This ICCV paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

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

Tianrui Lou[1,2] **Xiaojun Jia**[3] **Siyuan Liang**[4] **Jiawei Liang**[1] **Ming Zhang**[5] **Yanjun Xiao**[6] **Xiaochun Cao**[1,2,*]

[1]*Sun Yat-Sen University* [2]*Peng Cheng Laboratory* [3]*Nanyang Technological University* [4]*National University of Singapore*

[5]*National Key Laboratory of Science and Technology on Information System Security* [6]*Nsfocus*

{loutianrui, jiaxiaojunqaq, pandaliang521}@gmail.com liangjw57@mail2.sysu.edu.cn

zm_stiss@163.com xiaoyanjun@nsfocus.com caoxiaochun@mail.sysu.edu.cn

Abstract

Physical adversarial attack methods expose the vulnerabilities of deep neural networks and pose a significant threat to safety-critical scenarios such as autonomous driving. Camouflage-based physical attack is a more promising approach compared to the patch-based attack, offering stronger adversarial effectiveness in complex physical environments. However, most prior work relies on mesh priors of the target object and virtual environments constructed by simulators, which are time-consuming to obtain and inevitably differ from the real world. Moreover, due to the limitations of the backgrounds in training images, previous methods often fail to produce multi-view robust adversarial camouflage and tend to fall into sub-optimal solutions. Due to these reasons, prior work lacks adversarial effectiveness and robustness across diverse viewpoints and physical environments. We propose a physical attack framework based on 3D Gaussian Splatting (3DGS), named PGA, which provides rapid and precise reconstruction with few images, along with photo-realistic rendering capabilities. Our framework further enhances cross-view robustness and adversarial effectiveness by preventing mutual and self-occlusion among Gaussians and employing a min-max optimization approach that adjusts the imaging background of each viewpoint, helping the algorithm filter out non-robust adversarial features. Extensive experiments validate the effectiveness and superiority of PGA. Our code is available at: https://github.com/TRLou/PGA.

1. Introduction

Despite the remarkable success of deep neural networks (DNNs) in various fields, such as computer vision [15] and natural language processing [9, 52], the emergence of adversarial attacks highlights the vulnerability of DNNs. Although digital attacks [4, 11, 12, 14, 22–24, 39, 42] targeting various tasks have raised concerns, physical attacks deployed in the real world pose even greater threats, stifling the use of DNNs in safety-critical domains such as autonomous driving [3, 5, 56], security surveillance [28, 29, 33–37, 43, 58, 59], and remote sensing [32, 38, 57]. We focus on physical attacks in autonomous driving, primarily targeting vehicle detection.

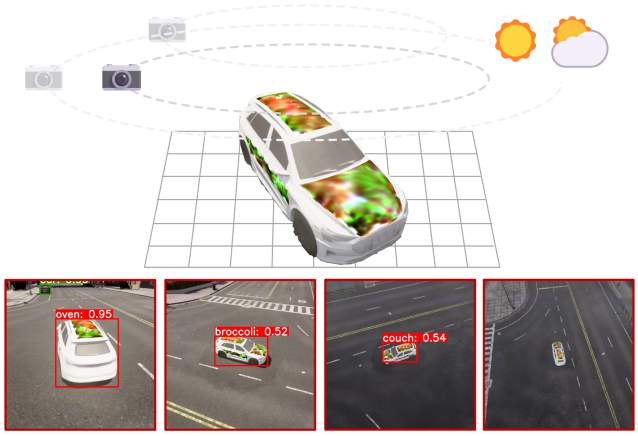


Figure 1. Visualization of multi-view robust adversarial camouflage generated by PGA, which effectively causes the victim detector to miss detections or misclassify the object across various environmental settings, including different shooting distances, pitch angles, azimuth angles, and weather conditions.

Physical attacks are often crafted in the digital domain and subsequently implemented by altering the physical properties of the target, such as patch application [2, 8, 17] or camouflage deployment [49, 50, 53–55, 62, 64]. The main challenge of physical attacks lies in minimizing the degradation of adversarial effectiveness when the generated adversarial camouflage is transferred from the digital domain to the physical domain, primarily due to environmental factors such as shooting distances, pitch angles, azimuth angles, and weather conditions.

Since adversarial camouflage offers higher robustness across different environmental settings compared to adver-

sarial patches, it has become a more prevalent research direction. Unlike adversarial patches, which only require pixel-level addition to the image of the target object during optimization, adversarial camouflage involves more complex shape-conforming computations. Zhang et al. [62] and Wu et al. [60] estimate the image of camouflage applied to the target object through black-box methods, using a neural approximation function and a genetic algorithm, respectively. Furthermore, to enhance adversarial effectiveness and robustness, a series of subsequent works [49, 50, 53–55, 64] develop and employ differentiable neural renderers to render images based on the target object's mesh. The differentiability of renderers allows for more precise white-box computation of adversarial camouflage.

Despite the success of previous methods to some extent, the robustness and adversarial effectiveness of the generated camouflage in physical environments remain limited due to the following two main reasons. Firstly, these methods rely heavily on prior mesh information of the target object and virtual environments constructed by simulators such as CARLA [6], which inevitably exhibit significant discrepancies from the real physical world. Secondly, prior works usually apply only simple augmentations to backgrounds and viewpoints. The limited backgrounds in training images hinder the optimization of multi-view robust camouflage in the physical world, often resulting in sub-optimal solutions and leading to low robustness and universality.

In this paper, we propose a multi-view robust physical 3DGS-based attack method (PGA), which employs 3DGS as the differentiable rendering pipeline. Thanks to the excellent reconstruction capabilities of 3DGS, PGA can quickly and accurately reconstruct the target object and background scene using only a few images, without the need for manually constructing. Additionally, 3DGS enables fast, differentiable rendering from specified camera viewpoints, providing photo-realistic imaging results in the iterative attack process of PGA. Furthermore, we propose to enhance the cross-view robustness and adversarial effectiveness of PGA by several methods. Firstly, we address the issue of imaging inconsistency in adversarial camouflage across different viewpoints by preventing both mutual and self-occlusion among Gaussians. Secondly, to generate physically adversarial camouflage that is robust and universal across various viewpoints, we design a min-max optimization approach. Concretely, we first add pixel-level perturbations to the background of each viewpoint's rendered image to maximize the detection loss, and then optimize the camouflage to minimize the loss, thereby obtaining multi-view robust adversarial features. Finally, we incorporate several common techniques and regularization terms in the loss function to further enhance the physical performance and visual naturalness of the camouflage, including Expectation over Transformations (EoT) [1], Non-Printability Score (NPS) [45] and primary color regularization. Extensive experiments demonstrate that our attack framework outperforms state-of-the-art methods in both the digital and physical domains. Moreover, leveraging the features of 3DGS, our approach enables rapid modeling and effective attacks on various objects in autonomous driving scenarios and can be extended to attack tasks in infrared object detection, please refer to the supplementary material.

Our main contributions are in three aspects:

- We propose the first physical adversarial attack framework based on 3D Gaussian Splatting. Leveraging the precise and fast reconstruction capabilities of 3DGS, our PGA framework enables attacks on arbitrary objects in the physical world.
- We further enhance the cross-view robustness and adversarial effectiveness. Firstly, we solve cross-view imaging inconsistency of camouflage by preventing mutual occlusion and self-occlusion of 3DGS. Secondly, we propose a min-max optimization method to filter out multi-view non-robust adversarial features.
- Extensive experiments validate the superiority of our framework to the state-of-the-art physical attack methods.

2. Related Work

Physical Adversarial Attack. Most existing physical attack studies focus on autonomous driving scenarios, such as traffic sign detection [7, 8, 10, 47], pedestrian detection [18–20, 48, 51, 61], and vehicle detection [49, 50, 53–55, 60, 62, 63, 65]. Compared to the previous two scenarios, physical attacks on vehicle detection are more challenging, as adversarial perturbations must remain robust across varying view angles, distances, and weather conditions. Given that some studies have revealed the inadequacy of patch-based physical attacks in meeting the stringent robustness demands, researchers have opted to devise adversarial camouflage as an alternative. To obtain camouflage and iteratively enhance its adversarial capability, a differentiable rendering process is essential. As initial attempts, some studies employed black-box methods to estimate the rendering results. Concretely, Zhang et al. [62] proposed to train a neural approximation function to imitate the rendering process, and Wu et al. [60] computed optimal adversarial camouflage using a genetic algorithm. To leverage a white-box setting for enhanced adversarial capabilities, some studies [49, 50, 53–55] have focused on employing differentiable rendering method [25, 49]. Wang et al. [54] proposed to suppress both model and human attention to gain visual naturalness and robustness. Additionally, they later introduced further suppression of model-shared attention to enhance transferability [55]. To overcome partial occluded and long-distance issues, Wang et al. [53] optimized full-coverage vehicle camouflage. Suryanto et al. [49] designed a more photo-realistic renderer and integrated it into

the attack framework, effectively enhancing the robustness of the camouflage. Moreover, they improved robustness and universality by utilizing tri-planar mapping and making targets both misclassified and undetectable [50]. Zhou et al. [64] addressed the complexities of weather conditions in physical scenarios by enhancing the neural renderer to accurately project vehicle textures and render images with environmental features like lighting and weather, forming the foundation of the RAUCA attack framework.

3D Modeling for Physical Attacks. Most of the above works require obtaining the mesh model of the target object in advance, which is time-consuming and labor-intensive. Recently, some 3D representations have made it easier to model new objects and provide differentiable rendering pipelines that can be employed in physical attack frameworks, e.g. NeRF [40], 3D Gaussian Splatting [26]. Li et al. [31] modeled target vehicles as NeRFs and optimized adversarial patches, resulting in improved physical realism. Huang et al. [21] proposed a transferable targeted attack approach that uses a grid-based NeRF to reconstruct the target object's mesh, optimizing both texture and geometry simultaneously during iterations. Despite these attack methods eliminating the dependency on the target object's mesh information, they are often limited by inherent drawbacks of NeRF, such as slow rendering, low quality, and high memory requirements. In this paper, we resort to 3DGS, which can rapidly and accurately reconstruct the scene using numerous 3D Gaussian ellipsoids and easily perform differentiable, photo-realistic multi-view rendering, serving as the 3D representation of the target object to implement a physical attack framework.

3. Preliminaries

In this section, we will first provide a brief introduction to 3DGS. Then we will analyze the challenges of generating deployable and effective adversarial camouflage using 3DGS as a differentiable rendering pipeline in the physical attack framework.

3DGS reconstructs the scene by representing it with a large set of Gaussians $\mathcal{G} = \{\boldsymbol{g}_1, \boldsymbol{g}_2, ..., \boldsymbol{g}_N\}$, where N denotes the number of Gaussians. Each Gaussian $\boldsymbol{g}$ is characterized by its mean μ_g and anisotropic covariance $\boldsymbol{\Sigma}_g$, and can be mathematically represented as:

$$\boldsymbol{g}(\boldsymbol{x}) = exp(-\frac{1}{2}(\boldsymbol{x}-\mu_g)^T\boldsymbol{\Sigma}_g^{-1}(\boldsymbol{x}-\mu_g)), \qquad (1)$$

where the mean μ_g determines its central position, and the covariance $\boldsymbol{\Sigma}_g$ is defined by a scaling vector $\boldsymbol{s}_g \in \mathbb{R}^3$ and a quaternion $\boldsymbol{q}_g \in \mathbb{R}^4$ that encodes the rotation of $\boldsymbol{g}$. Besides, 3DGS uses an $\alpha_g \in [0, 1]$ to represent the opacity of $\boldsymbol{g}$ and describes the view-dependent surface color $\boldsymbol{c}_g$ through spherical harmonics coefficients $\boldsymbol{k}_g$. To reconstruct a new scene, 3DGS requires only a few images $\mathcal{I}$ from different viewpoints as training inputs. Starting from a point cloud initialized by SfM [46], it optimizes and adjusts the parameters $\{\mu_g, \boldsymbol{s}_g, \boldsymbol{q}_g, \alpha_g, \boldsymbol{k}_g\}$ of each $\boldsymbol{g}$ to make the rendering closely resemble the real images. After training, an image $\boldsymbol{I}_{\boldsymbol{\theta_c}}$ can be differentially rendered through a rasterizer $\mathcal{R}$ by splatting each 3D Gaussian $\boldsymbol{g}$ onto the image plane as a 2D Gaussian, with pixel values efficiently computed through alpha blending given a viewpoint $\boldsymbol{\theta_c}$ and a set $\mathcal{G}$, formulated as $\boldsymbol{I}_{\boldsymbol{\theta_c}} = \mathcal{R}(\boldsymbol{\theta_c}, \mathcal{G})$. Then, the rendered images $\mathcal{I}_r = \{\boldsymbol{I}_{\boldsymbol{\theta_{c1}}}, \boldsymbol{I}_{\boldsymbol{\theta_{c2}}}, ...\}$ from various viewpoints are fed into the target detector $\mathcal{F}(\cdot; \boldsymbol{\theta_f})$, parameterized by $\boldsymbol{\theta_f}$, for evaluation. The objective of our attack framework is to iteratively refine the attributes of the Gaussians $\mathcal{G}$ to mislead the detection results of $\mathcal{F}$, ultimately yielding robust adversarial Gaussians $\mathcal{G}'$ and camouflage $\boldsymbol{\mathcal{T}}$.

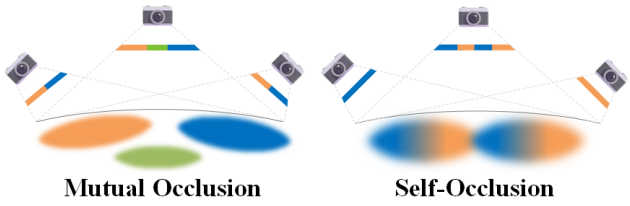


Figure 2. Illustration of mutual occlusion and self-occlusion issues in vanilla 3DGS that lead to cross-view inconsistencies.

Problem Analysis We resort to 3DGS to support the proposed attack framework, which brings numerous advantages, including rapid reconstruction of arbitrary scenes and fast, differentiable rendering capabilities. However, generating camouflage with strong adversarial effectiveness and robustness in the physical world remains a challenge due to two main reasons. **Firstly**, while vanilla 3DGS generally produces rendered images that align with the training set, discrepancies often exist between the represented 3D objects and their true values. Concretely, not all Gaussians are positioned accurately on the surface, leading to mutual-occlusion issues among the Gaussians when the viewpoint changes. Additionally, since 3DGS uses spherical harmonics with strong representational capabilities to describe surface color, the same Gaussian may exhibit vastly different colors due to self-occlusion when the viewpoint changes. The issues of mutual occlusion and self-occlusion result in significant inconsistencies in the rendered camouflage across different viewpoints, reducing adversarial effectiveness and hindering physical deployment. Please refer to Fig. 2. **Secondly**, in real-world scenarios, there are numerous factors affecting imaging results and detector performance, including shooting distance, angle, and weather conditions. During training, the limited variety of backgrounds makes it challenging to ensure that the generated adversarial camouflage is both universal and robust in real-world settings, leading traditional optimization methods to

often fall into suboptimal solutions. We address these two challenges individually and provide a detailed explanation in the following sections.

4. Methodology

We propose a novel physical attack framework based on 3D Gaussian Splatting named **PGA**. We first introduce the pipeline and formulation of our framework in Sec. 4.1, then followed by the proposed strategies to enhance physical adversarial effectiveness and robustness in Sec. 4.2.

4.1. Pipeline and Formulation of PGA Framework

The overall pipeline of our framework is shown in Fig 3. Our framework is composed of three components including a reconstruction module, a rendering module and an attack module.

Reconstruction module. Given a set of images $\mathcal{I} = \{\boldsymbol{I_1}, \boldsymbol{I_2}, ...\}$ from different viewpoints, we first reconstruct the Gaussians $\mathcal{G} = \{\boldsymbol{g_1}, \boldsymbol{g_2}, ..., \boldsymbol{g_N}\}$ of entire scene using the 3DGS training framework [26].

Rendering module. We select multiple camera viewpoints $\Theta = \{\boldsymbol{\theta_{c1}}, \boldsymbol{\theta_{c2}}, ...\}$ around the target object at varying distances, pitch angles, and azimuth angles, ensuring comprehensive coverage to facilitate the generation of physically robust adversarial camouflage. Then we obtain rendered images through rasterizer $\mathcal{R}$ provided by 3DGS:

$$\mathcal{I}_r = \mathcal{R}(\Theta, \mathcal{G}). \tag{2}$$

To ensure that adversarial perturbations are only added to the target object, we use SAM [27] to extract masks $\mathcal{M}$ from $\mathcal{I}_r$,

$$\mathcal{M} = \text{SAM}(\mathcal{I}_r, \mathcal{P}), \tag{3}$$

where $\mathcal{P}$ are prompts of the target object. We create a copy of the original rendered image $\mathcal{I}_r$ as $\mathcal{I}_{ori}$, and the final images to be detected can be expressed as:

$$\mathcal{I}_{\text{det}} = (\mathcal{I}_r \cdot \mathcal{M}) + (\mathcal{I}_{\text{ori}} \cdot (1 - \mathcal{M})) \tag{4}$$

Attack module. After calculating $\mathcal{I}_{det}$, we feed it into the victim object detector $\mathcal{F}$ to obtain the detection results:

$$\mathcal{B} = \mathcal{F}(\mathcal{I}_{\text{det}}; \boldsymbol{\theta_f}) = \{\boldsymbol{b_{\theta_{c1}}}, \boldsymbol{b_{\theta_{c2}}}, ...\}. \tag{5}$$

And the detection loss can be defined following [19] as:

$$\begin{aligned} \mathcal{L}_{\text{det}}(\mathcal{I}_{\text{det}}) &= \sum_{\boldsymbol{I}} \text{Conf}^{(\boldsymbol{I})}_{m^*}, \\ m^* &= \underset{m}{\text{argmax}} \text{IoU}(\boldsymbol{gt^{(I)}}, \boldsymbol{b}^{(\boldsymbol{I})}_m), \end{aligned} \tag{6}$$

where $\boldsymbol{I}$ is each input of a batch in $\mathcal{I}_{\text{det}}$, $\boldsymbol{b}_m$ is the m-th bounding box of detection results and Conf is corresponding confidence. $\mathcal{L}_{\text{det}}$ minimize the confidence score of the correct class in the box which has the maximum Intersection over Union (IoU) score with the ground truth $\boldsymbol{gt}$.

The optimization objective of the attack module can be formulated as:

$$\mathcal{G}' = \arg\min_{\mathcal{G}} \mathcal{L}_{\text{det}}(\mathcal{I}_{\text{det}}(\boldsymbol{\theta_c}, \mathcal{G})) \tag{7}$$

Considering the difficulty and feasibility of manipulating the shape of the target object in the physical domain, we only optimize the spherical harmonics coefficients k_g of the Gaussians $\mathcal{G}$, which represent the surface color, in an iterative attack process with a learning rate η:

$$\boldsymbol{k}^{t+1} = \boldsymbol{k}^t + \eta \nabla_{\boldsymbol{k}} \mathcal{L}_{\text{det}}(\mathcal{I}_{\text{det}}). \tag{8}$$

Upon completion of the iterative attack, the adversarial camouflage mesh $\boldsymbol{\mathcal{T}}$ can be derived from the optimized Gaussians $\mathcal{G}'$ following [13] and deployed in the physical environment.

4.2. Physical Adversarial Effectiveness and Robustness Enhancement

4.2.1. Improving Cross-Viewpoint Consistency

To tackle the issue of mutual occlusion, we facilitate the regularization terms from SuGaR [13] in the reconstruction module, aligning the Gaussians with the object surface and encouraging the Gaussians to reduce their opacity. These terms prevent the Gaussians from being optimized inside the object, ensuring that their surface color is not occluded by other Gaussians on the surface when the viewpoint changes.

Additionally, we observe that higher-order spherical harmonics provide Gaussians with strong representational power for surface color, causing different parts of a single Gaussian to exhibit vastly different colors. When the viewpoint changes, these colors can occlude each other. This phenomenon becomes especially evident during multi-view joint iterative attack optimization, as the optimizer tends to focus on refining the visible portions of each Gaussian from each viewpoint, resulting in significant local color variations. To address this self-occlusion problem, we propose optimizing only the zero-order term of the spherical harmonic coefficients $\langle \boldsymbol{k} \rangle_0$ during iterative attacks, ensuring uniform color changes across the surface of each Gaussian. With these two improvements, we can ensure that the same adversarial camouflage is optimized consistently during cross-view iterative optimization.

4.2.2. Multi-view Robust Adversarial Camouflage Optimization Method

Since it can be ragarded as Universal Adversarial Perturbation (UAP) [41] problem and the attack difficulty varies significantly across different viewpoints, we iteratively optimize the camouflage for each viewpoint in sequence. To avoid over-optimization on easier viewpoints, which could increase the difficulty of optimizing other viewpoints, we

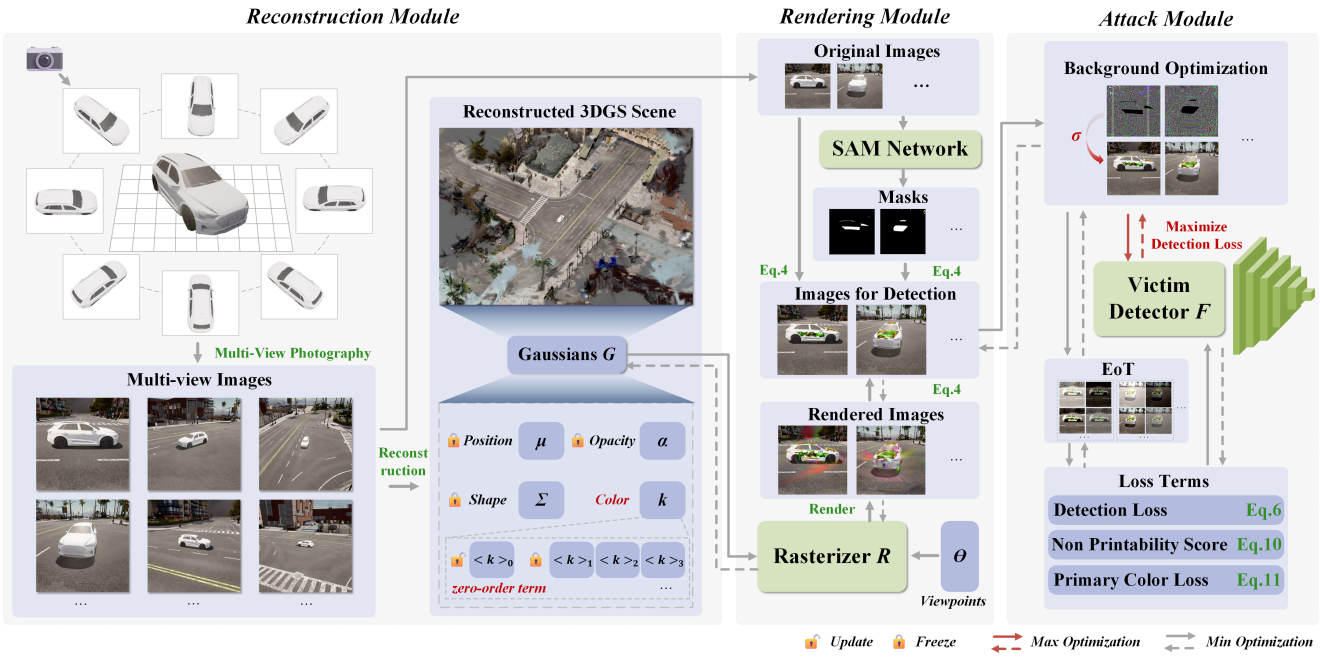


Figure 3. Demonstration of the framework of PGA. First, the reconstruction module captures multi-view images to build a 3DGS scene. Then the rendering module combines the clean background with the rendered adversarial camouflage to create the image for detection. Finally, the attack module applies a min-max optimization framework, first adding noise to the background to increase attack difficulty, then refining a multi-view robust camouflage with high adversarial effectiveness.

set an iteration limit for each viewpoint. Once the camouflage successfully attacks a given viewpoint, we skip the remaining iterations and proceed to optimize the next viewpoint.

Additionally, since the adversarial effectiveness of camouflage is affected by the background context features, we conduct a "counter adversarial attack" on the background to make the adversarial features more robust to background variations. Concretely, before each optimization iteration of the camouflage, we add point-wise noise $\boldsymbol{\sigma}$ to the background and optimize it iteratively using I-FGSM [30]. Note that the optimization stops once the detector can correctly detect the target object or the iteration limit is reached, as excessive interference would make the camouflage difficult to optimize. This process can be formulated as a min-max optimization problem:

$$\mathcal{G}' = \arg\min_{\mathcal{G}} \max_{\boldsymbol{\sigma}} \mathcal{L}_{\text{det}}(\mathcal{I}_{\text{det}}(\boldsymbol{\theta_c}, \mathcal{G}) + \boldsymbol{\sigma} \cdot (1 - \mathcal{M})) \quad \text{s.t. } ||\boldsymbol{\sigma}||_\infty \leq \epsilon, \tag{9}$$

where ϵ is a hyper-parameter denoting the budget of $\boldsymbol{\sigma}$.

4.2.3. Optimization Objective

In addition to addressing the key issues mentioned above, we employ several additional techniques within the physical 3DGS-based attack framework to further improve its adversarial effectiveness and imperceptibility in real-world scenarios. Firstly, we employ Expectation over Transformation (EoT) [1] in the optimization process, a technique widely used in various physical adversarial attack methods. Specifically, we apply a set of physical transformations, such as randomizing the scale, contrast, brightness, and adding noise, to enhance robustness. Secondly, we introduce Non-Printability Score (NPS) [45] to mitigate fabrication error:

$$\text{NPS} = \sum_{\hat{\boldsymbol{p}} \in \mathcal{C}(\mathcal{I}_{\text{det}})} \prod_{\boldsymbol{p}' \in P} |\hat{\boldsymbol{p}} - \boldsymbol{p}'|, \tag{10}$$

where P is a set of printable colors and $\mathcal{C}(\mathcal{I}_{det})$ is a set of RGB triples used in $\mathcal{I}_{det}$. Finally, to make the adversarial camouflage more imperceptible, we extract all background pixels from the training set, specifically $\mathcal{I}_{ori} \cdot (1 - \mathcal{M})$. Using K-means clustering, we group the background colors and select the top-k colors as the primary colors for the camouflage. During optimization, we add a regularization term to ensure that the camouflage remains close to the primary colors:

$$\mathcal{L}_{\text{clr}} = \frac{1}{|\Omega|} \sum_{(x,y) \in \Omega} \min_i ||\mathcal{I}_{\text{det}}(x, y) - \boldsymbol{c}_i||_2, \tag{11}$$

where (x, y) represents the position of pixel and $\Omega = \{(x, y) | \mathcal{M}(x, y) > 0\}$ is the set of pixel locations where the mask is non-zero. Further, we also constrain the L_2 norm

Table 1. Comparison results of AP@0.5(%) for different physical attack methods on the COCO datasets targeting different detection models under different distances and weathers. Note that the adversarial camouflage is generated using Faster R-CNN and evaluated for black-box transferability on YOLO-v5, Mask R-CNN and Deformable-DETR.

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
	DAS[54]	42.90	70.16	49.87	47.75	48.57	72.86	55.75	49.58	54.68
	FCA[53]	35.16	55.62	40.52	46.29	37.30	58.98	47.54	48.57	46.25
	DTA[49]	36.19	48.18	43.82	37.04	49.91	57.59	63.38	43.26	47.42
	ACTIVE[50]	32.44	45.61	44.35	41.59	38.42	51.16	51.05	49.83	44.31
	TAS[55]	43.31	65.59	58.32	43.64	47.24	68.35	57.76	45.50	53.71
	RAUCA[64]	21.71	46.94	31.90	36.54	27.85	56.01	36.50	39.79	37.16
	PGA	**4.52**	**39.10**	**10.62**	**28.31**	**5.60**	**46.99**	**16.67**	**35.90**	**23.46**
10	-	89.03	91.87	91.41	81.47	87.10	94.91	90.65	82.04	88.56
	DAS[54]	77.98	77.69	87.31	72.60	64.83	73.43	70.98	74.02	74.86
	FCA[53]	59.98	67.87	65.00	67.47	55.88	65.23	55.25	63.43	62.51
	DTA[49]	55.61	66.27	74.81	53.83	55.38	62.01	74.66	57.75	62.54
	ACTIVE[50]	59.00	68.94	71.67	52.41	60.02	69.81	64.46	61,79	63.76
	TAS[55]	53.85	69.41	80.56	55.21	53.86	75.57	68.34	52.25	63.63
	RAUCA[64]	18.88	56.70	31.00	44.85	21.74	59.37	34.29	47.17	39.25
	PGA	**1.40**	**45.53**	**8.44**	**30.89**	**0.71**	**48.18**	**8.53**	**30.54**	**21.78**
15	-	84.12	97.78	94.54	79.66	88.10	97.78	93.52	83.90	89.93
	DAS[54]	78.67	89.86	81.57	73.88	62.13	75.28	70.94	74.10	75.80
	FCA[53]	66.37	77.80	76.58	69.56	61.97	69.74	69.07	73.05	70.52
	DTA[49]	57.17	72.47	73.78	61.65	55.17	64.94	65.60	66.65	64.68
	ACTIVE[50]	53.58	78.98	60.16	60.54	57.56	68.40	58.77	69.50	63.44
	TAS[55]	55.79	70.57	67.21	67.25	65.23	68.34	73.32	68.28	67.00
	RAUCA[64]	37.80	63.32	58.27	44.69	38.46	64.97	46.19	56.73	51.30
	PGA	**1.95**	**52.96**	**9.40**	**29.58**	**7.16**	**59.86**	**12.24**	**31.10**	**25.53**
20	-	86.50	96.81	91.99	83.37	86.60	98.89	92.35	85.08	90.20
	DAS[54]	68.67	88.47	78.52	76.14	60.62	69.47	65.95	70.69	72.32
	FCA[53]	64.23	71.53	78.88	72.99	58.87	63.60	66.96	73.72	68.85
	DTA[49]	48.99	76.44	74.89	70.40	58.14	65.48	70.14	68.99	66.68
	ACTIVE[50]	39.70	70.77	64.31	67.28	50.47	65.02	57.00	70.66	60.65
	TAS[55]	67.20	84.92	85.21	70.25	57.33	74.42	71.13	64.54	71.88
	RAUCA[64]	37.29	59.34	59.07	48.60	32.84	55.57	42.89	60.39	49.50
	PGA	**1.85**	**43.95**	**14.60**	**23.14**	**5.40**	**41.42**	**14.63**	**20.83**	**20.73**

distance of the spherical harmonics coefficients before and after the attack to be as small as possible. Thus, the overall loss can now be reformulated as:

$$\begin{aligned}\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{det}}(T(\mathcal{I}_{\text{det}}(\boldsymbol{\theta_c}, \mathcal{G}) + \boldsymbol{\sigma} \cdot (1 - \mathcal{M}))) \\ + \lambda(\text{NPS} + \mathcal{L}_{\text{clr}} + ||\langle \boldsymbol{k} \rangle_0 - \langle \boldsymbol{k} \rangle_0^{\text{ori}}||_2)\end{aligned} \tag{12}$$

where T is transformations of EoT and λ is hyper-parameter and $\langle \boldsymbol{k} \rangle_0^{\text{ori}}$ presents initial values before attack. Meanwhile, the overall optimization objective and the iterative update process of the spherical harmonics coefficients can be reformulated separately as:

$$\mathcal{G}' = \arg\min_{\mathcal{G}} \max_{\boldsymbol{\sigma}} \mathcal{L}_{\text{total}}, \tag{13}$$

$$\langle \boldsymbol{k}^{t+1} \rangle_0 = \langle \boldsymbol{k}^t \rangle_0 + \eta \nabla_{\langle \boldsymbol{k} \rangle_0} \mathcal{L}_{\text{total}}. \tag{14}$$

5. Experiments

In this section, we first illustrate the experimental settings and implementation details. We then demonstrate the superiority and effectiveness of PGA through digital domain experiments, including extensive qualitative and quantitative comparisons of attack performance and ablation studies. Furthermore, we conduct physical domain experiments, presenting outstanding qualitative and quantitative results of the generated camouflage on a 1:24 scale toy car and a 1:1 scale real vehicle.

5.1. Experimental Setup

Datasets. To comprehensively validate the effectiveness of our attack method, we construct datasets for both the digital domain and the physical domain.

For the **digital domain dataset**, we use the CARLA simulation environment [6] based on Unreal Engine 4, a popular open-source simulator for autonomous driving scenarios, to construct high-fidelity and photo-realistic urban scenes. The test set for each attack method is created by capturing images with a camera positioned around the vehicle deployed with the corresponding adversarial camouflage. We select two kinds of weather (sunny and cloudy), four distances ($5m, 10m, 15m, 20m$) and five camera pitch angles ($20°, 30°, 40°, 50°, 60°$) because COCO-

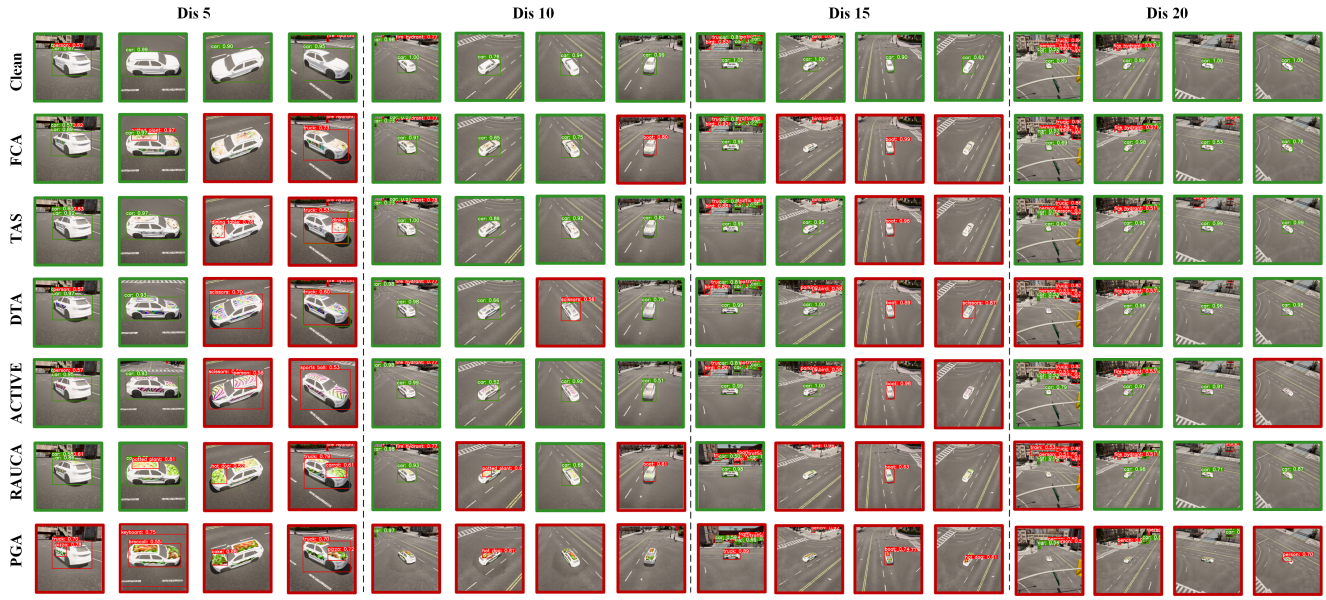


Figure 4. Visualization comparison of multi-view detection results in the digital world. Green-bordered images indicate correct detection of the target vehicle, while red-bordered images indicate either undetected targets or detection with incorrect classification.

pretrained detection models inherently have poor performance at greater distances or larger pitch angles, making those settings less informative for evaluation. For each setting, we conduct 360° surrounding photography at 10° intervals , resulting in 1440 images totally. For the **physical domain dataset**, we deploy a 1:1 scale real vehicle, GOLF Sportsvan, then capture a rotating video using a drone, and extract 282 images to input into the PGA framework to generate camouflage. The camouflage is deployed using stickers. Subsequently, we employ a drone to record videos and extract images analogous to the digital domain dataset to construct a test dataset. We further deploy the adversarial camouflage of PGA and other SOTA methods on a 1:24 scale Audi Q5 model car to conduct additional qualitative and quantitative experiments across diverse scenarios.

Target Models. We select commonly used detection model architectures for the experiments, including one-stage detectors: YOLO-v5; two-stage detectors: Faster R-CNN [44] and Mask R-CNN [16]; as well as transformer-based detectors: Deformable-DETR [66], with all models pre-trained on the COCO dataset.

Compared Methods. We select six state-of-the-art physical adversarial attack methods as our baseline for comparison, including DAS [54], FCA [53], DTA [49], ACTIVE [50], TAS [55] and RAUCA [64].

Evaluation Metrics. To evaluate the effectiveness of various attack methods on detection models, we use AP@0.5(%), following [49, 50, 64], which is a standard measure capturing both recall and precision at a detection IoU threshold of 0.5.

Table 2. Comparison of detection results for different attack methods at various pitch angles, specifically reporting the average AP@0.5 on Faster R-CNN for distances from 5m to 20m under both sunny and cloudy weather conditions.

Angle	20°	30°	40°	50°	60°	Average
-	91.30	87.00	88.04	78.70	65.46	82.10
FCA[53]	61.19	60.61	65.34	52.84	27.54	53.50
DTA[49]	65.11	62.37	57.71	53.09	25.72	52.80
ACTIVE[50]	56.94	60.70	65.27	43.64	11.52	47.61
RAUCA[64]	46.36	43.69	46.72	23.47	9.63	33.97
PGA	**21.01**	**4.62**	**4.11**	**3.90**	**0.00**	**6.73**

Table 3. Ablation study results of various techniques applied in the PGA attack framework.

Cons.	Min-Max	Faster R-CNN	Yolo-v5*	Mask R-CNN*	D-DETR*	Average
	✓	8.05	50.38	16.33	34.50	27.32
✓		10.23	54.40	20.56	36.82	30.50
✓	✓	**3.57**	**47.24**	**11.89**	**28.78**	**22.87**

5.2. Digital Experiments

In this section, we provide a comprehensive comparison of PGA and SOTA methods, demonstrating the advantages of PGA. In these experiments, Faster R-CNN is used as the victim model for white-box attacks, with the adversarial camouflage transferred to other three detectors (marked with *, and * represents the same meaning throughout) to evaluate transferability. Note that we primarily use partial coverage camouflage in this section, following [54, 55]. This setting is more challenging due to the reduced optimization space, yet we adopt it because it greatly facilitates real-world deployment and significantly lowers deployment costs. Additionally, we provide comparative experiments using full-coverage camouflage, where PGA still outperforms other methods; please refer to the Appendix.

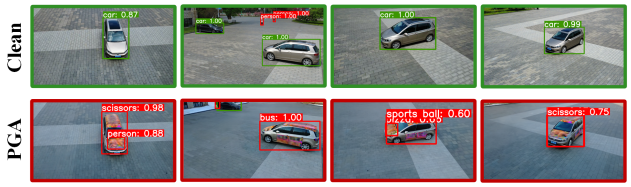


Figure 5. Visualization results from physical experiments on a 1:1 real car. We deploy the PGA adversarial camouflage using stickers and capture images from multiple viewpoints with a drone.

Digital World Attack. We compare the digital attack performance of PGA with SOTA methods across multiple weather conditions, distances, and viewpoints. Although PGA can directly reconstruct and attack using real photos, for a fair comparison with mainstream vehicle physical attack methods, we sample clean vehicle images in CARLA, reconstruct the 3D scene, and then conduct the PGA attack. The results in Tab. 1 show that PGA achieves the best attack performance in all settings, indicating that the generated adversarial camouflage possesses high adversarial strength, high multi-view robustness, and strong transferability.

In addition, we conduct comparative experiments on camera pitch angles in Tab. 2, and the results indicate that PGA consistently outperforms at all angles. We select views ranging from 20° to 60° because detectors pre-trained on COCO perform poorly at higher bird's-eye view angles.

Visualization. We present visualizations in Fig. 4 comparing the detection results of PGA with other attack methods in the digital domain. The results indicate that, compared to other SOTA methods, PGA exhibits superior adversarial effectiveness and multi-view robustness across various distances, pitch angles and azimuth angles. More results under different lighting and weather conditions are provided in the supplementary material.

Ablation Study. We conduct an ablation experiment focusing on the two techniques in PGA, including multi-view camouflage consistency (Cons.) and the min-max optimization framework, with the results shown in Tab. 3. It is apparent that using all two techniques simultaneously achieves the best attack performance. More ablation experiments are provided in the supplementary materials.

5.3. Physical Experiments

1:24 Physical Experiment. We deploy adversarial camouflage generated by various SOTA methods as well as PGA on a 1:24 scale toy car. Images are captured from multiple viewpoints at distances of 50cm and 100cm to construct a physical scene dataset, which is subsequently evaluated using multiple detectors. Quantitative results are provided in Tab. 4, and qualitative results are shown in Fig. 6. These results indicate that PGA's photo-realistic modeling capability and multi-view adversarial robustness can effectively simulate physical environments and mitigate the degradation in adversarial camouflage performance caused by these conditions.

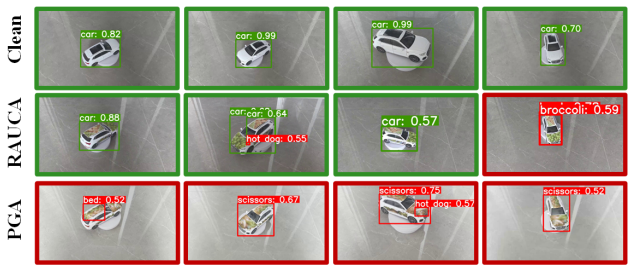


Figure 6. Visualization results from physical experiments on a 1:24 scale simulated car. We compare the attack visualization outcomes of clean samples, RAUCA [64] camouflage samples, and PGA camouflage samples from multiple viewpoints.

Table 4. Comparison results of AP@0.5(%) under physical settings. We deploy adversarial textures generated by different attack methods on a 1:24 scale toy car and capture images from multiple viewpoints at distances of 50cm and 100cm to construct a physical scene dataset for detection.

Dis	Method	Faster R-CNN	YOLO-v5*	Mask R-CNN*	D-DETR*	Average
50cm	-	86.12	90.71	85.36	89.25	87.86
	FCA[53]	66.41	61.37	58.55	59.43	61.44
	DTA[49]	55.58	57.49	56.12	60.98	58.79
	ACTIVE[50]	39.45	52.38	47.31	45.95	46.27
	RAUCA[64]	28.86	50.67	32.09	35.14	36.69
	PGA	**20.94**	**50.25**	**22.35**	**21.25**	**28.69**
100cm	-	90.19	92.95	89.32	93.02	91.37
	FCA[53]	44.16	48.95	49.08	50.24	48.10
	DTA[49]	50.81	48.11	53.02	51.81	50.93
	ACTIVE[50]	40.10	52.35	45.39	49.28	46.78
	RAUCA[64]	34.61	44.14	35.55	34.70	37.25
	PGA	**21.77**	**41.82**	**23.92**	**25.54**	**28.26**

1:1 Physical Experiment. We also apply the PGA framework for 3DGS modeling and adversarial camouflage generation on a 1:1 real vehicle. During drone-based image capture, easy-to-deploy calibration stickers are used to help SAM segment the camouflage areas. We conduct attack on Faster R-CNN, where AP@0.5(%) decreases from 88.48 to 25.67, with qualitative results presented in Fig. 5. Results show that with little manual effort and simple tools (a camera and some printed stickers), PGA can effectively reconstruct and attack real cars, posing a significant threat to autonomous driving safety.

6. Conclusion

In this paper, we propose a novel physical attack framework based on 3D Gaussian Splatting named PGA. Further, we improve the physical adversarial effectiveness and multi-view robustness by improving the cross-viewpoint consistency of the camouflage and using a multi-view robust min-max adversarial camouflage optimization method. Experiments prove that PGA can effectively attack arbitrary objects in both digital and physical domains, even in infrared modality. We hope our work can inspire efforts to improve true robustness in the physical world.

Acknowledgment. Supported by ①Shenzhen Science and Technology Program(KJZD20240903095730039), ②the CCF-NSFOCUS 'Kunpeng' Research Fund(CCF-NSFOCUS 2024003), ③Shenzhen Science and Technology Program(JCYJ20210324102204012) and ④the Fundamental Research Funds for the Central Universities, Sun Yat-sen University under Grants No. 23xkjc010.

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, 5

[2] Tom B Brown, Dandelion Mané, Aurko Roy, Martín Abadi, and Justin Gilmer. Adversarial patch. *arXiv preprint arXiv:1712.09665*, 2017. 1

[3] Yulong Cao, S Hrushikesh Bhupathiraju, Pirouz Naghavi, Takeshi Sugawara, Z Morley Mao, and Sara Rampazzi. You can't see me: Physical removal attacks on {lidar-based} autonomous vehicles driving frameworks. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 2993–3010, 2023. 1

[4] 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

[5] Yao Deng, Xi Zheng, Tianyi Zhang, Chen Chen, Guannan Lou, and Miryung Kim. An analysis of adversarial attacks and defenses on autonomous driving models. In *2020 IEEE international conference on pervasive computing and communications (PerCom)*, pages 1–10. IEEE, 2020. 1

[6] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. Carla: An open urban driving simulator. In *Conference on robot learning*, pages 1–16. PMLR, 2017. 2, 6

[7] Ranjie Duan, Xingjun Ma, Yisen Wang, James Bailey, A Kai Qin, and Yun Yang. Adversarial camouflage: Hiding physical-world attacks with natural styles. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1000–1008, 2020. 2

[8] Kevin Eykholt, Ivan Evtimov, Earlence Fernandes, Bo Li, Amir Rahmati, Chaowei Xiao, Atul Prakash, Tadayoshi Kohno, and Dawn Song. Robust physical-world attacks on deep learning visual classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1625–1634, 2018. 1, 2

[9] Deng-Ping Fan, Ge-Peng Ji, Peng Xu, Ming-Ming Cheng, Christos Sakaridis, and Luc Van Gool. Advances in deep concealed scene understanding. *Visual Intelligence*, 1(1):16, 2023. 1

[10] Weiwei Feng, Baoyuan Wu, Tianzhu Zhang, Yong Zhang, and Yongdong Zhang. Meta-attack: Class-agnostic and model-agnostic physical adversarial attack. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 7787–7796, 2021. 2

[11] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*, 2014. 1

[12] Jindong Gu, Xiaojun Jia, Pau de Jorge, Wenqain Yu, Xinwei Liu, Avery Ma, Yuan Xun, Anjun Hu, Ashkan Khakzar, Zhijiang Li, et al. A survey on transferability of adversarial examples across deep neural networks. *arXiv preprint arXiv:2310.17626*, 2023. 1

[13] 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. 4

[14] 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*, pages 764–772, 2023. 1

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1

[16] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. 7

[17] Yu-Chih-Tuan Hu, Bo-Han Kung, Daniel Stanley Tan, Jun-Cheng Chen, Kai-Lung Hua, and Wen-Huang Cheng. Naturalistic physical adversarial patch for object detectors. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7848–7857, 2021. 1

[18] Zhanhao Hu, Siyuan Huang, Xiaopei Zhu, Fuchun Sun, Bo Zhang, and Xiaolin Hu. Adversarial texture for fooling person detectors in the physical world. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13307–13316, 2022. 2

[19] Zhanhao Hu, Wenda Chu, Xiaopei Zhu, Hui Zhang, Bo Zhang, and Xiaolin Hu. Physically realizable natural-looking clothing textures evade person detectors via 3d modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16975–16984, 2023. 4

[20] Lifeng Huang, Chengying Gao, Yuyin Zhou, Cihang Xie, Alan L Yuille, Changqing Zou, and Ning Liu. Universal physical camouflage attacks on object detectors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 720–729, 2020. 2

[21] Yao Huang, Yinpeng Dong, Shouwei Ruan, Xiao Yang, Hang Su, and Xingxing Wei. Towards transferable targeted 3d adversarial attack in the physical world. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24512–24522, 2024. 3

[22] Xiaojun Jia, Xingxing Wei, Xiaochun Cao, and Xiaoguang Han. Adv-watermark: A novel watermark perturbation for adversarial examples. In *Proceedings of the 28th ACM international conference on multimedia*, pages 1579–1587, 2020. 1

[23] Xiaojun Jia, Sensen Gao, Qing Guo, Ke Ma, Yihao Huang, Simeng Qin, Yang Liu, and Xiaochun Cao. Semantic-aligned adversarial evolution triangle for high-transferability vision-language attack. *arXiv preprint arXiv:2411.02669*, 2024.

[24] Xiaojun Jia, Sensen Gao, Simeng Qin, Tianyu Pang, Chao Du, Yihao Huang, Xinfeng Li, Yiming Li, Bo Li, and Yang Liu. Adversarial attacks against closed-source mllms via feature optimal alignment. *arXiv preprint arXiv:2505.21494*, 2025. 1

[25] Hiroharu Kato, Yoshitaka Ushiku, and Tatsuya Harada. Neural 3d mesh renderer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2

[26] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023. 3, 4

[27] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023. 4

[28] Dehong Kong, Siyuan Liang, and Wenqi Ren. Environmental matching attack against unmanned aerial vehicles object detection. *arXiv preprint arXiv:2405.07595*, 2024. 1

[29] Dehong Kong, Siyuan Liang, Xiaopeng Zhu, Yuansheng Zhong, and Wenqi Ren. Patch is enough: naturalistic adversarial patch against vision-language pre-training models. *Visual Intelligence*, 2(1):1–10, 2024. 1

[30] Alexey Kurakin, Ian J Goodfellow, and Samy Bengio. Adversarial examples in the physical world. In *Artificial intelligence safety and security*, pages 99–112. Chapman and Hall/CRC, 2018. 5

[31] Leheng Li, Qing Lian, and Ying-Cong Chen. Adv3d: generating 3d adversarial examples in driving scenarios with nerf. *arXiv preprint arXiv:2309.01351*, 2023. 3

[32] Jiawei Lian, Shaohui Mei, Shun Zhang, and Mingyang Ma. Benchmarking adversarial patch against aerial detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–16, 2022. 1

[33] Siyuan Liang, Xingxing Wei, Siyuan Yao, and Xiaochun Cao. Efficient adversarial attacks for visual object tracking. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVI 16*, 2020. 1

[34] Siyuan Liang, Xingxing Wei, and Xiaochun Cao. Generate more imperceptible adversarial examples for object detection. In *ICML 2021 Workshop on Adversarial Machine Learning*, 2021.

[35] Siyuan Liang, Longkang Li, Yanbo Fan, Xiaojun Jia, Jingzhi Li, Baoyuan Wu, and Xiaochun Cao. A large-scale multiple-objective method for black-box attack against object detection. In *European Conference on Computer Vision*, 2022.

[36] Siyuan Liang, Baoyuan Wu, Yanbo Fan, Xingxing Wei, and Xiaochun Cao. Parallel rectangle flip attack: A query-based black-box attack against object detection. *arXiv preprint arXiv:2201.08970*, 2022.

[37] Siyuan Liang, Wei Wang, Ruoyu Chen, Aishan Liu, Boxi Wu, Ee-Chien Chang, Xiaochun Cao, and Dacheng Tao. Object detectors in the open environment: Challenges, solutions, and outlook. *arXiv preprint arXiv:2403.16271*, 2024. 1

[38] Aishan Liu, Jun Guo, Jiakai Wang, Siyuan Liang, Renshuai Tao, Wenbo Zhou, Cong Liu, Xianglong Liu, and Dacheng Tao. {X-Adv}: Physical adversarial object attacks against x-ray prohibited item detection. In *32nd USENIX Security Symposium (USENIX Security 23)*, 2023. 1

[39] Tianrui Lou, Xiaojun Jia, Jindong Gu, Li Liu, Siyuan Liang, Bangyan He, and Xiaochun Cao. Hide in thicket: Generating imperceptible and rational adversarial perturbations on 3d point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24326–24335, 2024. 1

[40] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 3

[41] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, Omar Fawzi, and Pascal Frossard. Universal adversarial perturbations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1765–1773, 2017. 4

[42] Liang Muxue, Chuan Wang, Siyuan Liang, Aishan Liu, Zeming Liu, Liang Yang, and Xiaochun Cao. Adversarial instance attacks for interactions between human and object. 1

[43] Kien Nguyen, Tharindu Fernando, Clinton Fookes, and Sridha Sridharan. Physical adversarial attacks for surveillance: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 2023. 1

[44] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015. 7

[45] Mahmood Sharif, Sruti Bhagavatula, Lujo Bauer, and Michael K Reiter. Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition. In *Proceedings of the 2016 acm sigsac conference on computer and communications security*, pages 1528–1540, 2016. 2, 5

[46] Noah Snavely, Steven M. Seitz, and Richard Szeliski. Photo tourism. *ACM Transactions on Graphics*, page 835–846, 2006. 3

[47] Dawn Song, Kevin Eykholt, Ivan Evtimov, Earlence Fernandes, Bo Li, Amir Rahmati, Florian Tramer, Atul Prakash, and Tadayoshi Kohno. Physical adversarial examples for object detectors. In *12th USENIX workshop on offensive technologies (WOOT 18)*, 2018. 2

[48] Jialiang Sun, Wen Yao, Tingsong Jiang, Donghua Wang, and Xiaoqian Chen. Differential evolution based dual adversarial camouflage: Fooling human eyes and object detectors. *Neural Networks*, 163:256–271, 2023. 2

[49] 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. 1, 2, 6, 7, 8

[50] Naufal Suryanto, Yongsu Kim, Harashta Tatimma Larasati, Hyoeun Kang, Thi-Thu-Huong Le, Yoonyoung Hong, Hun-

min 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. 1, 2, 3, 6, 7, 8

[51] Simen Thys, Wiebe Van Ranst, and Toon Goedemé. Fooling automated surveillance cameras: adversarial patches to attack person detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 0–0, 2019. 2

[52] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017. 1

[53] 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. 1, 2, 6, 7, 8

[54] Jiakai Wang, Aishan Liu, Zixin Yin, Shunchang Liu, Shiyu Tang, and Xianglong Liu. Dual attention suppression attack: Generate adversarial camouflage in physical world. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8565–8574, 2021. 2, 6, 7

[55] Jiakai Wang, Xianglong Liu, Zixin Yin, Yuxuan Wang, Jun Guo, Haotong Qin, Qingtao Wu, and Aishan Liu. Generate transferable adversarial physical camouflages via triplet attention suppression. *International Journal of Computer Vision*, pages 1–17, 2024. 1, 2, 6, 7

[56] Ningfei Wang, Yunpeng Luo, Takami Sato, Kaidi Xu, and Qi Alfred Chen. Does physical adversarial example really matter to autonomous driving? towards system-level effect of adversarial object evasion attack. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4412–4423, 2023. 1

[57] Xiaofei Wang, Shaohui Mei, Jiawei Lian, and Yingjie Lu. Fooling aerial detectors by background attack via dual-adversarial-induced error identification. *IEEE Transactions on Geoscience and Remote Sensing*, 2024. 1

[58] Zhibo Wang, Siyan Zheng, Mengkai Song, Qian Wang, Alireza Rahimpour, and Hairong Qi. advpattern: Physical-world attacks on deep person re-identification via adversarially transformable patterns. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8341–8350, 2019. 1

[59] Xingxing Wei, Siyuan Liang, Ning Chen, and Xiaochun Cao. Transferable adversarial attacks for image and video object detection. *arXiv preprint arXiv:1811.12641*, 2018. 1

[60] Tong Wu, Xuefei Ning, Wenshuo Li, Ranran Huang, Huazhong Yang, and Yu Wang. Physical adversarial attack on vehicle detector in the carla simulator. *arXiv preprint arXiv:2007.16118*, 2020. 2

[61] Kaidi Xu, Gaoyuan Zhang, Sijia Liu, Quanfu Fan, Mengshu Sun, Hongge Chen, Pin-Yu Chen, Yanzhi Wang, and Xue Lin. Adversarial t-shirt! evading person detectors in a physical world. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*, pages 665–681. Springer, 2020. 2

[62] Yang Zhang, Hassan Foroosh, Philip David, and Boqing Gong. Camou: Learning physical vehicle camouflages to adversarially attack detectors in the wild. In *International Conference on Learning Representations*, 2018. 1, 2

[63] Yu Zhang, Zhiqiang Gong, Yichuang Zhang, Kangcheng Bin, Yongqian Li, Jiahao Qi, Hao Wen, and Ping Zhong. Boosting transferability of physical attack against detectors by redistributing separable attention. *Pattern Recognition*, 138:109435, 2023. 2

[64] 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. 1, 2, 3, 6, 7, 8

[65] Heran Zhu and Dazhong Rong. Multiview consistent physical adversarial camouflage generation through semantic guidance. In *2024 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2024. 2

[66] Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv preprint arXiv:2010.04159*, 2020. 7