Preprints.org

Article

Not peer-reviewed version

Does Adversarial Camouflage Really Work on Real Objects? An Empirical Study of Full-Coverage Camouflage on a Real Vehicle

Xingyu Di , Wei Cai * , Xin Wang , Zhongjie Yin , Haoran Jia

Posted Date: 30 March 2026

doi: 10.20944/preprints202603.2394.v1

Keywords: physical adversarial attack; adversarial camouflage; object detection; validation

Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 March 2026 doi:10.20944/preprints202603.2394.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Does Adversarial Camouflage Really Work on Real Objects? An Empirical Study of Full-Coverage Camouflage on a Real Vehicle

Xingyu Di, Wei Cai *, Xin Wang , Zhongjie Yin and Haoran Jia

[1] Xi'an Research Institute of High Technology
* Correspondence: caiwei_bu@163.com
† Current address: 710025 Xi'an, Shaanxi Province, People's Republic of China.

Abstract

The robustness of vision-language agents in real-world environments depends critically on the reliability of their underlying object detectors. Adversarial camouflage has emerged as a promising approach for executing multi-view attacks against these detectors, yet its effectiveness on full-scale, complex real-world objects remains largely unverified. Existing physical validations are predominantly limited to scaled models, leaving a significant gap in understanding real-world threats. Building upon prior digital simulations and scaled-model experiments, this study presents the first systematic quantitative evaluation of full-coverage adversarial camouflage applied to an actual vehicle. We transfer textures generated in the digital domain to a real vehicle and conduct extensive outdoor tests under varying lighting conditions and viewing angles, including aerial perspectives. The attack performance is benchmarked against multiple mainstream detectors. Our results reveal a discrepancy between digital and physical effectiveness. While the camouflage exhibits a measurable attack capability in the physical world, its impact is significantly attenuated by factors including texture transfer loss, environmental interference, and detector robustness. By providing empirical data and a detailed analysis of these limiting factors, this work offers actionable insights for designing more resilient vision-language perception systems against physical-world adversarial threats.

Keywords: physical adversarial attack; adversarial camouflage; object detection; validation

1. Introduction

With the advancement of deep learning in recent years, deep neural networks have been widely applied in computer vision. By learning from large-scale labeled data, object detection models can effectively extract features of target objects from input images and accurately predict their categories and locations. These models serve as the foundational technology for various computer vision tasks, including instance segmentation, object tracking, and image captioning [1]. Compared with tasks that only require determining the existence of specific targets, object detection exhibits higher complexity. This complexity stems from additional requirements: beyond verifying object presence, detection demands precise localization via bounding boxes and accurate classification into categories. Furthermore, object detection must often address complex scenarios involving multiple objects, varying sizes, occlusions, and changes in orientation. Distinct from traditional techniques that rely on handcrafted features such as threshold-based edge detection or contour analysis [2], modern intelligent object detection leverages computer vision theories and cutting-edge artificial intelligence algorithms represented by deep learning, which automatically locate and classify objects in images or videos [3].

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 March 2026 doi:10.20944/preprints202603.2394.v1

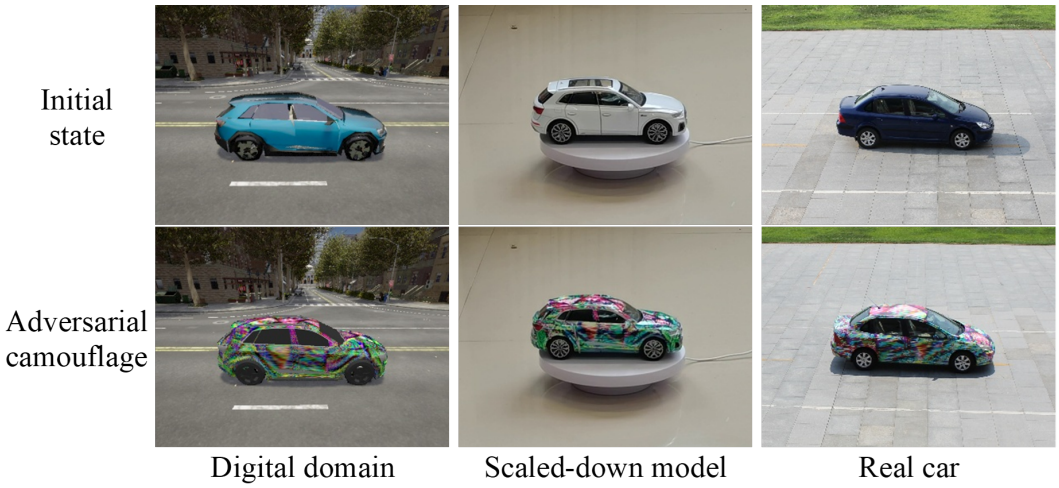

Figure 1. Effect of adversarial camouflage from digital to physical domain. Columns 1–3 show digital simulation, scaled-down model, and real vehicle effects respectively; rows 1–2 present vehicles without perturbation and with adversarial camouflage texture in different forms.

However, deep learning-based algorithms are susceptible to targeted adversarial attacks such as adversarial examples. Adding imperceptible tiny perturbations to input examples can cause well-trained models to fail, leading them to output incorrect predictions with high confidence [4]. This vulnerability stems from the intrinsic nature of deep learning models, which often over-rely on certain features of input data while lacking deep understanding of the data's inherent structure and logic. As research has progressed, researchers have found that such attacks can be implemented not only in the digital domain but also in the physical domain. This poses significant security risks to deep learning models deployed in real-world physical scenarios, including autonomous driving systems[2,5,6] and face recognition systems [7–9].

These object detectors are not merely standalone systems. They constitute the perceptual foundation for a new generation of vision-language agents (VL agents), such as autonomous vehicles and embodied robots, which must interpret and interact with the physical world. The security of these agents, therefore, hinges critically on the robustness of their underlying perception models. Early-stage physical adversarial attacks primarily targeted image classification tasks [10] and were predominantly implemented in the form of patches [11–15]. With the rapid advancement of autonomous driving technology, research on adversarial attacks targeting object detection systems in driving scenarios has gained significant momentum. Among the various attack paradigms developed, many adversarial camouflage attack methods have emerged, including CAMOU [16], DAS [17], FCA [18], DTA [19], and ACTIVE [20].

However, some of these attack methods have only been validated in digital simulation environments. Even for those verified on model vehicles, their effectiveness lacks sufficient persuasiveness. Regarding validation efforts conducted on real vehicles, the adversarial perturbations are primarily implemented in the form of stickers [21]or strips[22]. Notably, studies that validate the effectiveness of full-coverage adversarial camouflage textures on real vehicles remain scarce in the publicly available literature. This scarcity can be attributed to the high costs and substantial time and effort requirements of such schemes, as well as the difficulty in ensuring consistency of external factors such as scene conditions, weather, and lighting during validation. Motivated by this research gap and building upon reference [23], we conduct a full-coverage test of an adversarial camouflage texture on a real vehicle. The primary objective of this work is to investigate the effectiveness of full-coverage adversarial camouflage textures in real-world scenarios on a real object, moving beyond model-based validations as shown in Figure 1. The key contributions of our work are summarized as follows:

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 March 2026 doi:10.20944/preprints202603.2394.v1

- To the best of our knowledge, based on a comprehensive review of publicly available literature, this work represents the first attempt to quantitatively validate the effectiveness of full-coverage adversarial camouflage on real-world vehicles in the physical domain. We provide a scheme to transfer adversarial camouflage textures constructed in the digital domain onto a real-world vehicle, even though this vehicle differs from those employed during the training phase.
- We systematically validate the effectiveness of adversarial camouflage on a real vehicle across diverse real-world conditions. We conduct a detailed analysis of experimental results from both digital and physical domains, with primary focus on the adversarial camouflage applied to real vehicles.
- We provide empirical data and actionable insights for designing more robust vision-language perception systems by identifying and analyzing the key factors that undermine physical attack effectiveness, including texture transfer loss, environmental interference, and detector robustness.

2. Related Work

Adversarial examples are data samples that cause deep neural networks to make erroneous predictions with high confidence after applying human-imperceptible perturbations [24]. Based on implementation domain, adversarial attacks can be categorized into digital attacks and physical attacks. Digital attacks apply pixel-level perturbations in digital space and are mainly divided into gradient-based attacks (e.g., FGSM [25], PGD [26], MI-FGSM [27]) and optimization-based attacks (e.g., C&W [28], DeepFool [29]). These attacks can effectively bypass deep learning models, with some adversarial examples exhibiting cross-model transferability.

Physical attacks extend adversarial attacks to real-world scenarios, requiring deployment that addresses environmental noise, natural transformations, and physical space constraints [30]. Methods such as expectation over transformations [31] and robust physical perturbations [32] incorporate physical factors to enhance perturbation robustness. Early physical attacks were primarily implemented as adversarial patches [33], but single-view patches are susceptible to occlusions, viewing angle changes, and illumination variations.

To implement robust multi-view attacks, researchers have constructed adversarial perturbations on 3D models. CAMOU proposes iterative optimization for adversarial textures on 3D vehicles with complex shapes. MeshAdv [34] adopts the neural mesh renderer [35] as a differentiable renderer, using 3D shapes as the core carrier for adversarial perturbations. DAS disperses model attention from target regions to non-target regions via connected graphs, generating adversarial camouflage with semantic correlation to scenario context. FCA extends texture generation from local stickers to full-coverage textures on vehicle bodies. Cai et al. [23] proposed a data augmentation-based method that improves adversarial camouflage robustness. DTA and ACTIVE employ differentiable transformation networks to learn scene attributes from photorealistic renderers, enhancing image authenticity. With advances in 3D reconstruction, TT3D [36] leverages grid-based NeRF for targeted 3D adversarial examples with transferability. AdvNeRF [37] generates 3D adversarial meshes deceiving both visual and LiDAR perception. PGA [38] employs 3D Gaussian splatting for rapid scene reconstruction while addressing multi-view camouflage inconsistency. Our work focuses on evaluating adversarial camouflage textures on real objects to explore practical implications.

3. Digital to Physical Domain Mapping

3.1. Overview

Conducting adversarial camouflage experiments directly on real vehicles involves numerous practical challenges that greatly limit experimental scalability. First, the extremely high cost of full-vehicle wrapping restricts how many experimental iterations can be completed under standard research budgets. Second, complicated operational procedures such as vehicle preparation, texture application, and outdoor data collection lead to significant time consumption. Third, difficulties in controlling environmental variables including lighting, weather, and background clutter reduce the

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 March 2026 doi:10.20944/preprints202603.2394.v1

controllability and reproducibility of experiments, which in turn lowers the statistical reliability of the results. Moreover, a critical challenge arises from the mismatch between the 3D vehicle models used during adversarial texture training and the physical vehicle available for real-world testing. In our case, the adversarial camouflage was originally optimized for a specific vehicle model during the digital training phase (as detailed in reference [23]), but the actual test vehicle (Peugeot 307 sedan) differs in geometric structure, surface topology, and UV parameterization. This domain gap necessitates a systematic texture transfer methodology that preserves adversarial effectiveness while adapting to the target vehicle's physical characteristics.

To address these challenges, we develop a complete digital-to-physical domain mapping framework illustrated in Figure 2. The framework comprises two primary modules: (1) an image processing module for cross-vehicle texture transfer, which handles the geometric adaptation between different vehicle models; (2) a super-resolution reconstruction module that upscales low-resolution texture patterns to high-fidelity, print-ready designs matching real vehicle dimensions. This systematic approach enables the practical deployment of adversarial camouflage on physical objects while minimizing information loss during the domain transfer process.

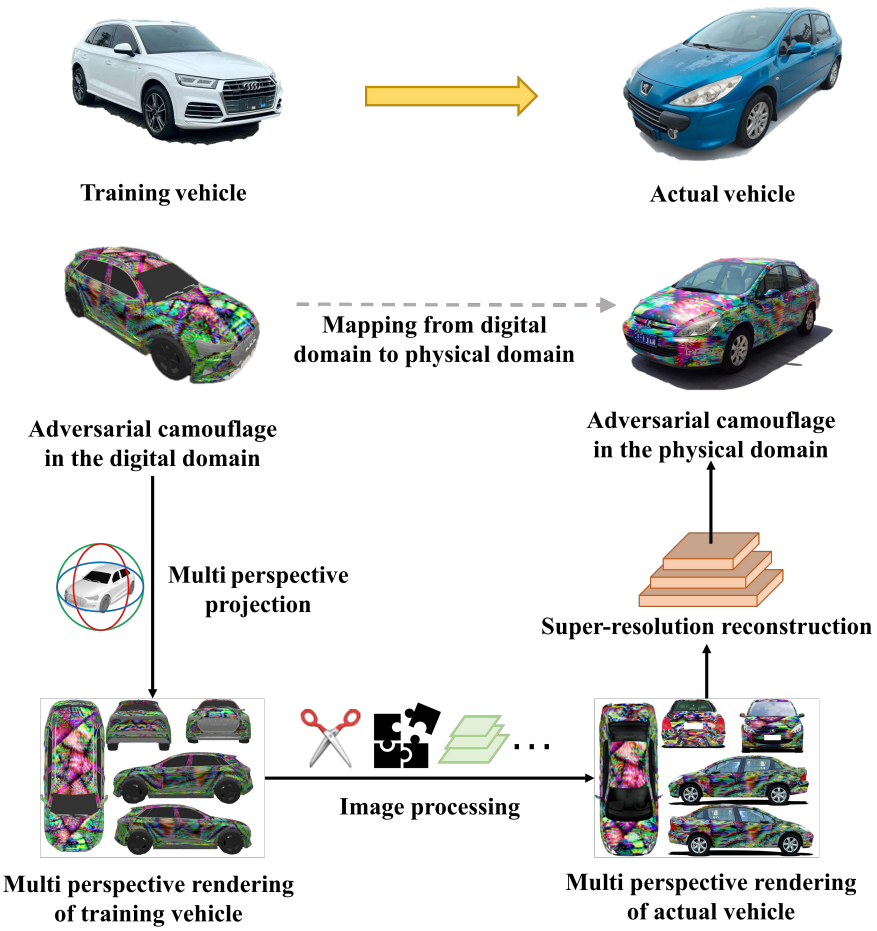

Figure 2. Framework of adversarial camouflage texture transfer to different vehicles. This framework comprises two core modules. The image processing module transfers adversarial camouflage textures from one vehicle model to another, and the super-resolution reconstruction module reconstructs small-sized scaled design drawings into high-definition real vehicle proportional design drawings.

3.2. Texture Transfer

The texture transfer process addresses the fundamental challenge of adapting adversarial patterns optimized for one vehicle geometry to a physically different target vehicle. This adaptation must preserve both the adversarial information encoded in the texture and the visual continuity required for

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 March 2026 doi:10.20944/preprints202603.2394.v1

seamless application. We decompose this problem into three sequential operations performed in order: semantic alignment, regional scaling, and seam optimization.

Semantic Alignment. To establish correspondence between the source vehicle model (used during training) and the target vehicle (physical test platform), we perform preprocessing that aligns key semantic regions. This involves identifying and matching functional components (hood, doors, roof, trunk) between the two vehicle geometries and establishing consistent UV coordinate mappings in these regions. The alignment process standardizes both the directional orientation and proportional scaling of corresponding surface patches. For each semantic region, the UV coordinate transformation is expressed as:

$$\mathbf{u}_t = D \cdot S \cdot \mathbf{u}_s + \mathbf{t} \tag{1}$$

where $\mathbf{u}_t$ and $\mathbf{u}_s$ represent the UV coordinate vectors in the target and source vehicle models respectively, D is the directional consistency matrix (typically encoding rotation or reflection to align texture orientation), $S = \text{diag}(s_u, s_v)$ is the diagonal scaling matrix with factors s_u and s_v for horizontal and vertical directions, and t is the translation offset vector that accounts for region displacement. This formulation ensures that texture patterns maintain proper orientation and proportional scaling when mapped to the target vehicle's UV space.

Regional Scaling. Different vehicle components (e.g., doors, roof panels) exhibit varying dimensional ratios between the source and target vehicles. To maintain pattern integrity while adapting to these geometric differences, we partition the texture into semantic regions and apply independent similarity transformations to each region. For each region, we employ a transformation that combines translation, rotation, and uniform scaling:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = s \cdot R(\theta) \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \tag{2}$$

where $R(\theta)$ is the 2×2 rotation matrix with angle θ, s is the uniform scaling factor, and $(t_x, t_y)$is the translation vector. This region-wise approach prevents global distortions that would otherwise occur from attempting to fit the entire texture with a single transformation.

Seam Optimization. After regional scaling, discontinuities inevitably emerge at the boundaries between differently transformed regions. To ensure visual continuity in the final texture, we employ Poisson image fusion for seamless boundary blending. This technique solves for a smooth intensity field that respects the gradient structure of the source regions while matching boundary conditions from the target image. The optimization is formulated as solving the Poisson equation:

$$\nabla^2 I = \nabla \cdot \mathbf{v} \quad \text{within region } \Omega \tag{3}$$

$$I = I_{\text{target}} \quad \text{on boundary } \partial\Omega \tag{4}$$

where $\mathbf{v}$ represents the gradient field from the source texture regions, and I_{target} specifies the boundary values from the target vehicle texture. This approach produces imperceptible transitions between regions while preserving the fine-grained adversarial patterns within each region.

3.3. Super-Resolution Reconstruction

The adversarial textures generated during the digital training phase are optimized at resolutions corresponding to the neural renderer's output dimensions, which are typically designed for efficient gradient computation rather than high-fidelity printing. Furthermore, the texture patterns generated for scale model validation (as in reference [23]) have spatial dimensions approximately 1:24 relative to a full-size vehicle. Directly scaling these low-resolution patterns to real vehicle dimensions would result in severe quality degradation—visible pixelation, loss of fine texture details, and blurred adversarial perturbations—all of which compromise both visual fidelity and attack effectiveness. To address this resolution gap, we adopt SwinIR [39], a state-of-the-art image super-resolution model based

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 March 2026 doi:10.20944/preprints202603.2394.v1

on Swin Transformer architecture. SwinIR effectively reconstructs high-frequency details through hierarchical feature learning with shifted window attention mechanisms, making it particularly suitable for preserving the fine-grained adversarial perturbations essential to our attack. The super-resolution pipeline can be formulated as:

$$I_{\text{high-res}} = \text{SwinIR}(\text{Upsample}(I_{\text{transferred}}, S_{vh})) \tag{5}$$

where $I_{\text{transferred}}$ is the texture obtained from the transfer process (Section 3.2), $\text{Upsample}(\cdot, S_v h)$ performs bicubic upsampling to match the physical scale $S_v h$ of the target vehicle (dimensions required for actual wrapping), and $\text{SwinIR}(\cdot)$ applies deep learning-based super-resolution for detail reconstruction and artifact suppression. This two-stage approach: first upsampling to target resolution, then applying learned refinement, produces the final high-resolution texture $I_{\text{high-res}}$ suitable for professional printing at true vehicle scale. This process is critical for maintaining adversarial information integrity when transitioning from digital simulation to physical fabrication.

3.4. Printability Constraints

Consumer-grade printers operate within limited color gamuts defined by their CMYK ink systems, which cannot reproduce all colors representable in digital RGB space. Adversarial textures optimized purely in the digital domain may contain colors outside this printable range, leading to unintended color shifts during physical fabrication that degrade attack effectiveness. To ensure accurate color reproduction, we incorporate the Non-Printable Score (NPS) metric [33] into our texture validation pipeline. The NPS quantifies how far each pixel's color deviates from the nearest printable color in the printer's gamut. For a given texture T with pixel set p_{texture}, the score is computed as:

$$\text{NPS} = \sum_{p_{\text{texture}} \in T} \min_{c_{\text{printable}} \in C} \| p_{\text{texture}} - c_{\text{printable}} \| \tag{6}$$

where C represents the set of colors achievable by the target printer, and $|| \cdot ||$ denotes the Euclidean distance in color space (typically CIE Lab space for perceptual uniformity). A high NPS indicates substantial deviation from printable colors, signaling high risk of color distortion during fabrication. Conversely, a low NPS confirms that the texture lies predominantly within the printer's reproducible range.

In our implementation, we evaluate the NPS of the transferred and super-resolved texture before finalizing the print design. This validation step provides quality assurance that the physical manifestation will closely match the digital design, thereby minimizing adversarial information loss during the digital-to-physical transition. When necessary, color adjustments can be made to reduce the NPS while preserving adversarial effectiveness to the extent possible.

4. Experiments

4.1. Experimental Setup

We apply the adversarial camouflage texture to a Peugeot 307 sedan as the test platform. The texture carrier is PVC matte color-change film, selected specifically for its low-reflectivity properties that reduce specular highlights under natural sunlight. Specular reflection would otherwise introduce spurious high-frequency noise that dilutes adversarial patterns, potentially compromising attack effectiveness.

Data Collection. All experiments were conducted under outdoor natural lighting conditions to capture authentic environmental variability. We designed a data collection plan to ensure comprehensive coverage of viewing angles and distances. The vehicle was positioned with its front orientation sequentially aligned to four cardinal directions (north, south, east, west). For each orientation, we captured images from multiple viewpoints: (1)Ground-level circular sampling: Images were captured while circumnavigating the vehicle at two distance levels (close range and far range); (2)Elevated

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 March 2026 doi:10.20944/preprints202603.2394.v1

viewpoint: Additional images were captured from a rooftop position at approximately 2 story elevation to simulate overhead surveillance scenarios. Figure 3 shows the shooting settings in a real scene. A total of 394 images were collected from multiple angles, including both clean samples (vehicle without adversarial camouflage) and adversarial samples (vehicle with adversarial camouflage). Examples of the collected images showing the vehicle before and after the application of the full-coverage adversarial camouflage are presented in Figure 4.

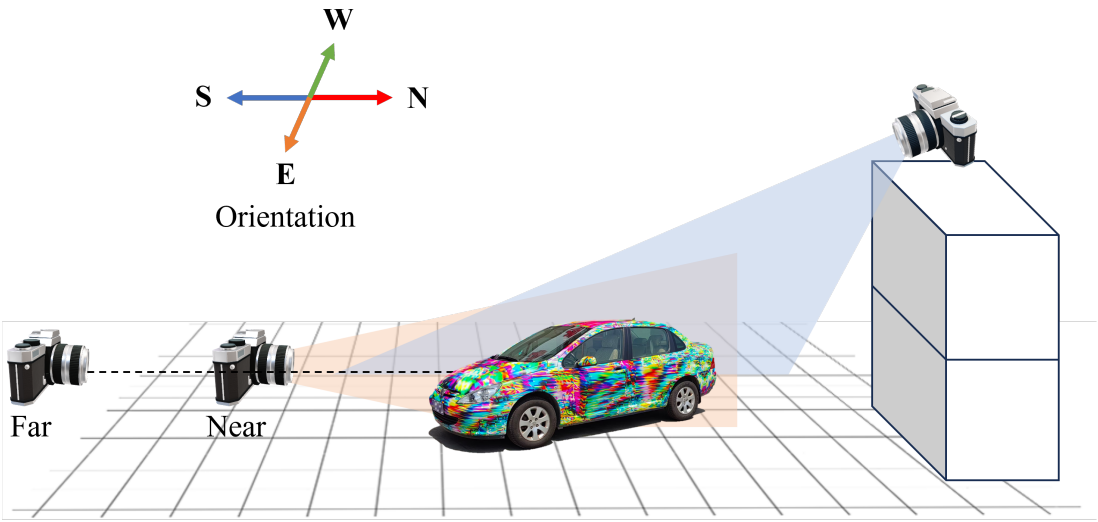

Figure 3. Collect images of the real vehicle with adversarial camouflage in real scenarios. The front of the vehicle faces east, south, west and north respectively.

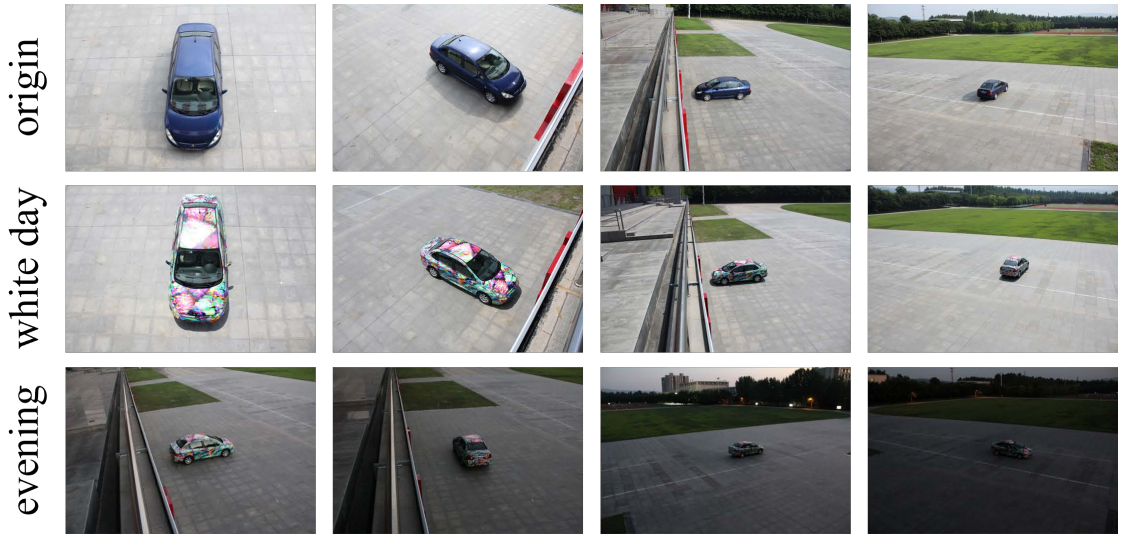

Figure 4. Dataset collected from the real vehicle before and after the adversarial camouflage attached.

Evaluated Detectors. We selected five state-of-the-art object detection models spanning different architectural paradigms: YOLOv5, YOLOv8 and YOLO11(representing single-stage anchor-based detectors at different maturity levels) [40], SSD (multi-scale feature pyramid approach) [41], Faster R-CNN (two-stage region-based method) [42], and RT-DETR (transformer-based detection) [43]. All models were evaluated using publicly available pre-trained weights from the MS COCO dataset [44] without fine-tuning, ensuring that results reflect the models' general object detection capabilities rather than domain-specific adaptations. This detector selection provides broad coverage of contemporary detection paradigms and facilitates insights into the transferability of our adversarial camouflage across architectural families. The confidence threshold is set to 0.25, and the non-maximum suppression (NMS) threshold is set to 0.5.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 March 2026 doi:10.20944/preprints202603.2394.v1

4.2. Evaluation Metrics and Main Results

4.2.1. Evaluation Metrics

Due to the temporal gap between capturing clean and adversarial sample sets (approximately one week, accounting for film production and installation), as well as inherent environmental variability (dynamic lighting, camera positioning uncertainties), perfect pixel-by-pixel correspondence between clean and adversarial images is not achievable. Consequently, we adopt aggregate detection performance metrics that characterize overall model behavior across each dataset:

$$P@0.5 = \frac{TP}{TP + FP} \tag{7}$$

where TP is the count of true positive detections and FP is the count of false positive detections.

$$\text{MR} = \frac{FN}{TP + FN} \tag{8}$$

where FN is the count of false negatives (missed detections).

Average Precision (AP) is defined as the area enclosed under the precision-recall (P-R) curve in object detection tasks, and serves as a metric for quantifying the overall detection accuracy of the model at different recall levels. These metrics provide complementary perspectives on attack impact: P@0.5 captures the detector's tendency to produce spurious or mislocalized detections, while MR directly quantifies evasion success. Higher MR and lower AP on adversarial examples (relative to clean samples) indicate effective adversarial perturbation.

4.2.2. Main Results

Table 1. Comparison results before and after adding adversarial camouflage to the real vehicle.

Detector	AP_{50} (%) ↓		MR (%) ↑	
	Origin	Adversarial	Origin	Adversarial
YOLO Series				
YOLOv5	98.80	94.06 (-4.74)	3.87	17.00 (+13.13)
YOLOv8	98.67	95.93 (-2.74)	4.66	10.50 (+5.84)
YOLO11	99.04	97.21 (-1.83)	4.39	7.35 (+2.96)
Other Architectures				
SSD	99.50	98.50 (-1.00)	14.10	12.70 (-1.40)
Faster R-CNN	97.40	92.00 (-5.40)	13.60	10.40 (-3.20)
RT-DETR	99.50	98.43 (-1.07)	0.00	2.50 (+2.50)

Tab. 1 demonstrates architecture-dependent effectiveness of adversarial camouflage on real vehicles. All detectors show reduced AP_{50}, with declines ranging from 1.00% (SSD) to 5.40% (Faster R-CNN), confirming the adversarial impact. For missing rate, YOLO-series detectors exhibit clear vulnerability: YOLOv5 shows the most severe degradation (MR increasing from 3.87% to 17.00%), while newer versions (YOLOv8, YOLO11) demonstrate progressively improved robustness. This aligns with the adversarial texture being optimized against YOLOv3. RT-DETR shows moderate robustness with only a 2.50 percentage point MR increase, likely due to its transformer-based architecture relying on global features rather than local textures. Notably, SSD and Faster R-CNN show decreased MR despite reduced AP_{50}. This counterintuitive result can be attributed to their high baseline MR, architectural differences in feature dependencies, and the one-week temporal gap between clean and adversarial image captures. These findings highlight that adversarial transferability is highly architecture-dependent, with implications for robust multi-model detection systems.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 March 2026 doi:10.20944/preprints202603.2394.v1

(**a**)

(**b**)

Figure 5. The effect of successful and failed attacks on real vehicles in physical domain.

Figure 5 illustrates representative cases of successful and failed attacks. In successful attack cases, detectors exhibit two distinct failure modes: (1) complete detection failure, where the vehicle remains undetected; or (2) semantic misclassification into other categories such as 'vase', 'skateboard', or 'suitcase'. Notably, uncovered vehicle components including windows, wheels, and headlights occasionally serve as residual detection cues, leading to misclassification into related vehicle categories rather than complete evasion. This observation suggests that achieving full-coverage adversarial camouflage on all vehicle surfaces remains challenging while maintaining vehicle functionality.

In failed attack cases, detectors successfully localize the vehicle despite the adversarial texture, though with noticeably reduced confidence scores (e.g., confidence decreasing from 0.90 to 0.74 in Figure 5b). This indicates that adversarial patterns disrupt learned feature representations to some extent, yet the vehicle's overall geometric silhouette combined with uncovered functional components provide sufficient visual cues for detection. The preserved detection capability in these cases aligns with our finding that certain architectures (particularly RT-DETR and newer YOLO versions) demonstrate greater robustness to texture-based perturbations.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 March 2026 doi:10.20944/preprints202603.2394.v1

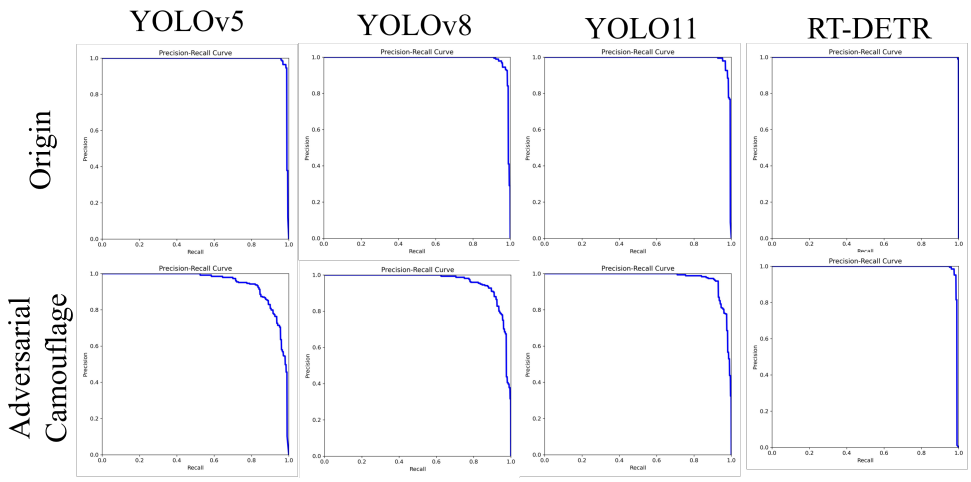

Figure 6. Precision-recall curves comparing detector performance on vehicles with and without adversarial camouflage. The reduced curve area for all detectors under camouflage demonstrates consistent attack effectiveness, with YOLOv5 showing the most significant degradation.

Figure 6 presents precision-recall curves(PR curves) that quantitatively confirm the performance degradation across all tested detectors. The reduced area under the PR curves for adversarial camouflaged vehicles demonstrates consistent attack impact, with YOLOv5 exhibiting the most pronounced degradation.

4.3. Evaluation of Digital-to-Physical Domain Performance

Table 2. Digital-Physical Adversarial Camouflage Performance.

		YOLOv5			YOLOv8			RT-DETR		
		P@0.5 ↓	AP_{50-95} ↓	MR↑	P@0.5 ↓	AP_{50-95} ↓	MR↑	P@0.5 ↓	AP_{50-95} ↓	MR↑
Digital	Ori	78.39	59.52	28.78	77.14	57.55	29.20	83.43	70.71	13.09
	Adv	52.12	23.31	59.50	50.73	22.88	63.20	57.49	24.61	63.27
Model	Ori	97.40	63.40	31.56	90.31	60.17	39.66	99.84	57.04	32.78
	Adv	45.13	46.81	19.84	41.83	35.04	47.29	98.58	34.43	62.78
Real	Ori	98.83	88.10	3.87	97.34	87.85	4.66	97.59	88.30	0.00
	Adv	93.44	86.14	17.00	91.46	88.40	10.50	97.98	90.72	2.50

We systematically evaluate adversarial camouflage effectiveness across three experimental settings: digital simulation, scaled physical models, and full-scale real vehicles. Tab. 2 summarizes the comparative results. Our experiments reveal a progressive degradation of attack efficacy as adversarial camouflage transitions from digital to physical domains. In digital simulation, the camouflage achieves substantial disruption of detector performance, with miss rates increasing from baseline values of approximately 28-30% to 59-63% across YOLO-series detectors. Physical model experiments demonstrate intermediate attack effectiveness, suggesting successful domain transfer despite geometric and material constraints. However, real vehicle deployment exhibits markedly attenuated performance, with YOLOv5 miss rate increasing only from 3.87% to 17.00%, representing a reduction in attack magnitude of approximately 60% compared to digital simulation.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 March 2026 doi:10.20944/preprints202603.2394.v1

4.4. Evaluation Under Varying Lighting Conditions

Table 3. Attack performance of adversarial camouflage in evening.

	YOLOv5		Faster R-CNN		SSD	
	P@0.5↓	MR↑	P@0.5↓	MR↑	P@0.5↓	MR↑
Adversarial Camouflage	96.4	7.6	87.7	0.5	96.5	0

To assess the influence of ambient illumination, we collected 200 additional images under diminishing natural light from late afternoon to evening twilight. Representative images are shown in Figure 4.

Tab.3 presents evening lighting results. Note that evening metrics use P@0.5, while daytime results in Tab. 1 report AP_{50}; despite this difference, performance trends remain comparable. Under evening lighting, detectors show substantial recovery compared to their daytime adversarial performance. Both YOLOv5 and Faster R-CNN exhibit improved precision and significantly reduced miss rates. SSD demonstrates complete recovery with miss rate dropping to 0. These patterns indicate that diminished illumination weakens adversarial camouflage effectiveness. Reduced lighting prevents the camera from capturing fine-grained adversarial texture details, causing detectors to rely more on coarse geometric features such as vehicle silhouettes, wheels, and windows that are more robust to texture perturbations.

4.5. Drone Aerial Evaluation

Using a DJI Mini 4 Pro drone, we captured 26 overhead images to assess attack performance from aerial perspectives. Representative detection results appear in Figure 7.

(a) YOLOv5 (b) YOLOv9 (c) Faster R-CNN

(d) SSD (e) DETR

Figure 7. Aerial view detection results vary across different detectors.

From overhead viewpoints, detectors produce striking misclassification patterns. YOLOv5 labels the camouflaged vehicle as 'snowboard', YOLOv9 identifies it as 'cell phone', Faster R-CNN detects 'bottle', while DETR also outputs 'cell phone'. Such unrelated category assignments suggest the camouflage successfully interferes with high-level semantic recognition of vehicles viewed from above. The enhanced vulnerability from overhead angles relates to viewing geometry. Top-down perspectives show mainly the roof surface, which carries full adversarial texture coverage, while hiding ground-level features like wheels and grilles that normally aid identification, detectors then rely primarily on top-surface top-surface texture we specifically optimized for attack. Ground-level views, by contrast,

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 March 2026 doi:10.20944/preprints202603.2394.v1

expose unmodified components such as windows, wheels, and bumpers that help maintain detection or at least produce related misclassifications like 'truck' or 'bus'.

Attack effectiveness clearly depends on viewing angle, with overhead positions proving especially problematic. The fact that single-stage, two-stage, and transformer-based detectors all show consistent misclassification behavior points to a fundamental vulnerability rather than an architecture-specific weakness. This matters particularly for aerial surveillance applications where top-down viewing dominates.

5. Discussion

Experimental results from previous sections indicate that although adversarial camouflage on real vehicles in the physical world exhibits a measurable attack effect, this effect is not pronounced and is substantially weaker than attack effects observed in digital domain simulation experiments and model car experiments in laboratory environments as reported in reference [23]. Analysis suggests that this outcome may result from information loss during processing, environmental noise interference, and the strong robustness of detection weights. This performance degradation reveals fundamental challenges in translating digitally-optimized perturbations to physical deployments, with critical implications for vision-based autonomous systems.

The digital-to-physical translation bottleneck stems from multiple factors. Vehicle model discrepancies necessitate texture transfer operations that degrade adversarial information through stretching, cropping, and seam optimization. Super-resolution reconstruction from 1:24 model scale to full-scale vehicles introduces clarity degradation and geometric distortion. More critically, physical instantiation encounters intrinsic limitations: natural illumination variations, material reflectance properties, printing quantization, and environmental noise substantially alter how adversarial textures appear to camera sensors versus simulated renderings. Moreover, detectors trained on MS COCO, a large-scale dataset containing hundreds of thousands of instances across diverse scenes and viewpoints, may inherently exhibit considerable robustness. This may suggest that large-scale heterogeneous training could naturally confer resistance to physical attacks, as models might develop feature representations favoring geometric and structural cues over texture-based patterns.

These findings directly address safety challenges for VL agents operating in physical environments. Unlike digital prompt injections targeting language interfaces, physical camouflage attacks target the perception module grounding agent behavior in real-world observations. The view-dependent vulnerability patterns we observe, especially the pronounced overhead susceptibility, have direct relevance for autonomous vehicles, aerial drones, and robotic systems requiring reliable multi-view perception.

The substantial sim-to-real gap suggests that while digital-domain attacks remain highly effective, real-world deployment substantially attenuates adversarial impacts through environmental interference and model robustness. For emerging VL agents in safety-critical applications, ensuring robust visual perception against intentional physical manipulation becomes essential.

6. Conclusions

In this work, we advanced full-coverage adversarial camouflage from simulation to physical domain verification of real objects for the first time, tested the actual effectiveness of adversarial camouflage on a real vehicle, and collected and evaluated them in various scenarios. Experiments have shown that adversarial camouflage can also affect the performance of object detection models in the real physical world, but this effect is not as significant as in digital domain experiments due to inconsistencies in vehicle models, natural environmental light changes, and imaging losses. We hope our experimental results and analysis provide scheme references and data support for future research on adversarial camouflage applications in the real physical world and further promote the design of robust perception systems for real-world intelligent systems.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 March 2026 doi:10.20944/preprints202603.2394.v1

Author Contributions: Conceptualization, W.C. and X.D.; methodology, X.D.; software, X.W.; validation, X.D. and H.J.; formal analysis, Z.Y.; investigation, X.D.; resources, W.C.; data curation, H.J.; writing—original draft preparation, X.D.; writing—review and editing, X.D.; visualization, Z.Y.; supervision, W.C.; project administration, W.C.; funding acquisition, W.C. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Acknowledgments: The authors sincerely thank all readers for their attention.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zou, Z.; Chen, K.; Shi, Z.; Guo, Y.; Ye, J. Object Detection in 20 Years: A Survey. *Proceedings of the IEEE* **2023**, *111*, 257–276. https://doi.org/10.1109/JPROC.2023.3238524.
2. Zhao, Z.Q.; Zheng, P.; Xu, S.T.; Wu, X. Object Detection With Deep Learning: A Review. *IEEE Transactions on Neural Networks and Learning Systems* **2019**, *30*, 3212–3232. https://doi.org/10.1109/TNNLS.2018.2876865.
3. Jiao, L.; Zhang, R.; Liu, F.; Yang, S.; Hou, B.; Li, L.; Tang, X. New Generation Deep Learning for Video Object Detection: A Survey. *IEEE Transactions on Neural Networks and Learning Systems* **2022**, *33*, 3195–3215. https://doi.org/10.1109/TNNLS.2021.3053249.
4. Szegedy, C.; Zaremba, W.; Sutskever, I.; Bruna, J.; Erhan, D.; Goodfellow, I.; Fergus, R. Intriguing Properties of Neural Networks, 2014, [arXiv:cs/1312.6199]. https://doi.org/10.48550/arXiv.1312.6199.
5. Deng, Y.; Zheng, X.; Zhang, T.; Chen, C.; Lou, G.; Kim, M. An Analysis of Adversarial Attacks and Defenses on Autonomous Driving Models. In Proceedings of the 2020 IEEE International Conference on Pervasive Computing and Communications (PerCom), March 2020, pp. 1–10. https://doi.org/10.1109/PerCom45495.2020.9127389.
6. Zhang, C.; Zhou, L.; Xu, X.; Wu, J.; Liu, Z. Adversarial Attacks of Vision Tasks in the Past 10 Years: A Survey. *ACM Comput. Surv.* **2025**, *58*, 52:1–52:42. https://doi.org/10.1145/3743126.
7. Hu, C.; Li, Y.; Feng, Z.; Wu, X. Toward Transferable Attack via Adversarial Diffusion in Face Recognition. *IEEE Transactions on Information Forensics and Security* **2024**, *19*, 5506–5519. https://doi.org/10.1109/TIFS.2024.3402167.
8. Singh, I.; Araki, T.; Kakizaki, K. Powerful Physical Adversarial Examples Against Practical Face Recognition Systems. In Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW), Waikoloa, HI, USA, January 2022; pp. 301–310. https://doi.org/10.1109/WACVW54805.2022.00036.
9. Vakhshiteh, F.; Nickabadi, A.; Ramachandra, R. Adversarial Attacks Against Face Recognition: A Comprehensive Study. *IEEE Access* **2021**, *9*, 92735–92756. https://doi.org/10.1109/ACCESS.2021.3092646.
10. Zhu, Y.; Zhao, Y.; Hu, Z.; Luo, T.; He, L. A Review of Black-Box Adversarial Attacks on Image Classification. *Neurocomputing* **2024**, *610*, 128512. https://doi.org/10.1016/j.neucom.2024.128512.
11. Brown, T.B.; Mané, D.; Roy, A.; Abadi, M.; Gilmer, J. Adversarial Patch, 2018, [arXiv:cs/1712.09665]. https://doi.org/10.48550/arXiv.1712.09665.
12. Hu, Y.C.T.; Chen, J.C.; Kung, B.H.; Hua, K.L.; Tan, D.S. Naturalistic Physical Adversarial Patch for Object Detectors. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, October 2021; pp. 7828–7837. https://doi.org/10.1109/ICCV48922.2021.00775.
13. Wang, Y.; Lv, H.; Kuang, X.; Zhao, G.; Tan, Y.a.; Zhang, Q.; Hu, J. Towards a Physical-World Adversarial Patch for Blinding Object Detection Models. *Information Sciences* **2021**, *556*, 459–471. https://doi.org/10.1016/j.ins.2020.08.087.
14. Wei, X.; Yu, J.; Huang, Y. Infrared Adversarial Patches with Learnable Shapes and Locations in the Physical World. *International Journal of Computer Vision* **2024**, *132*, 1928–1944. https://doi.org/10.1007/s11263-023-01963-y.
15. Zhu, X.; Li, X.; Li, J.; Wang, Z.; Hu, X. Fooling Thermal Infrared Pedestrian Detectors in Real World Using Small Bulbs. *Proceedings of the AAAI Conference on Artificial Intelligence* **2021**, *35*, 3616–3624. https://doi.org/10.1609/aaai.v35i4.16477.
16. Zhang, Y.; Foroosh, H.; David, P.; Gong, B. CAMOU: Learning Physical Vehicle Camouflages to Adversarially Attack Detectors in the Wild. In Proceedings of the International Conference on Learning Representations, September 2018.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 March 2026 doi:10.20944/preprints202603.2394.v1

17. Wang, J.; Liu, A.; Yin, Z.; Liu, S.; Tang, S.; Liu, X. Dual Attention Suppression Attack: Generate Adversarial Camouflage in Physical World. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, June 2021; pp. 8561–8570. https://doi.org/10.1109/CVPR46437.2021.00846.
18. Wang, D.; Jiang, T.; Sun, J.; Zhou, W.; Gong, Z.; Zhang, X.; Yao, W.; Chen, X. FCA: Learning a 3D Full-Coverage Vehicle Camouflage for Multi-View Physical Adversarial Attack. *Proceedings of the AAAI Conference on Artificial Intelligence* **2022**, *36*, 2414–2422. https://doi.org/10.1609/aaai.v36i2.20141.
19. Suryanto, N.; Kim, Y.; Kang, H.; Larasati, H.T.; Yun, Y.; Le, T.T.H.; Yang, H.; Oh, S.Y.; Kim, H. DTA: Physical Camouflage Attacks Using Differentiable Transformation Network. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, June 2022; pp. 15284–15293. https://doi.org/10.1109/CVPR52688.2022.01487.
20. Suryanto, N.; Kim, Y.; Larasati, H.T.; Kang, H.; Le, T.T.H.; Hong, Y.; Yang, H.; Oh, S.Y.; Kim, H. ACTIVE: Towards Highly Transferable 3D Physical Camouflage for Universal and Robust Vehicle Evasion. In Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, October 2023; pp. 4282–4291. https://doi.org/10.1109/ICCV51070.2023.00397.
21. Zhu, X.; Liu, Y.; Hu, Z.; Li, J.; Hu, X. Infrared Adversarial Car Stickers. In Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024, pp. 24284–24293. https://doi.org/10.1109/CVPR52733.2024.02292.
22. Sun, C.; Sun, J.; Zhang, X.; Li, Y.; Bai, Q.; Sun, H. Physical Strip Attack for Object Detection in Optical Remote Sensing. *IEEE Transactions on Geoscience and Remote Sensing* **2024**, *62*, 1–11. https://doi.org/10.1109/TGRS.2024.3434430.
23. Cai, W.; Di, X.; Jiang, X.; Wang, X.; Gao, W. Vehicle Robust Adversarial Texture Generation Based on Data Augmentation. *Xi Tong Gong Cheng Yu Dian Zi Ji Shu/Systems Engineering and Electronics* **2025**, *47*, 1757–1767. https://doi.org/10.12305/j.issn.1001-506X.2025.06.04.
24. Ren, K.; Zheng, T.; Qin, Z.; Liu, X. Adversarial Attacks and Defenses in Deep Learning. *Engineering* **2020**, *6*, 346–360. https://doi.org/10.1016/j.eng.2019.12.012.
25. Goodfellow, I.J.; Shlens, J.; Szegedy, C. Explaining and Harnessing Adversarial Examples, 2015, [arXiv:stat/1412.6572]. https://doi.org/10.48550/arXiv.1412.6572.
26. Madry, A.; Makelov, A.; Schmidt, L.; Tsipras, D.; Vladu, A. Towards Deep Learning Models Resistant to Adversarial Attacks, 2019, [arXiv:stat/1706.06083]. https://doi.org/10.48550/arXiv.1706.06083.
27. Dong, Y.; Liao, F.; Pang, T.; Su, H.; Zhu, J.; Hu, X.; Li, J. Boosting Adversarial Attacks with Momentum. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, June 2018; pp. 9185–9193. https://doi.org/10.1109/CVPR.2018.00957.
28. Carlini, N.; Wagner, D. Towards Evaluating the Robustness of Neural Networks. In Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP), May 2017, pp. 39–57. https://doi.org/10.1109/SP.2017.49.
29. Moosavi-Dezfooli, S.M.; Fawzi, A.; Frossard, P. DeepFool: A Simple and Accurate Method to Fool Deep Neural Networks. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, June 2016; pp. 2574–2582. https://doi.org/10.1109/CVPR.2016.282.
30. Wei, H.; Tang, H.; Jia, X.; Wang, Z.; Yu, H.; Li, Z.; Satoh, S.; Van Gool, L.; Wang, Z. Physical Adversarial Attack Meets Computer Vision: A Decade Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2024**, *46*, 9797–9817. https://doi.org/10.1109/TPAMI.2024.3430860.
31. Athalye, A.; Engstrom, L.; Ilyas, A.; Kwok, K. Synthesizing Robust Adversarial Examples. In Proceedings of the Proceedings of the 35th International Conference on Machine Learning. PMLR, July 2018, pp. 284–293.
32. Eykholt, K.; Evtimov, I.; Fernandes, E.; Li, B.; Rahmati, A.; Xiao, C.; Prakash, A.; Kohno, T.; Song, D. Robust Physical-World Attacks on Deep Learning Visual Classification. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, June 2018; pp. 1625–1634. https://doi.org/10.1109/CVPR.2018.00175.
33. Sharif, M.; Bhagavatula, S.; Bauer, L.; Reiter, M.K. Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition. In Proceedings of the Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna Austria, October 2016; pp. 1528–1540. https://doi.org/10.1145/2976749.2978392.
34. Xiao, C.; Yang, D.; Li, B.; Deng, J.; Liu, M. MeshAdv: Adversarial Meshes for Visual Recognition. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2019, pp. 6891–6900. https://doi.org/10.1109/CVPR.2019.00706.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 March 2026 doi:10.20944/preprints202603.2394.v1

35. Kato, H.; Ushiku, Y.; Harada, T. Neural 3D Mesh Renderer. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, June 2018; pp. 3907–3916. https://doi.org/10.1109/CVPR.2018.00411.
36. Huang, Y.; Dong, Y.; Ruan, S.; Yang, X.; Su, H.; Wei, X. Towards Transferable Targeted 3D Adversarial Attack in the Physical World. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 24512–24522.
37. Zhang, B.; Li, J.; Shi, Y.; Han, Y.; Hu, Q. AdvNeRF: Generating 3D Adversarial Meshes With NeRF to Fool Driving Vehicles. *IEEE Transactions on Information Forensics and Security* **2025**, *20*, 9673–9684. https://doi.org/10.1109/TIFS.2025.3609180.
38. Lou, T.; Jia, X.; Liang, S.; Liang, J.; Zhang, M.; Xiao, Y.; Cao, X. 3D Gaussian Splatting Driven Multi-View Robust Physical Adversarial Camouflage Generation. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2025, pp. 28752–28762.
39. Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; Timofte, R. SwinIR: Image Restoration Using Swin Transformer. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, October 2021; pp. 1833–1844. https://doi.org/10.1109/ICCVW54120.2021.00210.
40. Jocher, G.; Qiu, J.; Chaurasia, A. Ultralytics YOLO, 2023.
41. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the Computer Vision – ECCV 2016; Leibe, B.; Matas, J.; Sebe, N.; Welling, M., Eds., Cham, 2016; pp. 21–37.
42. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2017**, *39*, 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.
43. Zhao, Y.; Lv, W.; Xu, S.; Wei, J.; Wang, G.; Dang, Q.; Liu, Y.; Chen, J. DETRs Beat YOLOs on Real-time Object Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 16965–16974.
44. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In Proceedings of the Computer Vision – ECCV 2014; Fleet, D.; Pajdla, T.; Schiele, B.; Tuytelaars, T., Eds., Cham, 2014; pp. 740–755. https://doi.org/10.1007/978-3-319-10602-1_48.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.