Supplementary Material for "Physical 3D Adversarial Attacks against Monocular Depth Estimation in Autonomous Driving"

Table 1. Weather parameter settings in Carla to simulate cloudy, sunny, rainy, and foggy.

Parameters	Weather conditions			
	cloudy	sunny	rainy	foggy
sun_azimuth_angle	300	300	-1	300
sun_altitude_angle	45	90	45	45
cloudiness	30	10	100	100
precipitation	0	0	100	0
precipitation_deposits	0	0	90	0
fog_density	1	0	3	15

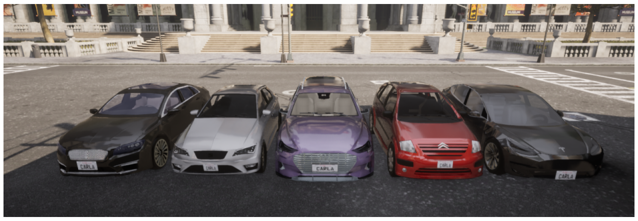


Figure 1. Five types of cars used in our evaluations.

A. Implementation Details

A.1. Experimental Settings

Scenes. We generate the datasets for training and evaluating attack methods using the Carla simulator. For autonomous driving scenarios, we selected a variety of background environments, including urban roads, highways, country roads, etc. To improve attack robustness in bad weather conditions, we choose four kinds of weather: foggy, rainy, sunny, and cloudy. The parameter settings for controlling the weather in CARLA are shown in Table 1.

Cars. We select multiple types of cars in the Carla simulator. Lincoln MKZ2017, Seat Leon, Audi Etron, and Citreon C3 are four types of cars used for effectiveness and robustness evaluation. Additionally, Tesla Model 3 is used for ablation studies. All these cars are visualized in Figure 1.

Camera settings. For attack robustness, we take images with the RGB camera sensor in Carla at a random angle of 0-360°, within a 3-15m distance range and a 0.5-2m height range, as shown in Figure 2. We take 8400 photorealistic images for attack texture generation (210 scenes × 4 weather conditions × 10 camera positions). For a single texture evaluation, we take a total of 6124 images (discrete part: 30 scenes × 4 weather conditions × 10 camera positions × 4 cars, and continuous part: 4 weather conditions × 331 frames).

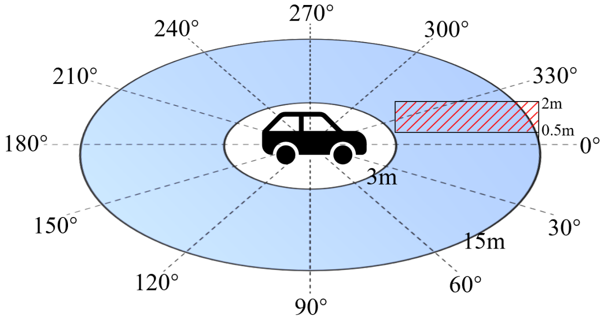


Figure 2. Camera position settings in Carla.

Table 2. Transformations distribution.

Transformations	Parameters	Remark
Rotation	$\pm 20°$	Camera Simulation
Noise	± 0.1	Random Noise
Brightness	± 0.2	Illumination
Contrast	$[0.9, 1.1]$	Camera Parameters
Scale	$[0.25, 1.25]$	Distance/Resize

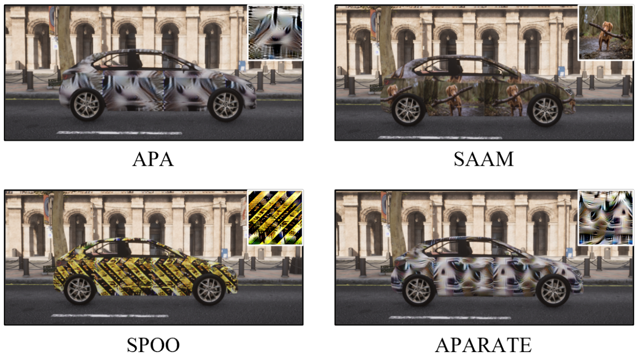

APA
SAAM
SPOO
APARATE

Figure 3. Adversarial cars with their initial patches (top right) in Carla.

A.2. Compared Attacks

We provide detailed parameter settings of compared methods: APA [6], SPOO [1], APARATE [3], SAAM [4].

Patch-oriented attacks. APA and SAAM are patch-oriented methods that fool MDE models into estimating an incorrect depth for the regions where the patterns are placed, independent of objects. APA is the first to attack MDE models in real scenes. SAAM leverages a semantic constraint to ensure the stealthiness of the generated adversarial patch. Data augmentation techniques used in the

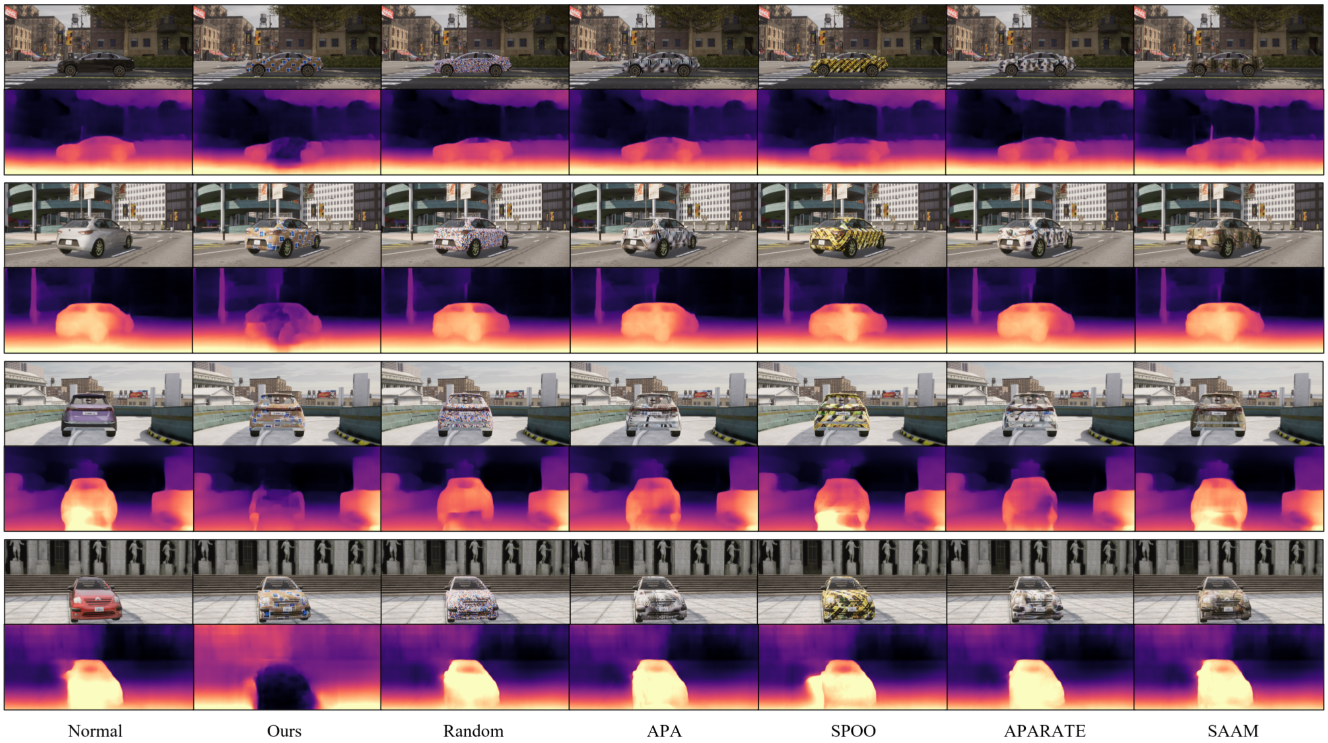


Figure 4. Comparisons of different attacks with various target vehicles and camera poses. The camera distance from the first row to the last is 9m, 7m, 5m, and 3m, respectively, and the camera viewpoints are lateral, lateral rear, rear, and front, respectively.

above methods mainly include rotation, brightness, contrast, etc. Detailed parameter settings are listed in Table 2.

Object-oriented attacks. SPOO and APARATE are object-oriented methods expected to affect the depth estimation of the whole target region. SPOO is the first to consider stealth for attacking MDE models in real scenes. APARATE designs a penalized loss function to enlarge the affected region. We use the same transformations in Table 2. For the above four attacks, initial patches with the size of 512×512 retrained on Monodepth2 [2] and the corresponding adversarial images in Carla are shown in Figure 3.

B. Evaluation Details

Various camera Poses and target vehicles. We evaluate attack methods on Monodepth2 with various camera poses, for distance and rotation. Figure 4 shows the attack performance of different attack methods on four target vehicles, at various distances and camera viewpoints. Our method outperforms the remaining attack methods on different target vehicles.

Various weather conditions. We evaluate attack methods on Monodepth2 under various weather conditions. Figure 5 shows the attack performance on the Audi Etron under four weather conditions. Our method outperforms the remaining attack methods under various weather conditions.

Various target objects. We evaluate attack methods on pedestrians, trucks, and buses against Monodepth2. Figure 6 shows the attack performance under default weather conditions. Considering the different sizes of the target objects, the texture size is correspondingly scaled by different factors.

Real-world evaluations. We provide detailed samples to show our camouflage texture performance in the real world. Figure 7 shows that the adversarial car can deceive Monodepth2, regardless of viewpoints.

Impact on a downstream task. Following the experiments in [1], we evaluate the impact of our attack on a point cloud-based 3D object detection model, PointPillars [5], and use Detection Rate as the metric to evaluate our method on 3D object detection. We collect 8 videos with 331 frames in Carla, where the observer drives sideways from the target vehicle, simulating the scenario of an autonomous vehicle. The target vehicles are covered by normal or camouflage textures under the four preset weather mentioned above. As shown in Figure 8, the benign vehicle can be correctly detected with a 3D bounding box. In contrast, the pseudo-Lidar point cloud of the camouflaged vehicles is severely distorted, and camouflaged vehicles are not detected. Table 3 shows the detection rate of the normal vehicle and

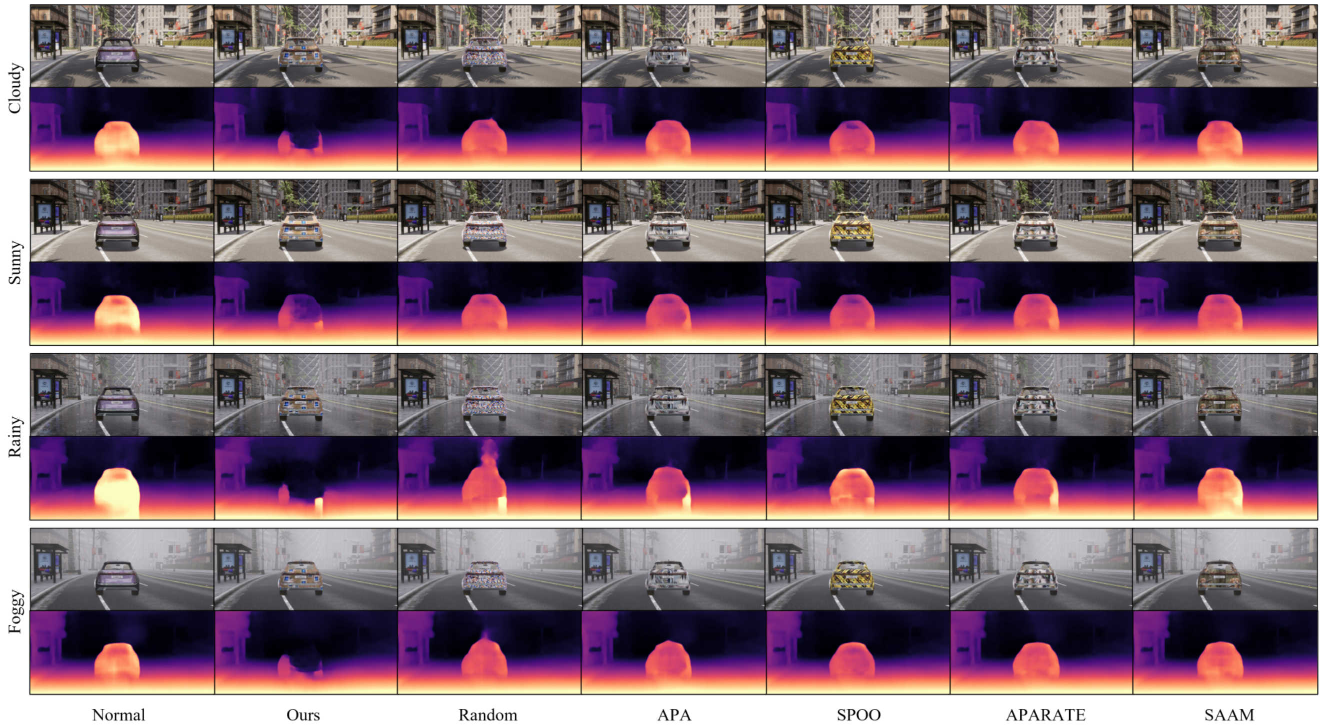


Figure 5. Comparisons of different attacks with various weather conditions: cloudy, sunny, rainy, and foggy. The camera viewpoint is the rear and the distance is 6m.

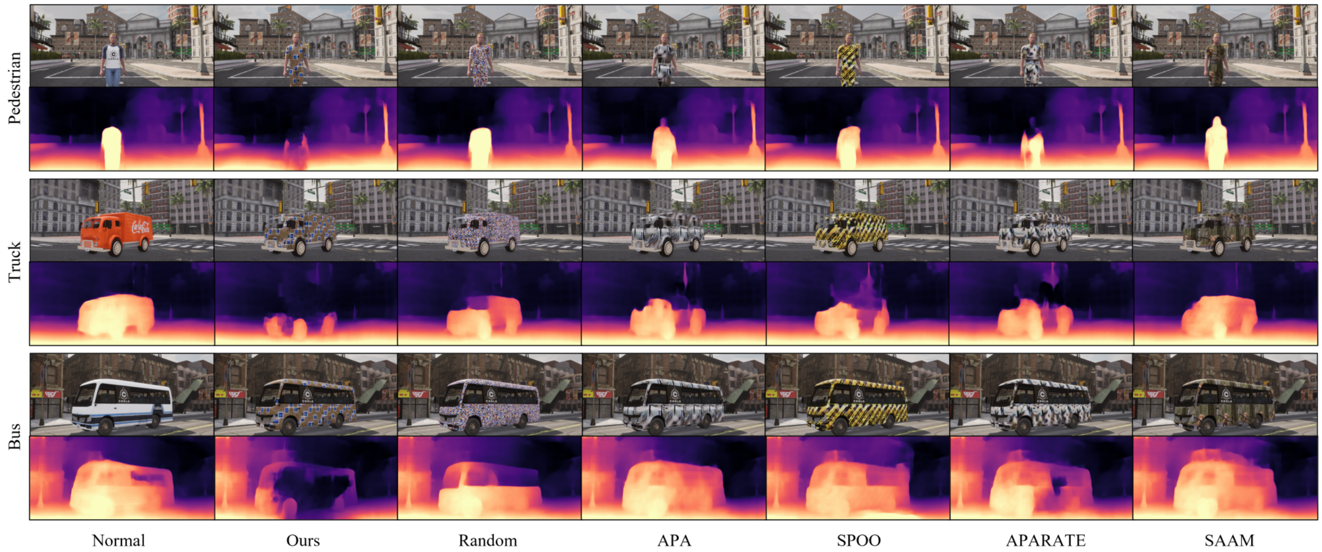


Figure 6. Comparisons of different attacks with various target objects: pedestrian, truck, and bus. Considering the different sizes of the target objects, the texture size is scaled by 0.5, 1.3, and 1.8, respectively.

camouflaged vehicles detected under different weather conditions.

References

[1] Zhiyuan Cheng, James Liang, Hongjun Choi, Guanhong Tao, Zhiwen Cao, Dongfang Liu, and Xiangyu Zhang. Physical attack on monocular depth estimation with optimal adversarial patches. In *Computer Vision – ECCV 2022*, pages 514–532,

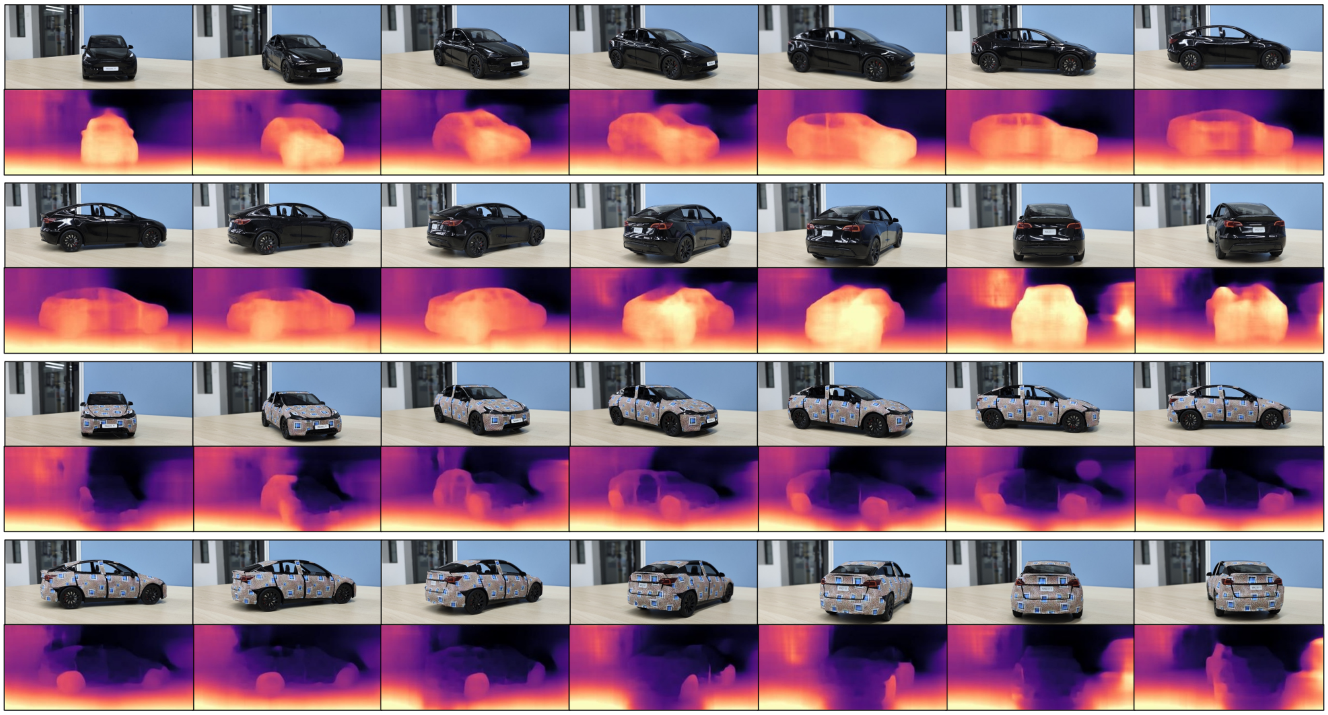
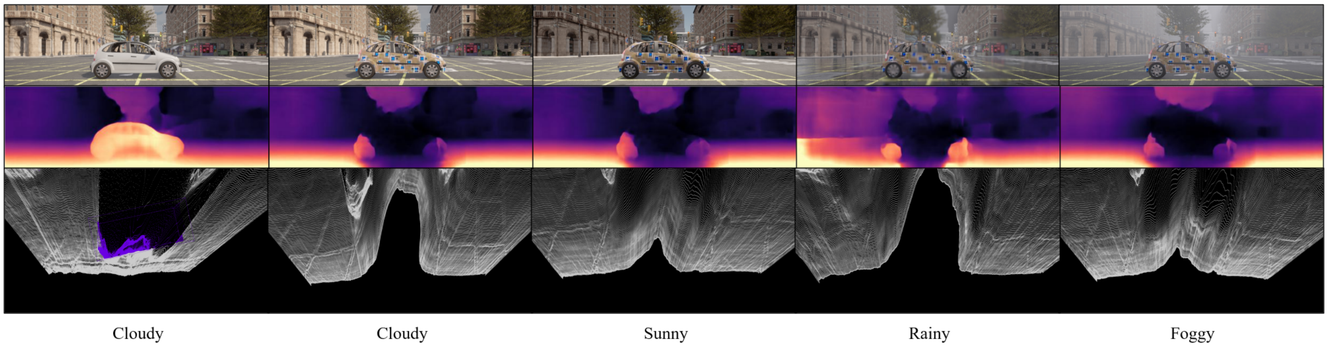


Figure 7. Real-world evaluations with normal and our adversarial scaled car models.



Cloudy

Cloudy

Sunny

Rainy

Foggy

Figure 8. Our attack against 3D object detection with various weather conditions.

Table 3. Our attack against 3D object detection with various weather conditions. Values are detection rate.

Methods	Weather conditions			
	cloudy	sunny	rainy	foggy
Normal	82.18%	86.40%	80.66%	78.25%
Ours	3.93%	5.74%	4.23%	2.72%

Cham, 2022. Springer Nature Switzerland. 1, 2

[2] Clément Godard, Oisin Mac Aodha, Michael Firman, and Gabriel J. Brostow. Digging into self-supervised monocular depth prediction. 2019. 2

[3] Amira Guesmi, Muhammad Abdullah Hanif, Ihsen Alouani, and Muhammad Shafique. Aparate: Adaptive adversarial patch for cnn-based monocular depth estimation for autonomous navigation, 2023. 1

[4] Amira Guesmi, Muhammad Abdullah Hanif, Bassem Ouni, and Muhammad Shafique. Saam: Stealthy adversarial attack on monoculor depth estimation. *ArXiv*, abs/2308.03108, 2023. 1

[5] Alex H. Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders for object detection from point clouds. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12689–12697, 2018. 2

[6] Koichiro Yamanaka, Ryutaroh Matsumoto, Keita Takahashi, and Toshiaki Fujii. Adversarial patch attacks on monocular

depth estimation networks. *IEEE Access*, 8:179094–179104, 2020. 1