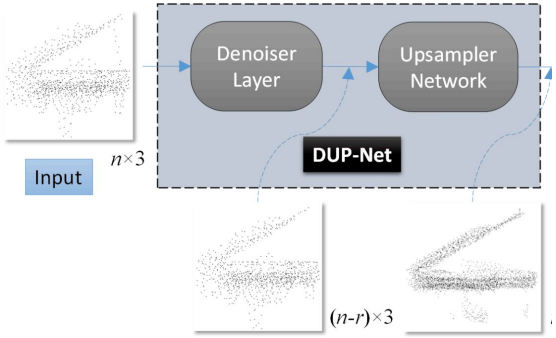
Exploring Deep Point-Cloud Robustness

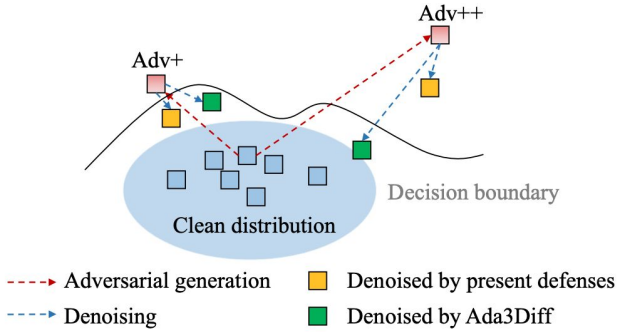
Hang Zhou
Simon Fraser University

Overview

Defense

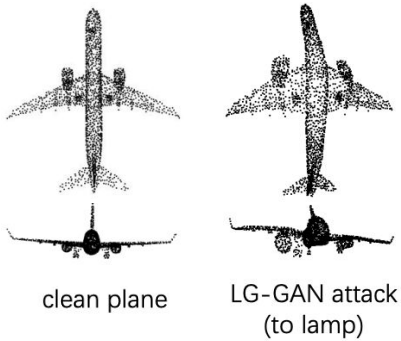


DUP-Net (ICCV 2019)

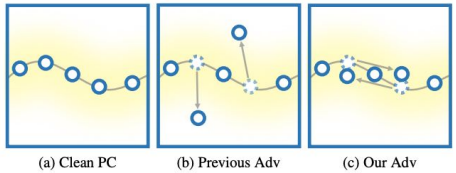


Ada3Diff (arXiv)

Attack



LG-GAN (CVPR 2020)



SI-Adv (CVPR 2022)

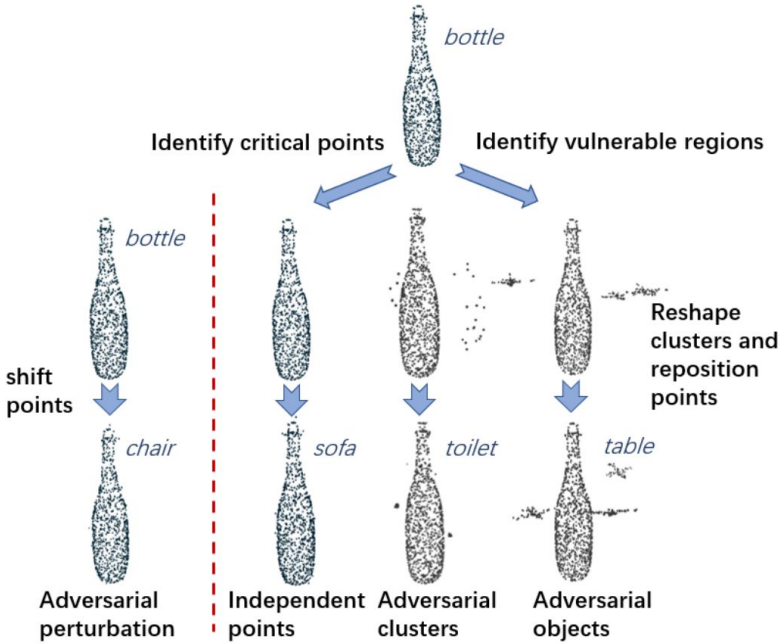
Motivation

DUP-Net

Point-cloud Carlini-Wagner attack

Facts:

- Optimization based attack
- Loss function too loose



[1] Nicholas Carlini and David Wagner, Towards evaluating the robustness of neural networks, S&P 2017
[2] Chong Xiang et al., Generating 3D adversarial point clouds, CVPR 2019

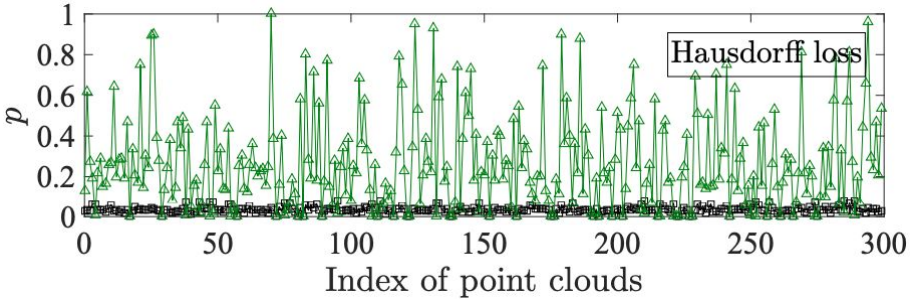
Statistic outlier removal

kNN outlier trim

- Non-differentiability

$$d_i = \frac{1}{k} \sum_{\mathbf{x}_j \in knn(\mathbf{x}_i, k)} \|\mathbf{x}_i - \mathbf{x}_j\|_2, \quad i = 1, \ldots, n.$$

$$\mathbf{X}' = \{\mathbf{x}_i | d_i < \bar{d} + \alpha \cdot \sigma\}.$$



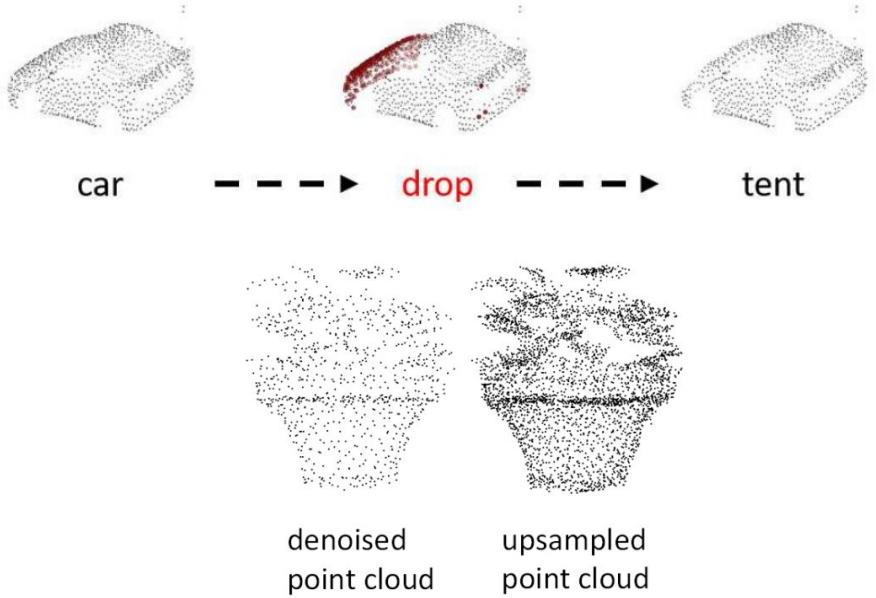
[1] Radu B. Rusu et al., Towards 3D point cloud based object maps for household environments, Robotics and Autonomous Systems 2008

Model-driven upsampler networks

Point-dropping attack

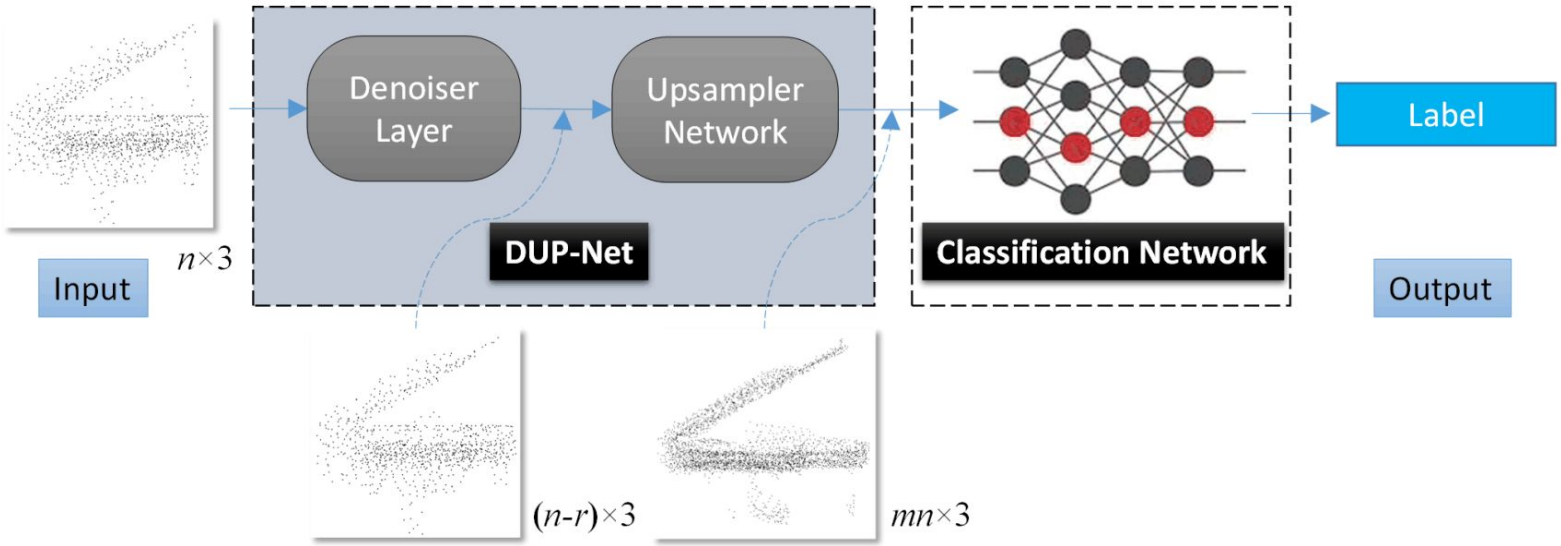
- critical subset

Upsampler network



[1] Charles Qi et al., PointNet: Deep learning on point sets for 3D classification and segmentation, CVPR 2017
[2] Tianhang Zheng et al., PointCloud saliency maps, ICCV 2019
[3] Lequan Yu et al., PU-Net: Point cloud upsampling network, CVPR 2018

Networks and loss functions



Victim models:

- PointNet
- PointNet++

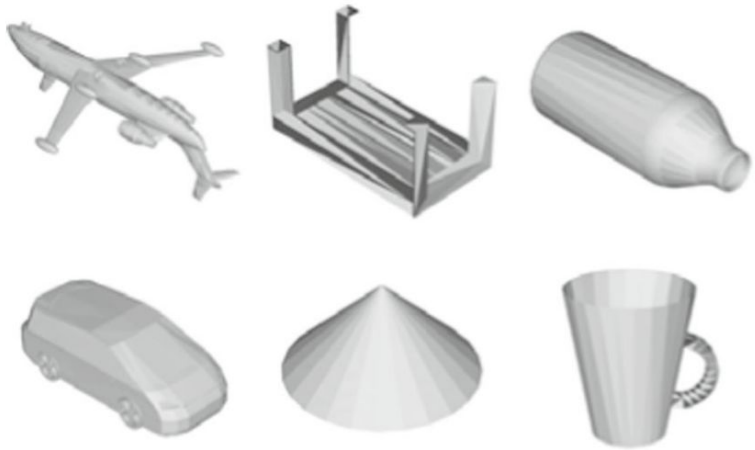
Datasets

Training on:

- ModelNet40, Visionair

Testing on:

- ModelNet40



[1] Zhirong Wu et al., 3D ShapeNets: A deep representation for volumetric shapes, CVPR 2015
[2] Lequan Yu et al., PU-Net: Point cloud upsampling network, CVPR 2018

Defense performance

Models	Target [4]	Defense (SRS)	Defense (SOR) (ours)	Defense (PU-Net) (ours)	Defense (DUP-Net) (ours)
Clean point cloud	**88.3%**	83.0%	86.5%	87.5%	86.3%
Adv (C&W + l_2 loss) [34]	0.7%	64.7%	81.4%	23.9%	**84.5%**
Adv (C&W + Hausdorff loss) [34]	12.7%	58.8%	59.8%	17.6%	**62.7%**
Adv (C&W + Chamfer loss) [34]	11.8%	59.5%	59.1%	18.0%	**61.4%**
Adv (C&W + 3 clusters) [34]	0.7%	**92.0%**	-	-	87.6%
Adv (C&W + 3 objects) [34]	2.7%	**92.4%**	-	-	68.4%
Adv (dropping 50 points) [42]	75.5%	68.1%	71.3%	**76.1%**	73.9%
Adv (dropping 100 points) [42]	63.2%	56.4%	60.0%	**67.7%**	64.3%
Adv (dropping 150 points) [42]	50.4%	45.0%	48.6%	**57.7%**	54.4%
Adv (dropping 200 points) [42]	39.1%	35.1%	36.8%	**48.1%**	43.7%

Motivation

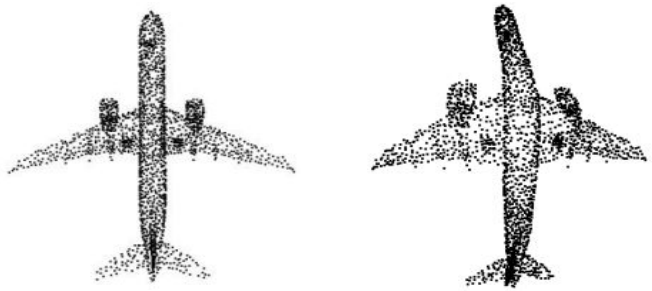
LG-GAN

Attack types

- Optimization-based
- Gradient-based
- Generative model-based

Advantages

- Outlier-less
- Flexible of attack category
- Deformation-based



clean plane

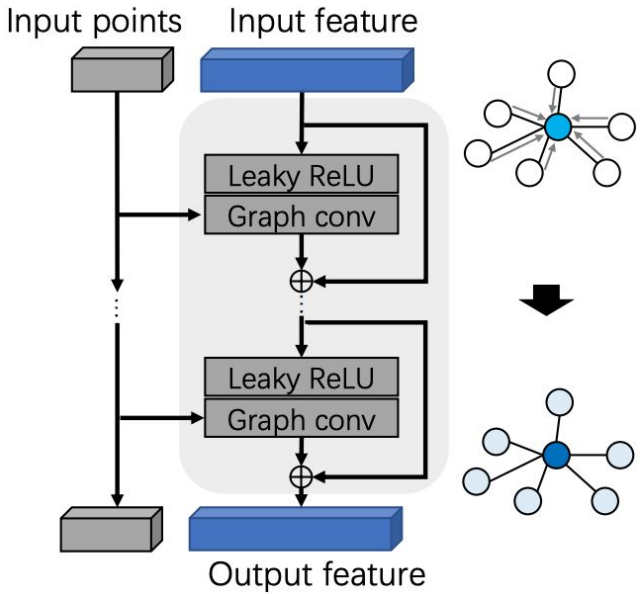
LG-GAN attack
(to lamp)

Generator and discriminator networks

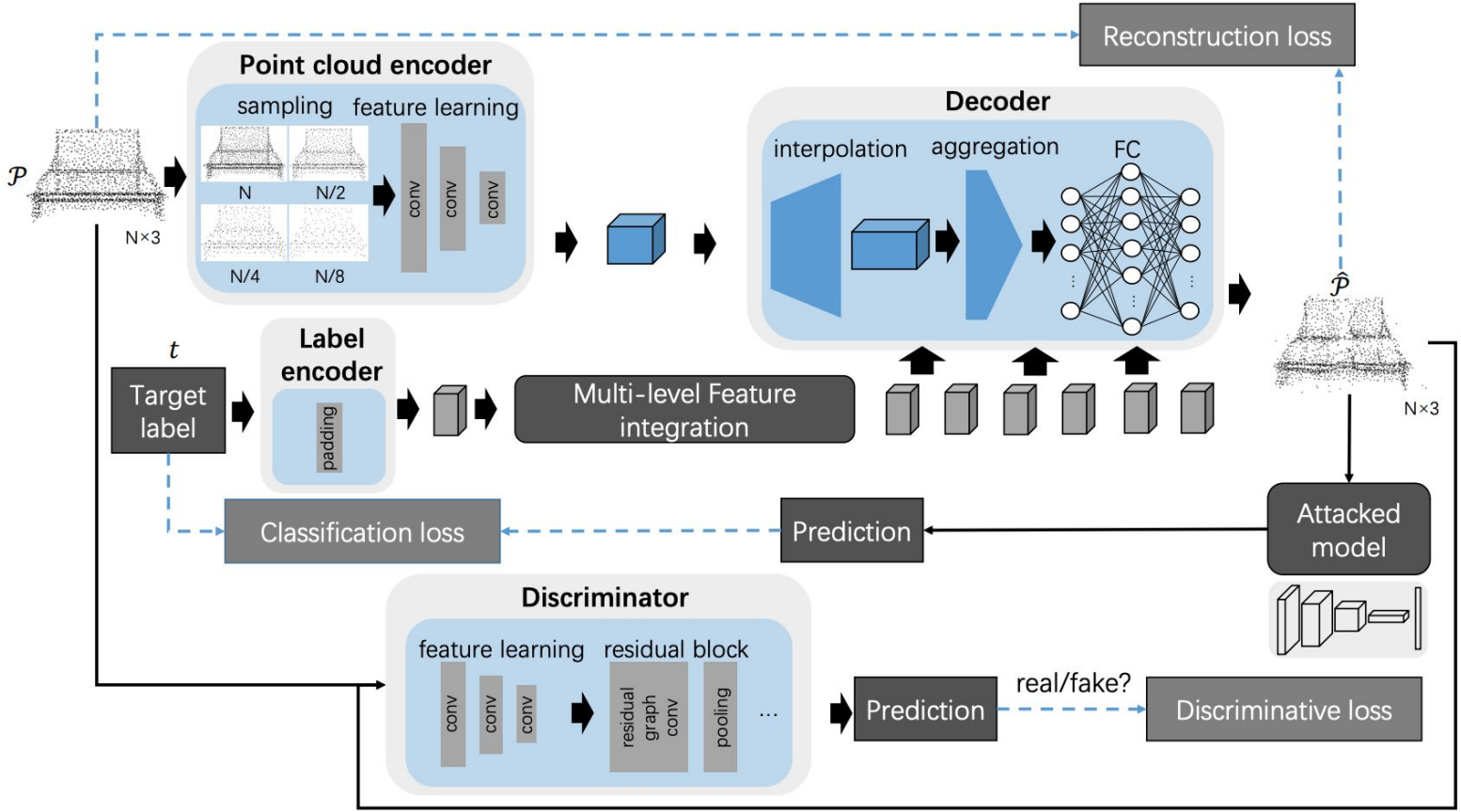
Label-guided conditional generator network

- Hierarchical point feature learning
- Feature decoding and label concatenation

Graph discriminator network



Networks and loss functions



Attack performance

	Target [5]	Defense (SRS) [43]	Defense (DUP-Net) [43]	ℓ_2 dist (meter)	Chamfer dist (meter)	Time (second)
C&W + ℓ_2 [36]	100	0	0	**0.01**	0.006	40.80
C&W + Hausdorff [36]	100	0	0	—	0.005	42.67
C&W + Chamfer [36]	100	0	0	—	**0.005**	43.73
C&W + 3 clusters [36]	94.7	2.7	0	—	0.120	52.00
C&W + 3 objects [36]	97.3	3.1	0	—	0.064	58.93
FGSM [20, 38]	12.2	5.2	2.8	0.15	0.129	0.082
IFGM [20, 38]	73.0	14.5	3.3	0.31	0.132	0.275
LG + Chamfer (ours)	96.1	75.4	13.9	0.63	0.137	0.037
single-layered LG-GAN (ours)	97.6	80.2	37.8	0.27	0.032	0.053
LG (ours)	97.1	85.0	72.0	0.25	0.028	**0.033**
LG-GAN (ours)	**98.3**	**88.8**	**84.8**	0.35	0.038	0.040

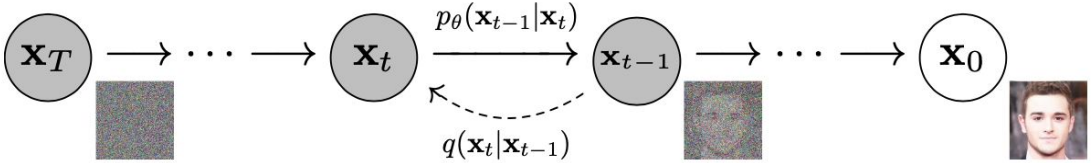
Motivation

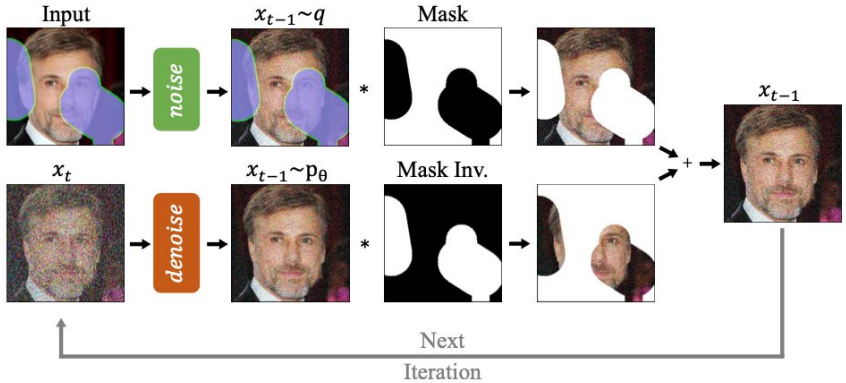
Ada3Diff

Denoising diffusion model

Defend-by-denoise

- noise density-aware (adaptive)
- distortion estimation
- Approximated by the distance of a point subset to the nearby plane

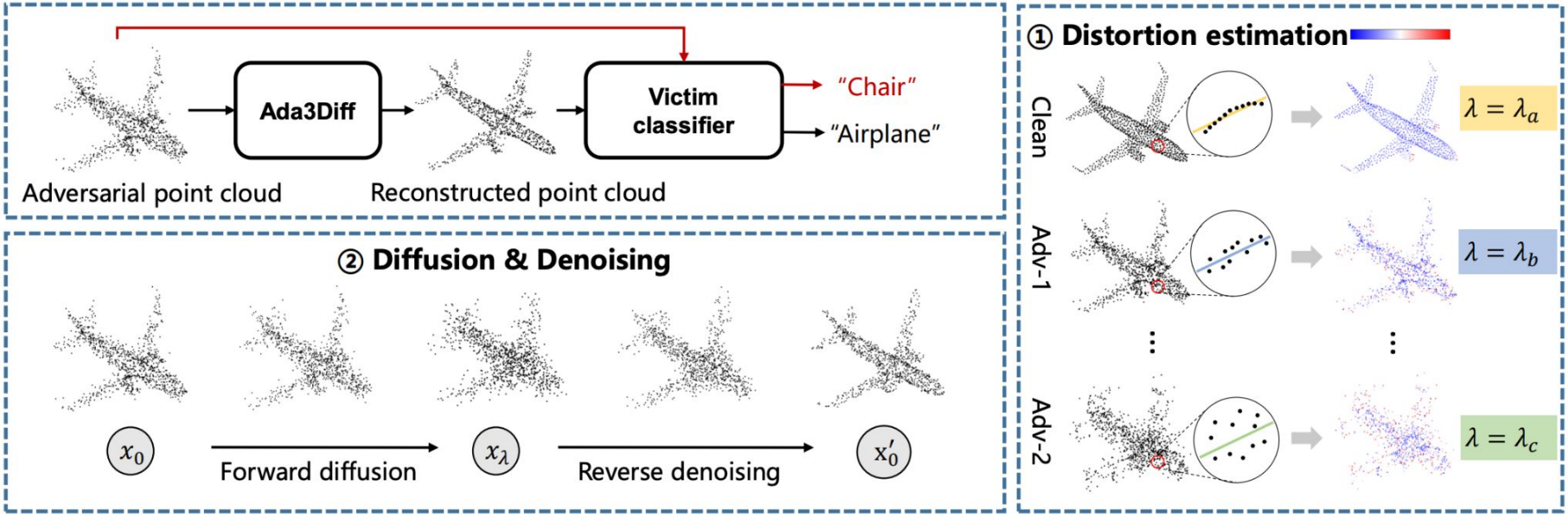


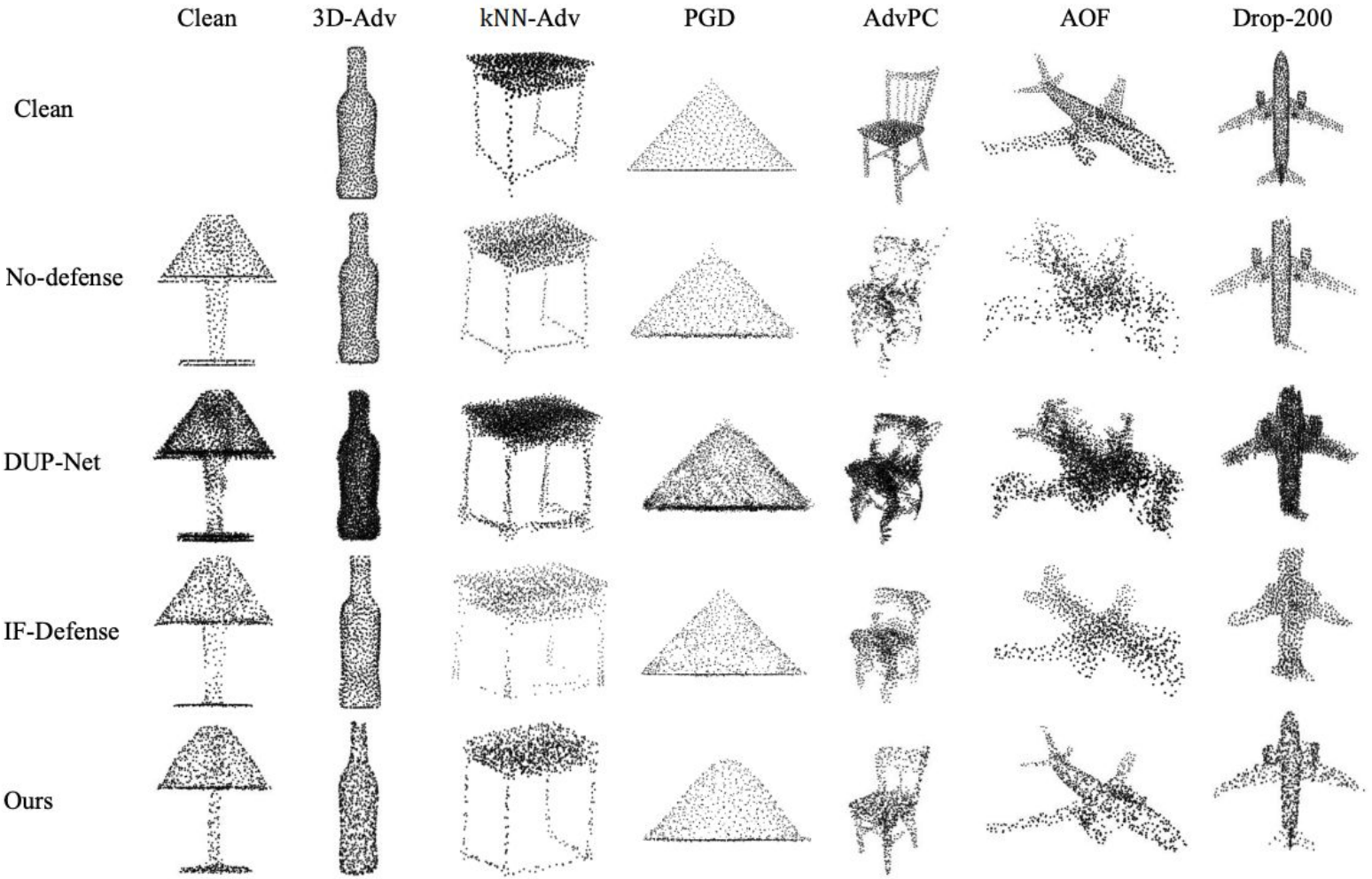


[1] Jonathan Ho et al., Denoising diffusion probabilistic models, NeurIPS 2020

[2] Andreas Lugmayr et al., RePaint: Inpainting using denoising diffusion probabilistic models, CVPR 2022

Framework


Ada3Diff
Victim classifier
"Chair"
"Airplane"
Adversarial point cloud
Reconstructed point cloud
② Diffusion & Denoising
x_0
Forward diffusion
x_λ
Reverse denoising
x'_0
① Distortion estimation
Clean
Adv-1
Adv-2
$\lambda = \lambda_a$
$\lambda = \lambda_b$
$\lambda = \lambda_c$


Clean
3D-Adv
kNN-Adv
PGD
AdvPC
AOF
Drop-200
Clean
No-defense
DUP-Net
IF-Defense
Ours

Quantitative results

Attacks	Clean	3D-Adv	kNN-Adv	PGD	AdvPC	AOF	Drop-200
CD ($\times 10^{-4}$)	0	0.59	1.20	6.35	17.2	21.3	11.0
No defense	**92.3**	0	7.21	0	0.85	0	75.2
SRS [35]	86.8	84.9	82.6	74.8	26.8	18.2	57.7
SOR [20]	91.0	**88.1**	78.2	64.0	16.9	8.27	**78.8**
DUP-Net [39]	88.8	88.0	85.3	79.5	49.4	31.6	74.2
IF-Defense [31]	86.7	86.6	85.7	83.7	62.3	47.3	77.6
Ada3Diff	88.4	87.7	**87.2**	**87.6**	**85.9**	**85.4**	78.1

Table 1. Comparison of robust classification accuracy on ModelNet40 under different attacks.

Motivation SI-Adv

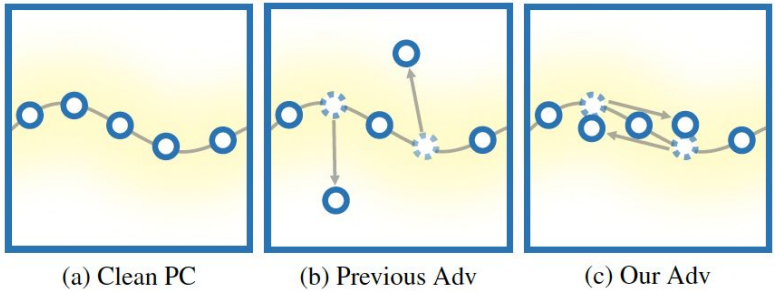
Black-box attack: more practical but challenging

Shape-invariant

- normal estimation

$$\mathcal{C}_{\boldsymbol{p}_i} = \sum_{\boldsymbol{q} \in \mathcal{N}_{\boldsymbol{p}_i}} (\boldsymbol{q} - \boldsymbol{p}_i) \otimes (\boldsymbol{q} - \boldsymbol{p}_i)$$

- Modification along the surface
- Max-margin logit loss



(a) Clean PC (b) Previous Adv (c) Our Adv

$$\mathcal{L}(\mathcal{P}, t; \theta_w) = \max\left([\mathcal{H}_w(\mathcal{P})]_t - \max_{j \neq t}[\mathcal{H}_w(\mathcal{P})]_j, 0\right) \qquad \mathcal{P} = \{\boldsymbol{R}_i^\top \boldsymbol{p}_i' - \boldsymbol{T}_i\}_{i=1}^N$$

[1] Hugues Hoppe et al., Surface reconstruction from unorganized points, SIGGRAPH 1992
[2] Nicholas Carlini and David Wagner, Towards evaluating the robustness of neural networks, S&P 2017

Algorithm

Algorithm 1: Shape-invariant Query-based Attack

Input: point-cloud input $(\mathcal{P}, l)$, black-box model $\mathcal{H}_b$, surrogate model $\mathcal{H}_w$ and step size ϵ.
Output: adversarial point cloud $\tilde{\mathcal{P}}$

Initialize the perturbation $\delta = 0$
Initialize the prediction pool $\mathbf{p}_c = [\mathcal{H}_b(\mathcal{P})]_c$
Transform $\mathcal{P}$ to $\mathcal{P}'$ ◁ Eq.(8)
Compute gradient map $\mathcal{G}$ of $\mathcal{P}'$ on $\mathcal{H}_w$ ◁ Eq.(10)
Compute and rank sensitivity map $\mathcal{S}$ ◁ Eq.(13)
while $l = \arg\max_c \mathbf{p}_c$ ***and*** $\mathcal{S} \neq \varnothing$ **do**
    Pick top ranked $q \in \mathcal{S}$ and $\mathcal{S} = \mathcal{S} \backslash \{q\}$
    Get its direction $\theta = \arctan(g_{i2}/g_{i1})$
    Compute basis $q = q \cdot (\cos\theta, \sin\theta, 0)$
    for $\alpha \in \{\epsilon, -\epsilon\}$ **do**
        Reverse $\mathcal{P}' + \delta + \alpha q$ to $\mathcal{P}_{inp}$ ◁ Eq.(8)
        Get prediction $\mathbf{p}'_c = [\mathcal{H}_b(\mathcal{P}_{inp})]_c$
        if $\mathbf{p}'_l < \mathbf{p}_l$ **then**
            Update $\delta = \delta + \alpha q$
            New prediction pool $\mathbf{p}_c = \mathbf{p}'_c$
        end
    end
end
return $\tilde{\mathcal{P}} = \mathcal{P}_{inp}$

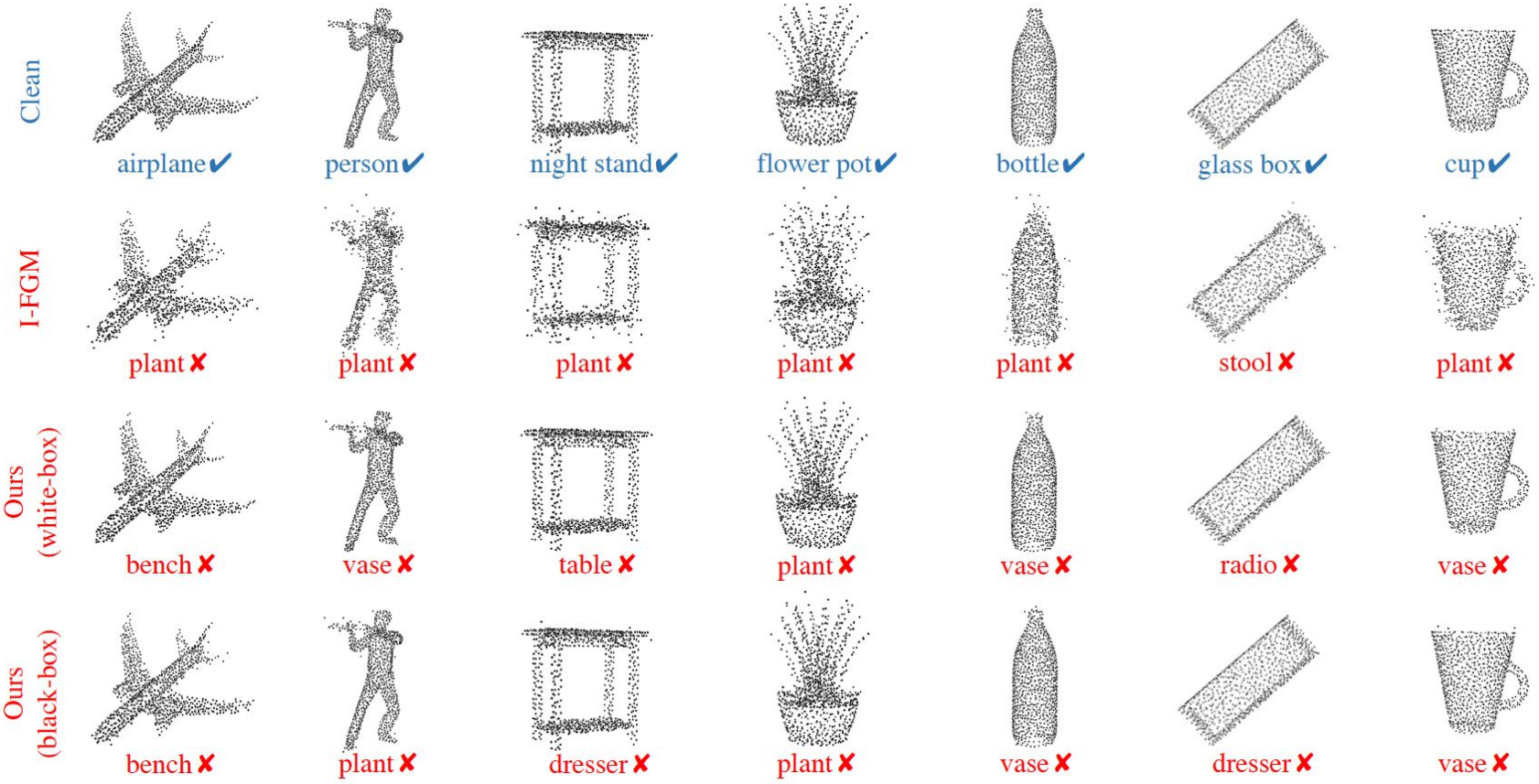
Experimental setup

Attacked models

- PoinetNet, PointNet++ (MSG), DGCNN, PAConv, SimpleView and CurveNet

[1] Charles Qi et al., PointNet: Deep learning on point sets for 3D classification and segmentation, CVPR 2017
[2] Charles Qi et al., PointNet++: Deep hierarchical feature learning on point sets in a metric space, NeurIPS 2017
[3] Yue Wang et al., Dynamic graph CNN for learning on point clouds, SIGGRAPH 2019
[4] PAConv: Position adaptive convolution with dynamic kernel assembling on point clouds, CVPR 2021
[5] Ankit Goyal et al., Revisiting point cloud shape classification with a simple and effective baseline, ICML 2021
[6] Tiange Xiang et al., Walk in the cloud: Learning curves for point clouds shape analysis, ICCV 2021

Quantitative results



Qualitative results

Attack	Defense	PointNet [6]				DGCNN [31]				CurveNet [36]			
		ASR↑ (%)	CD↓ (10^{-4})	HD↓ (10^{-2})	A.T↓ (s)	ASR↑ (%)	CD↓ (10^{-4})	HD↓ (10^{-2})	A.T↓ (s)	ASR↑ (%)	CD↓ (10^{-4})	HD↓ (10^{-2})	A.T↓ (s)
I-FGM [15]	-	100.0	6.96	3.04	1.17	99.5	16.85	2.39	2.02	100.0	15.61	2.89	**10.77**
MI-FGM [9]	-	99.5	35.99	4.16	1.31	95.9	120.10	4.99	2.14	99.0	119.17	5.10	10.88
PGD [21]	-	**100.0**	7.00	3.05	**1.17**	99.4	16.77	2.40	**2.02**	100.0	15.32	2.86	10.78
3d-Adv [34]	-	99.9	3.25	2.11	4.94	**100.0**	10.12	2.46	18.73	100.0	7.48	3.51	116.24
AdvPC [17]	-	99.8	16.57	3.43	2.90	98.8	14.35	1.48	8.10	99.9	18.36	2.78	64.76
Ours	-	99.8	**2.15**	**2.04**	1.32	99.9	**6.33**	**1.27**	3.87	**100.0**	**6.25**	**1.99**	21.53

Table 1. Quantitative comparison between our white-box shape-invariant attack and existing white-box attacks in terms of attack success rate (ASR), Chamfer distance (CD), Hausdorff distance (HD) and average time budget for each adversarial point cloud generation (A.T), where CD is multiplied by 10^4 and HD is multiplied by 10^2 for better comparison.

Takeaway

- Point-cloud defense
  Feature-aware
- Point-could attack
  Attack with addtional inputs, generative model, scene datasets
- Mesh
  Mesh attack, neural network for mesh steganalysis, other security-related 3D problems