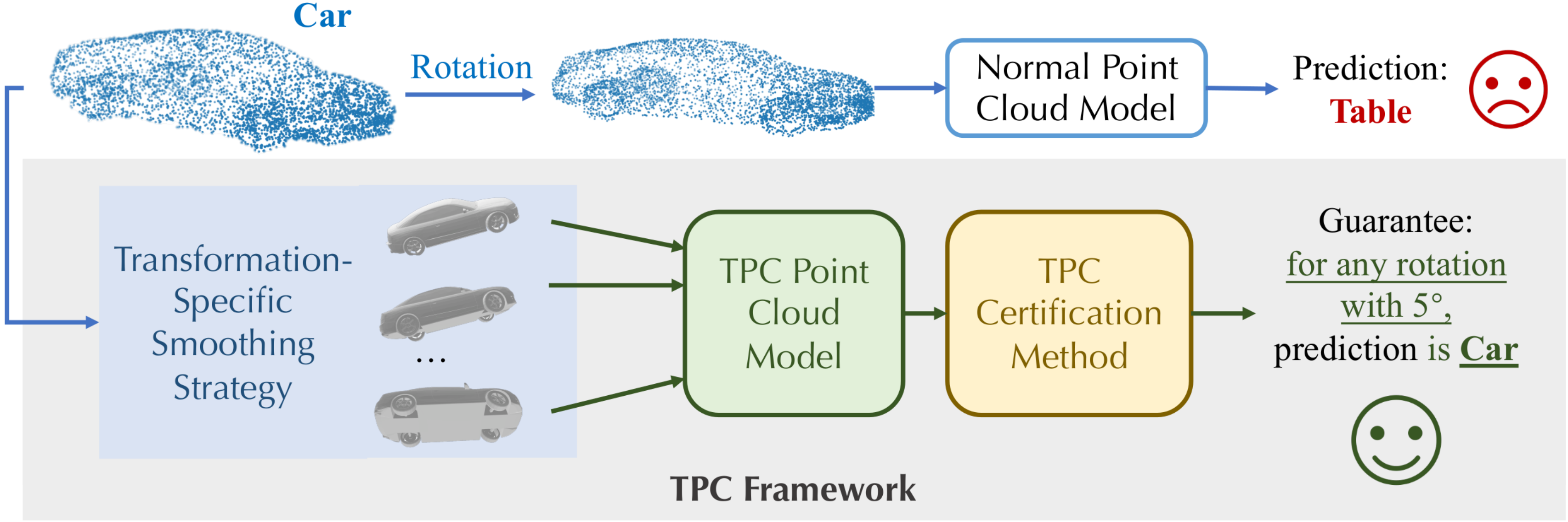


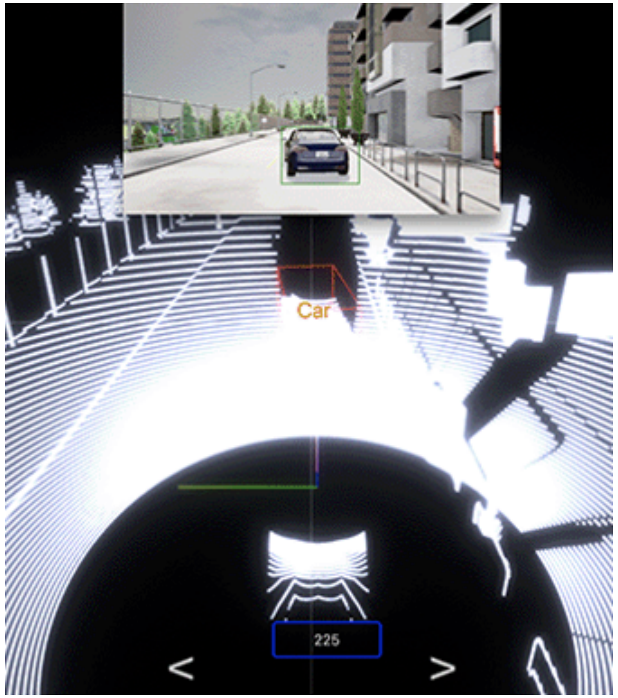
TPC: Transformation-Specific Smoothing for Point Cloud Models

Wenda Chu[1], Linyi Li[2], Bo Li[2]
(presenter)

[1] IIIS, Tsinghua University
[2] University of Illinois Urbana-Champaign

Robustness Vulnerabilities

- Point cloud models widely used in autonomous driving
- Autonomous driving has shown vulnerable to adversarial perturbations
  - Not only ℓ_p-bounded perturbations, but also semantic attacks
- We propose TPC framework that **boosts** certified robustness of **large point cloud models** against various **semantic attacks**



Simulation shows a rotated car cannot be detected.

Green: car detected Red: car not detected

Background: Randomized Smoothing

- Idea: add random noise to a base classifier $h : \mathcal{X} \to \mathcal{Y}$
  - The smoothed classifier has easy-to-compute robustness certification

$$g(x) := \arg\max_{y \in \mathcal{Y}} \mathbb{E}_\epsilon \Pr[h(x+\epsilon) = y], \epsilon \sim \mathcal{N}(0, \sigma^2 I).$$

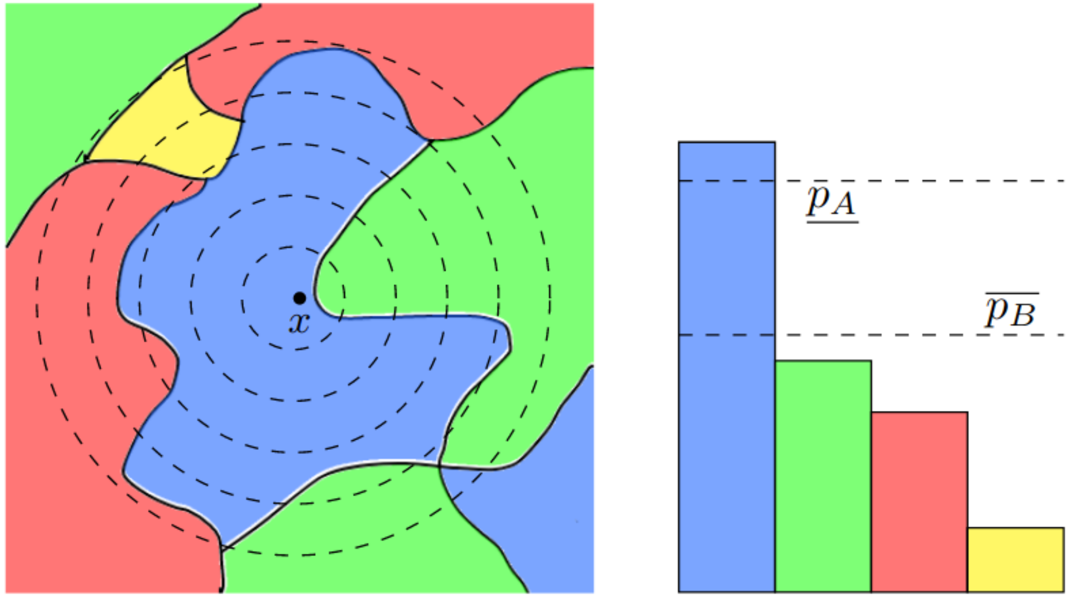


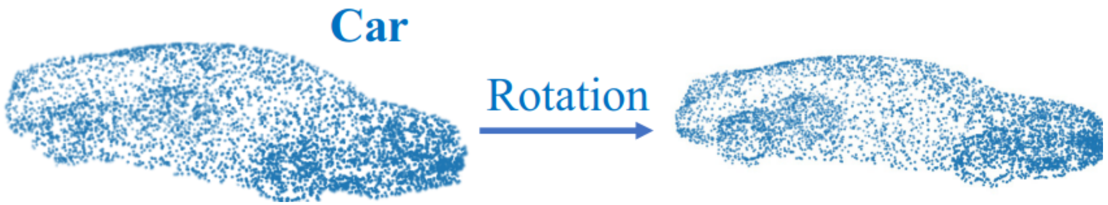
Image source: [Cohen et al. 2019]

Threat Model

- Adversary: applies parameterized transformations to point clouds

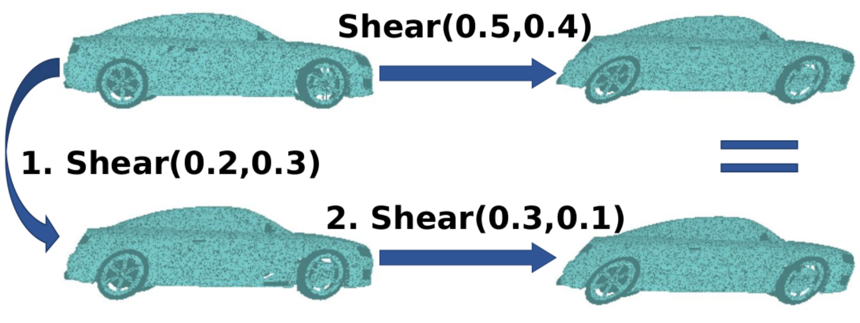
Point cloud Parameter space

- Transformation $\phi : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$
- Including: rotation, twist, shear, taper, L_p norm noise, etc.

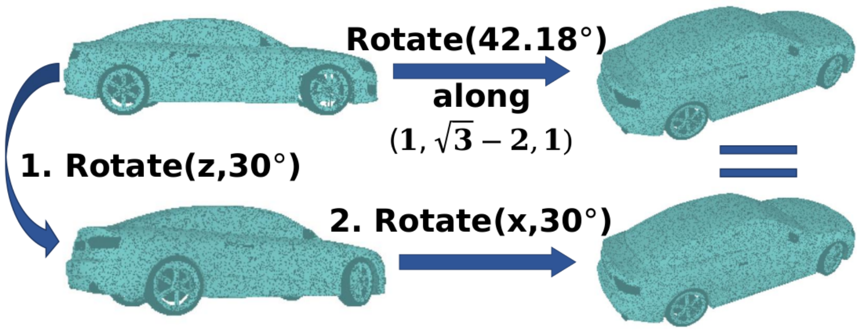


Transformation Taxonomy

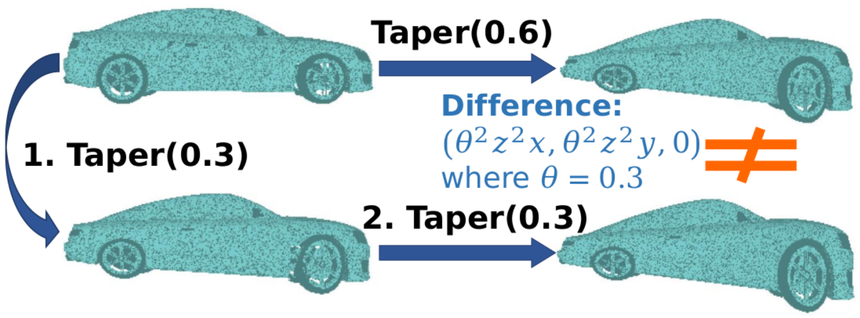
Transformations $\phi : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ fall into three categories:



(a) Additive



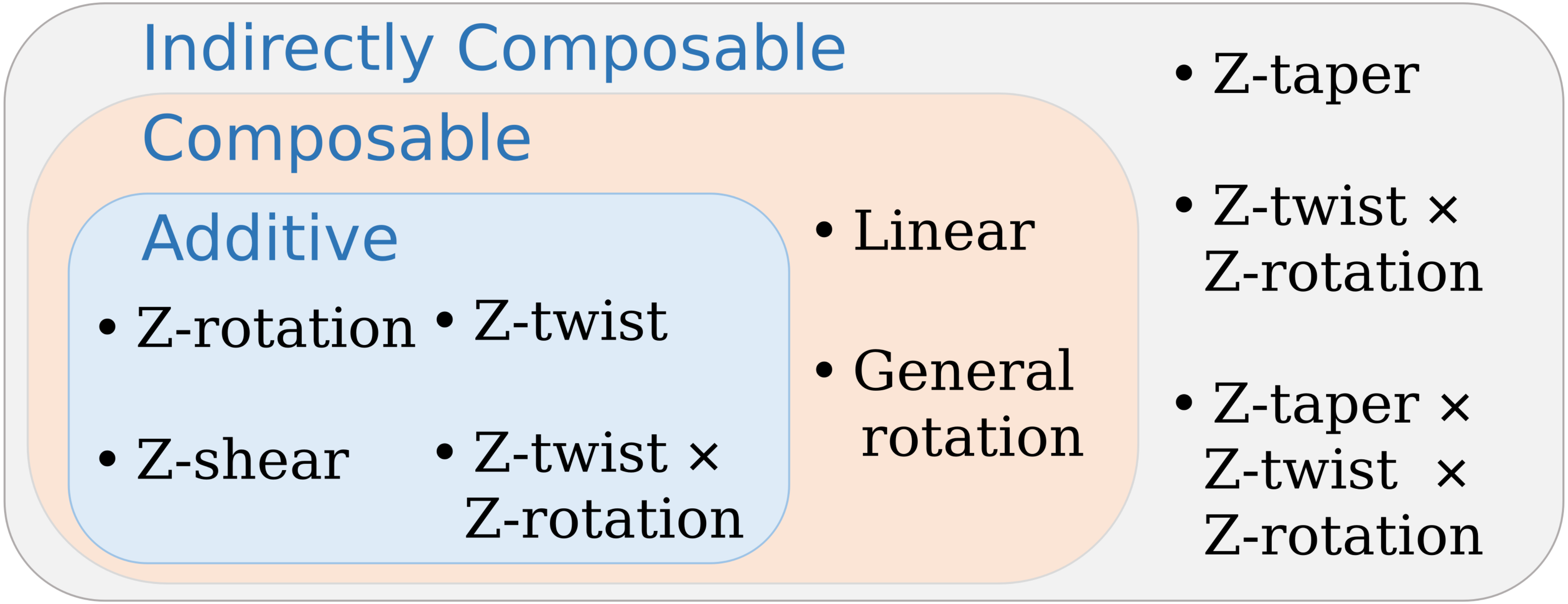
(b) Composable



(c) Indirectly composable

Transformation Taxonomy

Transformations $\phi : \mathcal{X} \times \mathcal{Z} \rightarrow \mathcal{X}$ nto three categories:



Certification Strategy

- Additive $\phi(\phi(x,\alpha),\beta) = \phi(x,\alpha+\beta)$

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

- Composable
  - e.g., Linear transformations, $\phi(x,\alpha) = (I+\alpha)x, \quad \alpha \in \mathbb{R}^{3\times 3}$

$$\|\alpha\|_F \le R, \quad R = \frac{\sigma\Big(\Phi^{-1}(\tilde{p}_A) - \Phi^{-1}(1-\tilde{p}_A)\Big)}{2+\sigma\Big(\Phi^{-1}(\tilde{p}_A) - \Phi^{-1}(1-\tilde{p}_A)\Big)}.$$

Certification Strategy

- Indirectly composable
  - Draw N samples $\{\alpha_i\}$ from the parameter space
  - Check that $p_A^{(i)} \geq p_B^{(i)}$ for each transformed point cloud $\phi(x, \alpha_i)$
  - Interpolation error

$$\begin{aligned}\mathcal{M}_{\mathcal{Z}} &:= \max_{\alpha \in \mathcal{Z}} \min_{1 \leq i \leq N} \mathcal{M}(\alpha, \alpha_i) \\ &= \max_{\alpha \in \mathcal{Z}} \min_{1 \leq i \leq N} \|\phi(x, \alpha) - \phi(x, \alpha_i)\|_2\end{aligned}$$

  - Guaranteed to be robust if

$$\mathcal{M}_{\mathcal{Z}} \leq \frac{\sigma}{2} \min_{1 \leq i \leq N} \left(\Phi^{-1}(p_A^{(i)}) - \Phi^{-1}(p_B^{(i)}) \right)$$

Empirical Results

Dataset: **ModelNet40** [Wu et al. 2015]
Architecture: **PointNet** [Qi et al. 2017]
Baseline: **DeepG3D** [Lorenz et al. 2021]

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

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

(b) Certified accuracy of TPC under $\theta = \pm 180^\circ$

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

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

Summary

TPC	DSRS

- Robustness certification for randomized smoothing

TPC

- Significantly tighter certification against semantic perturbations
- For large-scale point cloud models
- Core idea: transformation-specific smoothing

DSRS

- Theoretically tight certification against ℓ_p perturbations
- For arbitrary classifiers
- Core idea: double sampling

Paper: https://arxiv.org/abs/2201.12733
Code & Model & Data: github.com/Qianhewu/Point-Cloud-Smoothing
Poster: Hall E 211 (6:30 – 8:30 PM Today)

Paper: arxiv.org/abs/2206.07912
Code & Model & Data: github.com/lllylly/DSRS
Poster: Hall E 213 (6:30 – 8:30 PM Today)