3D Gaussian Splatting Driven Multi-View Robust Physical Adversarial Camouflage Generation

Supplementary Material

In the manuscript, we detail the proposed PGA attack framework and provide a wealth of experimental results. In the supplementary materials, we add more details and results, specifically including:

- We introduce the detailed hyper-parameter settings for PGA.
- We show that PGA can be extended to perform object detection attacks in infrared scenarios.
- We demonstrate that PGA can attack objects other than vehicles in autonomous driving scenarios, proving its ability to model and attack arbitrary objects and exposing the vulnerability of DNNs in autonomous driving.
- We present more ablation study results for comprehensive understanding.
- We provide additional qualitative and quantitative comparisons of PGA with other SOTA methods under full-coverage camouflage settings.
- We present the comparison of visualization results under the setting of targeted attacks.
- We provide additional visualization results of PGA attacks using photos captured from real-world scenes as input, aimed at generating adversarial T-shirts that prevent object detectors from recognizing pedestrians.
- We supplement visualization comparisons in cloudy scenarios within the digital domain.

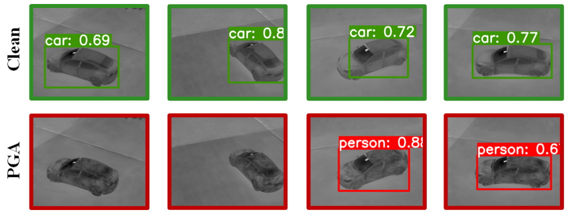

Figure 1. Visualization of clean samples and multi-view PGA adversarial camouflage results in infrared modality.

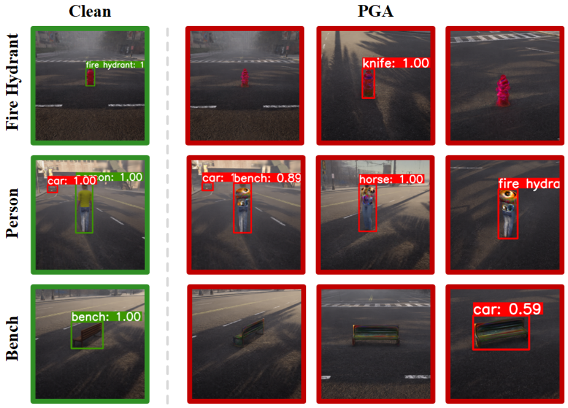

Figure 2. Visualization of clean samples and multi-view PGA adversarial camouflage results for three common objects in autonomous driving scenarios.

1. Detailed Settings for PGA

The proposed PGA is implemented based on the C&W attack framework [2], which uses a hyper-parameter λ to balance the trade-off relationship between adversarial strength and imperceptibility. In our method, the hyper-parameter λ is manually set to 0.1. The impact of different λ values on attack performance and visualization results is demonstrated in the ablation experiments in the next section. In PGA, our primary objective is to iteratively update the zero-order spherical harmonics coefficients of the 3D Gaussians $\langle \boldsymbol{k} \rangle_0$. Concretely, we use the Adam optimizer [4] to optimize $\langle \boldsymbol{k} \rangle_0$, with the learning rate η set to a default value of 0.005. For the reconstruction component in PGA, we sample 200–300 images of the target object from different viewpoints (conveniently extracted continuously from a video), covering various shooting distances, pitch angles, and azimuth angles. The reconstruction follows the parameter settings in SuGaR [3]. For the min-max optimization framework, as described in the main text, we add pixel-level noise to the background to maximize the loss and iteratively optimize the noise. Specifically, we perform 10 iterations during the maximization process, with the total perturbation budget ϵ set to 0.001 and the step size α for each attack step set to 0.0002. The overall noise perturbation magnitude is deliberately kept small to avoid excessive difficulty that could hinder the optimization of the camouflage. For the minimization process, we perform 20 iterations to ensure better optimization convergence. For the overall PGA attack, we use 100 epochs to optimize the final adversarial camouflage.

Implementation Details. We implement our framework and reproduce all the DNN models with PyTorch, and report the results on a workstation with an Intel Xeon Gold 6226R CPU@2.90Hz and 64GB of memory using a single RTX 3090 GPU.

2. Infrared Detector Attack.

In recent studies, target detection in the infrared modality has been widely applied, typically for monitoring critical

Table 1. Comparison results of AP@0.5 (%) for clean samples and those with PGA adversarial camouflages, demonstrating the effectiveness of PGA attacks on three types of objects commonly encountered in autonomous driving scenarios.

Object	Method	AP@0.5(%)			
		Faster RCNN	YOLO-v5*	Mask RCNN*	Average
Person	ORI	100.00	100.00	100.00	100.00
	PGA(Ours)	**12.65**	**81.02**	**19.62**	**37.76**
Fire Hydrant	ORI	67.08	93.67	85.44	82.06
	PGA(Ours)	**0.00**	**56.96**	**5.69**	**20.88**
Bench	ORI	68.29	75.94	72.13	72.12
	PGA(Ours)	**0.00**	**24.05**	**5.20**	**9.75**

areas and facilities under low-light or nighttime conditions using infrared cameras, as well as for providing auxiliary decision-making in autonomous driving, among other applications. Therefore, investigating the robustness of object detectors in the infrared modality is of significant importance. Since PGA employs 3DGS for scene modeling, it can directly utilize multi-view images captured by infrared cameras to quickly and accurately model and render scenes, enabling it to generate adversarial camouflage tailored for the infrared modality. To validate the effectiveness of PGA under the infrared modality, we use a drone equipped with an infrared lens to capture multi-view images of a real vehicle, reconstruct the 3D infrared scene, and conduct the PGA attack. Unlike visible light modality attacks, we apply a unidirectional adversarial perturbation by lowering the local temperature of the target object, resulting in darker areas in infrared imaging. This approach allows for practical deployment by affixing aerogel at corresponding positions. The attack results, shown in Fig. 1, confirm the feasibility of PGA under the infrared modality.

3. Attack Common Objects in Autonomous Driving Scenarios.

Considering that autonomous driving scenarios demand the highest safety, we select three common street objects: fire hydrants, pedestrians, and benches, to reveal the vulnerability of DNNs. Following the experimental settings in the main text, we place these objects in the CARLA simulator, capture about 200 images from multiple viewpoints, and feed them into the PGA physical attack framework. Within PGA, we perform fast, photo-realistic 3DGS modeling for each object, followed by adversarial camouflage generation. The results in Tab 1 demonstrate that the camouflage generated by PGA successfully attacks these objects, significantly reducing AP@0.5(%), thereby proving PGA's high applicability in AD scenarios and its effectiveness in attacking arbitrary objects. Visualization results for the three objects are shown in Fig. 2.

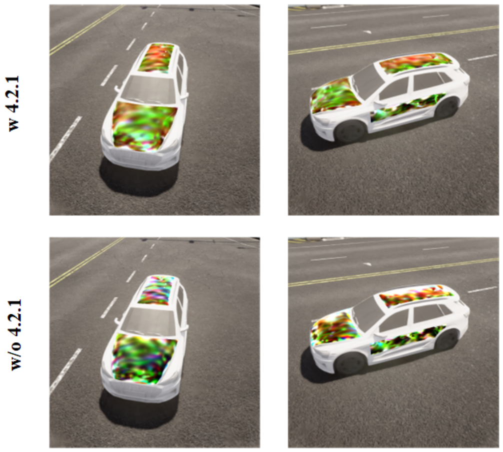

Figure 3. Visualization comparison of whether the cross-view camouflage consistency techniques described in Section 4.2.1 are applied. It is obvious that the camouflage below shows significant inconsistencies across viewpoints.

4. Supplement to Ablation Study

In this subsection, we conduct ablation experiments on key hyper-parameters of the PGA attack framework, including the learning rate of the spherical harmonics coefficients optimizer, the weight of the imperceptibility regularization term λ, and the perturbation budget upper limit ϵ in the min-max optimization. Quantitative comparison results are provided for each in Tab. 2, Tab. 3 and Tab. 4 respectively. Parameters marked with an asterisk in the table indicate the default values used in the main text.

Additionally, we compare the visualization results of PGA attacks with and without the techniques described in Section 4.2.1 of the main text, as shown in Fig. 3. It is obvious that the camouflage without the techniques in Section 4.2.1 exhibits significant inconsistencies when viewpoints change, leading to challenges in physical-world deployment and sub-optimal adversarial effectiveness and robustness.

Moreover, as mentioned in the main text, we employ a common technique in the physical attack domain: Expectation over Transformation (EoT) [1]. Our findings indicate that using EoT can slightly enhance PGA's adversarial performance and multi-view robustness. Ablation experiments show that incorporating EoT reduces the AP@50 on Faster R-CNN from 5.48 to 3.57.

5. Comparison under Full-Coverage Settings

We compare the digital attack performance of PGA with SOTA methods (including FCA [7], DTA [5], ACTIVE [6] and RAUCA [8]) across multiple weather conditions, distances, and viewpoints under the full-coverage camouflage setting. The results in Tab. 5 show that PGA achieves the best attack performance with full-coverage camouflage, in-

Table 2. Ablation experiments on different learning rate η values in PGA. This table provides comparison results of AP@0.5(%) for detection results, averaged across different distances and view angles.

η	Faster-RCNN	Yolo-v5*	Mask-RCNN*
0.0001	9.89	55.27	23.75
0.0005	12.80	59.78	29.35
0.001	5.55	54.73	19.40
0.005*	**2.53**	**52.55**	**18.07**
0.01	4.16	52.63	18.23
0.05	11.36	56.36	31.72

Table 3. Ablation experiments on different λ values in PGA. This table provides comparison results of AP@0.5(%) for detection results, averaged across different distances and view angles.

λ	Faster-RCNN	Yolo-v5*	Mask-RCNN*
0.001	3.26	52.78	16.09
0.01	5.83	52.24	25.87
0.1*	**2.53**	**52.55**	**18.07**
1	87.97	90.14	90.39
10	88.81	90.52	90.96

Table 4. Ablation experiments on different ϵ values in PGA. This table provides comparison results of AP@0.5(%) for detection results, averaged across different distances and view angles.

ϵ	Faster-RCNN	Yolo-v5*	Mask-RCNN*
0.0001	2.98	54.18	22.88
0.0005	2.53	**52.48**	18.79
0.001*	**2.53**	52.55	**18.07**
0.005	3.64	53.22	18.08
0.01	4.36	54.15	18.68

dicating that the generated adversarial camouflage exhibits high adversarial strength, high multi-view robustness, and high transferability. For visualization results, please refer to Fig. 4.

6. Performance of Targeted Attack

In the physical world, compared to untargeted attacks, targeted attacks pose a more severe threat, as deliberately falsifying detection results can lead to more significant safety issues. Our PGA attack framework can be easily extended to targeted attacks by simply replacing the detection loss with a targeted attack version. Concretely, the optimization objective of reducing the confidence of the original class is replaced to increase the confidence of the target class. In Figure 5, we present the visualization results of the targeted attack version of PGA before and after the attack. Three target classes are set: person, truck, and broccoli. The visualization demonstrates that the PGA attack can generate multi-view robust targeted adversarial camouflage.

7. Supplement to Visualization Result

In this section, we provide additional visualization results, comparing the detection results of PGA and other SOTA physical attack methods under multi-view and multi-distance settings in a cloudy environment, see Fig. 6

8. Performance of Real-world Attack

In the main text, we introduce the reconstruction and rendering modules of PGA attack framework, which leverage 3DGS to reconstruct and render the target object. Unlike other physical attack methods, PGA can directly model and attack using a set of real-world photos of the target object without requiring prior mesh information or manual construction of virtual environments. To demonstrate the feasibility of directly attacking real-world objects with PGA, we provide an additional visualization experiments involving a human subject to generate an adversarial T-shirt. Refer to Figure 7 for details.

Table 5. Comparison results of P@0.5(%) for different physical attack methods under full-coverage settings on the COCO dataset, targeting different detection models across various distances and weather conditions. Note that the adversarial camouflage is generated using Faster R-CNN and evaluated for black-box transferability on YOLO-v5 and Mask R-CNN.

Dis	Method	Sunny				Cloudy				Average
		Faster R-CNN	YOLO-v5*	Mask R-CNN*	D-DETR*	Faster R-CNN	YOLO-v5*	Mask R-CNN*	D-DETR*	
5	-	71.86	70.57	73.18	79.76	72.37	73.47	76.06	72.52	73.72
	FCA-F[7]	25.69	57.41	31.12	40.48	12.87	48.24	18.61	33.12	33.44
	DTA-F[5]	16.94	29.46	21.90	24.73	13.37	36.63	24.15	33.76	25.12
	ACTIVE-F[6]	4.45	21.05	12.80	15.33	6.09	14.90	10.43	12.45	12.19
	RAUCA-F[8]	1.35	19.26	10.91	14.09	4.48	13.65	11.75	15.49	11.37
	PGA-F	**0.00**	**11.44**	**0.00**	**8.41**	**1.38**	**9.89**	**2.90**	**5.44**	**4.93**
10	-	89.03	91.87	91.41	81.47	87.10	94.91	90.65	82.04	88.56
	FCA-F	52.84	70.74	55.65	60.34	28.45	45.44	29.78	22.32	45.70
	DTA-F	27.95	41.62	30.56	43.19	9.24	20.80	15.33	18.28	25.87
	ACTIVE-F	12.02	22.50	18.25	23.12	8.09	16.32	11.82	13.99	15.76
	RAUCA-F	0.22	17.38	1.20	18.64	6.50	12.67	11.50	10.20	9.79
	PGA-F	**0.00**	**12.55**	**0.00**	**8.27**	**2.63**	**10.08**	**4.53**	**10.45**	**6.06**
15	-	84.12	97.78	94.54	79.66	88.10	97.78	93.52	83.90	89.93
	FCA-F	46.59	79.72	64.10	66.23	18.48	49.16	29.88	40.10	49.28
	DTA-F	41.46	54.56	48.24	46.38	18.51	29.02	22.61	24.94	35.72
	ACTIVE-F	13.72	25.08	21.34	24.02	8.16	28.40	15.42	23.26	19.93
	RAUCA-F	5.75	16.46	12.49	15.49	4.22	**14.80**	10.56	13.81	11.70
	PGA-F	**0.00**	**10.72**	**0.00**	**6.41**	**2.08**	18.27	**5.42**	**10.45**	**6.67**
20	-	86.50	96.81	91.99	83.37	86.60	98.89	92.35	85.08	90.20
	FCA-F	38.61	74.35	50.10	64.32	30.85	42.20	33.76	36.24	46.30
	DTA-F	32.40	58.36	39.60	37.80	17.01	28.60	26.80	27.67	33.53
	ACTIVE-F	8.54	16.01	10.52	14.91	12.64	19.65	13.48	16.89	14.08
	RAUCA-F	3.91	17.50	5.08	12.45	5.88	16.23	14.55	14.81	11.30
	PGA-F	**0.00**	**10.66**	**0.00**	**4.69**	**3.83**	**14.10**	**5.65**	**8.23**	**5.90**

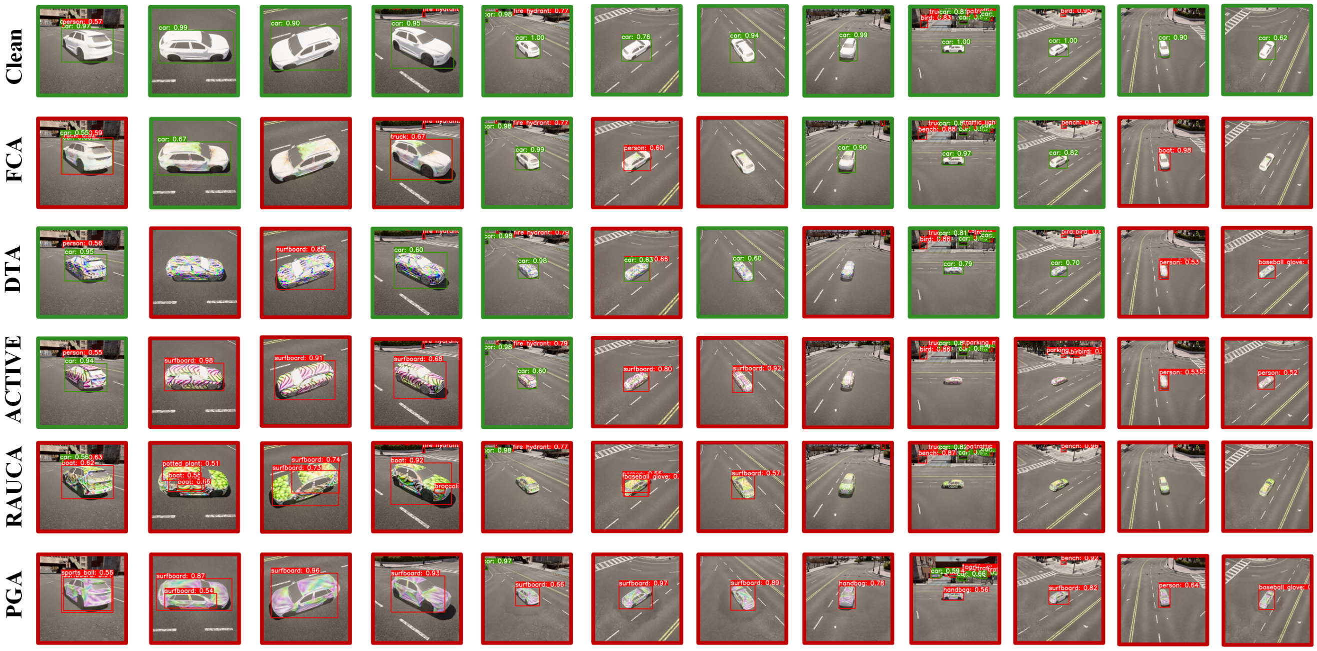

Figure 4. Visualization comparison of multi-view detection results in the digital domain using full-coverage adversarial camouflage. Green-bordered images indicate correct detection of the target vehicle, while red-bordered images indicate either missed detections or detections with incorrect classifications.

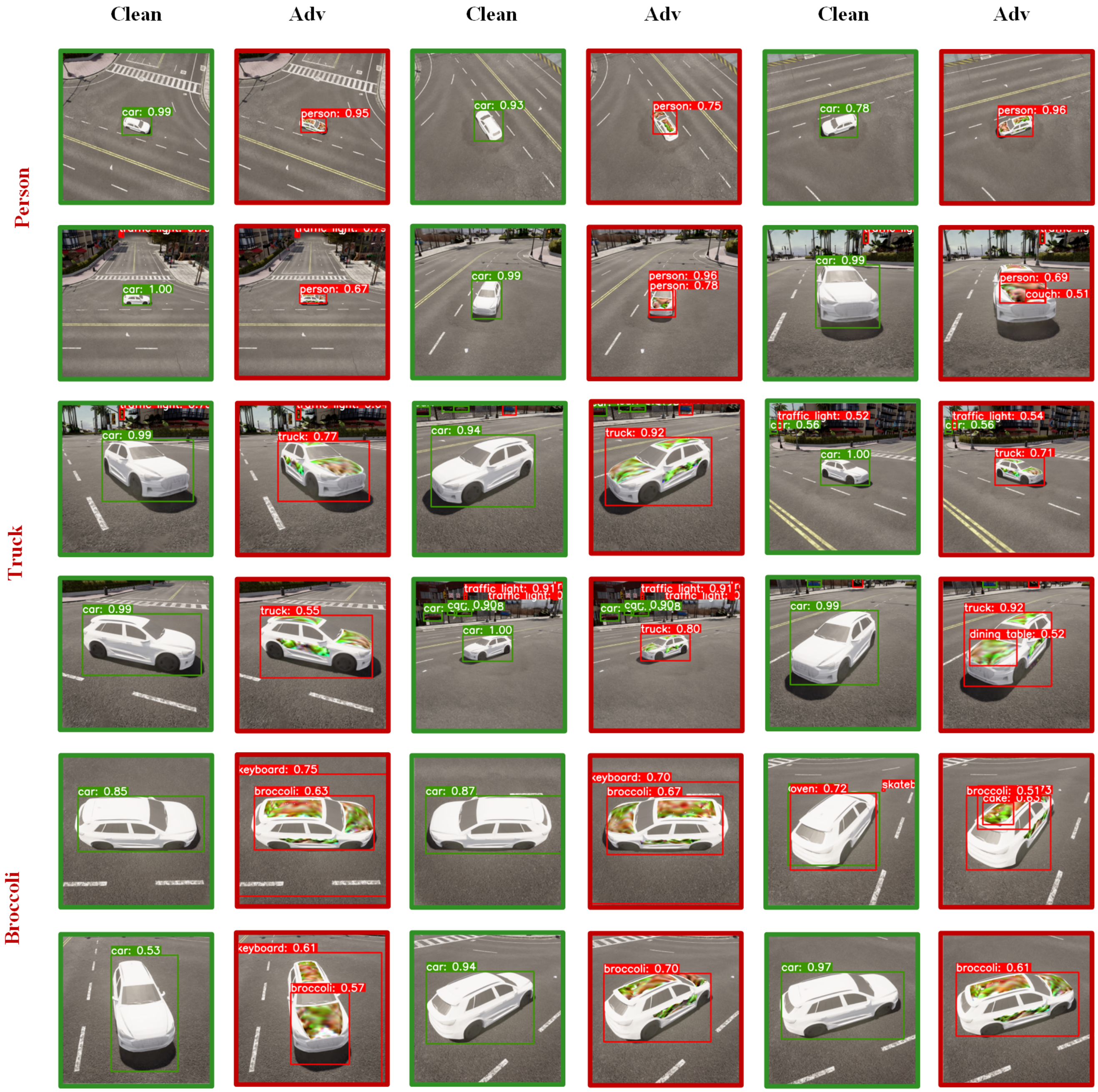

Figure 5. Visualization comparison of multi-view detection results in the digital world. Green-bordered images indicate correct detection of the target vehicle, while red-bordered images indicate either undetected targets or detection with incorrect classification.

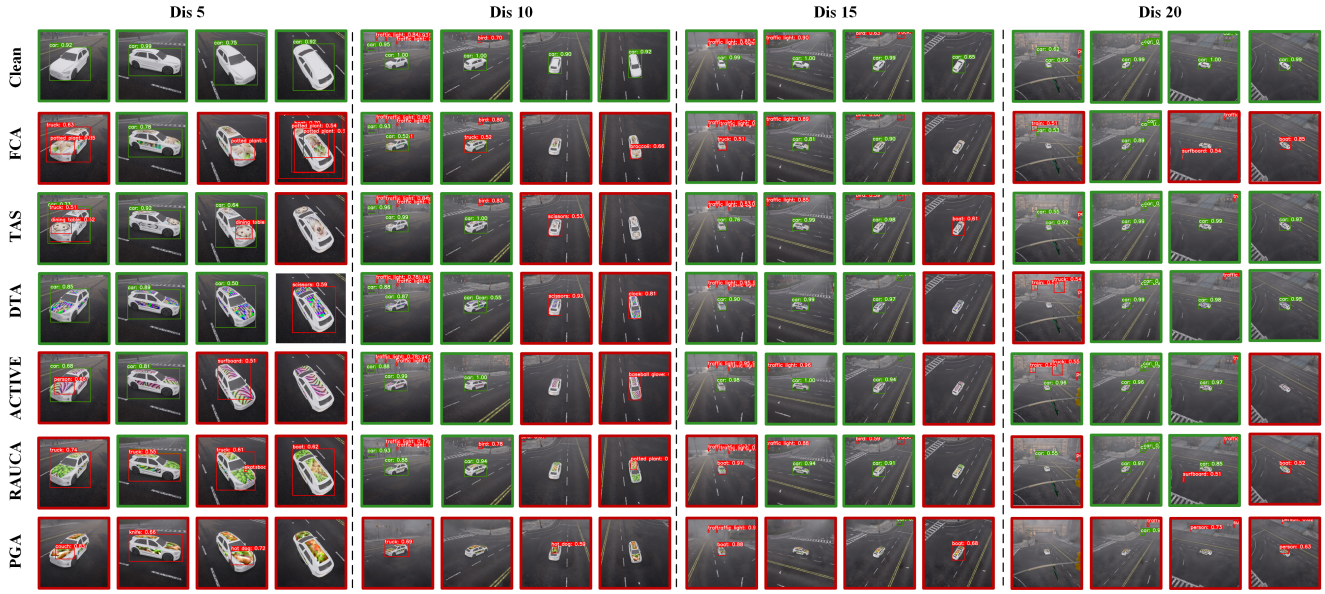

Figure 6. Visualization comparison of multi-view detection results in the digital domain under cloudy weather conditions. Green-bordered images indicate correct detection of the target vehicle, while red-bordered images indicate either undetected targets or detection with incorrect classification.

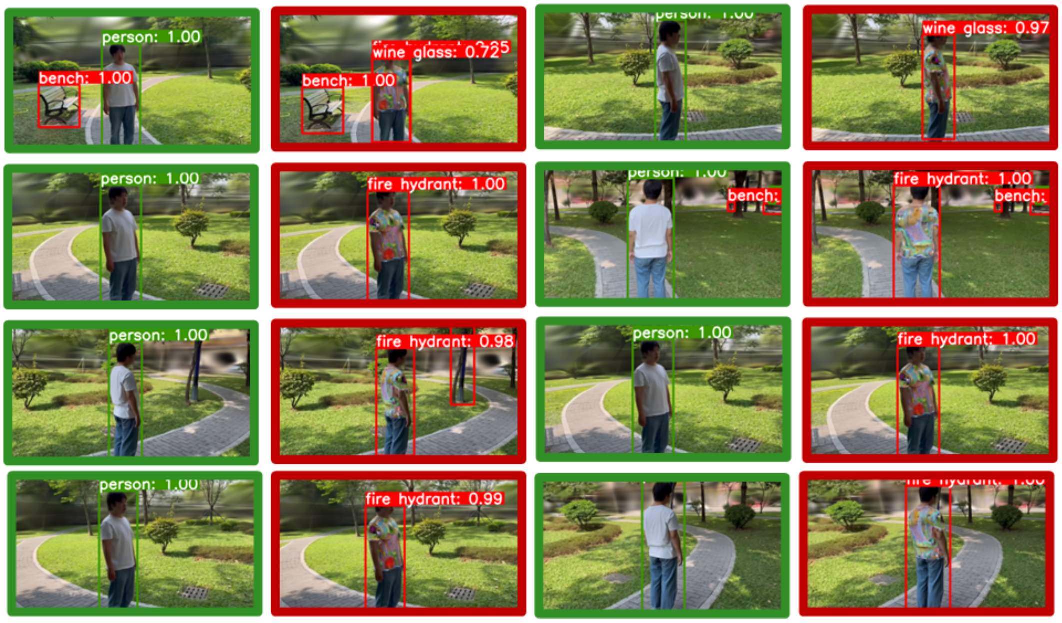

Figure 7. Visualization comparison of real-world objects before and after the PGA attack. The first four rows compare the detection results of a person wearing and not wearing the adversarial T-shirt, while the last four rows compare the detection results of a car with and without adversarial camouflage.

References

[1] Anish Athalye, Logan Engstrom, Andrew Ilyas, and Kevin Kwok. Synthesizing robust adversarial examples. In *International conference on machine learning*, pages 284–293. PMLR, 2018. 2

[2] Nicholas Carlini and David Wagner. Towards evaluating the robustness of neural networks. In *2017 ieee symposium on security and privacy (sp)*, pages 39–57. Ieee, 2017. 1

[3] Antoine Guédon and Vincent Lepetit. Sugar: Surface-aligned gaussian splatting for efficient 3d mesh reconstruction and high-quality mesh rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5354–5363, 2024. 1

[4] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 1

[5] Naufal Suryanto, Yongsu Kim, Hyoeun Kang, Harashta Tatimma Larasati, Youngyeo Yun, Thi-Thu-Huong Le, Hunmin Yang, Se-Yoon Oh, and Howon Kim. Dta: Physical camouflage attacks using differentiable transformation network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15305–15314, 2022. 2, 4

[6] Naufal Suryanto, Yongsu Kim, Harashta Tatimma Larasati, Hyoeun Kang, Thi-Thu-Huong Le, Yoonyoung Hong, Hunmin Yang, Se-Yoon Oh, and Howon Kim. Active: Towards highly transferable 3d physical camouflage for universal and robust vehicle evasion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4305–4314, 2023. 2, 4

[7] Donghua Wang, Tingsong Jiang, Jialiang Sun, Weien Zhou, Zhiqiang Gong, Xiaoya Zhang, Wen Yao, and Xiaoqian Chen. Fca: Learning a 3d full-coverage vehicle camouflage for multi-view physical adversarial attack. In *Proceedings of the AAAI conference on artificial intelligence*, pages 2414–2422, 2022. 2, 4

[8] Jiawei Zhou, Linye Lyu, Daojing He, and Yu Li. Rauca: A novel physical adversarial attack on vehicle detectors via robust and accurate camouflage generation. *arXiv preprint arXiv:2402.15853*, 2024. 2, 4