Towards a 3D Transfer-based Black-box Attack via Critical Feature Guidance

Supplementary Material

In the supplementary material, we provide additional experiments and results about our proposed CFG. First, more experimental results are provided by choosing more models as source models (Section A). Second, we show more visualization samples of our observations (Section B). Finally, we upload code as part of our supplementary material, and provide some brief explanations of the code file (Section C).

A. Additional Experiments

A.1. Transfer-based Attack on state-of-the-art models with multiple source models on ModelNet40

To better validate the performance of our attack method, we use multiple models as source models to generate the adversarial point clouds. After selecting the source model, we apply our attack algorithm to it, aiming at generating adversarial point clouds. The generated adversarial point clouds would be used to attack state-of-the-art models (e.g., PointConv [1], PointCNN [2], CurveNet [3], PCT [4], PT [5] and Point-PN [6]).

As shown in Table 1, PointNet, PointNet++ (MSG), PointNet++ (MSG), and DGCNN as the source models to generate the adversarial point clouds. Overall, our attack method maintains the leading performance relative to the comparative attack methods against these state-of-the-art models. We can find that for these state-of-the-art models, the adversarial point clouds generated by traditional 3D-Adv and GeoA3 attack methods have struggled to exhibit transferability, and the attack success rate has dropped to single digits. In such a difficult scenario, our attack method can still remain a strong threat.

A.2. Transfer-based attack on classical models and state-of-the-art models with multiple source models on ScanObjectNN

To better validate the generality of our attack method, we do the same experiments under the harder real dataset ScanObjectNN. As shown in Table 2, 3, we selected 3D-adv, GeoA3, and PF-Attack as baseline attack methods, and in general, our method outperforms all baseline attack methods in improving the transferability of the adversarial point clouds for all these classical and state-of-the-art classification models.

B. Visualization Results of Point Cloud Importance Maps

To better validate our observations, as shown in Fig. 1 and Fig. 2, we provide additional visualization results on ModelNet40 and ScanObjectNN datasets, respectively.

C. Code

The code we provide is complete, and you can run it directly after configuring your environment. The code file contains some core runtime files, i.e., first is the attack code (attack.py), you can run this code to generate the corresponding adversarial point clouds; second is the evaluation code (evaluate.py), you can run this code to evaluate the effect of the adversarial point clouds among different models; and finally is the defense code (defense.py), which you can run to evaluate how well the adversarial point clouds resists the defense.

References

[1] Wenxuan Wu, Zhongang Qi, and Li Fuxin. Pointconv: Deep convolutional networks on 3d point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9621–9630, 2019. 1

[2] Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. Pointcnn: Convolution on x-transformed points. *Advances in Neural Information Processing Systems*, pages 820–830, 2018. 1

[3] Tiange Xiang, Chaoyi Zhang, Yang Song, Jianhui Yu, and Weidong Cai. Walk in the cloud: Learning curves for point clouds shape analysis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 915–924, 2021. 1

[4] Meng-Hao Guo, Jun-Xiong Cai, Zheng-Ning Liu, Tai-Jiang Mu, Ralph R Martin, and Shi-Min Hu. Pct: Point cloud transformer. *Computational Visual Media*, 7:187–199, 2021. 1

[5] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip HS Torr, and Vladlen Koltun. Point transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16259–16268, 2021. 1

[6] Renrui Zhang, Liuhui Wang, Yali Wang, Peng Gao, Hongsheng Li, and Jianbo Shi. Parameter is not all you need: Starting from non-parametric networks for 3d point cloud analysis. *arXiv preprint arXiv:2303.08134*, 2023. 1

[7] Chong Xiang, Charles R Qi, and Bo Li. Generating 3d adversarial point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9136–9144, 2019. 2

[8] Yuxin Wen, Jiehong Lin, Ke Chen, CL Philip Chen, and Kui Jia. Geometry-aware generation of adversarial point clouds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(6):2984–2999, 2020. 2

[9] Bangyan He, Jian Liu, Yiming Li, Siyuan Liang, Jingzhi Li, Xiaojun Jia, and Xiaochun Cao. Generating transferable 3d adversarial point cloud via random perturbation factorization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 764–772, 2023. 2

Table 1. Transfer-based attack on state-of-the-art models on ModelNet40. Measure performance in terms of attack success rate (%). Adversarial point clouds are generated on multiple source models.

Source Model	Attack Method	$\epsilon = 0.18$						$\epsilon = 0.45$					
		PointConv	PointCNN	CurveNet	PCT	PT	Point-PN	PointConv	PointCNN	CurveNet	PCT	PT	Point-PN
PointNet	3D-Adv [7]	0.8	1.0	1.8	0	0.2	1.2	0.8	0.1	1.8	0	1.0	0.7
	GeoA3 [8]	10.3	2.1	5.5	3.3	18.3	5.5	14.9	5.5	6.7	7.5	19.6	10.2
	PF-Attack [9]	50.3	29.5	22.7	23.4	49.9	26.1	53.2	33.4	24.8	28.7	52.3	40.8
	Ours	**68.1**	**32.7**	**29.2**	**30.4**	**65.6**	**42.5**	**81.3**	**43.4**	**47.5**	**48.6**	**84.7**	**61.5**
PointNet++ (MSG)	3D-Adv [7]	1.3	1.0	2.6	1.8	1.4	5.7	1.3	0.5	2.6	1.8	1.8	8.7
	GeoA3 [8]	3.8	4.4	0.9	0	7.2	6.9	4.1	4.7	1.7	0	7.1	4.2
	PF-Attack [9]	50.4	**22.2**	23.1	16.6	50.0	28.9	52.6	25.5	22.9	16.0	57.7	35.3
	Ours	**56.9**	21.0	**26.5**	**26.4**	**66.0**	**35.1**	**71.6**	**26.2**	**46.8**	**47.5**	**77.6**	**61.3**
PointNet++ (SSG)	3D-Adv [7]	1.2	1.3	0.8	0	0.8	4.2	1.7	1.3	0.8	0	2.9	5.5
	GeoA3 [8]	4.1	6.0	3.4	0.4	2.5	7.6	3.3	5.6	2.9	0.4	3.8	7.2
	PF-Attack [9]	38.8	21.8	20.2	17.2	43.3	26.7	43.6	24.0	20.3	19.4	50.2	33.7
	Ours	**61.8**	**25.5**	**31.2**	**27.5**	**73.3**	**39.0**	**80.6**	**39.6**	**49.5**	**50.5**	**84.8**	**58.4**
DGCNN	3D-Adv [7]	2.9	3.0	5.5	0.4	4.6	8.9	2.9	3.4	5.9	1.3	8.3	5.9
	GeoA3 [8]	9.1	8.1	3.8	0	27.5	10.6	11.2	7.7	3.8	0	27.5	11.0
	PF-Attack [9]	62.4	29.7	23.6	26.7	61.7	34.9	69.9	34.1	35.7	37.8	74.8	52.9
	Ours	**70.1**	**45.2**	**35.4**	**33.5**	**75.5**	**48.4**	**76.5**	**62.1**	**47.8**	**44.4**	**85.1**	**62.9**

Table 2. Transfer-based attack on classical models on ScanObjectNN. Measure performance in terms of attack success rate (%). The number in bold indicates the best.

Source Model	Attack Method	$\epsilon = 0.18$				$\epsilon = 0.45$			
		PointNet	PointNet++ (MSG)	PointNet++ (SSG)	DGCNN	PointNet	PointNet++ (MSG)	PointNet++ (SSG)	DGCNN
PointNet	3D-Adv [7]	100	7.5	7.5	9.0	100	7.3	7.5	9.0
	GeoA3 [8]	100	18.3	20.1	18.2	100	17.8	19.5	17.8
	PF-Attack [9]	100	50.4	54.6	40.3	100	51.4	54.2	40.8
	Ours	100	**66.5**	**68.0**	**58.3**	100	**69.1**	**70.3**	**58.4**
PointNet++ (MSG)	3D-Adv [7]	5.4	100	33.0	18.0	4.8	100	33.2	16.9
	GeoA3 [8]	11.1	100	41.9	27.3	11.0	100	40.8	26.6
	PF-Attack [9]	38.3	100	63.2	44.8	40.0	100	64.3	44.7
	Ours	**43.2**	100	**73.9**	**52.9**	**47.6**	100	**74.4**	**55.8**
PointNet++ (SSG)	3D-Adv [7]	3.7	22.1	100	15.0	4.0	22.1	100	15.1
	GeoA3 [8]	6.3	29.8	100	23.1	6.1	29.1	100	22.7
	PF-Attack [9]	37.7	55.2	100	42.5	39.5	55.0	100	42.5
	Ours	**40.9**	**69.7**	100	**56.0**	**45.2**	**71.4**	100	**58.3**
DGCNN	3D-Adv [7]	5.3	14.3	16.5	100	5.3	14.3	16.6	100
	GeoA3 [8]	6.4	19.8	24.9	100	6.0	19.9	24.7	100
	PF-Attack [9]	38.6	53.5	57.7	100	40.5	55.5	58.9	100
	Ours	**41.9**	**60.1**	**63.9**	100	**49.9**	**67.1**	**68.1**	100

Table 3. Transfer-based attack on state-of-the-art models on ScanObjectNN. Measure performance in terms of attack success rate (%). Adversarial point clouds are generated on multiple source models.

Source Model	Attack Method	$\epsilon = 0.18$						$\epsilon = 0.45$					
		PointConv	PointCNN	CurveNet	PCT	PT	Point-PN	PointConv	PointCNN	CurveNet	PCT	PT	Point-PN
PointNet	3D-Adv [7]	9.3	10.6	10.0	8.6	12.8	27.3	9.2	2.1	10.0	9.0	13.2	27.2
	GeoA3 [8]	21.7	20.6	18.8	17.3	27.3	40.4	21.4	20.8	18.3	16.7	27.0	41.9
	PF-Attack [9]	50.1	45.0	47.1	45.6	51.9	68.8	47.9	45.5	46.4	47.1	51.8	69.5
	Ours	**61.7**	**50.9**	**64.5**	**63.6**	**67.0**	**80.0**	**63.8**	**53.7**	**66.5**	**66.8**	**68.9**	**81.8**
PointNet++ (MSG)	3D-Adv [7]	17.4	14.2	17.7	16.4	25.8	41.6	17.2	14.6	17.6	16.2	25.0	41.3
	GeoA3 [8]	30.1	24.7	23.2	24.5	37.3	48.8	29.6	22.9	22.9	24.0	37.0	50.5
	PF-Attack [9]	54.7	48.3	51.7	50.6	56.0	72.1	54.9	49.4	52.6	52.3	55.7	72.0
	Ours	**60.5**	**48.8**	**57.9**	**59.7**	**66.6**	**75.9**	**63.1**	**50.3**	**59.4**	**63.3**	**67.2**	**76.0**
PointNet++ (SSG)	3D-Adv [7]	15.4	12.7	15.6	14.6	22.5	38.2	15.0	12.4	15.3	14.1	22.2	38.8
	GeoA3 [8]	26.2	20.3	19.0	21.0	31.3	48.1	26.5	20.8	19.0	21.6	30.3	47.8
	PF-Attack [9]	52.3	46.3	49.7	50.8	53.7	70.3	52.1	46.9	49.3	50.8	55.5	71.9
	Ours	**61.3**	**47.7**	**55.5**	**61.1**	**64.8**	**73.5**	**63.1**	**50.1**	**57.4**	**62.5**	**67.8**	**74.1**
DGCNN	3D-Adv [7]	24.1	16.8	17.2	17.8	28.3	46.3	24.0	15.6	17.2	17.6	28.4	45.5
	GeoA3 [8]	29.4	23.2	20.0	21.2	32.6	50.8	29.6	21.5	19.9	21.5	33.6	52.0
	PF-Attack [9]	54.4	49.8	50.3	50.4	57.7	70.3	54.6	51.4	50.2	52.9	57.6	70.9
	Ours	**60.4**	**52.8**	**58.9**	**62.1**	**63.5**	**75.8**	**64.7**	**56.6**	**63.9**	**65.9**	**67.8**	**76.7**

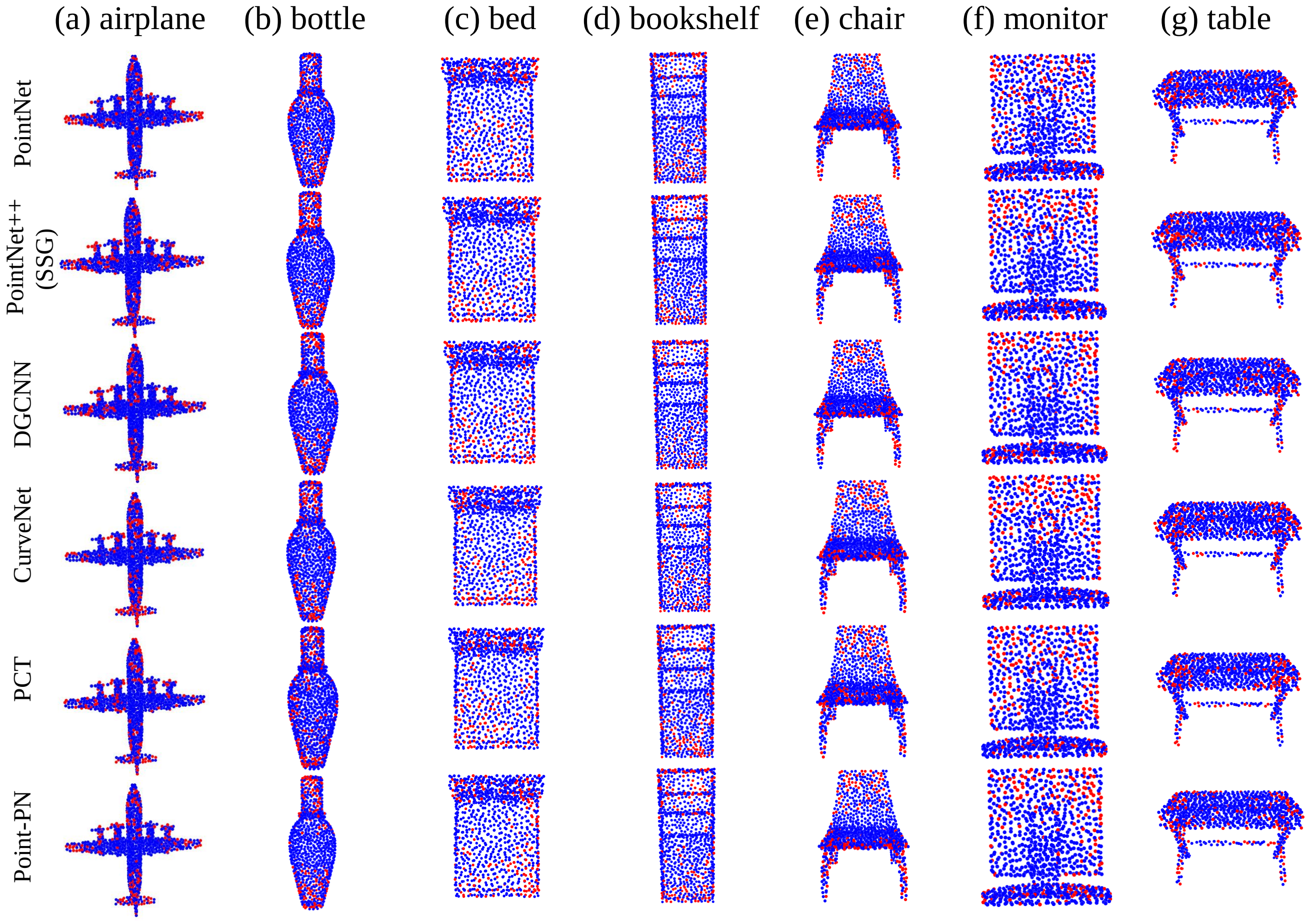

Figure 1. Visualization results of point cloud importance maps obtained from different 3D models on ModelNet40. Red color represents important and blue color represents unimportant.

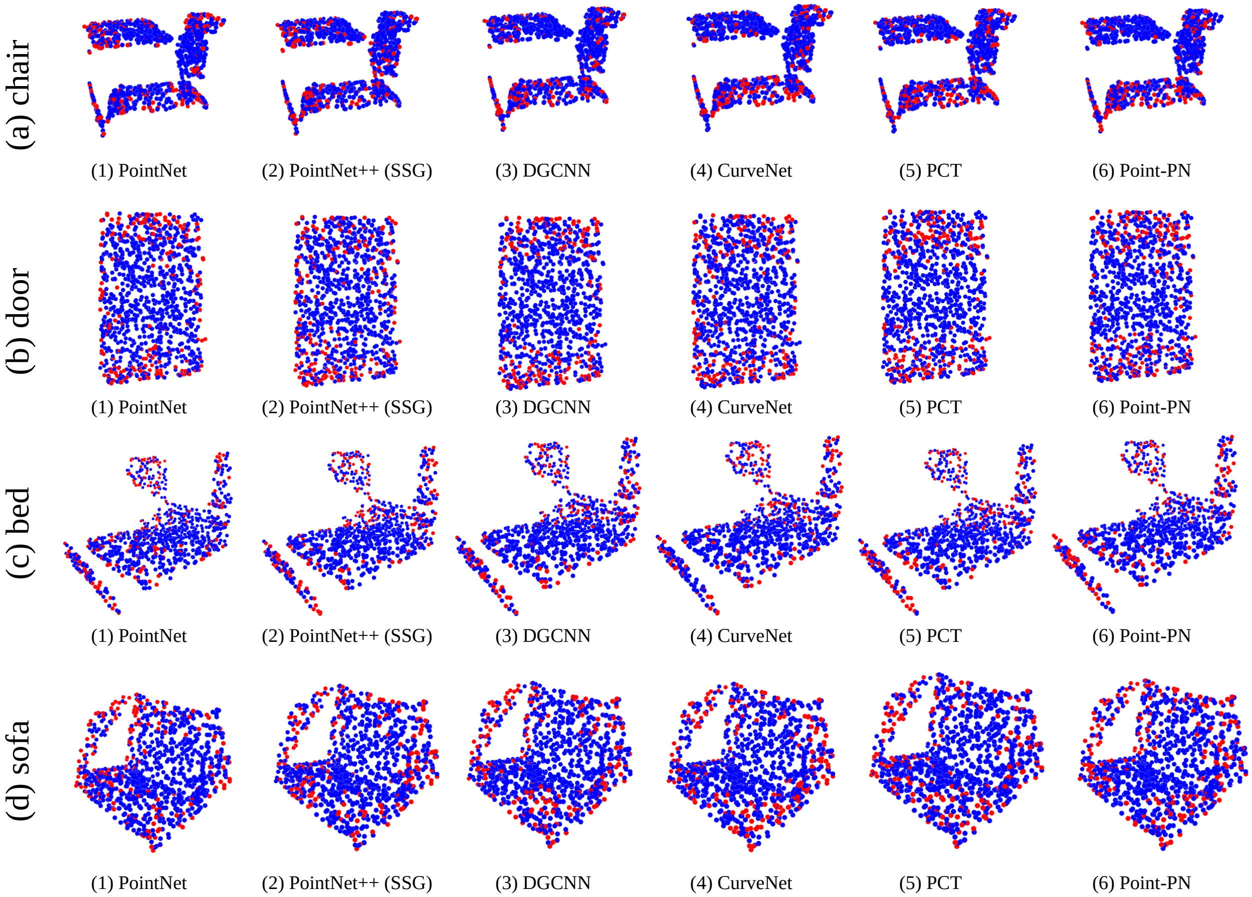

Figure 2. Visualization results of point cloud importance maps obtained from different 3D models on ScanObjectNN. The red color represents important, and blue color represents unimportant.