Proceedings of the 2025 International Conference on Machine Learning and Cybernetics, 13-15 July, 2025

2025 International Conference on Machine Learning and Cybernetics (ICMLC) | 979-8-3315-8736-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICMLC66258.2025.11280227

BALANCING REALISM AND ATTACK EFFICACY: ADVERSARIAL TEXTURE GENERATION FROM AUTHENTIC CLOTHING PATTERNS

DINGKUN ZHOU[1], PATRICK P. K. CHAN[2,*]

[1] School of Future Technologies, South China University of Technology, China
[2] Shien-Ming Wu School of Intelligent Engineering, South China University of Technology, China
E-MAIL: jackkun818@gmail.com, patrickchan@scut.edu.cn

Abstract:

Adversarial attacks against object detectors have often been limited by the use of visually conspicuous or synthetic patterns, reducing their practicality in real-world scenarios. In this work, a novel framework is proposed to generate naturalistic adversarial textures by transforming authentic Southeast Asian clothing designs through a structured pipeline. The method is built upon 2D-to-3D UV mapping, dual K-means clustering for pattern simplification, and differentiable Voronoi-based optimization to enable gradient-based attacks while preserving visual realism. The effectiveness of the approach is evaluated using dynamic human models rendered with 360° viewpoint rotations and varying body geometries based on the SMPL framework. Detection confidence from YOLOv3 is consistently suppressed, with an average value of 0.3966 observed across five body models. Strong generalization is demonstrated under diverse poses and viewpoints, confirming the robustness of the adversarial textures. These findings suggest that visually plausible garments can be exploited to achieve stealthy adversarial effects, and future work may explore extensions to handle challenging lighting conditions and real-time applications.

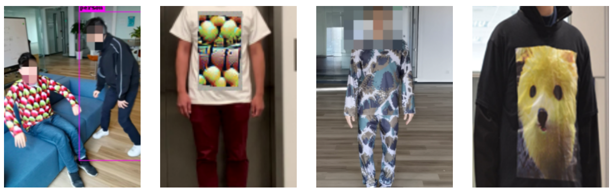
Keywords:

Human detection, Physical-world adversarial attack, Natural-looking textures

1. Introduction

Person detection systems have been extensively deployed in surveillance, access control, and security monitoring applications due to their high accuracy and reliability. These systems, often powered by deep learning models such as Faster R-CNN [13] and YOLO [14], play a crucial role in public safety and automated threat detection. However, recent research [15, 16] has revealed a significant vulnerability. Individuals can evade detection by wearing specially designed adversarial clothing that interfere with the machine learning model's perception [1, 9]. These adversarial examples exploit the weaknesses in visual feature extraction [17] and classification [20], allowing the person to remain undetected by the system. These evasion techniques pose a serious security threat, particularly in critical scenarios, where reliable person detection is essential for safety and operational integrity.

However, the adversarial textures generated by existing evasion methods are often characterized by patterns that appear noticeably unnatural and visually distinct from typical clothing. As illustrated in Figure 1, irregular shapes and colors are commonly featured, causing the designs to stand out from conventional apparel and making them easily noticeable and likely to attract suspicion from human observers.



(a) (b) (c) (d)

FIGURE 1. Unnatural clothing textures generated by existing evasion methods including adversarial patches [9], adversarial T-shirts [1], adversarial textures [3], and universal physical camouflage [4].

In this study, an evasion method is proposed to generate more natural adversarial textures with the goal of reducing visual suspicion. The adversarial patterns are designed to resemble traditional Southeast Asian clothing motifs. A pretrained 2D-to-3D ResNet module is employed to map 2D images into initial 3D mesh textures. Subsequently, a second-order dimensionality reduction algorithm based on K-means is introduced to compress pixel-level image information into a trainable parameter tensor. On this basis, an improved control-point algorithm is applied, which balances the distribution of control points on the mesh with a smoothness constraint. The texture is then updated to ensure that the final design possesses strong adversarial properties while preserving the realistic appearance characteristic of natural clothing.

The key contributions of this work are as follows:

- A novel pipeline is proposed to generate adversarial textures by transforming authentic Southeast Asian clothing patterns, offering a visually natural alternative to previous synthetic or manually designed methods.
- The approach maintains strong attack effectiveness while enhancing robustness through training on diverse human poses and multi-viewpoint rendering, ensuring reliability across real-world conditions.
- Experimental results demonstrate that the generated textures effectively suppress detection confidence and remain visually indistinguishable from natural clothing, overcoming the conspicuousness issue of earlier methods.

The remainder of this paper is structured as follows. Section 2 provides a review of prior physical-world adversarial attacks on person-detection systems and explains the motivation for developing visually plausible garments. In Section 3, we describe our proposed pipeline in detail, which includes 2D-to-3D texture mapping, differentiable Voronoi texture optimization, and pose-diversified training. The experimental setup is outlined in Section 4, followed by both quantitative and qualitative results that illustrate the robustness of our natural-looking adversarial clothing under various poses, viewpoints, and body shapes. We conclude the paper in Section 5, where current limitations are discussed and potential directions for future research are proposed.

2. Related Work

Physical adversarial attacks against person detection systems can be broadly categorized into patch-based and texture-based approaches. Patch-based methods focus on attaching localized adversarial patterns, such as printed patches or clothing designs, to the subject's body. In contrast, texture-based methods aim to generate full-surface adversarial textures that integrate more naturally with garments. This study focuses on texture-based attacks due to their greater potential for maintaining visual realism.

Patch-based attacks were initially demonstrated by optimizing printed patterns to reduce object detector confidence under real-world conditions [9]. Subsequent improvements incorporated fabric deformation modeling using Thin Plate Spline (TPS) transformations to enhance robustness across different body poses and viewpoints, demonstrating effectiveness against both YOLOv2 and Faster R-CNN [1, 13]. More advanced strategies, such as Universal Physical Camouflage (UPC), extended adversarial patterns across entire garments by jointly attacking region proposal and classification stages while applying semantic constraints to improve visual plausibility [4]. Recent work by Cheng et al. addresses the "self-coupling" problem in existing patches through the DePatch approach, which uses decoupled block designs to maintain effectiveness even when parts of the patch are occluded or damaged [10]. Although these approaches achieve high evasion rates, they often result in conspicuous patterns that draw human attention and may fail under significant viewpoint changes.

Texture-based attacks have emerged as a more promising direction for combining adversarial robustness with visual integration. Recent methods such as AdvTexture [7] and TC-EGA [3] use learnable generators and adversarial losses to produce textures that remain effective under geometric transformations and physical deformations. Techniques like AdvCaT [6] introduce topology-aware mappings and camouflage-inspired patterns using Voronoi diagrams and Gumbel-softmax color assignment to maintain consistency on 3D garments. While these methods improve both effectiveness and realism, many still produce high-contrast, repetitive, or military-style patterns that appear unnatural in everyday scenarios.

These limitations highlight the ongoing challenge of balancing adversarial strength with visual plausibility. To address this, our work proposes a novel approach that generates adversarial textures based on authentic Southeast Asian clothing patterns, aiming to produce designs that are both effective in evading detection and indistinguishable from natural attire.

3 Method

This section presents our adversarial camouflage texture generation pipeline, which transforms authentic clothing patterns into visually natural yet adversarially effective textures. The method consists of five main components: (1) mapping 2D clothing images to 3D mesh textures (Section 3.1), (2) extracting and optimizing control points for differentiable texture manipulation (Section 3.2), (3) generating adversarial textures using soft Voronoi mapping and Gumbel-Softmax sampling (Section 3.3), (4) incorporating diverse human poses to improve robustness (Section 3.4), and (5) defining the adversarial loss function for detection confidence suppression (Section 3.5). The overall pipeline is illustrated in Figure 2.

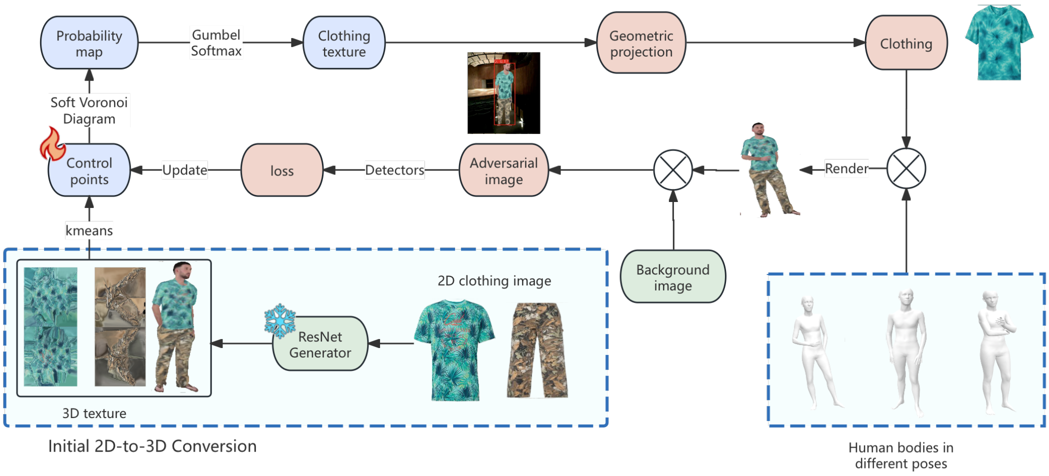


FIGURE 2. Pipeline for generating natural-looking adversarial camouflage textures.

3.1 2D-to-3D Texture Mapping

To project real clothing patterns onto 3D human models, we first perform clothing region extraction and mesh registration, followed by UV-image correspondence generation and texture sampling.

Garment regions is first extracted from 2D images using GrabCut segmentation [11]. These masks are then used to guide non-rigid deformation of a standard SMPL mesh to fit the 2D contours. Registration is performed through two-stage optimization. A rigid alignment step is followed by per-vertex deformation with smoothness and coupling constraints. The total energy function is defined as:

$$E_{\text{total}} = \lambda_{\text{sil}} E_{\text{sil}} + \lambda_{\text{pose}} E_{\text{pose}} + \lambda_{\text{shape}} E_{\text{shape}} + \lambda_{\text{lap}} E_{\text{lap}} + \lambda_{\text{coup}} E_{\text{coup}} \quad (1)$$

where E_{sil}, E_{pose}, E_{shape}, E_{lap}, and E_{coup} represent the energy terms for silhouette consistency, pose prior, shape prior, mesh smoothness, and vertex coupling regularity, respectively. Each corresponding weight λ controls the relative importance of its associated term in the overall optimization.

Using the registered mesh, UV-to-image mappings are computed by identifying triangle face membership and barycentric coordinates for each UV point. The 3D positions are interpolated as:

$$P_{\text{3D}}(u, v) = w_0 V_0 + w_1 V_1 + w_2 V_2 \quad (2)$$

These 3D points are projected back to 2D image space via a virtual camera model, providing dense mappings suitable for neural rendering. The final UV texture map is generated by sampling colors from the 2D clothing image using the computed correspondences. Occluded or missing regions are filled using repair masks to ensure complete and continuous texture coverage.

3.2 Control Point Extraction and Optimization

To enable differentiable texture manipulation, we compress high-resolution textures into a set of representative control points using a dual K-means clustering strategy. Pixels are grouped by color into coordinate sets S_k for each color c_k:

$$S_k = (x_i, y_i) \mid I(x_i, y_i) = c_k \quad (3)$$

For each group, spatial K-means clustering is performed to find control point centroids μ_j, minimizing intra-cluster variance:

$$\arg\min_{U} \sum_{j=1}^{N\_k} \sum_{\mathbf{p} \in U_j} |\mathbf{p} - \mu_j|^2 \quad (4)$$

If insufficient control points are obtained, additional points are interpolated between existing ones:

$$\mu_{\text{new}} = \frac{1}{2}(\mu_a + \mu_b) \quad (5)$$

This results in a consistent and differentiable control point representation for each color.

3.3 Differentiable Texture Generation

With extracted control points, adversarial textures are synthesized using soft Voronoi diagrams and Gumbel-Softmax sampling. Each color is associated with N_P control points forming a tensor $\mathcal{T} \in \mathbb{R}^{N_C \times N_P \times 2}$. For each pixel x, the distance-based weight is:

$$w_i(x) = \sum_{j=1}^{N\_P} \exp(-\alpha |x - b_{ij}|^2) \tag{6}$$

Normalized probabilities across all colors are computed as:

$$p_i(x) = \frac{w_i(x)}{\sum_{k=1}^{N_C} w_k(x)} \tag{7}$$

A mean filter is applied to smooth spatial transitions. To discretely assign colors while preserving differentiability, Gumbel-Softmax [12] is applied with interpolated randomness:

$$u_i^{\text{mix}} = \lambda u_i^{\text{fix}} + (1 - \lambda) u_i^{\text{train}} \tag{8}$$

The sampled logit is computed as:

$$\tilde{p} * i(x) = \frac{\exp((\alpha_i(x))/\tau)}{\sum * j = 1^{N_C} \exp((\alpha_j(x))/\tau)} \tag{9}$$

where $\alpha_i(x) = g_i + \ln p_i(x)$. The final pixel color is obtained by weighted combination of color embeddings. Adversarial textures are optimized by minimizing detection confidence:

$$L_{\text{adv}} = \sum_x \text{Conf}_{\hat{i}(x)} \tag{10}$$

Gradients are backpropagated to update control point positions and trainable variables, enabling end-to-end texture optimization.

3.4 Pose Diversification for Robustness

To ensure robustness under real-world pose variations, we augment training using a set of 3D human models posed in typical configurations. Unlike prior works that evaluate under limited conditions [9, 2, 3], we render textures on human models in multiple postures using Blender, enhancing the generalizability of the adversarial effect.

3.5 Adversarial Loss Function

The adversarial loss function for detection confidence suppression is formulated as:

$$L_{\det} = \sum_x \text{Conf}(x)_{i^*}, \quad i^* = \arg\max_i \text{IoU}\big(\text{gt}(x), b_i(x)\big) \tag{11}$$

where $\text{Conf}(x)_{i^*}$ represents the confidence score output by the YOLO detector for the ground truth bounding box that has the highest IoU with predicted box $b_i(x)$ at pixel location x.

(a) Pose 1 (b) Pose 2 (c) Pose 3 (d) Pose 4 (e) Pose 5

FIGURE 3. The human poses considered in training.

4 Experiments

We evaluate texture robustness and effectiveness through dynamic video validation experiments across continuous viewpoints and multiple human models.

4.1 Experimental Setup

The experiments are conducted on an Ubuntu 24.04 system equipped with an NVIDIA RTX 5090 GPU to accelerate training. The model is configured with $N_P = 600$ control points for texture representation, and a Gumbel-Softmax temperature parameter of $\tau = 0.1$ is used to enable differentiable discrete sampling.

To simulate realistic camera motion, we generate dynamic test videos covering 360° horizontal viewpoints. To evaluate generalization capability, the generated textures are applied to five distinct SMPL models exhibiting diverse body shapes and poses. The primary evaluation metric is the detection confidence score produced by YOLOv3 [14], with an attack considered successful when the confidence falls below 50%.

4.2 Experimental Results and Discussion

4.2.1 Visual Comparison

In contrast to prior approaches [9, 2, 3] that produce visually conspicuous patterns, our method leverages a systematic pipeline to transform authentic Southeast Asian clothing textures into effective adversarial patterns. As shown in Figure 4, the generated textures closely resemble the original garments, preserving visual realism while inducing strong adversarial effects. This demonstrates the effectiveness of our approach in balancing attack performance with visual naturalness, successfully mitigating the trade-off encountered in earlier works.

TABLE 1. Results of Our Model on Different Poses

Model	Average Confidence	Average IoU
Pose 1	0.5376	0.6935
Pose 2	0.4164	0.7362
Pose 3	0.2685	0.6773
Pose 4	0.2668	0.8468
Pose 5	0.4939	0.6386
Average	0.3966	0.7185

(a) Original Image (b) Adversarial Texture

FIGURE 4. Comparison between the original image (a) and the adversarial texture (b) generated using our pipeline.

4.2.2 Influence on Detection Confidence

Table 1 presents the average detection confidence across five body models observed over 360° horizontal viewpoints. The results indicate that our generated textures consistently suppress YOLOv3 detection confidence below 0.55 across all models. Notably, the strongest suppression is observed on Body 2 and Body 3, both falling below 0.3, highlighting the method's ability to generalize across varying body geometries. Detailed quantitative results are summarized in Table 1, where our approach achieves an overall average detection confidence of 0.3966. The lowest detection confidences are recorded on Body 3 (0.2668) and Body 2 (0.2685), further confirming the robustness of the attack across different model configurations.

Figure 5 illustrates the variation in detection confidence across 360° rotations for all body models. While occasional peaks are observed at specific angles, particularly when the front or back silhouette of the body is most prominently exposed, the detection confidence remains below the 0.5 threshold for the majority of viewpoints. This indicates that our generated textures are not only effective in reducing detectability overall but also exhibit strong robustness to changes in viewpoint, consistently deceiving the object detector under dynamic conditions. The observed fluctuations in the confidence curves highlight the intricate interplay between the adversarial effect, the object's 3D structure, and the viewing angle. These insights point to promising directions for further refinement and optimization of adversarial texture generation.

5 Conclusion

This paper presents a novel method for adversarial texture generation that effectively mitigates the trade-off between attack success and visual naturalness. By transforming authentic Southeast Asian clothing textures through a systematic pipeline that integrates 2D-to-3D mapping, dual K-means clustering, and differentiable Voronoi optimization, our method produces adversarial patterns that remain visually plausible.

Extensive experiments demonstrate the robustness of our approach across varying human poses and 360° viewpoints, achieving an average YOLOv3 detection confidence of 0.3966 across five distinct body models. Future work will focus on extending the method to handle extreme lighting conditions and meeting the demands of real-time deployment.

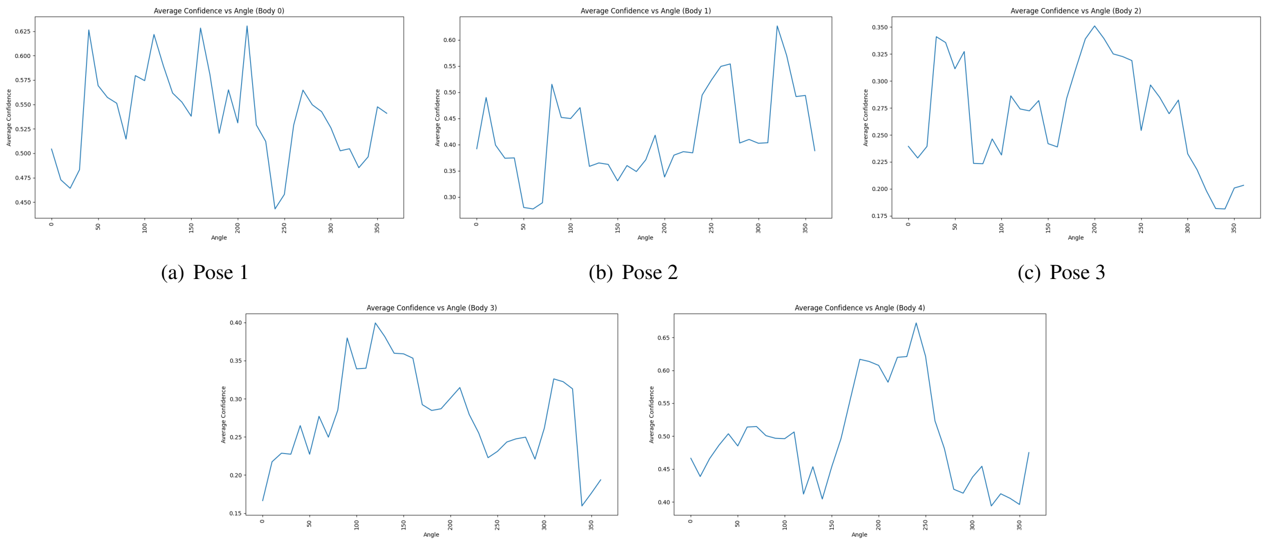
References

[1] 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", arXiv preprint arXiv:1910.11099v3 [cs.CV], July 6, 2020.

[2] 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 (ICCV), 2021, pp. 7828–7837.

[3] 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 (CVPR), 2022, pp. 13297–13306.

[4] Lifeng Huang, Chengying Gao, Yuyin Zhou, Cihang Xie, Alan L. Yuille, Changqing Zou, and Ning Liu, "Universal



(a) Pose 1 (b) Pose 2 (c) Pose 3

(d) Pose 4 (e) Pose 5

FIGURE 5. Detection Confidence Variation Across 360° Viewpoints for Different Poses

Physical Camouflage Attacks on Object Detectors", arXiv preprint arXiv:1909.04326v2 [cs.CV], April 21, 2020.

[5] Aymen Mir, Thiemo Alldieck, and Gerard Pons-Moll, "Learning to Transfer Texture from Clothing Images to 3D Humans," *arXiv preprint arXiv:2003.02050v2 [cs.CV]*, 30 Mar 2020.

[6] 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 (CVPR)*, 2023.

[7] Guixu Lin, Muyao Niu, Qingtian Zhu, Zhengwei Yin, Zhuoxiao Li, Shengfeng He, and Yinqiang Zheng, "Adversarial Attacks on Event-Based Pedestrian Detectors: A Physical Approach," in *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence (AAAI-25)*, 2025.

[8] Anish Athalye, Nicholas Carlini, and David Wagner, "Obfuscated Gradients Give a False Sense of Security: Circumventing Defenses to Adversarial Examples", International Conference on Machine Learning (ICML), pp. 274-283, 2018.

[9] Simen Thys, Wiebe Van Ranst, and Toon Goedemé, "Fooling automated surveillance cameras: adversarial patches to attack person detection," in *Proceedings of the CVPR Workshops on Adversarial Vision Challenge*, 2019.

[10] Mingda Cheng, Haoxuan Qu, Zhiyuan Cheng, Xiaodan Li, and Weiming Zhang, "DePatch: Towards Robust Adversarial Patch for Evading Person Detectors in the Real World," arXiv preprint arXiv:2405.09327, 2024.

[11] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake, ""GrabCut" - Interactive Foreground Extraction using Iterated Graph Cuts," in *ACM SIGGRAPH 2004 Papers*, pp. 309–314, 2004.

[12] Eric Jang, Shixiang Gu, and Ben Poole, "Categorical Reparameterization with Gumbel-Softmax," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2017.

[13] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *Advances in Neural Information Processing Systems (NIPS)*, pp. 91–99, 2015.

[14] Joseph Redmon and Ali Farhadi, "YOLOv3: An Incremental Improvement," arXiv preprint arXiv:1804.02767, 2018.

[15] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus, "Intriguing Properties of Neural Networks," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2014.

[16] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy, "Explaining and Harnessing Adversarial Examples," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2015.

[17] Andrew Ilyas, Shibani Santurkar, Dimitris Tsipras, Logan Engstrom, Brandon Tran, and Aleksander Madry, "Adversarial Examples Are Not Bugs, They Are Features," in *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 125–136, 2019.

[18] Logan Engstrom, Dimitris Tsipras, Ludwig Schmidt, and Aleksander Madry, "A Rotation and a Translation Suffice: Fooling CNNs with Simple Transformations," in *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 1743–1753, 2019.

[19] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, and Pascal Frossard, "DeepFool: A Simple and Accurate Method to Fool Deep Neural Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2574–2582, 2016.

[20] Nicholas Carlini and David A. Wagner, "Towards Evaluating the Robustness of Neural Networks," in *Proceedings of the IEEE Symposium on Security and Privacy (S&P)*, pp. 39–57, 2017.