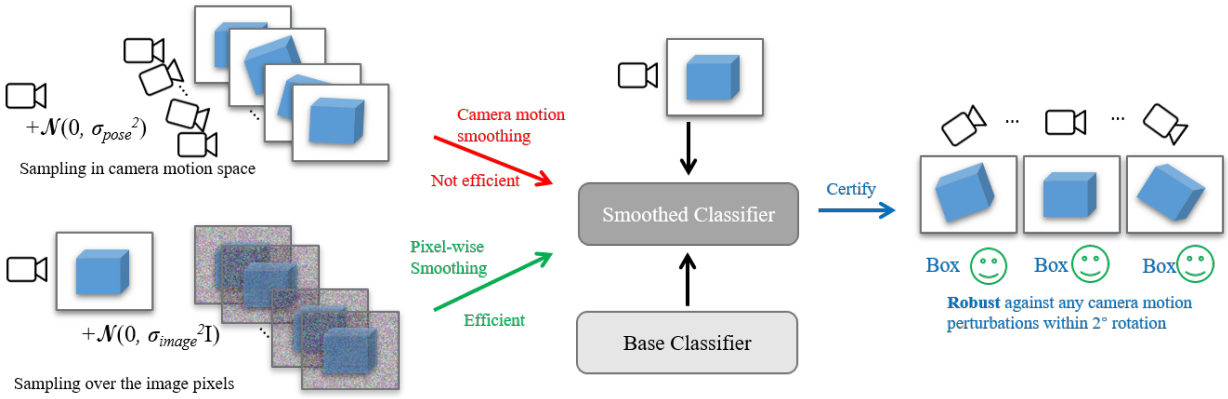

Figure 1: Overview of the robustness certification using pixel-wise smoothing (green) to avoid non-efficient sampling in the camera motion space with too many projected frames required (red).

Engstrom et al., 2019, Hendrycks and Dietterich, 2018, Kanbak et al., 2018, Liu et al., 2018], presents challenges due to the complex optimization landscape involved in adversarial attacks. Recent literature aims to provide the robustness guarantee against semantic transformations [Hao et al., 2022, Li et al., 2021, Ruoss et al., 2021, Alfarra et al., 2021, Balunović et al., 2019], with either function relaxations-based deterministic guarantees [Balunović et al., 2019, Mohapatra et al., 2020, Lorenz et al., 2021, Ruoss et al., 2021, Yang et al., 2022] or random smoothing based high-confident probabilistic guarantees [Fischer et al., 2020, Li et al., 2021, Alfarra et al., 2021, Chu et al., 2022, Hao et al., 2022]. However, the robustness against projective transformation induced by sensor movement is rarely studied in the literature, while we believe it is a commonly seen perturbation source in practical applications. Recent work [Hu et al., 2022b] proposes camera motion smoothing to certify such robustness by leveraging smoothing distribution in the camera motion space. It is known to be tight as a resolvable transformation [Li et al., 2021, Chu et al., 2022, Hao et al., 2022], which has the overlapped domain and support with smoothing distribution in the camera motion space. However, it requires high computational resources for Monte Carlo sampling with image projections as well as the expensive prior of the dense point cloud as an oracle. The above limitations motivate us to study a more efficient robustness certification method.

3 BACKGROUND AND PROBLEM FORMULATION

3.1 3D-2D Projective Transformation

Following the literature in computer vision and graphics [Hartley and Zisserman, 2003, Shinya and Forgue, 1991], image projection can be obtained through the intrinsic matrix $\mathbf{K}$ of the camera and the extrinsic matrix of camera pose $\alpha = (\theta, t) \in \mathcal{Z}$ given dense 3D point cloud $\mathbb{P} \subset \mathbb{R}^3$ based on direct rasterization with z-buffering. In this way, each 3D point $P \in \mathbb{P}$ can be projected to a 2D position on the image plane through the 3D-2D projection function $\rho : \mathbb{P} \times \mathcal{Z} \to \mathbb{R}^2$ and pixel-wise depth function $D : \mathbb{P} \times \mathcal{Z} \to \mathbb{R}$.

Definition 3.1 (3D-2D position projection)**.** For any 3D point $P = (X, Y, Z) \in \mathbb{P} \subset \mathbb{R}^3$ under the camera coordinate with the camera intrinsic matrix $\mathbf{K}$, based on the camera motion of $\alpha = (\theta, t) \in \mathcal{Z} \subset \mathbb{R}^6$ with rotation matrix $\mathbf{R} = \exp(\theta^\wedge) \in SO(3)$ and translation vector $t \in \mathbb{R}^3$, define the projection function $\rho : \mathbb{P} \times \mathcal{Z} \to \mathbb{R}^2$ and the depth function $D : \mathbb{P} \times \mathcal{Z} \to \mathbb{R}$ as

$$[\rho(P, \alpha), 1]^\top = \frac{1}{D(P, \alpha)} \mathbf{K}\mathbf{R}^{-1}(P - t) \tag{1}$$

$$D(P, \alpha) = [0, 0, 1]\mathbf{R}^{-1}(P - t) \tag{2}$$

As defined in [Hu et al., 2022b], given the K-channel colored 3D point cloud $V \in \mathcal{V}$, the 2D colored image $x \in \mathcal{X}$ can be obtained through 3D-2D projective transformation $O : \mathcal{V} \times \mathcal{Z} \to \mathcal{X}$, as shown in Figure 2. Based on this, define the 2D projective transformation $\phi : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ given the relative camera motion $\alpha \in \mathcal{Z}$. Specifically, if only one coordinate of α is non-zero, then it is denoted as one-axis relative camera motion $\alpha \in \mathcal{S}$.

Definition 3.2 (3D-2D K-channel pixel-wise projection)**.** Given the position projection function $\rho : \mathbb{P} \times \mathcal{Z} \to \mathbb{R}^2$ and the depth function $D : \mathbb{P} \times \mathcal{Z} \to \mathbb{R}$ with K-channel 3D point cloud $V \in \mathcal{V} : \mathbb{P} \times [0, 1]^K$, for camera motion $\alpha \in \mathcal{Z}$, define the 3D-2D pixel projection function as $O : \mathcal{V} \times \mathcal{Z} \to \mathcal{X}$ to return K-channel

Table 1: Table of notations for domain and function symbols

Domain Symbols	Meanings	Function Symbols	Meanings
$\mathcal{X} \subset [0,1]^K \times \mathbb{Z}^2$	2D image with K channels	$\rho : \mathbb{P} \times \mathcal{Z}(\text{or } \mathcal{S}) \to \mathbb{R}^2$	3D-2D position projection function
$\mathcal{Z} \subset \mathbb{R}^6$	Parameterized camera motion space	$D : \mathbb{P} \times \mathcal{Z}(\text{or } \mathcal{S}) \to \mathbb{R}$	Pixel-wise depth function
$\mathcal{S} \subset \mathbb{R} \times \mathbf{0}^5$	One-axis relative camera pose	$\phi : \mathcal{X} \times \mathcal{Z}(\text{or } \mathcal{S}) \to \mathcal{X}$	2D projective transformation function
$\mathbb{P} \subset \mathbb{R}^3$	3D points	$O : \mathcal{V} \times \mathcal{Z}(\text{or } \mathcal{S}) \to \mathcal{X}$	3D-2D pixel projection function
$\mathcal{V} \subset \mathbb{R}^3 \times [0,1]^K$	3D points with K channels	$p : \mathcal{Y} \mid \mathcal{X} \to \mathbb{R}$	Score function in base classifier
$\mathcal{Y} \subset \mathbb{Z}$	Label space for classification	$q : \mathcal{Y} \mid \mathcal{X}; \varepsilon \to \mathbb{R}$	Score function in ε-smoothed classifier

image $x \in \mathcal{X}$, $x = O(V, \alpha)$ where

$$x_{k,r,s} = O(V,\alpha)_{k,r,s} = V_{P^*_\alpha, k} \tag{3}$$

$$\text{where } P^*_\alpha = \underset{\{P \in \mathbb{P} \mid \lfloor \rho(P,\alpha) \rfloor = (r,s)\}}{\operatorname{argmin}} D(P,\alpha) \tag{4}$$

where $\lfloor \cdot \rfloor$ is the floor function. Specifically, if $x = O(V, 0)$, given the relative camera pose $\alpha \in \mathcal{Z}$, define the 2D projective transformation $\phi : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ as $\phi(x, \alpha) = O(V, \alpha)$.

3.2 Threat Model and Certification Goal

We formulate the camera motion perturbation as an adversarial attack for the image classifier $g : \mathcal{X} \to \mathcal{Y}$, where there exists some camera pose $\alpha \in \mathcal{Z}$ under which the captured image $\phi(x, \alpha)$ fools the classifier g, making a wrong prediction of object label. Specifically, if the whole dense point cloud V is known as prior, we can obtain the image projection $\phi : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ directly through 3D-2D projective transformation $O : \mathcal{V} \times \mathcal{Z} \to \mathcal{X}$, i.e.

$$x = \phi(x, 0) = O(V, 0), \phi(x, \alpha) = O(V, \alpha) \tag{5}$$

The certification goal is to provide robustness guarantees for vision models against all camera motion perturbations within a certain radius in the camera pose space through 3D-2D projective transformation. We formulate the goal as given the 2D projective transformation ϕ, finding a set $\mathcal{Z}_{\text{adv}} \subseteq \mathcal{Z}_\phi$ for a classifier g such that,

$$g(x) = g(\phi(x, \alpha)), \forall \alpha \in \mathcal{Z}_{\text{adv}}. \tag{6}$$

4 PIXEL-WISE SMOOTHING CERTIFICATION METHOD

4.1 Pixel-wise Smoothed Classifier with Projection

To achieve the certification goal (6), we adopt the stochastic classifier based on the randomized smoothing [Cohen et al., 2019] over pixel level to construct the pixel-wise smoothed classifier, instead of the expensive smoothed classifier using smoothing distribution in camera motion space [Hu et al., 2022b]. Combining the image projection discussed above, we present the definition of smoothed classifier below, by adding zero-mean pixel-wise Gaussian noise to each projected image and taking the empirical mean as the smoothed prediction.

Definition 4.1 (ε-smoothed classifier with 2D image projection)**.** Let $\phi : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ be a 2D projective transformation parameterized with the one-axis relative camera motion $\alpha \in \mathcal{S} \subset \mathcal{Z}$ and $\varepsilon \in \mathcal{X} \sim \mathcal{P}_\varepsilon$ as the smoothing distribution. Let $x = \phi(x, 0)$ be under the original camera pose and $h : \mathcal{X} \to \mathcal{Y}$ be a base classifier $h(x) = \operatorname{argmax}_{y \in \mathcal{Y}} p(y \mid x)$. Under the one-axis relative camera pose $\alpha \in \mathcal{S}$, we define the ε-smoothed classifier $g : \mathcal{X} \to \mathcal{Y}$ as

$$g(x; \alpha; \varepsilon) = \underset{y \in \mathcal{Y}}{\operatorname{argmax}}\, q(y \mid \phi(x, \alpha); \varepsilon) \tag{7}$$

$$\text{where } q(y \mid \phi(x, \alpha); \varepsilon) = \mathbb{E}_{\varepsilon \sim \mathcal{P}_\varepsilon} p(y \mid \phi(x, \alpha) + \varepsilon) \tag{8}$$

Remark 4.2. We remark that the definition of smoothed classifier 4.1 is also applicable to the general 6-DoF translation and rotation of camera motion ($\alpha \in \mathcal{Z}$), but we mainly focus on the one-axis relative camera motion ($\alpha \in \mathcal{S} \subset \mathcal{Z}$) to make it consistent with the previous work [Hu et al., 2022b].

Based on the pixel-wise smoothing, we then introduce the uniform partitions in camera motion space to fully cover all the pixel values within the camera motion perturbation radius. Specifically, we derive the upper bound of the partition interval (PwS) and its Lipschitz-based approximation (PsW-L), which is further used to relax the prior from the entire dense point cloud to the one-frame point cloud in the projection oracle (PwS-OF).

4.2 PwS: Certification via Pixel-wise Smoothing with Prior of Entire Point Cloud

Given the entire dense colored point cloud $V : \mathbb{P} \times [0,1]^K$ for the image projection, the projection function $O(V, \alpha)$ at each pixel $(r, s) \in \mathbb{Z}^2$ is a piecewise constant function w.r.t α, as shown in Figure 3. This is because the projected pixel value is determined by the target 3D point $P^* \in \mathbb{P}$ which has the least projected depth on

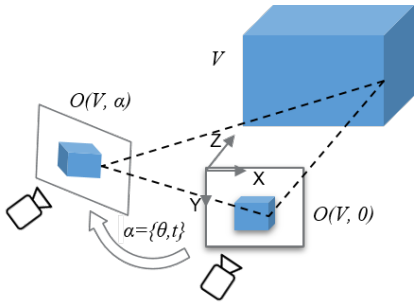

Figure 2: Image projection and coordinate framework in camera motion space.

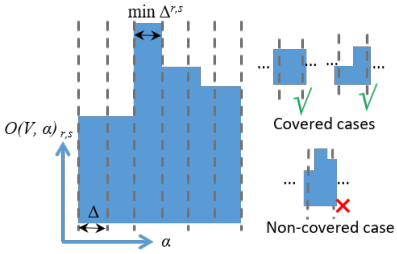

Figure 3: Uniform partitions of Δ in camera motion α to fully cover all the pixel values $O(V,\alpha)_{r,s}$.

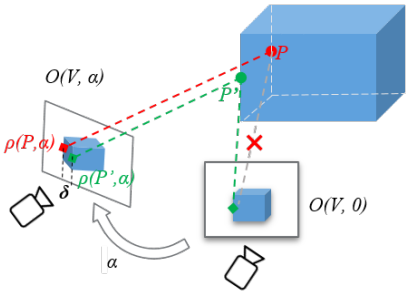

Figure 4: Projection from P' as the one-frame point and unknown P in entire points with δ-convexity.

the pixel (r,s) for any camera pose within the motion interval $\mathbb{U}_{P^*,r,s}$, which is defined as *consistent camera motion interval*, showing the set of views for which the same 3D scene point projects to a given pixel.

Definition 4.3 (Consistent camera motion interval)**.** Given the 3D points $\mathbb{P} \subset \mathbb{R}^3$ and the projection function $\rho : \mathbb{P} \times \mathcal{Z} \to \mathbb{R}^2$ and the depth function $D : \mathbb{P} \times \mathcal{Z} \to \mathbb{R}$, for any $P^* \in \mathbb{P}$ projected on (r,s) with the least depth values, define the camera motion set $\mathbb{U}_{P^*,r,s} \subset \mathcal{Z}$ as the consistent camera motion interval,

$$\mathbb{U}_{P^*,r,s} = \{\alpha \mid P^* = \underset{\{P\in\mathbb{P}|\lfloor\rho(P,\alpha)\rfloor=(r,s)\}}{\operatorname{argmin}} D(P,\alpha)\} \quad (9)$$

Specifically, for one-axis consistent camera motion interval $\mathbb{U}_{P,r,s} \subset \mathcal{S}$, in the absence of ambiguity, we regard $\mathbb{U}_{P,r,s}$ as a subset of one-dimensional real number field based on non-zero coordinate in $\mathcal{S}$ in the following mathematical notations.

Since all the intervals of the piecewise constant function $O(V,\alpha)_{r,s}$ correspond to different 3D points projected to pixel (r,s) as the camera motion α varies within one-axis camera motion perturbation $\mathcal{S}$, we introduce Lemma 4.4, which demonstrates that for any given pixel (r,s), an upper bound $\Delta^{r,s}$ exists for the *consistent camera motion interval* $\mathbb{U}_{P,r,s}$, regardless of $P \in \mathbb{P}$. This upper bound ensures that all 3D-2D projections will fall within the projections of the endpoints of $\Delta^{r,s}$ for any camera motion that falls within $\Delta^{r,s}$. In this scenario, we describe the projection function $O(V,\alpha)_{r,s}$ as being *fully covered* by this consistent interval upper bound $\Delta^{r,s}$, as illustrated in Figure 3. The proof of Lemma 4.4 can be located in Appendix Section B.

Lemma 4.4 (Upper bound of fully-covered motion interval)**.** *Given the projection from entire 3D point $V \in \mathcal{V} : \mathbb{P} \times [0,1]^K$ along one-axis translation or rotation and the consistent camera motion interval $\mathbb{U}_{P,r,s}$ for any $P \in \mathbb{P}$ projected on (r,s), define the interval $\Delta^{r,s}$ as,*

$$\Delta^{r,s} = \min_{P\in\mathbb{P}}\{\sup \mathbb{U}_{P,r,s} - \inf \mathbb{U}_{P,r,s}\} \quad (10)$$

then for any projection on (r,s) under camera motion $u \in \bigcup_{P\in\mathbb{P}} \mathbb{U}_{P,r,s}$, we have $\forall\ 0 \le \Delta^* \le \Delta^{r,s}$

$$O(V, u+\Delta^*)_{r,s} \in \{O(V,u)_{r,s}, O(V,u+\Delta^{r,s})_{r,s}\}$$

Based on $\Delta^{r,s}$ in Lemma 4.4 for each pixel, we adopt uniform partitions over one-axis camera motion space $\mathcal{S} \subseteq \mathcal{Z}$, resulting in the robustness certification in Theorem 4.5. The key to the proof is to upper bound the projection error [Li et al., 2021, Chu et al., 2022] in Equation (11), which can be done through the fully-covered image partitions. The full proof of Theorem 4.5 can be found in Appendix Section B.

Theorem 4.5 (Certification with fully-covered partitions)**.** *For the image projection from entire 3D point cloud $V \in \mathcal{V} : \mathbb{P} \times [0,1]^K$, let ϕ be one-axis rotation or translation with parameters in $\mathcal{S} \subseteq \mathcal{Z}$ and uniformly partitioned $\{\alpha_i\}_{i=1}^N \subseteq \mathcal{S}$ with interval Δ_α, where*

$$\Delta_\alpha \le \min_{r,s} \Delta^{r,s} = \min_{r,s} \min_{P\in\mathbb{P}}\{\sup \mathbb{U}_{P,r,s} - \inf \mathbb{U}_{P,r,s}\}$$

Let $y_A, y_B \in \mathcal{Y}$, $\varepsilon \sim \mathcal{N}(0,\sigma^2\mathbb{I}_d)$ and suppose that for any i, the ε-smoothed classifier defined by $q(y \mid x;\varepsilon) := \mathbb{E}(p(y \mid x+\varepsilon))$ has class probabilities that satisfy with top-2 classes y_A, y_B as $q(y_A \mid \phi(x,\alpha_i);\varepsilon) \ge p_A^{(i)} \ge p_B^{(i)} \ge \max_{y\neq y_A} q(y \mid \phi(x,\alpha_i);\varepsilon)$,

then it is guaranteed that $\forall \alpha \in \mathcal{S}$: $y_A = \operatorname{argmax}_y q(y \mid \phi(x,\alpha);\varepsilon)$ if

$$\begin{aligned}\max_{1\le i\le N-1} \sqrt{\frac{1}{2}\sum_{k,r,s}(O(V,\alpha_i)_{krs} - O(V,\alpha_{i+1})_{krs})^2} \\ < \frac{\sigma}{2}\min_{1\le i\le N}\left(\Phi^{-1}\left(p_A^{(i)}\right) - \Phi^{-1}\left(p_B^{(i)}\right)\right).\end{aligned} \quad (11)$$

4.3 PwS-L: Certification via Pixel-wise Smoothing with Lipschitz-approximated Partitions

In more general cases without knowing the prior of the entire point cloud, the projected pixels within the *consistent camera motion interval* $\mathbb{U}_{P,r,s}$ are easier to find compared to $\mathbb{U}_{P,r,s}$ itself using Definition 4.3. In this case, we propose to approximate the upper bound of the fully-covered interval $\Delta^{r,s}$ as $\Delta_L^{r,s}$ by leveraging the Lipschitz property of projection oracle, as shown in Lemma 4.6.

Lemma 4.6 (Approximated upper bound of fully-covered interval)**.** *For the projection from entire 3D point $V \in \mathcal{V} : \mathbb{P} \times [0,1]^K$, given the one-axis monotonic position projection function ρ in l_∞ norm over camera motion space, if the Lipschitz-based interval $\Delta_L^{r,s}$ is defined as*

$$\Delta_L^{r,s} = \min_{P \in \mathbb{P}} \frac{1}{L_P} \mid \rho(P, \sup \mathbb{U}_{P,r,s}) - \rho(P, \inf \mathbb{U}_{P,r,s}) \mid_\infty$$

where L_P is the Lipschitz constant for projection function ρ given 3D point P,

$$L_P = \max\{\max_{\alpha \in \mathcal{S}} |\frac{\partial \rho_1(P,\alpha)}{\partial \alpha}|, \max_{\alpha \in \mathcal{S}} |\frac{\partial \rho_2(P,\alpha)}{\partial \alpha}|\} \quad (12)$$

then it holds that $\Delta_L^{r,s} \leq \Delta^{r,s} = \min_{P \in \mathbb{P}}\{\sup \mathbb{U}_{P,r,s} - \inf \mathbb{U}_{P,r,s}\}$.

Based on the monotonicity of projection over camera motion and Theorem 4.5, Theorem 4.7 below holds for more general cases but with a smaller approximated fully-covered interval $\Delta_L^{r,s}$ and more uniform partitions N as a trade-off compared to Theorem 4.5. The full proof of Lemma 4.6 and Theorem 4.7 can be found in Appendix Section C.

Theorem 4.7 (Certification with approximated partitions)**.** *For the projection from entire 3D point cloud $V \in \mathcal{V} : \mathbb{P} \times [0,1]^K$ through the monotonic position projection function ρ in l_∞ norm over camera motion space, let ϕ be one-axis rotation or translation with parameters in $\mathcal{S} \subseteq \mathcal{Z}$ and uniformly partitioned $\{\alpha_i\}_{i=1}^N \subseteq \mathcal{S}$ with interval Δ_α, where*

$$\Delta_\alpha \leq \min_{r,s} \min_{P \in \mathbb{P}} \frac{1}{L_P} |\rho(P, \sup \mathbb{U}_{P,r,s}) - \rho(P, \inf \mathbb{U}_{P,r,s})|_\infty$$

$$L_P = \max\{\max_{\alpha \in \mathcal{S}} |\frac{\partial \rho_1(P,\alpha)}{\partial \alpha}|, \max_{\alpha \in \mathcal{S}} |\frac{\partial \rho_2(P,\alpha)}{\partial \alpha}|\} \quad (13)$$

Let $y_A, y_B \in \mathcal{Y}$, $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_d)$ and suppose that for any i, the ε-smoothed classifier defined by $q(y \mid x; \varepsilon) := \mathbb{E}(p(y \mid x + \varepsilon))$ has class probabilities that satisfy with top-2 classes y_A, y_B as $q(y_A \mid \phi(x, \alpha_i); \varepsilon) \geq p_A^{(i)} \geq p_B^{(i)} \geq \max_{y \neq y_A} q(y \mid \phi(x, \alpha_i); \varepsilon)$, then it is guaranteed that $\forall \alpha \in \mathcal{S}$: $y_A = \text{argmax}_y\, q(y \mid \phi(x, \alpha); \varepsilon)$ if

$$\max_{1 \leq i \leq N-1} \sqrt{\frac{1}{2} \sum_{k,r,s} (O(V, \alpha_i)_{krs} - O(V, \alpha_{i+1})_{krs})^2}$$
$$< \frac{\sigma}{2} \min_{1 \leq i \leq N} \left(\Phi^{-1}\left(p_A^{(i)}\right) - \Phi^{-1}\left(p_B^{(i)}\right) \right) \quad (14)$$

4.4 PwS-OF: Certification via Pixel-wise Smoothing with One-Frame Point Cloud as Prior

In this section, we extend our discussion to a more generalized scenario where only a one-frame dense point cloud is known. Such data can be conveniently acquired through various practical methods such as stereo vision [Gennery, 1977], depth cameras [Izadi et al., 2011], or monocular depth estimation [Eigen et al., 2014]. Before delving into the Lipschitz properties of the projection oracle based on a single-frame point cloud, we begin by defining the image projection derived from the one-frame point cloud with δ-convexity, showing how close the pixels projected from entire 3D points are to the pixels projected from the one-frame point cloud, as shown in Figure 4.

Definition 4.8 (3D projection from one-frame point cloud with δ-convexity)**.** Given a one-frame point cloud $\mathbb{P}_1 \subset \mathbb{R}^3$, define the K-channel image $x \in \mathcal{X} : [0,1]^K \times \mathbb{Z}^2$ as

$$x_{k,r,s} = O(V, 0)_{k,r,s} = V_{P_1,k} \quad (15)$$
$$\text{where } P_1 \in \mathbb{P}_1 \text{ s.t. } \lfloor \rho(P_1, \alpha) \rfloor = (r, s) \quad (16)$$

Define δ-convexity that with the minimal $\delta > 0$, for any new point $P \notin \mathbb{P}_1$ and any $\alpha \in \mathcal{S}$, there exists $P' \in \mathbb{P}_1$ where $\mid \rho(P, \alpha) - \rho(P', \alpha) \mid_\infty \leq \delta$, it holds that $D(P, \alpha) \geq D(P', \alpha)$.

Following the Lipschitz-based approximation of the upper bound of the fully-covered interval in Section 4.3, we further approximate the fully-covered upper bound of the interval based on δ-convexity point cloud in Lemma 4.9, followed by the certification in Theorem 4.10. Note that the finite constant C_δ in Lemma 4.9 and Theorem 4.10 can be expressed in the closed form in Appendix Lemma D.2 together with the full proof in Appendix Section D.

Lemma 4.9 (Approximated upper bound of interval from one-frame point cloud)**.** *Given the projection from one-frame 3D point cloud $V_1 \in \mathcal{V}_1 : \mathbb{P}_1 \times [0,1]^K$ and unknown entire point cloud $V \in \mathcal{V} : \mathbb{P} \times [0,1]^K$ with δ-convexity, if the one-axis position projection function ρ is monotonic in l_∞ norm over camera motion space, there exists a finite constant $C_\delta \geq 0$ such that if the*

interval $\Delta_{OF}^{r,s}$ *is defined as,*

$$\Delta_{OF}^{r,s} = \min_{P_1 \in \mathbb{P}_1} \frac{1}{L_{P_1} + C_\delta} \min_{P_1 \in \mathbb{P}_1} (| \rho(P_1, \sup \mathbb{U}_{P_1,r,s}) - \rho(P_1, \inf \mathbb{U}_{P_1,r,s}) |_\infty - 2\delta) \quad (17)$$

where L_{P_1} *is the Lipschitz constant for projection function* ρ *given 3D point* P_1*,*

$$L_{P_1} = \max\{\max_{\alpha \in \mathcal{S}} |\frac{\partial \rho_1(P_1, \alpha)}{\partial \alpha}|, \max_{\alpha \in \mathcal{S}} |\frac{\partial \rho_2(P_1, \alpha)}{\partial \alpha}|\} \quad (18)$$

then for any $P \in \mathbb{P}$*, it holds that* $\Delta_{OF}^{r,s} \leq \Delta^{r,s} = \min_{P \in \mathbb{P}}\{\sup \mathbb{U}_{P,r,s} - \inf \mathbb{U}_{P,r,s}\}$

Theorem 4.10 (Certification with approximated partitions from one-frame point cloud)**.** *For the projection from one-frame 3D point cloud* $V_1 \in \mathcal{V}_1 : \mathbb{P}_1 \times [0,1]^K$ *and unknown entire point cloud* $V \in \mathcal{V} : \mathbb{P} \times [0,1]^K$ *with* δ*-convexity, let* ϕ *be the one-axis rotation or translation with parameters in* $\mathcal{S} \subseteq \mathcal{Z}$ *with the monotonic projection function* ρ *in* l_∞ *norm over* $\mathcal{S}$*, we have uniformly partitioned* $\{\alpha_i\}_{i=1}^N \subseteq \mathcal{S}$ *under interval* Δ_α *with finite constant* $C_\delta \geq 0$ *where*

$$\Delta_\alpha \leq \min_{r,s} \min_{P_1 \in \mathbb{P}_1} \frac{1}{L_{P_1} + C_\delta} \min_{P_1 \in \mathbb{P}_1} (| \rho(P_1, \sup \mathbb{U}_{P_1,r,s}) - \rho(P_1, \inf \mathbb{U}_{P_1,r,s}) |_\infty - 2\delta) \quad (19)$$

$$L_{P_1} = \max\{\max_{\alpha \in \mathcal{S}} |\frac{\partial \rho_1(P_1, \alpha)}{\partial \alpha}|, \max_{\alpha \in \mathcal{S}} |\frac{\partial \rho_2(P_1, \alpha)}{\partial \alpha}|\} \quad (20)$$

Let $y_A, y_B \in \mathcal{Y}$*,* $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_d)$ *and suppose that for any* i*, the* ε*-smoothed classifier defined by* $q(y \mid x; \varepsilon) := \mathbb{E}(p(y \mid x + \varepsilon))$ *has class probabilities that satisfy with top-2 classes* y_A, y_B *as* $q(y_A \mid \phi(x, \alpha_i); \varepsilon) \geq p_A^{(i)} \geq p_B^{(i)} \geq \max_{y \neq y_A} q(y \mid \phi(x, \alpha_i); \varepsilon)$*,*

then it is guaranteed that $\forall \alpha \in \mathcal{S}$*:* $y_A = \text{argmax}_y\, q(y \mid \phi(x, \alpha); \varepsilon)$ *if*

$$\max_{1 \leq i \leq N-1} \sqrt{\frac{1}{2} \sum_{k,r,s} (\phi(x, \alpha_i)_{krs} - \phi(x, \alpha_{i+1})_{krs})^2} < \frac{\sigma}{2} \min_{1 \leq i \leq N} \left(\Phi^{-1}\left(p_A^{(i)}\right) - \Phi^{-1}\left(p_B^{(i)}\right)\right) \quad (21)$$

5 EXPERIMENTS

In this section, we aim to answer two questions: the first one is can we avoid too much "camera shaking" in the certification — prevent sampling in camera motion space through the proposed method with much fewer projected frames required? The second question we want to answer is what is the trade-off of the proposed method regarding efficiency and effectiveness compared to the resolvable baseline as an upper bound? We first get started with the experimental setup to answer these questions. The code is available at `https://github.com/HanjiangHu/pixel-wise-smoothing`. More details about experiments can be found in Appendix Section E.

5.1 Experimental Setup

Dataset and Smoothed Classifiers. We adopt the MetaRoom dataset [Hu et al., 2022b] for the experiment, which contains camera poses associated with the entire dense point cloud. The dataset contains 20 different indoor objects for classification with camera motion perturbations of translation and rotation along x, y, and z axes $(T_z, T_x, T_z, R_z, R_x, R_y)$. The default perception models are based on ResNet-50 and ResNet-101 is used for different model complexity. To enhance the robustness against pixel-wise perturbation, the base classifiers are fine-tuned with both zero-mean Gaussian data augmentation [Cohen et al., 2019] and pre-trained diffusion-based denoiser [Carlini et al., 2022], whose notation is with *-Diffusion*. The default variance σ^2 of pixel-wise smoothing is with $\sigma = 0.5$, while results for $\sigma = 0.25, 0.75$ are also shown in the ablation study. With these pixel-wise smoothed classifiers, the results of Theorem 4.5, 4.7, 4.10 are denoted as *PwS*, *PwS-L* and *PwS-OF*, respectively. All the experiments are conducted on an Ubuntu 20.04 server with NVIDIA A6000 and 512G RAM.

Baseline and Evaluation Metrics. We adopt the camera motion smoothing (CMS) method [Hu et al., 2022b] as the baseline on the MetaRoom dataset, which is based on the *resolvable* projection with the tightest certification as an upper bound. The base classifiers are kept the same for fair comparisons. Note that the diffusion-based denoiser [Carlini et al., 2022], designed for pixel-wise Gaussian denoising, is not applicable to the CMS baseline whose input is without Gaussian noise. We use the *number of required projected frames* as a hardware-independent metric to evaluate certification efficiency, as image capture is the most resource-intensive part of certifying against camera motion. To assess certification effectiveness, we report *certified accuracy* – the ratio of the test images that are both correctly classified and satisfy the guarantee condition in the certification theorems [Li et al., 2021, Chu et al., 2022, Hu et al., 2022b], fairly showing how well the certification goal is achieved under the given radius for different certification methods.

5.2 Efficient Certification Requiring Fewer Projected Frames

We first show that our proposed certification is efficient with much fewer projected frames required. Frome Table 2, it can be seen that for the 10k Monte Carlo

Table 2: Comparison of numbers of required projection frames and the percent ratio w.r.t. the resolvable baseline (100%) for the same 10k Monte Carlo sampling.

Num. of Required Projected Frames	CMS [Hu et al., 2022b]	PwS & PwS-Diffusion	PwS-L & PwS-L-Diffusion	PwS-OF & PwS-OF-Diffusion
T_z, 10mm radius	10k / 100%	3.5k / 35%	5.1k / 51%	5.9k / 59%
R_y, 0.25° radius	10k / 100%	3.4k / 34%	5.0k / 50%	6.1k / 61%

Table 3: Certified accuracy and percent w.r.t. the resolvable baseline as the upper bound (100%)

Radii along T_z	CMS [Hu et al., 2022b]	PwS-Diffusion	PwS-L-Diffusion	PwS-OF-Diffusion
10mm	0.491 / 100%	0.392 / 79.8%	0.392 / 79.8%	0.400 / 81.5%
20mm	0.475 / 100%	0.300 / 63.2%	0.292 / 61.5%	0.300 / 63.2%

Table 4: Comparison of certified accuracy along z-axis translation and y-axis rotation

Projection type	Radii along T_z		Radii along R_y	
	10mm	20mm	0.25°	0.5°
PwS	0.223	0.192	0.142	0.100
PwS-L	0.223	0.192	0.150	0.092
PwS-OF	0.231	0.192	0.150	0.108

sampling to certify z-axis translation T_z and y-axis rotation R_y, the baseline CMS [Hu et al., 2022b] needs 10k projected frames to construct the camera motion smoothed classifier, while our proposed certification only needs 30% - 60% projected frames as the partitioned images with pixel-wise smoothing of 10k Monte Carlo sampling. Besides, since the uniform partitioning can be done in the offline one-time manner, the proposed methods successfully avoid the impractical "camera shaking" dilemma in the robustness certification against camera motion.

We can also find that the certified accuracy of *PwS*, *PwS-L* and *PwS-OF* are very close in Table 4 and 3, because the only difference is that they use different numbers of partitioned images as shown in Table 2. This is consistent with the theoretical analysis that our proposed three certification theorems are theoretically supposed to have the same certified accuracy performance but with different numbers of projected frames as partitions.

5.3 The trade-off of Certified Accuracy and Efficiency

In this section, we present the trade-off regarding certified accuracy and the number of required projected frames compared to the tightest certification CMS [Hu et al., 2022b] as an upper bound as the resolvable case [Li et al., 2021, Hao et al., 2022]. In Table 3, although our proposed methods with data augmentation are looser than CMS due to the pixel-wise smoothing with much fewer projected frames and better efficiency, the diffusion-based *PwS* can remarkably boost the certified accuracy to achieve about 80% of the upper bound with only 30% of image projections in Table 2. Therefore, a significant trade-off can be seen by slightly sacrificing the certification effectiveness but saving a huge amount of image projection frames in certifications.

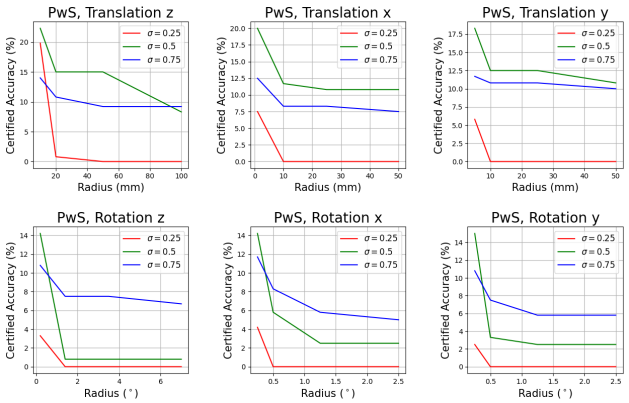

Figure 5: Certified accuracy of ResNet50 with smoothing variance $\sigma = 0.25, 0.5, 0.75$ under different radii along $T_z, T_x, T_y, R_z, R_x, R_y$.

5.4 Ablation Study

Influence of the variance of the smoothing distribution. As shown in Table 5, for all translation and rotation axes, the certified accuracy with $\sigma = 0.5$ is higher than that with $\sigma = 0.75$, showing the accuracy/robustness trade-off [Cohen et al., 2019]. However, the performance with $\sigma = 0.25$ is poor because the pixel-wise smoothing is too weak to cover the pro-

Table 5: Certified accuracy of different smoothing variance σ^2 in all projections for PwS

Projection and radius	T_z, 10mm	T_x, 5mm	T_y, 5mm	R_z, 0.7°	R_x, 0.25°	R_y, 0.25°
$\sigma = 0.25$	0.198	0.0	0.058	0.0	0.042	0.025
$\sigma = 0.5$	0.223	0.192	0.183	0.133	0.142	0.150
$\sigma = 0.75$	0.140	0.117	0.117	0.092	0.117	0.108

Table 6: Certified accuracy with different numbers of partitioned images in pixel-wise smoothing

Number of Partitions Radius 0.25°, y-axis rotation R_y	1000	2000	3000	4000	5000	6000	7000
ResNet50	0.083	0.117	0.142	0.142	0.150	0.150	0.150
ResNet101	0.092	0.092	0.117	0.125	0.125	0.125	0.125

Table 7: Certified accuracy of different model complexity

Translation T_z	10mm		100mm	
σ of PwS	0.5	0.75	0.5	0.75
ResNet50	0.223	0.140	0.142	0.091
ResNet101	0.150	0.140	0.108	0.042

jection errors in the certification condition (11).

Trend of certified accuracy along larger radii. From Figure 5, it can be seen that given the smoothing variance, the certified accuracy generally decreases as the perturbation radii increase, showing that it is more likely to have non-robust adversarial samples in larger perturbation of camera motion. However, the decay of certification is usually slower with larger radii, indicating that there exist some camera poses which are certifably robust against very large perturbation of camera motion. Besides, under the smaller perturbation radii, $\sigma = 0.5$ has the best performance, while certification with larger variance $\sigma = 0.75$ performs better when the perturbation radius of rotation axes goes larger.

Performance with different partitions. In this ablation, we discuss the influence of partition numbers corresponding to the theory in empirical experiments under different models. As shown in Table 6, it can be seen that as the partition number goes up, the certification performance tends to converge, where the partitioned interval is within the upper bound of the motion interval and is consistent with Lemma 4.4 and Table 2.

Performance under different model complexity. In Table 7, we compare the certified accuracy performance of different model complexity. We can find larger model complexity will be less certifiably robust with lower certified accuracy under various variances and perturbation radii, which is consistent [Hu et al., 2022b] with previous work due to overfitting.

6 CONCLUSION AND LIMITATIONS

In this work, we propose an efficient robustness certification framework against camera motion perturbation through pixel-wise smoothing. We introduce a new partitioning method to fully cover the projection errors through the pixel-wise smoothed classifier. Furthermore, we adopt the Lipschitz property of projection to approximate the partition intervals and extend our framework to the case of only requiring the one-frame point cloud. Extensive experiments show a significant trade-off of using only 30% projected frames but achieving 80% certified accuracy compared to the upper bound baseline.

For the potential limitations, our main theorems are mostly based on direct rasterization with z-buffering, which may differ from real-world imaging. However, we believe that our uniform partition method can also deal with the more complicated cases of existing interpolation errors based on [Li et al., 2021]. Besides, although we have relaxed the requirement of the entire point cloud in projection oracle in certification, we still need some prior point cloud and the assumption about the convexity of the scenes. Another limitation lies in that there is no real-world validation, especially in the setting of autonomous driving due to the lack of such outdoor dataset, although the MetaRoom dataset provides a realistic indoor environment, which also points out the future work for the community of computer vision and other applications. Regarding negative social impact, improved robustness in visual perception models might lead to increased surveillance capabilities and potential misuse, infringing on individual privacy and raising ethical concerns.

Acknowledgements

We thank Dr. Bo Li from the University of Chicago for the insightful discussion and feedback.

References

[Alfarra et al., 2021] Alfarra, M., Bibi, A., Khan, N., Torr, P. H., and Ghanem, B. (2021). Deformrs: Certifying input deformations with randomized smoothing. *arXiv preprint arXiv:2107.00996*.

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

[Carlini et al., 2022] Carlini, N., Tramer, F., Kolter, J. Z., et al. (2022). (certified!!) adversarial robustness for free! *arXiv preprint arXiv:2206.10550*.

[Chu et al., 2022] Chu, W., Li, L., and Li, B. (2022). Tpc: Transformation-specific smoothing for point cloud models. In *International Conference on Machine Learning*. PMLR.

[Cohen et al., 2019] Cohen, J., Rosenfeld, E., and Kolter, Z. (2019). Certified adversarial robustness via randomized smoothing. In *International Conference on Machine Learning*, pages 1310–1320. PMLR.

[Dathathri et al., 2020] Dathathri, S., Dvijotham, K., Kurakin, A., Raghunathan, A., Uesato, J., Bunel, R. R., Shankar, S., Steinhardt, J., Goodfellow, I., Liang, P. S., and Kohli, P. (2020). Enabling certification of verification-agnostic networks via memory-efficient semidefinite programming. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 5318–5331.

[Deng et al., 2009] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee.

[Dhariwal and Nichol, 2021] Dhariwal, P. and Nichol, A. (2021). Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34:8780–8794.

[Dreossi et al., 2018] Dreossi, T., Jha, S., and Seshia, S. A. (2018). Semantic adversarial deep learning. In *International Conference on Computer Aided Verification*, pages 3–26. Springer.

[Ehlers, 2017] Ehlers, R. (2017). Formal verification of piece-wise linear feed-forward neural networks. In *International Symposium on Automated Technology for Verification and Analysis*, pages 269–286. Springer.

[Eigen et al., 2014] Eigen, D., Puhrsch, C., and Fergus, R. (2014). Depth map prediction from a single image using a multi-scale deep network. *Advances in neural information processing systems*, 27.

[Engstrom et al., 2019] Engstrom, L., Tran, B., Tsipras, D., Schmidt, L., and Madry, A. (2019). Exploring the landscape of spatial robustness. In *International conference on machine learning*, pages 1802–1811. PMLR.

[Fischer et al., 2020] Fischer, M., Baader, M., and Vechev, M. T. (2020). Certified defense to image transformations via randomized smoothing. In *NeurIPS*.

[Gennery, 1977] Gennery, D. B. (1977). A stereo vision system for an autonomous vehicle. In *IJCAI*, pages 576–582.

[Goodfellow et al., 2014] Goodfellow, I. J., Shlens, J., and Szegedy, C. (2014). Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*.

[Hao et al., 2022] Hao, Z., Ying, C., Dong, Y., Su, H., Song, J., and Zhu, J. (2022). Gsmooth: Certified robustness against semantic transformations via generalized randomized smoothing. In *International Conference on Machine Learning*, pages 8465–8483. PMLR.

[Hartley and Zisserman, 2003] Hartley, R. and Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge university press.

[He et al., 2016] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

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

[Hosseini and Poovendran, 2018] Hosseini, H. and Poovendran, R. (2018). Semantic adversarial examples. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1614–1619.

[Hu et al., 2023a] Hu, H., Liu, C., and Zhao, D. (2023a). Robustness verification for perception models against camera motion perturbations. In *ICML Workshop on Formal Verification of Machine Learning (WFVML)*.

[Hu et al., 2022a] Hu, H., Liu, Z., Chitlangia, S., Agnihotri, A., and Zhao, D. (2022a). Investigating the impact of multi-lidar placement on object detection for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2550–2559.

[Hu et al., 2022b] Hu, H., Liu, Z., Li, L., Zhu, J., and Zhao, D. (2022b). Robustness certification of visual perception models via camera motion smoothing. In *6th Annual Conference on Robot Learning*.

[Hu et al., 2023b] Hu, H., Yang, B., Qiao, Z., Liu, S., Zhu, J., Liu, Z., Ding, W., Zhao, D., and Wang, H. (2023b). Seasondepth: Cross-season monocular depth prediction dataset and benchmark under multiple environments. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 11384–11389. IEEE.

[Izadi et al., 2011] Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., et al. (2011). Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 559–568.

[Jeong and Shin, 2020] Jeong, J. and Shin, J. (2020). Consistency regularization for certified robustness of smoothed classifiers. *Advances in Neural Information Processing Systems*, 33:10558–10570.

[Kanbak et al., 2018] Kanbak, C., Moosavi-Dezfooli, S.-M., and Frossard, P. (2018). Geometric robustness of deep networks: analysis and improvement. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4441–4449.

[Katz et al., 2017] Katz, G., Barrett, C., Dill, D. L., Julian, K., and Kochenderfer, M. J. (2017). Reluplex: An efficient smt solver for verifying deep neural networks. In *International conference on computer aided verification*, pages 97–117. Springer.

[Kong et al., 2023] Kong, L., Xie, S., Hu, H., Ng, L. X., Cottereau, B., and Ooi, W. T. (2023). Robodepth: Robust out-of-distribution depth estimation under corruptions. *Advances in Neural Information Processing Systems*, 36.

[Li et al., 2021] Li, L., Weber, M., Xu, X., Rimanic, L., Kailkhura, B., Xie, T., Zhang, C., and Li, B. (2021). Tss: Transformation-specific smoothing for robustness certification. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, pages 535–557.

[Li et al., 2020] Li, L., Xie, T., and Li, B. (2020). Sok: Certified robustness for deep neural networks. *arXiv preprint arXiv:2009.04131*.

[Li et al., 2023] Li, Y., Hu, H., Liu, Z., Xu, X., Zhao, D., and Huang, X. (2023). Influence of camera-lidar configuration on 3d object detection for autonomous driving. *arXiv preprint arXiv:2310.05245*.

[Liu et al., 2019] Liu, C., Arnon, T., Lazarus, C., Barrett, C., and Kochenderfer, M. J. (2019). Algorithms for verifying deep neural networks. *arXiv preprint arXiv:1903.06758*.

[Liu et al., 2018] Liu, H.-T. D., Tao, M., Li, C.-L., Nowrouzezahrai, D., and Jacobson, A. (2018). Beyond pixel norm-balls: Parametric adversaries using an analytically differentiable renderer. In *International Conference on Learning Representations*.

[Lorenz et al., 2021] Lorenz, T., Ruoss, A., Balunović, M., Singh, G., and Vechev, M. (2021). Robustness certification for point cloud models. *arXiv preprint arXiv:2103.16652*.

[Ma et al., 2018] Ma, X., Li, B., Wang, Y., Erfani, S. M., Wijewickrema, S., Schoenebeck, G., Song, D., Houle, M. E., and Bailey, J. (2018). Characterizing adversarial subspaces using local intrinsic dimensionality. In *International Conference on Learning Representations*.

[Madry et al., 2018] Madry, A., Makelov, A., Schmidt, L., Tsipras, D., and Vladu, A. (2018). Towards deep learning models resistant to adversarial attacks. In *International Conference on Learning Representations*.

[Mintun et al., 2021] Mintun, E., Kirillov, A., and Xie, S. (2021). On interaction between augmentations and corruptions in natural corruption robustness. *Advances in Neural Information Processing Systems*, 34:3571–3583.

[Mohapatra et al., 2020] Mohapatra, J., Weng, T.-W., Chen, P.-Y., Liu, S., and Daniel, L. (2020). Towards verifying robustness of neural networks against a family of semantic perturbations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 244–252.

[Müller et al., 2022] Müller, M. N., Makarchuk, G., Singh, G., Püschel, M., and Vechev, M. T. (2022).

Prima: general and precise neural network certification via scalable convex hull approximations. *Proc. ACM Program. Lang.*, 6(POPL):1–33.

[Pang et al., 2022] Pang, T., Lin, M., Yang, X., Zhu, J., and Yan, S. (2022). Robustness and accuracy could be reconcilable by (Proper) definition. In Chaudhuri, K., Jegelka, S., Song, L., Szepesvari, C., Niu, G., and Sabato, S., editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 17258–17277. PMLR.

[Pei et al., 2017] Pei, K., Cao, Y., Yang, J., and Jana, S. (2017). Towards practical verification of machine learning: The case of computer vision systems. *arXiv preprint arXiv:1712.01785*.

[Raghunathan et al., 2018a] Raghunathan, A., Steinhardt, J., and Liang, P. (2018a). Certified defenses against adversarial examples. *arXiv preprint arXiv:1801.09344*.

[Raghunathan et al., 2018b] Raghunathan, A., Steinhardt, J., and Liang, P. S. (2018b). Semidefinite relaxations for certifying robustness to adversarial examples. *Advances in Neural Information Processing Systems*, 31.

[Ruoss et al., 2021] Ruoss, A., Baader, M., Balunović, M., and Vechev, M. (2021). Efficient certification of spatial robustness. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 2504–2513.

[Salman et al., 2019a] Salman, H., Li, J., Razenshteyn, I., Zhang, P., Zhang, H., Bubeck, S., and Yang, G. (2019a). Provably robust deep learning via adversarially trained smoothed classifiers. *Advances in Neural Information Processing Systems*, 32.

[Salman et al., 2020] Salman, H., Sun, M., Yang, G., Kapoor, A., and Kolter, J. Z. (2020). Denoised smoothing: A provable defense for pretrained classifiers. *Advances in Neural Information Processing Systems*, 33:21945–21957.

[Salman et al., 2019b] Salman, H., Yang, G., Zhang, H., Hsieh, C.-J., and Zhang, P. (2019b). A convex relaxation barrier to tight robustness verification of neural networks. *Advances in Neural Information Processing Systems*, 32.

[Samangouei et al., 2018] Samangouei, P., Kabkab, M., and Chellappa, R. (2018). Defense-gan: Protecting classifiers against adversarial attacks using generative models. In *International Conference on Learning Representations*.

[Sayed and Brostow, 2021] Sayed, M. and Brostow, G. (2021). Improved handling of motion blur in online object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1706–1716.

[Shinya and Forgue, 1991] Shinya, M. and Forgue, M.-C. (1991). Interference detection through rasterization. *The Journal of Visualization and Computer Animation*, 2(4):132–134.

[Singh et al., 2019] Singh, G., Gehr, T., Püschel, M., and Vechev, M. (2019). An abstract domain for certifying neural networks. *Proceedings of the ACM on Programming Languages*, 3(POPL):41.

[Sun et al., 2022] Sun, T., Segu, M., Postels, J., Wang, Y., Van Gool, L., Schiele, B., Tombari, F., and Yu, F. (2022). Shift: A synthetic driving dataset for continuous multi-task domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21371–21382.

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

[Tjeng et al., 2018] Tjeng, V., Xiao, K. Y., and Tedrake, R. (2018). Evaluating robustness of neural networks with mixed integer programming. In *International Conference on Learning Representations*.

[Tramèr et al., 2018] Tramèr, F., Boneh, D., Kurakin, A., Goodfellow, I., Papernot, N., and McDaniel, P. (2018). Ensemble adversarial training: Attacks and defenses. In *6th International Conference on Learning Representations, ICLR 2018-Conference Track Proceedings*.

[Tramer et al., 2020] Tramer, F., Carlini, N., Brendel, W., and Madry, A. (2020). On adaptive attacks to adversarial example defenses. *Advances in Neural Information Processing Systems*, 33:1633–1645.

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

[Xiao et al., 2018] Xiao, C., Li, B., Zhu, J. Y., He, W., Liu, M., and Song, D. (2018). Generating adversarial examples with adversarial networks. In *27th International Joint Conference on Artificial Intelligence, IJCAI 2018*, pages 3905–3911. International Joint Conferences on Artificial Intelligence.

[Xu et al., 2022] Xu, C., Ding, W., Lyu, W., Liu, Z., Wang, S., He, Y., Hu, H., Zhao, D., and Li, B. (2022). Safebench: A benchmarking platform for safety evaluation of autonomous vehicles. *Advances in Neural Information Processing Systems*, 35:25667–25682.

[Yang et al., 2022] Yang, R., Laurel, J., Misailovic, S., and Singh, G. (2022). Provable defense against geometric transformations. *arXiv preprint arXiv:2207.11177*.

[Zhang et al., 2022a] Zhang, H., Wang, S., Xu, K., Li, L., Li, B., Jana, S., Hsieh, C.-J., and Kolter, J. Z. (2022a). General cutting planes for bound-propagation-based neural network verification. In *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*.

[Zhang et al., 2018] Zhang, H., Weng, T.-W., Chen, P.-Y., Hsieh, C.-J., and Daniel, L. (2018). Efficient neural network robustness certification with general activation functions. In *Advances in neural information processing systems*, pages 4939–4948.

[Zhang et al., 2022b] Zhang, J., Li, L., Zhang, C., and Li, B. (2022b). Care: Certifiably robust learning with reasoning via variational inference. *arXiv preprint arXiv:2209.05055*.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes. See Sec. 3 and 4.]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Not Applicable]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes. The code is at `https://github.com/HanjiangHu/pixel-wise-smoothing`.]

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   (b) Complete proofs of all theoretical results. [Yes. The detailed proofs are in the appendix.]
   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes, they are in the appendix and `https://github.com/HanjiangHu/pixel-wise-smoothing`.]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes, details are in the PDF of the appendix.]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Not Applicable]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes. It is in the appendix.]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Yes]
   (b) The license information of the assets, if applicable. [Yes]
   (c) New assets either in the supplemental material or as a URL, if applicable. [Yes]
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

A Preliminary Definitions and Theorems

A.1 Preliminary Definitions

Definition A.1 (3D-2D position projection, **restated** of Definition 3.1 and Definition 1 from [Hu et al., 2022b])**.** For any 3D point $P = (X, Y, Z) \in \mathbb{P} \subset \mathbb{R}^3$ under the camera coordinate with the camera intrinsic matrix $\mathbf{K}$, based on the camera motion of $\alpha = (\theta, t) \in \mathcal{Z} \subset \mathbb{R}^6$ with rotation matrix $\mathbf{R} = \exp(\theta^\wedge) \in SO(3)$ and translation vector $t \in \mathbb{R}^3$, define the projection function $\rho : \mathbb{P} \times \mathcal{Z} \to \mathbb{R}^2$ and the depth function $D : \mathbb{P} \times \mathcal{Z} \to \mathbb{R}$ as

$$[\rho(P, \alpha), 1]^\top = \frac{1}{D(P, \alpha)} \mathbf{K}\mathbf{R}^{-1}(P - t), \quad D(P, \alpha) = [0, 0, 1]\mathbf{R}^{-1}(P - t) \tag{22}$$

Definition A.2 (3D-2D K-channel pixel-wise projection, **restated** of Definition 3.2 and Definition 2 from [Hu et al., 2022b])**.** Given the position projection function $\rho : \mathbb{P} \times \mathcal{Z} \to \mathbb{R}^2$ and the depth function $D : \mathbb{P} \times \mathcal{Z} \to \mathbb{R}$ with K-channel 3D point cloud $V \in \mathcal{V} : \mathbb{P} \times [0, 1]^K$, for camera motion $\alpha \in \mathcal{Z}$, define the 3D-2D pixel projection function as $O : \mathcal{V} \times \mathcal{Z} \to \mathcal{X}$ to return K-channel image $x \in \mathcal{X}$, $x = O(V, \alpha)$ where

$$x_{k,r,s} = O(V, \alpha)_{k,r,s} = V_{P^*_\alpha, k}, \quad \text{where } P^*_\alpha = \underset{\{P \in \mathbb{P} | \lfloor \rho(P,\alpha) \rfloor = (r,s)\}}{\operatorname{argmin}} D(P, \alpha) \tag{23}$$

where $\lfloor \cdot \rfloor$ is the floor function. Specifically, if $x = O(V, 0)$, given the relative camera pose $\alpha \in \mathcal{Z}$, define the 2D projective transformation $\phi : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ as $\phi(x, \alpha) = O(V, \alpha)$.

Definition A.3 (ε-smoothed classifier with 2D image projection, **restated** of Definition 4.1)**.** Let $\phi : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ be a 2D projective transformation parameterized with the one-axis relative camera motion $\alpha \in \mathcal{S} \subset \mathcal{Z}$ and $\varepsilon \in \mathcal{X} \sim \mathcal{P}_\varepsilon$ as the smoothing distribution. Let $x = \phi(x, 0)$ be under the original camera pose and $h : \mathcal{X} \to \mathcal{Y}$ be a base classifier $h(x) = \operatorname{argmax}_{y \in \mathcal{Y}} p(y \mid x)$. Under the one-axis relative camera pose $\alpha \in \mathcal{S}$, we define the ε-smoothed classifier $g : \mathcal{X} \to \mathcal{Y}$ as

$$g(x; \alpha; \varepsilon) = \underset{y \in \mathcal{Y}}{\operatorname{argmax}}\, q(y \mid \phi(x, \alpha); \varepsilon), \text{where } q(y \mid \phi(x, \alpha); \varepsilon) = \mathbb{E}_{\varepsilon \sim \mathcal{P}_\varepsilon} p(y \mid \phi(x, \alpha) + \varepsilon) \tag{24}$$

Remark A.4. We remark that the definition of the smoothed classifier A.3 is also applicable to the general 6-DoF translation and rotation of camera motion ($\alpha \in \mathcal{Z}$), but we mainly focus on the one-axis relative camera motion ($\alpha \in \mathcal{S} \subset \mathcal{Z}$) to make it consistent with the previous work [Hu et al., 2022b].

A.2 Main Theorems for Transformation Specific Smoothing

Definition A.5. (Differentially resolvable 3D-2D projective transformation) Let $\phi : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ be a 2D projective transformation based on a 3D oracle $O : \mathcal{V} \times \mathcal{Z} \to \mathcal{X}$ with noise space $\mathcal{Z}_\phi$ and let $\psi : \mathcal{X} \times \mathcal{Z}_\psi \to \mathcal{X}$ be a resolvable 2D transformation with noise space $\mathcal{Z}_\psi$. We say that ϕ can be differentially resolved by ψ if for any $V \in \mathcal{V}$ there exists function $\delta_V : \mathcal{Z}_\phi \times \mathcal{Z}_\phi \to \mathcal{Z}_\psi$ such that for any $\alpha \in \mathcal{Z}_\phi$ and any $\beta \in \mathcal{Z}_\phi$,

$$\phi(O(V, 0), \alpha) = \psi(\phi(O(V, 0), \beta), \delta_V(\alpha, \beta)). \tag{25}$$

Specifically, the projective transformation from a discrete oracle can be resolved by the additive transformation by

$$\psi(x, \delta) = x + \delta, \delta_V(\alpha, \beta) = \phi(O(V, 0), \alpha) - \phi(O(V, 0), \beta)$$

Following the category of resolvable and non-resolvable or differentially resolvable transformation in literature [Li et al., 2021, Chu et al., 2022, Hao et al., 2022], we define the differentially resolvable property of 3D-2D projective transformation in Definition A.5. Therefore, the following theorems are applicable as the foundation of other theorems in this paper.

Theorem A.6 (Theorem 2 in [Li et al., 2021])**.** *Let $\phi : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ be a transformation that is resolved by $\psi : \mathcal{X} \times \mathcal{Z}_\psi \to \mathcal{X}$. Let $\varepsilon \sim \mathcal{P}_\varepsilon$ be a $\mathcal{Z}_\psi$-valued random variable and suppose that the smoothed classifier $g : \mathcal{X} \to \mathcal{Y}$ given by $q(y \mid x; \varepsilon) = \mathbb{E}(p(y \mid \psi(x, \varepsilon)))$ predicts $g(x; \varepsilon) = y_A = \operatorname{argmax}_y q(y \mid x; \varepsilon)$. Let $\mathcal{S} \subseteq \mathcal{Z}_\phi$ and $\{\alpha_i\}_{i=1}^N \subseteq \mathcal{Z}$ be a set of transformation parameters such that for any i, the class probabilities satisfy*

$$q(y_A \mid \phi(x, \alpha_i); \varepsilon) \geq p_A^{(i)} \geq p_B^{(i)} \geq \max_{y \neq y_A} q(y \mid \phi(x, \alpha_i); \varepsilon). \tag{26}$$

Then there exists a set $\Delta^* \subseteq \mathcal{Z}_\psi$ with the property that, if for any $\alpha \in \mathcal{S}, \exists \alpha_i$ with $\delta_x(\alpha, \alpha_i) \in \Delta^*$, then it is guaranteed that

$$q(y_A \mid \phi(x, \alpha); \varepsilon) > \max_{y \neq y_A} q(y \mid \phi(x, \alpha); \varepsilon). \tag{27}$$

The rigorous proof of Theorem A.6 can be found in [Li et al., 2021].

Theorem A.7 (Theorem 7 in [Chu et al., 2022], Corollary 2 in [Li et al., 2021])**.** *Let $\psi(x, \delta) = x + \delta$ and let $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_d)$. Furthermore, let ϕ be a transformation with parameters in $\mathcal{Z}_\phi \subseteq \mathbb{R}^m$ and let $\mathcal{S} \subseteq \mathcal{Z}_\phi$ and $\{\alpha_i\}_{i=1}^N \subseteq \mathcal{S}$. Let $y_A \in \mathcal{Y}$ and suppose that for any i, the ε-smoothed classifier defined by $q(y \mid x; \varepsilon) := \mathbb{E}(p(y \mid x+\varepsilon))$ has class probabilities that satisfy*

$$q(y_A \mid \phi(x, \alpha_i); \varepsilon) \geq p_A^{(i)} \geq p_B^{(i)} \geq \max_{y \neq y_A} q(y \mid \phi(x, \alpha_i); \varepsilon). \tag{28}$$

Then it is guaranteed that $\forall \alpha \in \mathcal{S}$: $y_A = \operatorname{argmax}_y q(y \mid \phi(x, \alpha); \varepsilon)$ if the maximum projective error

$$M_{\mathcal{S}} := \max_{\alpha \in \mathcal{S}} \min_{1 \leq i \leq N} \|\phi(x, \alpha) - \phi(x, \alpha_i)\|_2 \tag{29}$$

$$\textit{satisfies} \quad M_{\mathcal{S}} < R := \frac{\sigma}{2} \min_{1 \leq i \leq N} \left(\Phi^{-1}\left(p_A^{(i)}\right) - \Phi^{-1}\left(p_B^{(i)}\right) \right). \tag{30}$$

The Theorem A.7 is directly obtained if the projective transformation can be resolved by additive transformation, as shown in Definition A.5. We direct readers to [Li et al., 2021] for rigorous proof of Theorem A.7.

B Proofs in Certification via Pixel-wise Smoothing with Prior of Entire Point Cloud (PwS)

Here we present the proof of Lemma 4.4 and Theorem 4.5, which gives the certification using the fully-covered interval of uniform partitions and serves as the foundation of the following theorems.

Definition B.1 (Consistent camera motion interval, **restated** of Definition 4.3)**.** Given the 3D points $\mathbb{P} \subset \mathbb{R}^3$ and the projection function $\rho : \mathbb{P} \times \mathcal{Z} \to \mathbb{R}^2$ and the depth function $D : \mathbb{P} \times \mathcal{Z} \to \mathbb{R}$, for any $P^* \in \mathbb{P}$ projected on (r, s) with the least depth values, define the camera motion set $\mathbb{U}_{P^*,r,s} \subset \mathcal{Z}$ as the consistent camera motion interval,

$$\mathbb{U}_{P^*,r,s} = \{\alpha \mid P^* = \operatorname*{argmin}_{\{P \in \mathbb{P} \mid \lfloor \rho(P,\alpha) \rfloor = (r,s)\}} D(P, \alpha)\} \tag{31}$$

Specifically, for one-axis consistent camera motion interval $\mathbb{U}_{P,r,s} \subset \mathcal{S}$, in the absence of ambiguity, we regard $\mathbb{U}_{P,r,s}$ as a subset of one-dimensional real number field based on non-zero coordinate in $\mathcal{S}$ in the following mathematical notations.

B.1 Proof on Lemma 4.4: Upper bound of fully-covered motion interval

Lemma B.2 (Upper bound of fully-covered motion interval, **restated** of Lemma 4.4)**.** *Given the projection from entire 3D point $V \in \mathcal{V} : \mathbb{P} \times [0, 1]^K$ along one-axis translation or rotation and the consistent camera motion interval $\mathbb{U}_{P,r,s}$ for any $P \in \mathbb{P}$ projected on (r, s), define the interval $\Delta^{r,s}$ as,*

$$\Delta^{r,s} = \min_{P \in \mathbb{P}} \{\sup \mathbb{U}_{P,r,s} - \inf \mathbb{U}_{P,r,s}\} \tag{32}$$

then for any projection on (r, s) under camera motion $u \in \bigcup_{P \in \mathbb{P}} \mathbb{U}_{P,r,s}$, we have $\forall\ 0 \leq \Delta^* \leq \Delta^{r,s}$

$$O(V, u + \Delta^*)_{r,s} \in \{O(V, u)_{r,s}, O(V, u + \Delta^{r,s})_{r,s}\} \tag{33}$$

Proof. Considering the projection function ρ on r, s with any 3D point $P : \lfloor \rho(P, \alpha) \rfloor = (r, s)$ at camera motion $\alpha \in \mathcal{S}$ with the least depth, $\mathbb{U}_{P,r,s} \neq \emptyset$. With

$$\Delta^{r,s} = \min_{P \in \mathbb{P}} \{\sup \mathbb{U}_{P,r,s} - \inf \mathbb{U}_{P,r,s}\} = \min_{P \in \mathbb{P} \mid \lfloor \rho(P,\alpha) \rfloor = (r,s), \alpha \in \mathbb{U}_{P,r,s}} \sup \mathbb{U}_{P,r,s} - \inf \mathbb{U}_{P,r,s} \tag{34}$$

For the 3D projective oracle O on pixel (r,s), based on the definition of $O(V,u), O(V,u+\Delta^{r,s})$, we have $O(V,u)_{r,s} = V_{P_u}, O(V,u+\Delta^{r,s})_{r,s} = V_{P_{u+\Delta^{r,s}}}$, where

$$P_u = \underset{\{P\in\mathbb{P}|\lfloor\rho(P,u)\rfloor=(r,s)\}}{\operatorname{argmin}} D(P,u), P_{u+\Delta^{r,s}} = \underset{\{P\in\mathbb{P}|\lfloor\rho(P,u+\Delta^{r,s})\rfloor=(r,s)\}}{\operatorname{argmin}} D(P,u+\Delta^{r,s})$$

Suppose there exists $\Delta^* \in [0,\Delta^{r,s}]$ such that

$$O(V,u+\Delta^*)_{r,s} \neq O(V,u)_{r,s}, O(V,u+\Delta^{r,s})_{r,s} \neq O(V,u+\Delta^{r,s})_{r,s}$$

i.e., there exists $O(V,u+\Delta^*)_{r,s} = V_{P_{u+\Delta^*}}$ such that $P_u \neq P_{u+\Delta^*}, P_{u+\Delta^{r,s}} \neq P_{u+\Delta^*}$. In this case, according to the definition of $\mathbb{U}_{P_{u+\Delta^*},r,s}$, it holds that

$$\sup \mathbb{U}_{P_{u+\Delta^*},r,s} - \inf \mathbb{U}_{P_{u+\Delta^*},r,s} < \Delta^{r,s}$$

which contradicts with (34). Therefore, such $O(V,u+\Delta^*)_{r,s}$ does not exist and for any $0 \leq \Delta^* \leq \Delta^{r,s}$,

$$O(V,u+\Delta^*)_{r,s} \in \{O(V,u)_{r,s}, O(V,u+\Delta^{r,s})_{r,s}\}$$

which concludes the proof. □

B.2 Proof on Theorem 4.5: Certification with fully-covered partitions

Theorem B.3 (Certification with fully-covered partitions, **restated** of Theorem 4.5)**.** *For the image projection from entire 3D point cloud $V \in \mathcal{V}: \mathbb{P}\times[0,1]^K$, let ϕ be one-axis rotation or translation with parameters in $\mathcal{S} \subseteq \mathcal{Z}$ and uniformly partitioned $\{\alpha_i\}_{i=1}^N \subseteq \mathcal{S}$ with interval Δ_α, where*

$$\Delta_\alpha \leq \min_{r,s} \Delta^{r,s} = \min_{r,s} \min_{P\in\mathbb{P}} \{\sup \mathbb{U}_{P,r,s} - \inf \mathbb{U}_{P,r,s}\}$$

Let $y_A, y_B \in \mathcal{Y}$, $\varepsilon \sim \mathcal{N}(0,\sigma^2\mathbb{I}_d)$ and suppose that for any i, the ε-smoothed classifier defined by $q(y \mid x;\varepsilon) := \mathbb{E}(p(y \mid x+\varepsilon))$ has class probabilities that satisfy with top-2 classes y_A, y_B as $q(y_A \mid \phi(x,\alpha_i);\varepsilon) \geq p_A^{(i)} \geq p_B^{(i)} \geq \max_{y\neq y_A} q(y \mid \phi(x,\alpha_i);\varepsilon)$,

then it is guaranteed that $\forall \alpha \in \mathcal{S}$: $y_A = \operatorname{argmax}_y q(y \mid \phi(x,\alpha);\varepsilon)$ if

$$\max_{1\leq i\leq N-1} \sqrt{\frac{1}{2}\sum_{k,r,s}(O(V,\alpha_i)_{krs} - O(V,\alpha_{i+1})_{krs})^2} < \frac{\sigma}{2} \min_{1\leq i\leq N} \left(\Phi^{-1}\left(p_A^{(i)}\right) - \Phi^{-1}\left(p_B^{(i)}\right)\right). \tag{35}$$

Proof. According to Theorem A.7, we need to upper bound

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

with uniform partitions under interval of $\Delta_\alpha = \alpha_{i+1} - \alpha_i$.

$$\begin{aligned}
&\max_{\alpha\in\mathcal{S}} \min_{1\leq i\leq N} \|\phi(x,\alpha) - \phi(x,\alpha_i)\|_2 = \max_{\alpha\in\mathcal{S}} \min_{1\leq i\leq N} \|O(V,\alpha) - O(V,\alpha_i)\|_2 \\
\leq &\max_{1\leq i\leq N-1} \sqrt{\max_{\alpha_i\leq\alpha\leq\alpha_{i+1}} \min\{(O(V,\alpha) - O(V,\alpha_i))^2, (O(V,\alpha) - O(V,\alpha_{i+1}))^2\}} \\
= &\max_{1\leq i\leq N-1} \sqrt{\max_{\alpha_i\leq\alpha\leq\alpha_{i+1}} \min\{\sum_{k,r,s}(O(V,\alpha)_{krs} - O(V,\alpha_i)_{krs})^2, \sum_{k,r,s}(O(V,\alpha)_{krs} - O(V,\alpha_{i+1})_{krs})^2\}} \\
\leq &\max_{1\leq i\leq N-1} \sqrt{\max_{\alpha_i\leq\alpha\leq\alpha_{i+1}} \frac{1}{2}\{\sum_{k,r,s}(O(V,\alpha)_{krs} - O(V,\alpha_i)_{krs})^2 + \sum_{k,r,s}(O(V,\alpha)_{krs} - O(V,\alpha_{i+1})_{krs})^2\}} \\
= &\max_{1\leq i\leq N-1} \sqrt{\max_{\alpha_i\leq\alpha\leq\alpha_{i+1}} \frac{1}{2}\sum_{k,r,s}[(O(V,\alpha)_{krs} - O(V,\alpha_i)_{krs})^2 + (O(V,\alpha)_{krs} - O(V,\alpha_{i+1})_{krs})^2]}
\end{aligned}$$

For any $\alpha_i \le \alpha \le \alpha_{i+1}$ on the pixel (r,s),

$$\begin{aligned}\Delta_\alpha &\le \min_{r,s}\min_{P\in\mathbb{P}}\{\sup \mathbb{U}_{P,r,s} - \inf \mathbb{U}_{P,r,s}\}\\ &\le \min_{r,s}\min_{P\in\mathbb{P}\mid\lfloor\rho(P,\alpha)\rfloor=(r,s),\alpha\in\mathbb{U}_{P,r,s}} \sup \mathbb{U}_{P,r,s} - \inf \mathbb{U}_{P,r,s}\\ &\le \min_{P\in\mathbb{P}\mid\lfloor\rho(P,\alpha)\rfloor=(r,s),\alpha\in\mathbb{U}_{P,r,s}} \sup \mathbb{U}_{P,r,s} - \inf \mathbb{U}_{P,r,s}\end{aligned}$$

Based on Lemma B.2, we have

$$O(V,\alpha)_{r,s} \in \{O(V,\alpha_i)_{r,s}, O(V,\alpha_{i+1})_{r,s}\}$$

i.e., either $O(V,\alpha)_{r,s} = O(V,\alpha_i)_{r,s}$ or $O(V,\alpha)_{r,s} = O(V,\alpha_{i+1})_{r,s}$ holds. Therefore, for any k,r,s, we have

$$(O(V,\alpha)_{krs} - O(V,\alpha_i)_{krs})^2 + (O(V,\alpha)_{krs} - O(V,\alpha_{i+1})_{krs})^2 = (O(V,\alpha_i)_{krs} - O(V,\alpha_{i+1})_{krs})^2$$

Therefore,

$$\begin{aligned}&\max_{\alpha\in\mathcal{S}}\min_{1\le i\le N}\|\phi(x,\alpha)-\phi(x,\alpha_i)\|_2\\ \le&\max_{1\le i\le N-1}\sqrt{\max_{\alpha_i\le\alpha\le\alpha_{i+1}}\frac{1}{2}\sum_{k,r,s}[(O(V,\alpha)_{krs} - O(V,\alpha_i)_{krs})^2 + (O(V,\alpha)_{krs} - O(V,\alpha_{i+1})_{krs})^2]}\\ =&\max_{1\le i\le N-1}\sqrt{\frac{1}{2}\sum_{k,r,s}(O(V,\alpha_i)_{krs} - O(V,\alpha_{i+1})_{krs})^2}\end{aligned}$$

So if

$$\max_{1\le i\le N-1}\sqrt{\frac{1}{2}\sum_{k,r,s}(O(V,\alpha_i)_{krs} - O(V,\alpha_{i+1})_{krs})^2} < \frac{\sigma}{2}\min_{1\le i\le N}\left(\Phi^{-1}\left(p_A^{(i)}\right) - \Phi^{-1}\left(p_B^{(i)}\right)\right). \tag{36}$$

then it holds that

$$M_{\mathcal{S}} := \max_{\alpha\in\mathcal{S}}\min_{1\le i\le N}\|\phi(x,\alpha)-\phi(x,\alpha_i)\|_2 < R := \frac{\sigma}{2}\min_{1\le i\le N}\left(\Phi^{-1}\left(p_A^{(i)}\right) - \Phi^{-1}\left(p_B^{(i)}\right)\right). \tag{37}$$

which concludes the proof based on Theorem A.7. □

C Proofs in Certification via Pixel-wise Smoothing with Lipschitz-approximated Partitions (PwS-L)

In this section, we show the proof of Lemma 4.6 and Theorem 4.7, where the upper bound of the partitioning interval can be approximated through the Lipschitz property of the projection oracle.

C.1 Proof on Lemma 4.6: Approximated upper bound of fully-covered interval

Lemma C.1 (Approximated upper bound of fully-covered interval, **restated** from Lemma 4.6)**.** *For the projection from entire 3D point $V \in \mathcal{V} : \mathbb{P}\times[0,1]^K$, given the one-axis monotonic position projection function ρ in l_∞ norm over camera motion space, if the Lipschitz-based interval $\Delta_L^{r,s}$ is defined as*

$$\Delta_L^{r,s} = \min_{P\in\mathbb{P}}\frac{1}{L_P} \mid \rho(P,\sup\mathbb{U}_{P,r,s}) - \rho(P,\inf\mathbb{U}_{P,r,s}) \mid_\infty \tag{38}$$

where L_P is the Lipschitz constant for projection function ρ given 3D point P,

$$L_P = \max\{\max_{\alpha\in\mathcal{S}}|\frac{\partial\rho_1(P,\alpha)}{\partial\alpha}|, \max_{\alpha\in\mathcal{S}}|\frac{\partial\rho_2(P,\alpha)}{\partial\alpha}|\} \tag{39}$$

then it holds that $\Delta_L^{r,s} \le \Delta^{r,s} = \min_{P\in\mathbb{P}}\{\sup\mathbb{U}_{P,r,s} - \inf\mathbb{U}_{P,r,s}\}$.

Proof. Considering the projection function ρ on r, s with any 3D point $P : \lfloor \rho(P, \alpha) \rfloor = (r, s)$ at camera motion $\alpha \in \mathcal{S}$ with the least depth, $\mathbb{U}_{P,r,s} \neq \emptyset$ then the interval $\Delta_L^{r,s}$ satisfies

$$\begin{aligned}\Delta_L^{r,s} &= \min_{P\in\mathbb{P}} \frac{1}{L_P} \mid \rho(P, \sup \mathbb{U}_{P,r,s}) - \rho(P, \inf \mathbb{U}_{P,r,s}) \mid_\infty \\ &\leq \min_{\{P\in\mathbb{P}|\lfloor\rho(P,\alpha)\rfloor=(r,s),\alpha\in\mathcal{S}\}} \frac{1}{L_P} \mid \rho(P, \sup \mathbb{U}_{P,r,s}) - \rho(P, \inf \mathbb{U}_{P,r,s}) \mid_\infty \\ &\leq \frac{1}{L_P} \mid \rho(P, \sup \mathbb{U}_{P,r,s}) - \rho(P, \inf \mathbb{U}_{P,r,s}) \mid_\infty\end{aligned}$$

For the Lipschitz constant L_P for projection function ρ with point P, we have

$$\begin{aligned}L_P &= \max\{\max_{\alpha\in\mathcal{S}} |\frac{\partial\rho_1(P,\alpha)}{\partial\alpha}|, \max_{\alpha\in\mathcal{S}} |\frac{\partial\rho_2(P,\alpha)}{\partial\alpha}|\} \\ &= \max\{\max_{\alpha,\beta\in\mathcal{S}} \frac{|\rho_1(P,\alpha)-\rho_1(P,\beta)|}{|\alpha-\beta|}, \max_{\alpha,\beta\in\mathcal{S}} \frac{|\rho_2(P,\alpha)-\rho_2(P,\beta)|}{|\alpha-\beta|}\} \\ &= \max_{\alpha,\beta\in\mathcal{S}} \max\{\frac{|\rho_1(P,\alpha)-\rho_1(P,\beta)|}{|\alpha-\beta|}, \frac{|\rho_2(P,\alpha)-\rho_2(P,\beta)|}{|\alpha-\beta|}\} \\ &= \max_{\alpha,\beta\in\mathcal{S}} \frac{|\rho(P,\alpha)-\rho(P,\beta|_\infty}{|\alpha-\beta|}\end{aligned} \tag{40}$$

Therefore, we have

$$\begin{aligned}&\max_{\alpha\in\mathcal{S}} \frac{|\rho(P,\alpha+\Delta_L^{r,s})-\rho(P,\alpha)|_\infty}{\Delta_L^{r,s}} \\ \leq &\max_{\alpha,\beta\in\mathcal{S}} \frac{|\rho(P,\alpha)-\rho(P,\beta)|_\infty}{|\alpha-\beta|} = L_P \\ \leq &\frac{1}{\Delta_L^{r,s}} \mid \rho(P, \sup \mathbb{U}_{P,r,s}) - \rho(P, \inf \mathbb{U}_{P,r,s}) \mid_\infty\end{aligned}$$

Therefore, for any $u \in \mathcal{S}$

$$\mid \rho(P, u+\Delta_L^{r,s}) - \rho(P,u) \mid_\infty \leq \max_{\alpha\in\mathcal{S}} |\rho(P,\alpha+\Delta_L^{r,s})-\rho(P,\alpha)|_\infty \leq \mid \rho(P, \sup \mathbb{U}_{P,r,s}) - \rho(P, \inf \mathbb{U}_{P,r,s}) \mid_\infty$$

Based on the monotonicity of ρ in l_∞ norm over $\mathbb{U}_{P,r,s}$ given P, let $u = \inf \mathbb{U}_{P,r,s}$, we have $\Delta_L^{r,s} \leq \sup \mathbb{U}_{P,r,s} - \inf \mathbb{U}_{P,r,s}$ for any P and (r, s). Therefore, for pixel (r, s),

$$\Delta_L^{r,s} \leq \min_{P\in\mathbb{P}|\lfloor\rho(P,\alpha)\rfloor=(r,s),\alpha\in\mathbb{U}_{P,r,s}} \sup \mathbb{U}_{P,r,s} - \inf \mathbb{U}_{P,r,s} = \min_{P\in\mathbb{P}} \{\sup \mathbb{U}_{P,r,s} - \inf \mathbb{U}_{P,r,s}\}$$

which concludes the proof. □

C.2 Proof on Theorem 4.7: Certification with approximated partitions

Theorem C.2 (Certification with approximated partitions, **restated** of Theorem 4.7)**.** *For the projection from entire 3D point cloud $V \in \mathcal{V} : \mathbb{P} \times [0, 1]^K$ through the monotonic position projection function ρ in l_∞ norm over camera motion space, let ϕ be one-axis rotation or translation with parameters in $\mathcal{S} \subseteq \mathcal{Z}$ and uniformly partitioned $\{\alpha_i\}_{i=1}^N \subseteq \mathcal{S}$ with interval Δ_α, where*

$$\Delta_\alpha \leq \min_{r,s} \Delta_L^{r,s} = \min_{r,s} \min_{P\in\mathbb{P}} \frac{1}{L_P} \mid \rho(P, \sup \mathbb{U}_{P,r,s}) - \rho(P, \inf \mathbb{U}_{P,r,s}) \mid_\infty \tag{41}$$

$$L_P = \max\{\max_{\alpha\in\mathcal{S}} |\frac{\partial\rho_1(P,\alpha)}{\partial\alpha}|, \max_{\alpha\in\mathcal{S}} |\frac{\partial\rho_2(P,\alpha)}{\partial\alpha}|\} \tag{42}$$

Let $y_A, y_B \in \mathcal{Y}$, $\varepsilon \sim \mathcal{N}(0, \sigma^2\mathbb{I}_d)$ and suppose that for any i, the ε-smoothed classifier defined by $q(y \mid x; \varepsilon) := \mathbb{E}(p(y \mid x + \varepsilon))$ has class probabilities that satisfy with top-2 classes y_A, y_B as $q(y_A \mid \phi(x, \alpha_i); \varepsilon) \geq p_A^{(i)} \geq p_B^{(i)} \geq \max_{y\neq y_A} q(y \mid \phi(x, \alpha_i); \varepsilon)$,

then it is guaranteed that $\forall \alpha \in \mathcal{S}$: $y_A = \operatorname{argmax}_y q(y \mid \phi(x, \alpha); \varepsilon)$ *if*

$$\max_{1\le i\le N-1} \sqrt{\frac{1}{2}\sum_{k,r,s}(O(V,\alpha_i)_{krs} - O(V,\alpha_{i+1})_{krs})^2} < \frac{\sigma}{2} \min_{1\le i\le N} \left(\Phi^{-1}\left(p_A^{(i)}\right) - \Phi^{-1}\left(p_B^{(i)}\right)\right).$$

Proof. According to Lemma C.1, at each pixel (r, s) we have

$$\Delta_L^{r,s} \le \min_{P\in\mathbb{P}}\{\sup \mathbb{U}_{P,r,s} - \inf \mathbb{U}_{P,r,s}\}$$

Therefore,

$$\Delta_\alpha \le \min_{r,s} \Delta_L^{r,s} \le \min_{r,s}\min_{P\in\mathbb{P}}\{\sup \mathbb{U}_{P,r,s} - \inf \mathbb{U}_{P,r,s}\}$$

Now applying Theorem B.3, the theorem is directly proved as a corollary. □

D Proofs in Certification via Pixel-wise Smoothing with One-Frame Point Cloud as Prior (PwS-OF)

We finally give the proof of Lemma 4.9 and Theorem 4.10, where the certification using Lipschitz-based partitioning can be extended to the more general case with only one-frame point cloud known.

Definition D.1 (3D projection from one-frame point cloud with δ-convexity, **restated** of Definition 4.8)**.** Given a one-frame point cloud $\mathbb{P}_1 \subset \mathbb{R}^3$, define the K-channel image $x \in \mathcal{X} : [0,1]^K \times \mathbb{Z}^2$ as

$$x_{k,r,s} = O(V,0)_{k,r,s} = V_{P_1,k}, \text{ where } P_1 \in \mathbb{P}_1 \text{ s.t. } \lfloor \rho(P_1,\alpha) \rfloor = (r,s) \tag{43}$$

Define δ-convexity that with the minimal $\delta > 0$, for any new point $P \notin \mathbb{P}_1$ and any $\alpha \in \mathcal{S}$, there exists $P' \in \mathbb{P}_1$ where $\mid \rho(P,\alpha) - \rho(P',\alpha) \mid_\infty \le \delta$, it holds that $D(P,\alpha) \ge D(P',\alpha)$.

D.1 Lemma D.2 and Proof: Lipschitz relaxation for 1-axis translation or rotation

Lemma D.2 (Lipschitz relaxation for 1-axis translation or rotation)**.** *Given δ-convexity projection from one-frame 3D point cloud $V_1 \in \mathcal{V}_1 : \mathbb{P}_1 \times [0,1]^K$ and unknown complete point cloud $V \in \mathcal{V} : \mathbb{P} \times [0,1]^K$, if the projection function ρ is the 1-axis translation or rotation, for any $P \in \mathbb{P}\backslash\mathbb{P}_1$ there exists a finite constant $C_\delta \ge 0$ such that for any 1-axis camera motion $\alpha \in \mathcal{S} = [-b, b]$,*

$$L_P \le \max_{P'\in\mathbb{P}_1} L_{P'} + C_\delta$$

Specifically, for z-axis translation T_z over $\mathcal{S} = [-b, b]$,

$$C_\delta^{T_z} = \max_{(X',Y',Z')\in\mathbb{P}_1} \frac{\delta}{Z' - b}$$

for x-axis translation T_x over $\mathcal{S} = [-b, b]$,

$$C_\delta^{T_x} = 0$$

for y-axis translation T_y over $\mathcal{S} = [-b, b]$,

$$C_\delta^{T_y} = 0$$

for z-axis rotation R_z over $\mathcal{S} = [-b, b]$,

$$C_\delta^{R_z} = \max\{\frac{f_x}{f_y}, \frac{f_y}{f_x}\}\delta$$

for x-axis rotation R_x over $\mathcal{S} = [-b, b]$,

$$C_\delta^{R_x} = \frac{\delta^2}{f_y} + \max_{(X',Y',Z')\in\mathbb{P}_1} \max_{\theta=\pm b} \max\{\frac{\delta}{f_y}\frac{f_x|X'| + f_y|Y'\cos\theta + Z'\sin\theta|}{-Y'\sin\theta + Z'\cos\theta}, \frac{2\delta|Y'\cos\theta + Z'\sin\theta|}{-Y'\sin\theta + Z'\cos\theta}\}$$

for y-axis rotation R_y over $\mathcal{S} = [-b, b]$,

$$C_\delta^{R_y} = \frac{\delta^2}{f_y} + \max_{(X',Y',Z')\in\mathbb{P}_1} \max_{\theta=\pm b} \max\{\frac{2\delta|X'\cos\theta - Z'\sin\theta|}{X'\sin\theta + Z'\cos\theta}, \frac{\delta}{f_x}\frac{f_y|Y'| + f_x|X'\cos\theta - Z'\sin\theta|}{X'\sin\theta + Z'\cos\theta}\}$$

Proof. We prove this lemma through 6 cases of all 1-axis camera motion. 1) For the translation along z-axis, for any $P = (X, Y, Z) \in \mathbb{P} \backslash \mathbb{P}_1$ and any camera motion $\alpha = \{t_z\} \in [-b, b]$, we have

$$D_{T_z}(P, \alpha) = Z - t_z, \quad \rho_{T_z}(P, \alpha) = (\frac{f_x X + c_x(Z - t_z)}{Z - t_z}, \frac{f_y Y + c_y(Z - t_z)}{Z - t_z})$$

$$\begin{aligned} L_P &= \max\{\max_{\alpha \in \mathcal{S}} |\frac{\partial \rho_1(P, \alpha)}{\partial \alpha}|, \max_{\alpha \in \mathcal{S}} |\frac{\partial \rho_2(P, \alpha)}{\partial \alpha}|\} \\ &= \max\{\max_{t_z \in [-b,b]} \frac{f_x |X|}{(Z - t_z)^2}, \max_{t_z \in [-b,b]} \frac{f_y |Y|}{(Z - t_z)^2}\} \\ &= \max\{\frac{f_x |X|}{(Z - b)^2}, \frac{f_y |Y|}{(Z - b)^2}\} \quad (\text{by } a < b < Z) \end{aligned}$$

According to the definition of δ-convexity, there exists $P' = (X', Y', Z') \in \mathbb{P}_1$ where $\mid \rho(P, \alpha) - \rho(P', \alpha) \mid_\infty \leq \delta$, it holds that

$$D(P, \alpha) \geq D(P', \alpha)$$

So we have for any $t_z \in [-b, b]$

$$|\frac{f_x X}{Z - t_z} - \frac{f_x X'}{Z' - t_z}| \leq \delta, |\frac{f_y Y}{Z - t_z} - \frac{f_y Y'}{Z' - t_z}| \leq \delta, Z' - t_z < Z - t_z$$

$$\begin{aligned} L_P &= \max\{\frac{f_x |X|}{(Z - b)^2}, \frac{f_y |Y|}{(Z - b)^2}\} \\ &\leq \max\{\frac{1}{Z - b}(\frac{f_x |X'|}{Z' - b} + \delta), \frac{1}{Z - b}(\frac{f_y |Y'|}{Z' - b} + \delta)\} \\ &\leq \max\{\frac{1}{Z' - b}(\frac{f_x |X'|}{Z' - b} + \delta), \frac{1}{Z' - b}(\frac{f_y |Y'|}{Z' - b} + \delta)\} \\ &= \max\{\frac{f_x |X'|}{(Z' - b)^2}, \frac{f_y |Y'|}{(Z' - b)^2}\} + \frac{\delta}{Z' - b} \\ &\leq L_{P'} + \max_{(X', Y', Z') \in \mathbb{P}_1} \frac{\delta}{Z' - b} \\ &\leq \max_{P' \in \mathbb{P}_1} L_{P'} + C_\delta^{T_z} \end{aligned}$$

where

$$C_\delta^{T_z} = \max_{(X', Y', Z') \in \mathbb{P}_1} \frac{\delta}{Z' - b}$$

2) For the translation along x-axis, for any $P = (X, Y, Z) \in \mathbb{P} \backslash \mathbb{P}_1$ and any camera motion $\alpha = \{t_x\} \in [-b, b]$, we have

$$D_{T_x}(P, \alpha) = Z, \quad \rho_{T_x}(P, \alpha) = (\frac{f_x(X - t_x) + c_x Z}{Z}, \frac{f_y Y + c_y Z}{Z})$$

$$\begin{aligned} L_P &= \max\{\max_{\alpha \in \mathcal{S}} |\frac{\partial \rho_1(P, \alpha)}{\partial \alpha}|, \max_{\alpha \in \mathcal{S}} |\frac{\partial \rho_2(P, \alpha)}{\partial \alpha}|\} \\ &= \max\{\max_{t_x \in [-b,b]} \frac{f_x}{Z}, 0\} \\ &= \frac{f_x}{Z} \end{aligned}$$

According to the definition of δ-convexity, there exists $P' = (X', Y', Z') \in \mathbb{P}_1$ where $\mid \rho(P, \alpha) - \rho(P', \alpha) \mid_\infty \leq \delta$, it holds that

$$D(P, \alpha) \geq D(P', \alpha)$$

So we have $Z' < Z$

$$\begin{aligned} L_P &= \frac{f_x}{Z} \leq \frac{f_x}{Z'} = L_{P'} + 0 \\ &\leq \max_{P' \in \mathbb{P}_1} L_{P'} + C_\delta^{T_x} \end{aligned}$$

where

$$C_\delta^{T_x} = 0$$

3) For the translation along y-axis, for any $P = (X, Y, Z) \in \mathbb{P} \backslash \mathbb{P}_1$ and any camera motion $\alpha = \{t_y\} \in [-b, b]$, we have

$$D_{T_y}(P, \alpha) = Z, \quad \rho_{T_y}(P, \alpha) = (\frac{f_x X + c_x Z}{Z}, \frac{f_y(Y - t_y) + c_y Z}{Z})$$

$$\begin{aligned} L_P &= \max\{\max_{\alpha \in \mathcal{S}} |\frac{\partial \rho_1(P, \alpha)}{\partial \alpha}|, \max_{\alpha \in \mathcal{S}} |\frac{\partial \rho_2(P, \alpha)}{\partial \alpha}|\} \\ &= \max\{0, \max_{t_y \in [-b, b]} \frac{f_y}{Z}\} \\ &= \frac{f_y}{Z} \end{aligned}$$

According to the definition of δ-convexity, there exists $P' = (X', Y', Z') \in \mathbb{P}_1$ where $\mid \rho(P, \alpha) - \rho(P', \alpha) \mid_\infty \leq \delta$, it holds that

$$D(P, \alpha) \geq D(P', \alpha)$$

So we have $Z' < Z$

$$\begin{aligned} L_P &= \frac{f_y}{Z} \leq \frac{f_y}{Z'} = L_{P'} + 0 \\ &\leq \max_{P' \in \mathbb{P}_1} L_{P'} + C_\delta^{T_y} \end{aligned}$$

where

$$C_\delta^{T_y} = 0$$

4) For the rotation along z-axis, for any $P = (X, Y, Z) \in \mathbb{P} \backslash \mathbb{P}_1$ and any camera motion $\alpha = \{\theta\} \in [-b, b]$, we have

$$D_{R_z}(P, \alpha) = Z, \quad \rho_{R_z}(P, \alpha) = (\frac{f_x \cos\theta X + f_x \sin\theta Y}{Z} + c_x, \frac{f_y \cos\theta Y + f_y \sin\theta X}{Z} + c_y)$$

$$\begin{aligned} L_P &= \max\{\max_{\alpha \in \mathcal{S}} |\frac{\partial \rho_1(P, \alpha)}{\partial \alpha}|, \max_{\alpha \in \mathcal{S}} |\frac{\partial \rho_2(P, \alpha)}{\partial \alpha}|\} \\ &= \max\{\max_{\theta \in [-b, b]} \frac{|-X \sin\theta + Y \cos\theta|}{Z} f_x, \max_{\theta \in [-b, b]} \frac{|-Y \sin\theta + X \cos\theta|}{Z} f_y\} \\ &= \max\{\frac{|-X \sin\theta_1 + Y \cos\theta_1|}{Z} f_x, \frac{|-Y \sin\theta_2 + X \cos\theta_2|}{Z} f_y\} \end{aligned}$$

where

$$\theta_1 = \operatorname*{argmax}_{\theta \in [-b, b]} \frac{|-X \sin\theta + Y \cos\theta|}{Z} f_x, \theta_2 = \operatorname*{argmax}_{\theta \in [-b, b]} \frac{|-Y \sin\theta + X \cos\theta|}{Z} f_y$$

According to the definition of δ-convexity, there exists $P' = (X', Y', Z') \in \mathbb{P}_1$ where $\mid \rho(P, \alpha) - \rho(P', \alpha) \mid_\infty \leq \delta$, it holds that

$$D(P, \alpha) \geq D(P', \alpha)$$

So we have $Z' < Z$ and for any $\theta \in [-b, b]$

$$|\frac{f_x \cos\theta X + f_x \sin\theta Y}{Z} - \frac{f_x \cos\theta X' + f_x \sin\theta Y'}{Z'}| \leq \delta, |\frac{f_y \cos\theta Y + f_y \sin\theta X}{Z} - \frac{f_y \cos\theta Y' + f_y \sin\theta X'}{Z'}| \leq \delta$$

$$
\begin{aligned}
L_P =& \max\{\frac{|-X\sin\theta_1 + Y\cos\theta_1|}{Z} f_x, \frac{|-Y\sin\theta_2 + X\cos\theta_2|}{Z} f_y\} \\
\leq& \max\{\frac{|-X'\sin\theta_1 + Y'\cos\theta_1|}{Z'} f_x, \frac{|-Y'\sin\theta_2 + X'\cos\theta_2|}{Z'} f_y\} \\
&+ \max\{|-\sin\theta_1(\frac{X}{Z} - \frac{X'}{Z'}) + \cos\theta_1(\frac{Y}{Z} - \frac{Y'}{Z'})| f_x, |-\sin\theta_2(\frac{Y}{Z} - \frac{Y'}{Z'}) + \cos\theta_2(\frac{X}{Z} - \frac{X'}{Z'})| f_y\} \\
\leq& \max\{\max_{\theta\in[-b,b]} \frac{|-X'\sin\theta + Y'\cos\theta|}{Z'} f_x, \max_{\theta\in[-b,b]} \frac{|-Y'\sin\theta + X'\cos\theta|}{Z'} f_y\} \\
&+ \max\{|\frac{f_y\cos(-\theta_1)Y + f_y\sin(-\theta_1)X}{Z} - \frac{f_y\cos(-\theta_1)Y' + f_y\sin(-\theta_1)X'}{Z'}| \frac{f_x}{f_y}, \\
&|\frac{f_x\cos(-\theta_2)X + f_x\sin(-\theta_2)Y}{Z} - \frac{f_x\cos(-\theta_2)X' + f_x\sin(-\theta_2)Y'}{Z'}| \frac{f_y}{f_x}\} \\
\leq& L_{P'} + \max\{\frac{f_x}{f_y}, \frac{f_y}{f_x}\}\delta \\
\leq& \max_{P'\in\mathbb{P}_1} L_{P'} + C_\delta^{R_z}
\end{aligned}
$$

where

$$
C_\delta^{R_z} = \max\{\frac{f_x}{f_y}, \frac{f_y}{f_x}\}\delta
$$

5) For the rotation along x-axis, for any $P = (X, Y, Z) \in \mathbb{P}\backslash\mathbb{P}_1$ and any camera motion $\alpha = \{\theta\} \in [-b, b]$, we have

$$
\begin{aligned}
D_{R_x}(P, \alpha) &= -\sin\theta Y + \cos\theta Z \\
\rho_{R_x}(P, \alpha) &= (\frac{f_x X}{-Y\sin\theta + Z\cos\theta} + c_x, \frac{Y\cos\theta + Z\sin\theta}{-Y\sin\theta + Z\cos\theta} f_y + c_y)
\end{aligned}
$$

$$
\begin{aligned}
L_P &= \max\{\max_{\alpha\in\mathcal{S}} |\frac{\partial\rho_1(P,\alpha)}{\partial\alpha}|, \max_{\alpha\in\mathcal{S}} |\frac{\partial\rho_2(P,\alpha)}{\partial\alpha}|\} \\
&= \max\{\max_{\theta\in[-b,b]} \frac{|(Y\cos\theta + Z\sin\theta)X|}{(-Y\sin\theta + Z\cos\theta)^2} f_x, \max_{\theta\in[-b,b]} \frac{Y^2 + Z^2}{(-Y\sin\theta + Z\cos\theta)^2} f_y\} \\
&= \max\{\frac{|(Y\cos\theta_1 + Z\sin\theta_1)X|}{(-Y\sin\theta_1 + Z\cos\theta_1)^2} f_x, \frac{Y^2 + Z^2}{(-Y\sin\theta_2 + Z\cos\theta_2)^2} f_y\}
\end{aligned}
$$

where

$$
\theta_1 = \operatorname*{argmax}_{\theta\in[-b,b]} \frac{|(Y\cos\theta + Z\sin\theta)X|}{(-Y\sin\theta + Z\cos\theta)^2} f_x, \theta_2 = \operatorname*{argmax}_{\theta\in[-b,b]} \frac{Y^2 + Z^2}{(-Y\sin\theta + Z\cos\theta)^2} f_y
$$

According to the definition of δ-convexity, there exists $P' = (X', Y', Z') \in \mathbb{P}_1$ where $|\ \rho(P, \alpha) - \rho(P', \alpha)\ |_\infty \leq \delta$, it holds that

$$
D(P, \alpha) \geq D(P', \alpha)
$$

So we have, for any $\theta \in [-b, b]$,

$$
0 < -\sin\theta Y' + \cos\theta Z' < -\sin\theta Y + \cos\theta Z
$$

$$
|\frac{f_x X}{-Y\sin\theta + Z\cos\theta} - \frac{f_x X'}{-Y'\sin\theta + Z'\cos\theta}| \leq \delta, |\frac{Y\cos\theta + Z\sin\theta}{-Y\sin\theta + Z\cos\theta} f_y - \frac{Y'\cos\theta + Z'\sin\theta}{-Y'\sin\theta + Z'\cos\theta} f_y| \leq \delta
$$

$$
\begin{aligned}
L_P =& \max\{\frac{|(Y\cos\theta_1 + Z\sin\theta_1)X|}{(-Y\sin\theta_1 + Z\cos\theta_1)^2} f_x, \frac{Y^2+Z^2}{(-Y\sin\theta_2 + Z\cos\theta_2)^2} f_y\} \\
\leq& \max\{\frac{|(Y'\cos\theta_1 + Z'\sin\theta_1)X'|}{(-Y'\sin\theta_1 + Z'\cos\theta_1)^2} f_x, \frac{Y'^2+Z'^2}{(-Y'\sin\theta_2 + Z'\cos\theta_2)^2} f_y\} \\
&+ \max\{\frac{\delta^2}{f_y} + \frac{\delta}{f_y}\frac{f_x|X'| + f_y|Y'\cos\theta_1 + Z'\sin\theta_1|}{-Y'\sin\theta_1 + Z'\cos\theta_1}, \frac{\delta^2}{f_y} + \frac{2\delta|Y'\cos\theta_2 + Z'\sin\theta_2|}{-Y'\sin\theta_2 + Z'\cos\theta_2}\} \\
\leq& \max\{\max_{\theta\in[-b,b]} \frac{|(Y'\cos\theta + Z'\sin\theta)X|}{(-Y'\sin\theta + Z'\cos\theta)^2} f_x, \max_{\theta\in[-b,b]} \frac{Y'^2+Z'^2}{(-Y'\sin\theta + Z'\cos\theta)^2} f_y\} \\
&+ \max\{\max_{\theta\in[-b,b]} \frac{\delta^2}{f_y} + \frac{\delta}{f_y}\frac{f_x|X'| + f_y|Y'\cos\theta + Z'\sin\theta|}{-Y'\sin\theta + Z'\cos\theta}, \max_{\theta\in[-b,b]} \frac{\delta^2}{f_y} + \frac{2\delta|Y'\cos\theta + Z'\sin\theta|}{-Y'\sin\theta + Z'\cos\theta}\}
\end{aligned}
$$

It is easy to find that,

$$
\frac{\partial \frac{|X'|}{-Y'\sin\theta + Z'\cos\theta}}{\partial\theta} = \frac{|X|(Y'\cos\theta + Z'\sin\theta)}{(-Y'\sin\theta + Z'\cos\theta)^2}, \frac{\partial \frac{Y'\cos\theta + Z'\sin\theta}{-Y'\sin\theta + Z'\cos\theta}}{\partial\theta} = \frac{Y'^2+Z'^2}{(-Y'\sin\theta + Z'\cos\theta)^2}
$$

So the functions below are increasing if $Y'\cos\theta + Z'\sin\theta > 0$, and vice versa.

$$
\max_{\theta\in[-b,b]} \frac{|X'|}{-Y'\sin\theta + Z'\cos\theta} = \max\{\frac{|X'|}{-Y'\sin b + Z'\cos b}, \frac{|X'|}{Y'\sin b + Z'\cos b}\}
$$

$$
\max_{\theta\in[-b,b]} \frac{|Y'\cos\theta + Z'\sin\theta|}{-Y'\sin\theta + Z'\cos\theta} = \max\{\frac{Y'\cos b + Z'\sin b}{-Y'\sin b + Z'\cos b}, \frac{Y'\cos b - Z'\sin b}{Y'\sin b + Z'\cos b}\}
$$

$$
\begin{aligned}
L_P \leq& \max\{\max_{\theta\in[-b,b]} \frac{|(Y'\cos\theta + Z'\sin\theta)X|}{(-Y'\sin\theta + Z'\cos\theta)^2} f_x, \max_{\theta\in[-b,b]} \frac{Y'^2+Z'^2}{(-Y'\sin\theta + Z'\cos\theta)^2} f_y\} \\
&+ \max\{\max_{\theta\in[-b,b]} \frac{\delta^2}{f_y} + \frac{\delta}{f_y}\frac{f_x|X'| + f_y|Y'\cos\theta + Z'\sin\theta|}{-Y'\sin\theta + Z'\cos\theta}, \max_{\theta\in[-b,b]} \frac{\delta^2}{f_y} + \frac{2\delta|Y'\cos\theta + Z'\sin\theta|}{-Y'\sin\theta + Z'\cos\theta}\} \\
\leq& L_{P'} + \frac{\delta^2}{f_y} + \max_{(X',Y',Z')\in\mathbb{P}_1} \max_{\theta=\pm b} \max\{\frac{\delta}{f_y}\frac{f_x|X'| + f_y|Y'\cos\theta + Z'\sin\theta|}{-Y'\sin\theta + Z'\cos\theta}, \frac{2\delta|Y'\cos\theta + Z'\sin\theta|}{-Y'\sin\theta + Z'\cos\theta}\} \\
\leq& \max_{P'\in\mathbb{P}_1} L_{P'} + C_\delta^{R_x}
\end{aligned}
$$

where

$$
C_\delta^{R_x} = \frac{\delta^2}{f_y} + \max_{(X',Y',Z')\in\mathbb{P}_1} \max_{\theta=\pm b} \max\{\frac{\delta}{f_y}\frac{f_x|X'| + f_y|Y'\cos\theta + Z'\sin\theta|}{-Y'\sin\theta + Z'\cos\theta}, \frac{2\delta|Y'\cos\theta + Z'\sin\theta|}{-Y'\sin\theta + Z'\cos\theta}\}
$$

6) For the rotation along y-axis, for any $P = (X, Y, Z) \in \mathbb{P}\backslash\mathbb{P}_1$ and any camera motion $\alpha = \{\theta\} \in [-b, b]$, we have

$$
\begin{aligned}
D_{R_y}(P,\alpha) &= \sin\theta X + \cos\theta Z \\
\rho_{R_y}(P,\alpha) &= (\frac{X\cos\theta - Z\sin\theta}{X\sin\theta + Z\cos\theta} f_x + c_x, \frac{f_y Y}{X\sin\theta + Z\cos\theta} + c_y)
\end{aligned}
$$

$$
\begin{aligned}
L_P &= \max\{\max_{\alpha\in\mathcal{S}} |\frac{\partial\rho_1(P,\alpha)}{\partial\alpha}|, \max_{\alpha\in\mathcal{S}} |\frac{\partial\rho_2(P,\alpha)}{\partial\alpha}|\} \\
&= \max\{\max_{\theta\in[-b,b]} \frac{X^2+Z^2}{(X\sin\theta + Z\cos\theta)^2} f_x, \max_{\theta\in[-b,b]} \frac{|(X\cos\theta - Z\sin\theta)Y|}{(X\sin\theta + Z\cos\theta)^2} f_y\} \\
&= \max\{\frac{X^2+Z^2}{(X\sin\theta_1 + Z\cos\theta_1)^2} f_x, \frac{|(X\cos\theta_2 - Z\sin\theta_2)Y|}{(X\sin\theta_2 + Z\cos\theta_2)^2} f_y\}
\end{aligned}
$$

where

$$
\theta_1 = \operatorname*{argmax}_{\theta\in[-b,b]} \max_{\theta\in[-b,b]} \frac{X^2+Z^2}{(X\sin\theta + Z\cos\theta)^2} f_x, \theta_2 = \operatorname*{argmax}_{\theta\in[-b,b]} \frac{|(X\cos\theta - Z\sin\theta)Y|}{(X\sin\theta + Z\cos\theta)^2} f_y
$$

According to the definition of δ-convexity, there exists $P' = (X', Y', Z') \in \mathbb{P}_1$ where $\mid \rho(P, \alpha) - \rho(P', \alpha) \mid_\infty \leq \delta$, it holds that

$$D(P, \alpha) \geq D(P', \alpha)$$

So we have, for any $\theta \in [-b, b]$,

$$0 < \sin\theta X' + \cos\theta Z' < \sin\theta X + \cos\theta Z$$

$$|\frac{X\cos\theta - Z\sin\theta}{X\sin\theta + Z\cos\theta} f_x - \frac{X'\cos\theta - Z'\sin\theta}{X'\sin\theta + Z'\cos\theta} f_x| \leq \delta, |\frac{f_y Y}{X\sin\theta + Z\cos\theta} - \frac{f_y Y'}{X'\sin\theta + Z'\cos\theta}| \leq \delta$$

$$\begin{aligned}
L_P =& \max\{\frac{X^2 + Z^2}{(X\sin\theta_1 + Z\cos\theta_1)^2} f_x, \frac{|(X\cos\theta_2 - Z\sin\theta_2)Y|}{(X\sin\theta_2 + Z\cos\theta_2)^2} f_y\} \\
\leq& \max\{\frac{X'^2 + Z'^2}{(X'\sin\theta_1 + Z'\cos\theta_1)^2} f_x, \frac{|(X'\cos\theta_2 - Z'\sin\theta_2)Y'|}{(X'\sin\theta_2 + Z'\cos\theta_2)^2} f_y\} \\
&+ \max\{\frac{\delta^2}{f_x} + \frac{2\delta|X'\cos\theta_1 - Z'\sin\theta_1|}{X'\sin\theta_1 + Z'\cos\theta_1}, \frac{\delta^2}{f_x} + \frac{\delta}{f_x}\frac{f_y|Y'| + f_x|X'\cos\theta_2 - Z'\sin\theta_2|}{X'\sin\theta_2 + Z'\cos\theta_2}\} \\
\leq& \max\{\max_{\theta\in[-b,b]} \frac{X^2 + Z^2}{(X\sin\theta + Z\cos\theta)^2} f_x, \max_{\theta\in[-b,b]} \frac{|(X\cos\theta - Z\sin\theta)Y|}{(X\sin\theta + Z\cos\theta)^2} f_y\} \\
&+ \max\{\max_{\theta\in[-b,b]} \frac{\delta^2}{f_x} + \frac{2\delta|X'\cos\theta - Z'\sin\theta|}{X'\sin\theta + Z'\cos\theta}, \max_{\theta\in[-b,b]} \frac{\delta^2}{f_x} + \frac{\delta}{f_x}\frac{f_y|Y'| + f_x|X'\cos\theta - Z'\sin\theta|}{X'\sin\theta + Z'\cos\theta}\}
\end{aligned}$$

It is easy to find that,

$$\frac{\partial \frac{|Y'|}{X'\sin\theta + Z'\cos\theta}}{\partial\theta} = \frac{|X|(X'\cos\theta - Z'\sin\theta)}{(X'\sin\theta + Z'\cos\theta)^2}, \frac{\partial \frac{X'\cos\theta - Z'\sin\theta}{X'\sin\theta + Z'\cos\theta}}{\partial\theta} = \frac{X'^2 + Z'^2}{(X'\sin\theta + Z'\cos\theta)^2}$$

So the functions below are increasing if $X'\cos\theta - Z'\sin\theta > 0$, and vice versa.

$$\max_{\theta\in[-b,b]} \frac{|Y'|}{X'\sin\theta + Z'\cos\theta} = \max\{\frac{|Y'|}{X'\sin b + Z'\cos b}, \frac{|Y'|}{-X'\sin b + Z'\cos b}\}$$

$$\max_{\theta\in[-b,b]} \frac{X'\cos\theta - Z'\sin\theta}{X'\sin\theta + Z'\cos\theta} = \max\{\frac{X'\cos b - Z'\sin b}{X'\sin b + Z'\cos b}, \frac{X'\cos b + Z'\sin b}{-X'\sin b + Z'\cos b}\}$$

$$\begin{aligned}
L_P \leq& \max\{\max_{\theta\in[-b,b]} \frac{X^2 + Z^2}{(X\sin\theta + Z\cos\theta)^2} f_x, \max_{\theta\in[-b,b]} \frac{|(X\cos\theta - Z\sin\theta)Y|}{(X\sin\theta + Z\cos\theta)^2} f_y\} \\
&+ \max\{\max_{\theta\in[-b,b]} \frac{\delta^2}{f_x} + \frac{2\delta|X'\cos\theta - Z'\sin\theta|}{X'\sin\theta + Z'\cos\theta}, \max_{\theta\in[-b,b]} \frac{\delta^2}{f_x} + \frac{\delta}{f_x}\frac{f_y|Y'| + f_x|X'\cos\theta - Z'\sin\theta|}{X'\sin\theta + Z'\cos\theta}\} \\
\leq& L_{P'} + \frac{\delta^2}{f_x} + \max_{(X',Y',Z')\in\mathbb{P}_1} \max_{\theta=\pm b} \max\{\frac{2\delta|X'\cos\theta - Z'\sin\theta|}{X'\sin\theta + Z'\cos\theta}, \frac{\delta}{f_x}\frac{f_y|Y'| + f_x|X'\cos\theta - Z'\sin\theta|}{X'\sin\theta + Z'\cos\theta}\} \\
\leq& \max_{P'\in\mathbb{P}_1} L_{P'} + C_\delta^{R_y}
\end{aligned}$$

where

$$C_\delta^{R_y} = \frac{\delta^2}{f_y} + \max_{(X',Y',Z')\in\mathbb{P}_1} \max_{\theta=\pm b} \max\{\frac{2\delta|X'\cos\theta - Z'\sin\theta|}{X'\sin\theta + Z'\cos\theta}, \frac{\delta}{f_x}\frac{f_y|Y'| + f_x|X'\cos\theta - Z'\sin\theta|}{X'\sin\theta + Z'\cos\theta}\}$$

Therefore, there exists a finite constant $C_\delta \geq 0$ such that for any 1-axis camera motion $\alpha \in \mathcal{S} = [-b, b]$, $L_P \leq \max_{P'\in\mathbb{P}_1} L_{P'} + C_\delta$, and finding all such finite constants concludes the proof. □

D.2 Proof on Lemma 4.9: Approximated upper bound of the interval from one-frame point cloud

Lemma D.3 (Approximated upper bound of interval from one-frame point cloud, **restated** of Lemma 4.9)**.** *Given the projection from one-frame 3D point cloud $V_1 \in \mathcal{V}_1 : \mathbb{P}_1 \times [0, 1]^K$ and unknown entire point cloud*

$V \in \mathcal{V} : \mathbb{P} \times [0,1]^K$ with δ-convexity, if the one-axis position projection function ρ is monotonic in l_∞ norm over camera motion space, there exists a finite constant $C_\delta \geq 0$ such that if the interval $\Delta_{OF}^{r,s}$ is defined as,

$$\Delta_{OF}^{r,s} = \min_{P_1 \in \mathbb{P}_1} \frac{1}{L_{P_1} + C_\delta} \min_{P_1 \in \mathbb{P}_1} (\mid \rho(P_1, \sup \mathbb{U}_{P_1,r,s}) - \rho(P_1, \inf \mathbb{U}_{P_1,r,s}) \mid_\infty - 2\delta) \tag{44}$$

where L_{P_1} is the Lipschitz constant for projection function ρ given 3D point P_1,

$$L_{P_1} = \max\{\max_{\alpha \in \mathcal{S}} |\frac{\partial \rho_1(P_1, \alpha)}{\partial \alpha}|, \max_{\alpha \in \mathcal{S}} |\frac{\partial \rho_2(P_1, \alpha)}{\partial \alpha}|\} \tag{45}$$

then for any $P \in \mathbb{P}$, it holds that $\Delta_{OF}^{r,s} \leq \Delta^{r,s} = \min_{P \in \mathbb{P}}\{\sup \mathbb{U}_{P,r,s} - \inf \mathbb{U}_{P,r,s}\}$

Proof. Since we do not know the entire point cloud as an oracle, we categorize $P \in \mathbb{P}$ within known 3D point cloud $\mathbb{P}_1$ and unknown 3D points $\mathbb{P} \backslash \mathbb{P}_1$ and discuss the two cases respectively.

(1) If $P \in \mathbb{P}_1$, because $\mathbb{P}_1$ is reconstructed from one image, for any $\alpha \in \mathcal{S}, (r,s) \in \mathbb{G}$ the projection equation $\lfloor \rho(P, \alpha) \rfloor = (r,s)$ has only one unique solution P, i.e.

$$\{P \in \mathbb{P}_1 \mid \lfloor \rho(P, \alpha) \rfloor = (r,s), \alpha \in \mathcal{S}\} = \{P_1 \in \mathbb{P}_1\}$$

Besides, with $C_\delta \geq 0, \delta > 0$, we have

$$\begin{aligned}
\Delta_{OF}^{r,s} &= \min_{P_1 \in \mathbb{P}_1} \frac{1}{L_{P_1} + C_\delta} \min_{P_1 \in \mathbb{P}_1} (\mid \rho(P_1, \sup \mathbb{U}_{P_1,r,s}) - \rho(P_1, \inf \mathbb{U}_{P_1,r,s}) \mid_\infty - 2\delta) \\
&\leq \min_{P_1 \in \mathbb{P}_1} \frac{1}{L_{P_1} + C_\delta} (\mid \rho(P_1, \sup \mathbb{U}_{P_1,r,s}) - \rho(P_1, \inf \mathbb{U}_{P_1,r,s}) \mid_\infty - 2\delta) \\
&< \min_{P \in \mathbb{P}_1} \frac{1}{L_P} \mid \rho(P, \sup \mathbb{U}_{P,r,s}) - \rho(P, \inf \mathbb{U}_{P,r,s}) \mid_\infty
\end{aligned} \tag{46}$$

Then directly based on Lemma C.1, we have for any $P \in \mathbb{P}_1$, at pixel (r,s),

$$\Delta_{OF}^{r,s} \leq \min_{P \in \mathbb{P}}\{\sup \mathbb{U}_{P,r,s} - \inf \mathbb{U}_{P,r,s}\}$$

the proof is concluded.

(2) If $P \in \mathbb{P} \backslash \mathbb{P}_1$, suppose it is projected to on r, s through projection function ρ from 3D point $P : \lfloor \rho(P, \alpha) \rfloor = (r,s)$ over camera motion set $\mathbb{U}_{P,r,s} \neq \emptyset$, based on Lemma D.2, for the 1-axis translation or rotation projective function ρ, there exists a constant $C_\delta \geq 0$ such that

$$L_P \leq \max_{P_1 \in \mathbb{P}_1} L_{P_1} + C_\delta$$

Therefore, the interval $\Delta_{OF}^{r,s}$ satisfies

$$\begin{aligned}
\Delta_{OF}^{r,s} &= \min_{P_1 \in \mathbb{P}_1} \frac{1}{L_{P_1} + C_\delta} \min_{P_1 \in \mathbb{P}_1} (\mid \rho(P_1, \sup \mathbb{U}_{P_1,r,s}) - \rho(P_1, \inf \mathbb{U}_{P_1,r,s}) \mid_\infty - 2\delta) \\
&\leq \frac{1}{L_P} \min_{P_1 \in \mathbb{P}_1} (\mid \rho(P_1, \sup \mathbb{U}_{P_1,r,s}) - \rho(P_1, \inf \mathbb{U}_{P_1,r,s}) \mid_\infty - 2\delta)
\end{aligned}$$

For the Lipschitz constant L_P for projection function ρ with point $P \in \mathbb{P} \backslash \mathbb{P}_1$, we have

$$\begin{aligned}
L_P &= \max\{\max_{\alpha \in \mathcal{S}} |\frac{\partial \rho_1(P, \alpha)}{\partial \alpha}|, \max_{\alpha \in \mathcal{S}} |\frac{\partial \rho_2(P, \alpha)}{\partial \alpha}|\} \\
&= \max\{\max_{\alpha, \beta \in \mathcal{S}} \frac{|\rho_1(P, \alpha) - \rho_1(P, \beta)|}{|\alpha - \beta|}, \max_{\alpha, \beta \in \mathcal{S}} \frac{|\rho_2(P, \alpha) - \rho_2(P, \beta)|}{|\alpha - \beta|}\} \\
&= \max_{\alpha, \beta \in \mathcal{S}} \max\{\frac{|\rho_1(P, \alpha) - \rho_1(P, \beta)|}{|\alpha - \beta|}, \frac{|\rho_2(P, \alpha) - \rho_2(P, \beta)|}{|\alpha - \beta|}\} \\
&= \max_{\alpha, \beta \in \mathcal{S}} \frac{|\rho(P, \alpha) - \rho(P, \beta|_\infty}{|\alpha - \beta|}
\end{aligned} \tag{47}$$

Therefore, we have

$$\begin{aligned}
&\max_{\alpha\in\mathcal{S}} \frac{|\rho(P,\alpha+\Delta_{OF}^{r,s})-\rho(P,\alpha)|_\infty}{\Delta_{OF}^{r,s}} \\
\leq &\max_{\alpha,\beta\in\mathcal{S}} \frac{|\rho(P,\alpha)-\rho(P,\beta)|_\infty}{|\alpha-\beta|} = L_P \\
\leq &\frac{1}{\Delta_{OF}^{r,s}} \min_{P_1\in\mathbb{P}_1} (\mid \rho(P_1,\sup\mathbb{U}_{P_1,r,s})-\rho(P_1,\inf\mathbb{U}_{P_1,r,s}) \mid_\infty -2\delta)
\end{aligned}$$

From the Definition D.1, when the camera motion is at $\sup\mathbb{U}_{P,r,s}$, there exists $\mathbb{P}_{sup}\subset\mathbb{P}_1$ such that

$$\forall P'_{sup}\in\mathbb{P}_{sup}, \mid \rho(P,\sup\mathbb{U}_{P,r,s})-\rho(P'_{sup},\sup\mathbb{U}_{P,r,s}) \mid_\infty \leq \delta$$

Similarly when the camera motion is at $\inf\mathbb{U}_{P,r,s}$, there exists $\mathbb{P}_{inf}\subset\mathbb{P}_1$ such that

$$\forall P'_{inf}\in\mathbb{P}_{inf}, \mid \rho(P,\inf\mathbb{U}_{P,r,s})-\rho(P'_{inf},\inf\mathbb{U}_{P,r,s}) \mid_\infty \leq \delta$$

Now choose $P'\in\mathbb{P}_{inf}\cap\mathbb{P}_{sup}$ such that

$$\mid \rho(P',\sup\mathbb{U}_{P,r,s})-\rho(P',\inf\mathbb{U}_{P,r,s}) \mid_\infty \geq 2\delta$$

Therefore, based on the Absolute Value Inequalities, we have

$$\begin{aligned}
&\mid \rho(P,\sup\mathbb{U}_{P,r,s})-\rho(P,\inf\mathbb{U}_{P,r,s}) \mid_\infty \\
\geq &\mid\mid \rho(P',\sup\mathbb{U}_{P,r,s})-\rho(P',\inf\mathbb{U}_{P,r,s}) \mid_\infty -2\delta \mid \\
= &\mid \rho(P',\sup\mathbb{U}_{P,r,s})-\rho(P',\inf\mathbb{U}_{P,r,s}) \mid_\infty -2\delta
\end{aligned}$$

Since $\sup\mathbb{U}_{P,r,s}\notin\mathbb{U}_{P',r,s}$ and $,\inf\mathbb{U}_{P,r,s}\notin\mathbb{U}_{P',r,s}$, so we have

$$\sup\mathbb{U}_{P,r,s}-\inf\mathbb{U}_{P,r,s}\geq\sup\mathbb{U}_{P',r,s}-\inf\mathbb{U}_{P',r,s}$$

Based on the monotonicity of ρ in l_∞ norm given P, we have

$$\mid \rho(P',\sup\mathbb{U}_{P,r,s})-\rho(P',\inf\mathbb{U}_{P,r,s}) \mid_\infty \geq \mid \rho(P',\sup\mathbb{U}_{P',r,s})-\rho(P',\inf\mathbb{U}_{P',r,s}) \mid_\infty$$

Combining these inequities above, for any $u\in\mathcal{S}$,

$$\begin{aligned}
\mid \rho(P,u+\Delta_{OF}^{r,s})-\rho(P,u) \mid_\infty &\leq \max_{\alpha\in\mathcal{S}} |\rho(P,\alpha+\Delta_{OF}^{r,s})-\rho(P,\alpha)|_\infty \\
&\leq \min_{P_1\in\mathbb{P}_1} (\mid \rho(P_1,\sup\mathbb{U}_{P_1,r,s})-\rho(P_1,\inf\mathbb{U}_{P_1,r,s}) \mid_\infty -2\delta) \\
&\leq \mid \rho(P',\sup\mathbb{U}_{P',r,s})-\rho(P',\inf\mathbb{U}_{P',r,s}) \mid_\infty -2\delta \\
&\leq \mid \rho(P',\sup\mathbb{U}_{P,r,s})-\rho(P',\inf\mathbb{U}_{P,r,s}) \mid_\infty -2\delta \\
&\leq \mid \rho(P,\sup\mathbb{U}_{P,r,s})-\rho(P,\inf\mathbb{U}_{P,r,s}) \mid_\infty
\end{aligned}$$

Based on the monotonicity of ρ in l_∞ norm given P, let $u=\inf\mathbb{U}_{P,r,s}$, we have $\Delta_{OF}^{r,s}\leq\sup\mathbb{U}_{P,r,s}-\inf\mathbb{U}_{P,r,s}$ for any P and (r,s). Therefore, for pixel (r,s),

$$\Delta_{OF}^{r,s}\leq\min_{P\in\mathbb{P}}\{\sup\mathbb{U}_{P,r,s}-\inf\mathbb{U}_{P,r,s}\}$$

which concludes the proof. □

D.3 Proof on Theorem 4.10: Certification with approximated partitions from one-frame point cloud

Theorem D.4 (Certification with approximated partitions from one-frame point cloud, **restated** of Theorem 4.10)**.** *For the projection from one-frame 3D point cloud $V_1\in\mathcal{V}_1:\mathbb{P}_1\times[0,1]^K$ and unknown entire point cloud $V\in\mathcal{V}:\mathbb{P}\times[0,1]^K$ with δ-convexity, let ϕ be the one-axis rotation or translation with parameters in $\mathcal{S}\subseteq\mathcal{Z}$*

with the monotonic projection function ρ in l_∞ norm over $\mathcal{S}$, we have uniformly partitioned $\{\alpha_i\}_{i=1}^N \subseteq \mathcal{S}$ under interval Δ_α with finite constant $C_\delta \geq 0$ where

$$\Delta_\alpha \leq \min_{r,s} \Delta_{OF}^{r,s} = \min_{r,s} \min_{P_1 \in \mathbb{P}_1} \frac{1}{L_{P_1} + C_\delta} \min_{P_1 \in \mathbb{P}_1} (\mid \rho(P_1, \sup \mathbb{U}_{P_1,r,s}) - \rho(P_1, \inf \mathbb{U}_{P_1,r,s}) \mid_\infty - 2\delta)$$
$$L_{P_1} = \max\{\max_{\alpha \in \mathcal{S}} |\frac{\partial \rho_1(P_1, \alpha)}{\partial \alpha}|, \max_{\alpha \in \mathcal{S}} |\frac{\partial \rho_2(P_1, \alpha)}{\partial \alpha}|\} \tag{48}$$

Let $y_A, y_B \in \mathcal{Y}$, $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_d)$ and suppose that for any i, the ε-smoothed classifier defined by $q(y \mid x; \varepsilon) := \mathbb{E}(p(y \mid x + \varepsilon))$ has class probabilities that satisfy with top-2 classes y_A, y_B as $q(y_A \mid \phi(x, \alpha_i); \varepsilon) \geq p_A^{(i)} \geq p_B^{(i)} \geq \max_{y \neq y_A} q(y \mid \phi(x, \alpha_i); \varepsilon)$,

then it is guaranteed that $\forall \alpha \in \mathcal{S}$: $y_A = \operatorname{argmax}_y q(y \mid \phi(x, \alpha); \varepsilon)$ if

$$\max_{1 \leq i \leq N-1} \sqrt{\frac{1}{2} \sum_{k,r,s} (\phi(x, \alpha_i)_{krs} - \phi(x, \alpha_{i+1})_{krs})^2} < \frac{\sigma}{2} \min_{1 \leq i \leq N} \left(\Phi^{-1}\left(p_A^{(i)}\right) - \Phi^{-1}\left(p_B^{(i)}\right) \right).$$

Proof. According to Theorem A.7, we need to upper bound $\max_{\alpha \in \mathcal{S}} \min_{1 \leq i \leq N} \|\phi(x, \alpha) - \phi(x, \alpha_i)\|_2$ with uniform partitions under interval of $\Delta_\alpha = \alpha_{i+1} - \alpha_i$.

$$\begin{aligned}
&\max_{\alpha \in \mathcal{S}} \min_{1 \leq i \leq N} \|\phi(x, \alpha) - \phi(x, \alpha_i)\|_2 \\
\leq &\max_{1 \leq i \leq N-1} \sqrt{\max_{\alpha_i \leq \alpha \leq \alpha_{i+1}} \min\{(\phi(x, \alpha) - \phi(x, \alpha_i))^2, (\phi(x, \alpha) - \phi(x, \alpha_{i+1}))^2\}} \\
= &\max_{1 \leq i \leq N-1} \sqrt{\max_{\alpha_i \leq \alpha \leq \alpha_{i+1}} \min\{\sum_{k,r,s} (\phi(x, \alpha)_{krs} - \phi(x, \alpha_i)_{krs})^2, \sum_{k,r,s} (\phi(x, \alpha)_{krs} - \phi(x, \alpha_{i+1})_{krs})^2\}} \\
\leq &\max_{1 \leq i \leq N-1} \sqrt{\max_{\alpha_i \leq \alpha \leq \alpha_{i+1}} \frac{1}{2}\{\sum_{k,r,s} (\phi(x, \alpha)_{krs} - \phi(x, \alpha_i)_{krs})^2 + \sum_{k,r,s} (\phi(x, \alpha)_{krs} - \phi(x, \alpha_{i+1})_{krs})^2\}} \\
= &\max_{1 \leq i \leq N-1} \sqrt{\max_{\alpha_i \leq \alpha \leq \alpha_{i+1}} \frac{1}{2} \sum_{k,r,s} [(\phi(x, \alpha)_{krs} - \phi(x, \alpha_i)_{krs})^2 + (\phi(x, \alpha)_{krs} - \phi(x, \alpha_{i+1})_{krs})^2]}
\end{aligned}$$

According to Lemma D.3, for any $\alpha_i \leq \alpha \leq \alpha_{i+1}$ on the pixel (r, s),

$$\begin{aligned}
\Delta_\alpha \leq \min_{r,s} \Delta_{OF}^{r,s} &\leq \min_{r,s} \min_{P \in \mathbb{P}} \{\sup \mathbb{U}_{P,r,s} - \inf \mathbb{U}_{P,r,s}\} \\
&\leq \min_{P \in \mathbb{P}} \{\sup \mathbb{U}_{P,r,s} - \inf \mathbb{U}_{P,r,s}\}
\end{aligned}$$

Based on Lemma B.2, we have

$$O(V, \alpha)_{r,s} \in \{O(V, \alpha_i)_{r,s}, O(V, \alpha_{i+1})_{r,s}\}$$

i.e., either $O(V, \alpha)_{r,s} = O(V, \alpha_i)_{r,s}$ or $O(V, \alpha)_{r,s} = O(V, \alpha_{i+1})_{r,s}$ holds. With $O(V, \alpha) = \phi(x, \alpha)$ for any $\alpha \in \mathcal{S}$, for any k, r, s, we have

$$(\phi(x, \alpha)_{krs} - \phi(x, \alpha_i)_{krs})^2 + (\phi(x, \alpha)_{krs} - \phi(x, \alpha_{i+1})_{krs})^2 = (\phi(x, \alpha_i)_{krs} - \phi(x, \alpha_{i+1})_{krs})^2$$

Therefore,

$$\begin{aligned}
&\max_{\alpha \in \mathcal{S}} \min_{1 \leq i \leq N} \|\phi(x, \alpha) - \phi(x, \alpha_i)\|_2 \\
\leq &\max_{1 \leq i \leq N-1} \sqrt{\max_{\alpha_i \leq \alpha \leq \alpha_{i+1}} \frac{1}{2} \sum_{k,r,s} [(\phi(x, \alpha)_{krs} - \phi(x, \alpha_i)_{krs})^2 + (\phi(x, \alpha)_{krs} - \phi(x, \alpha_{i+1})_{krs})^2]} \\
= &\max_{1 \leq i \leq N-1} \sqrt{\frac{1}{2} \sum_{k,r,s} (\phi(x, \alpha_i)_{krs} - \phi(x, \alpha_{i+1})_{krs})^2}
\end{aligned}$$

So if

$$\max_{1\le i\le N-1}\sqrt{\frac{1}{2}\sum_{k,r,s}(\phi(x,\alpha_i)_{krs}-\phi(x,\alpha_{i+1})_{krs})^2}<\frac{\sigma}{2}\min_{1\le i\le N}\left(\Phi^{-1}\left(p_A^{(i)}\right)-\Phi^{-1}\left(p_B^{(i)}\right)\right). \tag{49}$$

then it holds that

$$M_{\mathcal{S}}:=\max_{\alpha\in\mathcal{S}}\min_{1\le i\le N}\|\phi(x,\alpha)-\phi(x,\alpha_i)\|_2<R:=\frac{\sigma}{2}\min_{1\le i\le N}\left(\Phi^{-1}\left(p_A^{(i)}\right)-\Phi^{-1}\left(p_B^{(i)}\right)\right). \tag{50}$$

which concludes the proof based on Theorem A.7. □

E More Experiment Details and Results

In this section, we present more details of the experiments, including experimental setup, model training, and more results and analysis.

E.1 Experimental Setup

Model training. For the model training, we first mix the training sets of all motion types, $T_z, T_x, T_y, R_z, R_x, R_y$, generating the whole training set for data augmentation. We adopt the ImageNet [Deng et al., 2009] pre-trained models with architectures of ResNet101 and ResNet50 [He et al., 2016] from torch vision and fine-tune them for 100 epochs using the six mixed motion augmented training data [Hu et al., 2022b] with a learning rate of 0.001 and consistency regularization [Jeong and Shin, 2020, Hu et al., 2022b]. For the base model with diffusion-based denoiser [Carlini et al., 2022], we adopt the pre-trained unconditional 256x256 diffusion model [Dhariwal and Nichol, 2021] as the frozen denoiser and only fine-tune the base classifier for 1 epoch to address the domain shift. To fairly compare the certification performance, we keep the base classifiers the same for the baseline and ours. Please note that diffusion-based denoiser [Carlini et al., 2022] does not apply to CMS baseline because the denoiser is designed for pixel-wise Gaussian denoising as the pre-trained model and can only apply to pixel-wise smoothing.

Details of the required projection frames. Since image capturing from the camera is the most non-efficient and time-consuming part of the certification against camera motion, we adopt the *number of required projected frames* in the certification as the fair hardware-independent metric to compare the certification efficiency. Theoretically, the number of required projection frames for the uniform partitions can be calculated based on Lemma 4.4, 4.6 and 4.9 for *PwS*, *PwS-L*, *PwS-OF* in the implementation of MetaRoom dataset. Empirically, it can be observed that the certification performance will converge as the partition number increases, as shown in Section 5.4. Therefore, to avoid redundant partitions in the implementation, we adopt a quantile of over 99% for the minimal fully-covered partition interval across all the pixels as the required offline projection frames and take the average over all images.

Table 8: Required projection frames as partitions and corresponding certified accuracy for different axes of rotation and translation, with ResNet-50 and $\sigma = 0.5$

Num. of Required Projected Frames / Certified Accuracy	PwS	PwS-L	PwS-OF
x-axis translation T_x, 5mm radius	3.1k / 0.192	5.0k / 0.192	6.1k / 0.192
y-axis translation T_y, 5mm radius	3.2k / 0.183	4.8k / 0.183	6.0k / 0.183
x-axis rotation R_x, 0.25° radius	3.8k / 0.183	5.3k / 0.192	6.5k / 0.192
z-axis rotation R_z, 0.7° radius	3.2k / 0.133	4.9k / 0.133	5.8k / 0.133

Metric of average certification time per image. In addition to the *number of required projected frames*, we also report the average certification time per image as an efficiency metric. We conduct all the experiments on the same machine with NVIDIA A6000 and 512G RAM so that it is fair to compare the time consumption of different methods. To calculate the average time per image, we collect the time duration for every test sample, starting from reading the camera pose and ending with deciding whether the certification condition holds or not using Monte Carlo based algorithm [Cohen et al., 2019]. We then find the mean of all the time duration and report it as the average certification time per image.

Table 9: Certified accuracy with different smoothing variance σ^2 in all projections under ResNet-101

Projection and radius	T_z, 10mm	T_x, 5mm	T_y, 5mm	R_z, 0.7°	R_x, 0.25°	R_y, 0.25°
$\sigma = 0.25$	0.175	0.0	0.075	0.0	0.058	0.025
$\sigma = 0.5$	0.150	0.167	0.133	0.117	0.142	0.125
$\sigma = 0.75$	0.140	0.100	0.133	0.075	0.117	0.100

Table 10: Certified accuracy under different smoothing variances and different radii

Radii along z-axis translation, T_z	10mm		100mm	
Smoothing variance σ^2, PwS	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 0.25$	$\sigma = 0.5$
ResNet50	0.198	0.223	0.0	0.142
ResNet101	0.175	0.150	0.0	0.108

Empirical robust accuracy and benign accuracy We further compare the empirical robust accuracy and benign accuracy of the smoothed classifier added with pixel-wise Gaussian smoothing distribution. Following the metric in previous work [Li et al., 2021, Hu et al., 2022b], we adopt 100-perturbed worst-case empirical robust accuracy, which is calculated as follows. By making 100 uniform perturbations with the attack radius in the camera motion space, we feed every motion-perturbed image into the smoothed classifier; if any of these 100 perturbed images are wrongly classified, we report the smoothed classifier is not robust at the test sample. The empirical robust accuracy is calculated as the ratio of robust test samples among all the test samples. The benign accuracy is the correctly-classified ratio of the test samples without any camera motion perturbation. The certified accuracy is at the same certified radius as the perturbation radius.

E.2 More Experimental Results

Projected frames required and certified accuracy for more axes projection. From Table 8, we can see that for each translation or rotation, the number of required projected frames of *PwS* is less than *PwS-L* and *PwS-OF*, which is consistent with the theoretical analysis of Lemma 4.4, 4.6 and 4.9. Also, the certified accuracy is very close due to the convergence and saturation of certification performance regarding partition number.

Influence of the variance of the smoothing distribution on ResNet-101. Similar to Table 5, the certified accuracy with $\sigma = 0.5$ is no better than that with $\sigma = 0.75$ due to the accuracy/robustness trade-off [Cohen et al., 2019] for all translation and rotation axes in Table 9. Different from Table 5, the performance with $\sigma = 0.25$ is better for z-axis translation than $\sigma = 0.5$ while much poorer for other axes, which shows that the more complex model is less sensitive to z-axis translation to cover the projection errors in the certification condition.

Performance under different model complexity with smaller variance. Different from Table 7, under smaller variance of pixel-wise smoothing in Table 10, larger model complexity can be more certifiably robust with lower certified accuracy, showing less influence of robust overfitting with smaller variance.

Performance with different partitions along z-axis translation. We report the influence of partition numbers in empirical experiments under different models along z-axis translation in Table 11. Similar to Table 6, the certification accuracy will be saturated as the partition number increases, which covers the required projected frames in Table 2.

Table 11: Certified accuracy with different numbers of partitioned images in pixel-wise smoothing under different radii

Number of Partitions							
Resnet-50, z-axis translation T_z	1000	2000	3000	4000	5000	6000	7000
10mm radius	0.183	0.192	0.217	0.223	0.223	0.223	0.231
20mm radius	0.167	0.175	0.183	0.192	0.192	0.192	0.200

Table 12: Certification time per image and perception ratio comparison with baseline CMS

Certification time (s / image)	CMS [Hu et al., 2022b]	PwS	PwS-L	PwS-OF
T_z, 10mm radius	2085.8 / 100%	172.9 / 8.3%	279.8 / 13.2%	479.5 / 23.0%
R_y, 0.25 radius	2035.4 / 100%	236.4 / 11.6%	398.7 / 19.6%	608.5 / 29.9%

Table 13: Comparison of certified accuracy and empirical robust accuracy. The pixel-wise smoothed classifier is based on ResNet-50 with $\sigma = 0.75$

Perturbation radii for T_z (mm)	Certified Accuracy	Empirical Robust Accuracy	Benign Accuracy
[-10, 10]	0.140	0.283	
[-20, 20]	0.117	0.275	0.308
[-100, 100]	0.091	0.258	
Perturbation radii for R_y (°)	Certified Accuracy	Empirical Robust Accuracy	Benign Accuracy
[-0.25, 0.25]	0.108	0.275	
[-0.5, 0.5]	0.092	0.258	0.316
[-2.5, 2.5]	0.058	0.117	

Comparison of Certification Time Efficiency. We compare the time efficiency for the baseline CMS [Hu et al., 2022b] and ours in Table 12. Note that all the baselines of CMS are with 10k Monte Carlo sampling as default [Hu et al., 2022b] for the comparison. We can see that our method has significantly less wall-clock time compared to the baseline in terms of certification time per image, due to requiring fewer projected frames offline.

Comparison with Empirical Robust Accuracy and Benign Accuracy. As shown in Table 13, we compare the certified accuracy, empirical robust accuracy, and benign accuracy for smoothed classifier of ResNet50 with $\sigma = 0.75$. It can be seen that the certified accuracy is lower than the empirical robustness accuracy, which is consistent with the claim in the previous work [Li et al., 2021, Hu et al., 2022b]. The empirical robust accuracy is lower than the benign accuracy, showing that even the pixel-wise smoothed classifier is not robust against the camera motion perturbation and needs certification to provable guarantee the robustness. Please note that we adopt the smoothed classifier, which requires the input image to be added pixel-wise zero-mean Gaussian noise as smoothing. This is the reason why the empirical robust accuracy and benign accuracy are quite low compared to that in the previous work [Hu et al., 2022b], owing to the accuracy/robustness trade-off of smoothing distribution [Cohen et al., 2019].