AdvGrasp: Adversarial Attacks on Robotic Grasping from a Physical Perspective

Xiaofei Wang[1,2], **Mingliang Han**[1], **Tianyu Hao**[3], **Cegang Li**[1], **Yunbo Zhao**[1,4*] and **Keke Tang**[3*]
[1]Department of Automation, University of Science and Technology of China
[2]SmartMore Corporation
[3]Cyberspace Institute of Advanced Technology, Guangzhou University
[4]Institute of Artificial Intelligence, Hefei Comprehensive National Science Center
wxf9545@mail.ustc.edu.cn, mlhan@mail.ustc.edu.cn, howty666@gmail.com,
lcg123@mail.ustc.edu.cn, ybzhao@ustc.edu.cn, tangbohutbh@gmail.com

Abstract

Adversarial attacks on robotic grasping provide valuable insights into evaluating and improving the robustness of these systems. Unlike studies that focus solely on neural network predictions while overlooking the physical principles of grasping, this paper introduces AdvGrasp, a framework for adversarial attacks on robotic grasping from a physical perspective. Specifically, AdvGrasp targets two core aspects: lift capability, which evaluates the ability to lift objects against gravity, and grasp stability, which assesses resistance to external disturbances. By deforming the object's shape to increase gravitational torque and reduce stability margin in the wrench space, our method systematically degrades these two key grasping metrics, generating adversarial objects that compromise grasp performance. Extensive experiments across diverse scenarios validate the effectiveness of AdvGrasp, while real-world validations demonstrate its robustness and practical applicability.

1 Introduction

Grasping serves as a fundamental mechanism for robots to interact with the physical world [Bicchi and Kumar, 2000; Lin *et al.*, 2022], enabling a wide range of applications from industrial automation [Cutkosky and others, 1989] to domestic services [Matheus and Dollar, 2010]. As a critical component of many safety-critical systems, even minor errors in robotic grasping can lead to severe and unpredictable consequences. Recent studies have shown that robotic grasping systems are vulnerable to various threats [Yaacoub *et al.*, 2022], including adversarial attacks [Szegedy *et al.*, 2014], where imperceptible perturbations to input data can cause erroneous decisions. Understanding and addressing these threats, including adversarial attacks, is essential for improving the robustness and reliability of robotic grasping systems.

Despite these developments, research on adversarial attacks in robotic grasping remains relatively limited. Alharthi and Brandão [2024] extended adversarial attack techniques from image-based classification neural networks to image-based grasp quality networks, generating adversarial examples that reduced predicted grasp quality. However, these attacks focus solely on exploiting vulnerabilities in grasp quality networks without addressing the physical grasping process directly. Fundamentally, this approach targets a neural network rather than the physical act of grasping. One limitation is that not all robotic grasping systems rely on neural networks for grasp quality evaluation, which limits the applicability of such attacks. Moreover, even in systems that do use these networks, their instability under adversarial conditions highlights the unreliability of their predictions. Namely, successfully attacking a grasp quality network does not necessarily translate to disrupting the physical grasp itself, see Fig. 1. This highlights the importance of developing adversarial attack strategies that directly target the grasping process, accounting for the physical interactions and challenges inherent in robotic grasping.

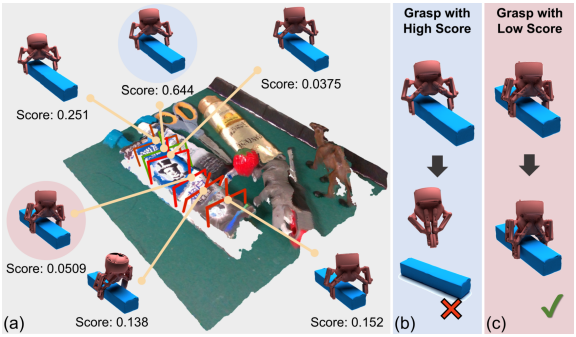

Figure 1: Illustration of the limitations of relying solely on neural network predictions. (a) A scenario from GraspNet-1Billion, where the provided baseline network, i.e., GraspNet, generates multiple grasps with quality scores [Fang *et al.*, 2020; Fang *et al.*, 2023]. However, (b) a predicted grasp with a high-quality score fails in execution, while (c) another grasp with a low-quality score succeeds.

Indeed, robotic grasping is a fundamentally complex physical process rather than a straightforward end-to-end task. It requires intricate interactions involving forces, torques, and friction, which are difficult to model solely with neural networks. Furthermore, a successful grasp involves multiple stages: overcoming gravity to lift the object and maintaining stability throughout transportation, often under the influ-

*Yunbo Zhao and Keke Tang are co-corresponding authors.

ence of external disturbances such as vibrations, collisions, or unexpected forces. These challenges underscore the limitations of adversarial attacks that fail to consider the physical realities of grasping and instead focus exclusively on neural network vulnerabilities. This raises a critical question: can adversarial attacks on robotic grasping be systematically formulated from a physical perspective, capturing these essential interactions and processes?

To address these challenges, we devise AdvGrasp, a systematic adversarial attack framework for robotic grasping from a physical perspective. Specifically, AdvGrasp focuses on two critical aspects of grasp performance: lift capability, which evaluates the gripper's ability to overcome gravity, and grasp stability, which measures its capacity to resist external disturbances. By subtly deforming the object's shape to increase the gravitational torque and reduce the stability margin in the wrench space, our method disrupts the grasp wrench equilibrium, reducing the physical effectiveness of grasping. To validate our approach, we introduce AdvGrasp-20, a comprehensive benchmark featuring 20 groups of objects with diverse shapes and multiple grasp configurations generated by both traditional and neural network-based methods for common two- and three-finger robotic grippers. Extensive experiments across various settings on this benchmark demonstrate the effectiveness of AdvGrasp. Besides, we also validate its performance in real-world scenarios to further confirm its practical applicability in physical robotic systems.

Overall, our contribution is summarized as follows:

- We present AdvGrasp-20, a benchmark with diverse object shapes and grasp configurations to standardize the evaluation of adversarial attacks on grasping.
- We propose a physical-aware adversarial attack framework for robotic grasping that systematically targets two critical metrics: lift capability and grasp stability.
- We validate the proposed framework through extensive experiments and real-world evaluations, demonstrating its effectiveness in degrading grasp performance.

2 Related Work

2.1 Security Issues in Robotics

As robotic systems are applied more broadly in diverse applications [Chen *et al.*, 2015; Siciliano *et al.*, 2008], their interconnected modules, such as communication networks, perception systems, and control mechanisms, have increasingly become targets for various attacks [Yaacoub *et al.*, 2022]. Network security threats, such as data injection and denial-of-service (DoS), have been shown to compromise localization and disrupt operations [Guerrero-Higueras *et al.*, 2018]. Similarly, tampering with sensor data or control signals in human-robot collaboration can lead to unsafe behaviors or task failures [Meng and Weitschat, 2021; Amaya-Mejía *et al.*, 2022]. Operating systems like the Robot Operating System (ROS) are also vulnerable. Dieber *et al.* [2017] identified key weaknesses in ROS, while Mazzeo and Staffa [2020] proposed methods to mitigate privileged-access attacks. Although robotic security has been widely studied, adversarial attacks specifically targeting robotic grasping remain largely unexplored. This paper aims to bridge this gap by systematically investigating such attacks.

2.2 Adversarial Attacks on Robotic Grasping

Adversarial attacks aim to generate perturbed examples that mislead neural networks into making incorrect predictions. Since the introduction of adversarial examples by Szegedy *et al.* [2014], these attacks have been extensively studied in computer vision [Goodfellow *et al.*, 2015; Dong *et al.*, 2018; Carlini and Wagner, 2017; Moosavi-Dezfooli *et al.*, 2016; Tang *et al.*, 2024c] and extended to domains such as 3D point cloud perception [Xiang *et al.*, 2019; Tang *et al.*, 2022; Tang *et al.*, 2023; Tang *et al.*, 2024a; Tang *et al.*, 2024b; Tang *et al.*, 2024d; Tang *et al.*, 2025a; Tang *et al.*, 2025b; Wang *et al.*, 2025], natural language processing [Morris *et al.*, 2020], and audio processing [Zheng *et al.*, 2021]. Recently, robotic grasping has also been investigated within the context of adversarial attacks.

Alharthi and Brandão [2024] extended adversarial attack techniques from image-based classification neural networks to image-based grasp quality networks. Their work demonstrated adversarial examples in robotic grasping by modifying the intensity of a single pixel or physically placing a barely visible spherical object. However, this approach focuses on attacking the grasp quality network itself, rather than addressing the grasping process as a whole. Grasping is inherently a complex physical interaction involving factors such as forces, torques, and friction. Therefore, effective adversarial attacks on robotic grasping must account for these physical characteristics to accurately capture the challenges of the task.

Wang *et al.* [2019] incorporated certain physical properties of grasping in their study, but their objective was to generate objects that are universally difficult to grasp from any angle. This objective significantly increases the complexity of optimization and often leads to adversarial objects with unrealistic and conspicuous deformations. In contrast, our work focuses on attacking specific grasps of an object by systematically investigating adversarial attacks from a physical perspective, while ensuring that the generated perturbations remain realistic and physically plausible.

3 Problem Statement

3.1 Preliminaries

Formally, a 3D object $\mathcal{O}$ is represented as a triangle mesh defined by a set of triangular faces $\mathcal{T}(\mathcal{O})$ and vertices $\mathcal{V}(\mathcal{O})$. A successful grasp on $\mathcal{O}$ is denoted as $\mathcal{GSP} = \{\{\mathbf{c}_i\}, \{\mathbf{f}_i\}\}$, where $\{\mathbf{c}_i\}$ are the contact points, and $\{\mathbf{f}_i\}$ are the forces exerted by the robotic fingers.

At each contact point $\mathbf{c}_i$, the force $\mathbf{f}_i$ can be decomposed into two components:

$$\mathbf{f}_i = \mathbf{f}_i^{\perp} + \mathbf{f}_i^{t},$$

where $\mathbf{f}_i^{\perp}$ is the normal component, and $\mathbf{f}_i^{t}$ is the tangential component. According to the Coulomb's law [Mason *et al.*, 1989], the tangential force is constrained by the normal force:

$$\|\mathbf{f}_i^{t}\| \leq \mu \|\mathbf{f}_i^{\perp}\|,$$

where μ is the coefficient of friction.

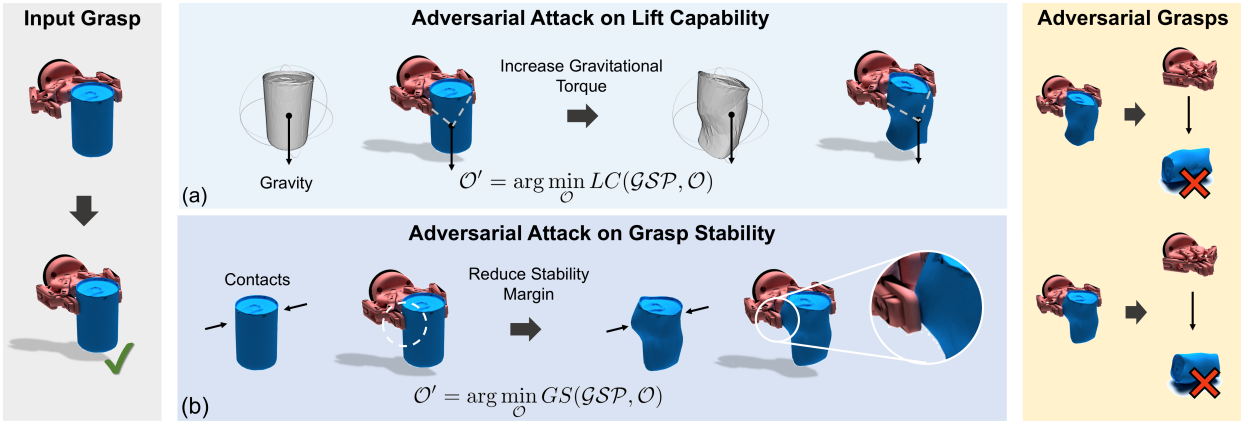

Figure 2: Illustration of AdvGrasp: Given a 3D object and its corresponding grasp configuration as inputs, AdvGrasp deforms the object in two key ways: (a) increasing gravitational torque to weaken lift capacity and (b) reducing the stability margin in the wrench space to compromise grasp stability, ultimately leading to grasp failure.

3.2 Stable Condition of A Grasp

Wrench. A wrench $\mathbf{w} \in \mathbb{R}^{1\times 6}$ is a vector that encapsulates both force and torque, defined as:

$$\mathbf{w} = (\mathbf{f},\ \tau)\,,$$

where $\mathbf{f} \in \mathbb{R}^{1\times 3}$ represents the force vector, and $\tau \in \mathbb{R}^{1\times 3}$ denotes the torque vector acting on the object.

Grasp Wrench Equilibrium. For a grasp $\mathcal{GSP}$ on an object $\mathcal{O}$ to be stable, the wrenches applied by the robotic gripper must counteract the external wrench $\mathbf{w}_{\text{external}}$, which may result from a single or multiple forces. This grasp wrench equilibrium is expressed as:

$$\sum \mathbf{w}_i + \mathbf{w}_{\text{external}} = \mathbf{0}, \tag{1}$$

where $\mathbf{w}_i$ is the wrench exerted by the gripper at contact point $\mathbf{c}_i$. In a coordinate system centered at the object's centroid $\mathbf{z}$, $\mathbf{w}_i$ is defined as:

$$\mathbf{w}_i = (\mathbf{f}_i,\ (\mathbf{c}_i - \mathbf{z}) \times \mathbf{f}_i + \tau_i)\,, \tag{2}$$

For soft finger contact models [Mahler *et al.*, 2018], the rotational torque τ_i satisfies:

$$\|\tau_i\| \leq \gamma \|\mathbf{f}_i^{\perp}\|,$$

where γ is the torsional friction coefficient.

This condition ensures that the external wrench is fully balanced by the combined wrenches applied at the contact points by the robotic gripper, thereby maintaining grasp stability.

3.3 Adversarial Attacks on Robotic Grasping

The goal of adversarial attacks on robotic grasping is to degrade the performance of a given grasp configuration $\mathcal{GSP}$ on an object $\mathcal{O}$ by perturbing it, producing an adversarial version $\mathcal{O}'$. These attacks can be approached from two core aspects of grasp functionality:

- **Lift Capability**: The gripper's ability to overcome gravity and successfully lift the object.
- **Grasp Stability**: The gripper's ability to maintain a secure hold under external disturbances.

By degrading these two core aspects, adversarial attacks effectively compromise robotic grasping performance.

4 Method

In this section, we introduce AdvGrasp, a systematic approach for generating adversarial objects that compromise the lift capability and grasp stability of robotic grasping. We first describe attack methods targeting these two grasp metrics individually, followed by the integrated framework and its implementation. Please refer to Fig. 2 for a demonstration.

4.1 Adversarial Attack on Lift Capability

Lift Capability of A Grasp. The lift capability of a grasp reflects the robotic gripper's ability to counteract gravitational forces and achieve wrench equilibrium. As outlined in Section 3.2, this equilibrium requires the combined wrenches at all contact points to balance the gravitational wrench.

The general condition for wrench equilibrium described in Eqn. 1 can be specified for gravity as:

$$\sum \mathbf{w}_i + \mathbf{w}_{gravity} = \sum (\mathbf{f}_i, (\mathbf{c}_i - \mathbf{z}) \times \mathbf{f}_i + \tau_i) + (m\mathbf{g}, \mathbf{0}) = \mathbf{0}, \tag{3}$$

where m is the mass of the object, and $\mathbf{g}$ is the gravitational acceleration vector.

To quantify lift capability, we adopt the metric as formulated in [Ferrari *et al.*, 1992]:

$$LC(\mathcal{GSP}, \mathcal{O}) = \max_{\mathcal{F}_\perp \in \mathcal{G}(\mathbf{w}_{gravity})} \frac{\|\mathbf{w}_{gravity}\|}{\|\mathcal{F}_\perp\|}, \tag{4}$$

where the set $\mathcal{G}(\mathbf{w}_{gravity})$ contains the normal forces $\mathcal{F}^{\perp}$ needed to balance the gravitational wrench $\mathbf{w}_{gravity}$.

Adversarial Strategy for Lift Capability. We aim to reduce the lift capability LC by perturbing the object's geometry. This optimization is formalized as:

$$\mathcal{O}' = \arg\min_{\mathcal{O}} LC(\mathcal{GSP}, \mathcal{O}), \tag{5}$$

where $\mathcal{O}'$ represents the adversarially modified object. By altering the geometry of $\mathcal{O}$, the gravitational torque is increased, necessitating a greater gripper force to maintain wrench equilibrium. This effectively reduces the object's lift capability LC, making it more challenging for the gripper to lift the object.

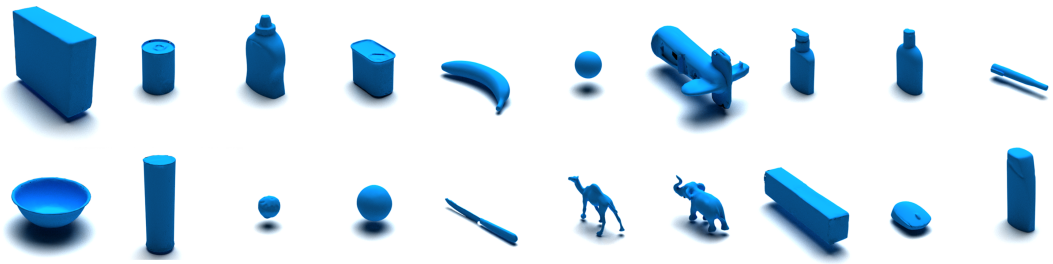

Figure 3: Visualization of the 20 objects in the AdvGrasp-20 benchmark. The first row includes: CRACKER BOX, TOMATO SOUP CAN, MUSTARD BOTTLE, POTTED MEAT CAN, BANANA, RACQUETBALL, TOY AIRPLANE, WASH SOUP, DABAO SOD, BAOKE MARKER. The second row includes: BOWL, CHIPS CAN, STRAWBERRY, ORANGE, KNIFE, CAMEL, LARGE ELEPHANT, DARLIE BOX, MOUSE, SHAMPOO.

4.2 Adversarial Attack on Grasp Stability

Grasp Stability of A Grasp. Grasp stability measures the gripper's ability to resist external disturbances and maintain secure contact with the object throughout the manipulation process. As formulated in Ferrari *et al.* [1992], the stability score is expressed as:

$$GS(\mathcal{GSP}, \mathcal{O}) = \min_{S \in \text{ConvexHull}(\{\mathcal{W}_i\})} \text{Dis}_{\text{p2s}}(\mathbf{0}, S), \quad (6)$$

where $\text{ConvexHull}(\{\mathcal{W}_i\})$ represents the convex hull of the wrenches $\{\mathcal{W}_i\}$ applied at all contact points, and Dis_{p2s} computes the distance between the origin and the convex hull.

Adversarial Strategy for Grasp Stability. We reduce the stability score GS by introducing deformations near the contact regions of the object, altering the normal vectors $\mathbf{n}_i$ at these points. The optimization objective is:

$$\mathcal{O}' = \arg\min_{\mathcal{O}} GS(\mathcal{GSP}, \mathcal{O}). \quad (7)$$

By introducing targeted deformations to the object's shape, the surface normals $\mathbf{n}_i$ are altered, which adjusts the wrench space and reduces the stability margin. This effectively decreases the grasp's resistance to external disturbances, lowering the stability score GS.

4.3 Unified AdvGrasp Framework

Our method provides a unified framework that simultaneously addresses lift capability and grasp stability, ensuring comprehensive adversarial modifications. The optimization objective is defined as:

$$\mathcal{O}' = \arg\min_{\mathcal{O}} LC(\mathcal{GSP}, \mathcal{O}) + \lambda_1 GS(\mathcal{GSP}, \mathcal{O}) + \lambda_2 \,\text{Lap}(\mathcal{O}), \quad (8)$$

where $\text{Lap}(\mathcal{O})$ ensures smooth geometric deformations by penalizing irregularities in the object shape:

$$\text{Lap}(\mathcal{O}) = \sum_{v \in \mathcal{V}(\mathcal{O})} \left(\frac{1}{|\,\text{Neigh}(v)|} \sum_{v' \in \text{Neigh}(v)} (v' - v) \right)^2. \quad (9)$$

Here, $\mathcal{V}(\mathcal{O})$ represents the vertices of $\mathcal{O}$, $\text{Neigh}(v)$ denotes the neighboring vertices of v, and λ_1 and λ_2 are hyperparameters that balance the contributions of grasp stability, lift capability, and deformation regularization, respectively.

4.4 Implementation of AdvGrasp

AdvGrasp generates adversarial objects $\mathcal{O}'$ for robotic grasping through the following three steps:

Bounding Box Initialization. An axis-aligned bounding box is constructed for the object $\mathcal{O}$, providing spatial constraints for defining deformation regions.

Control Point Placement. Control points are evenly distributed across the bounding box surfaces, creating a structured framework for guiding subsequent deformations.

Iterative Shape Deformation and Optimization. Using the control points as anchors, iterative perturbations are applied to the object's geometry following [Ju *et al.*, 2023]. In each iteration, the resolution of control points is progressively increased, allowing for finer adjustments.

The adversarial object $\mathcal{O}'$ is refined to maximize degradation of grasp performance, while ensuring physical plausibility and maintaining imperceptibility of the modifications.

5 AdvGrasp-20 Benchmark

To advance research in the emerging field of adversarial attacks on robotic grasping, we establish a comprehensive benchmark, AdvGrasp-20, designed to facilitate fair comparisons and promote further development in this area.

Object Selection. We select 20 representative objects with diverse shapes from the GraspNet-1Billion dataset [Fang *et al.*, 2020; Fang *et al.*, 2023], including common items such as CRACKER BOX, TOMATO SOUP CAN, and MUSTARD BOTTLE. Please refer to Fig. 3 for a visual demonstration.

Grasp Generation. For each object, we generate 5 grasps for a two-finger gripper using the analytic Dex-Net 2.0 framework [Mahler *et al.*, 2017] and 5 grasps using the deep learning-based GraspNet. In addition, we generate 5 grasps for a three finger gripper using deep learning-based GenDexGrasp [Li *et al.*, 2022].

Post-processing. To ensure the generated grasps are feasible, we further perform post-processing. Specifically, we validate the grasps in the PyBullet simulator using the Robotiq 2-finger 2F-85 gripper and the Robotiq 3-finger gripper. Only the grasps that successfully lift the objects are retained as final grasps, while unsuccessful ones are filtered out.

By including diverse objects and grasp strategies, AdvGrasp-20 provides a robust platform for systematically evaluating adversarial attacks on various grasping configurations. This benchmark not only enables fair comparisons but

Object	2-Finger Grasp								3-Finger Grasp							
	MinGF				MaxLM				MinGF				MaxLM			
	Origin	ALC	AGS	AdvGrasp	Origin	ALC	AGS	AdvGrasp	Origin	ALC	AGS	AdvGrasp	Origin	ALC	AGS	AdvGrasp
CRACKER BOX	32.3	**44.4**	34.8	43.6	2.8	**2.6**	2.6	2.7	16.8	11.2	8.4	12.2	2.8	1.7	2.7	**1.5**
TOMATO SOUP CAN	16.2	17.1	**33.0**	32.0	3.7	**2.6**	3.2	3.3	8.8	17.0	11.6	**29.4**	9.9	4.0	4.1	**3.5**
MUSTARD BOTTLE	22.6	**26.0**	12.8	20.7	3.9	3.9	**3.7**	3.8	12.0	13.2	10.0	**14.5**	5.7	**3.6**	4.6	3.8
POTTED MEAT CAN	10.6	**20.6**	11.6	11.9	3.9	3.8	3.8	**3.8**	10.6	**19.4**	9.6	10.0	4.6	3.8	**2.9**	3.2
BANANA	16.6	**36.7**	27.4	29.9	3.7	**1.9**	4.0	3.3	11.0	12.6	17.6	**20.1**	9.6	6.0	**2.1**	4.9
BOWL	32.5	**33.4**	27.2	31.3	3.4	3.1	**3.1**	3.2	5.8	**16.6**	6.0	7.8	11.8	11.2	7.1	**6.4**
CHIPS CAN	27.9	34.0	**34.7**	28.5	3.3	**2.4**	2.5	3.0	16.8	10.7	8.8	**11.3**	5.8	3.5	**2.4**	3.5
STRAWBERRY	8.3	31.7	22.8	**33.2**	4.3	**4.1**	4.3	4.1	10.8	12.2	**15.4**	11.8	6.4	**1.5**	2.0	2.1
ORANGE	25.4	**33.6**	26.3	27.8	2.4	**1.6**	2.4	1.8	10.8	**12.6**	10.8	10.8	4.9	**1.8**	1.9	2.1
KNIFE	9.9	30.4	37.9	**39.9**	4.0	4.2	3.7	**3.5**	11.0	11.6	11.2	**12.8**	4.0	4.0	3.7	**3.7**
RACQUETBALL	21.9	7.4	14.6	**20.8**	3.9	5.0	4.1	**2.1**	10.6	21.2	24.3	**36.0**	4.9	2.0	2.4	**1.6**
TOY AIRPLANE	29.2	11.9	**22.1**	14.6	3.6	4.6	3.5	**1.8**	22.0	49.6	17.9	**50.0**	3.7	3.7	**2.0**	2.0
WASH SOUP	8.5	**32.8**	24.9	26.0	3.0	2.9	**2.2**	3.1	12.2	**19.6**	13.8	14.2	3.3	1.4	**1.4**	1.8
DABAO SOD	27.9	**34.1**	32.5	32.8	3.7	**3.0**	3.2	3.7	10.8	26.2	**31.9**	14.0	3.4	**1.9**	2.6	2.6
BAOKE MARKER	17.4	26.4	23.8	**33.9**	3.9	**3.3**	3.9	3.9	18.2	18.0	16.2	**19.8**	2.8	**1.9**	2.0	2.1
CAMEL	15.5	19.8	**22.7**	9.7	4.0	**3.7**	3.9	3.9	6.4	**12.0**	5.3	10.4	7.2	6.5	**3.6**	4.3
LARGE ELEPHANT	16.8	**27.0**	23.8	19.5	3.9	4.0	**3.8**	3.9	12.2	**26.4**	19.6	23.4	5.3	2.6	2.9	**2.3**
DARLIE BOX	14.8	26.6	26.7	**26.8**	3.8	**2.6**	3.1	3.0	6.0	19.8	16.2	**24.6**	5.3	**1.4**	1.6	1.8
MOUSE	20.7	**27.3**	22.8	25.4	3.6	**1.5**	2.4	1.9	13.2	26.6	25.2	**27.0**	4.5	**4.2**	4.2	4.4
SHAMPOO	21.9	24.9	**29.6**	14.0	3.3	2.3	**1.9**	2.2	10.8	27.1	25.6	**38.8**	9.6	**3.1**	6.8	4.5

Table 1: Comparison of attack performance in counteracting gravity for two-finger and three-finger robotic grasping.

also fosters further innovation in the field of robotic grasping under adversarial conditions.

6 Experimental Results

6.1 Experimental Setup

Implementation. We implement our framework in Python using the PyBullet simulator [Coumans and Bai, 2016] as the simulation environment. All simulations employ the soft finger contact model [Mahler *et al.*, 2018], with the friction coefficient set to $\mu = 0.6$ and the torsional friction coefficient $\gamma = 0.3$. To manipulate the object's shape, each face's four edges are evenly divided into a square grid based on the cage size, initially set to 0.04. The vertices of these grids serve as control points, which act as anchors for shape manipulation. The optimization process utilizes a simulated annealing algorithm, with an initial temperature $T_0 = 1000$, a minimum temperature $T_{\min} = 10^{-5}$, a cooling rate $\alpha = 0.98$, and a perturbation scale $\epsilon = 0.05 \times$ cage size for control points. After completing each optimization cycle, the cage size is halved, and new control points are generated for further optimization. This process is repeated for a total of five cycles. To balance the contributions of lift capability, grasp stability, and shape regularization in AdvGrasp, we set the weighting parameters to $\lambda_1 = 10000$ and $\lambda_2 = 50$.

Our Attack Solutions. We consider three attack solutions for evaluation: adversarial attacks targeting only lift capability (**ALC**), adversarial attacks targeting only grasp stability (**AGS**), and the unified approach, **AdvGrasp**, which simultaneously considers both aspects.

Evaluation Metrics. Unlike the binary success criteria in image classification adversarial attacks, evaluating adversarial attacks on robotic grasping from a physical perspective is inherently more complex. To address this, **we define a grasp failure as a scenario where the object undergoes significant slippage**, explicitly characterized by a displacement exceeding 0.02 m or a rotation surpassing 10°. Based on this definition, we establish three metrics to evaluate the performance of adversarial attacks on robotic grasping:

- **Minimal Grasp Force (MinGF)**: The minimal force required by the gripper to lift a 1 kg object. Specifically, we start the grasp force at 50 N and decrease it by 0.2 N per step until significant slippage occurs.
- **Maximal Lifting Mass (MaxLM)**: The maximum weight the gripper can lift when each finger applies a maximum force of 50 N. Specifically, we start the object's weight at 1 kg and increase it by 0.1 kg per step until significant slippage occurs.
- **Maximal External Disturbance (MaxED)**: The maximum external force a grasp can withstand while holding a 1 kg object, with each finger of the gripper applying a maximum force of 50 N. We use Fibonacci sphere sampling [Larkins *et al.*, 2012] to generate 50 directions on a unit sphere. For each direction, a force is applied over a time interval of $0.008\,\mathrm{s}$, starting at 1 N and increasing by 1 N per step. The force is incremented only if no significant slippage occurs in all 50 directions.

Among these metrics, MinGF and MaxLM primarily evaluate the grasp configuration's ability to counteract gravity, while MaxED focuses on its resistance to external forces. Together, these metrics form a comprehensive framework for assessing the physical robustness of robotic grasp configurations under adversarial conditions.

6.2 Main Results and Analyses

Attack Performance on Counteracting Gravity. As shown in Tab. 1, for objects of the same type and mass (1 kg), three-finger grasps generally achieve a lower MinGF compared to two-finger grasps, indicating that three-finger grippers require less force to counteract gravity for the same object. Additionally, under the same condition where each fin-

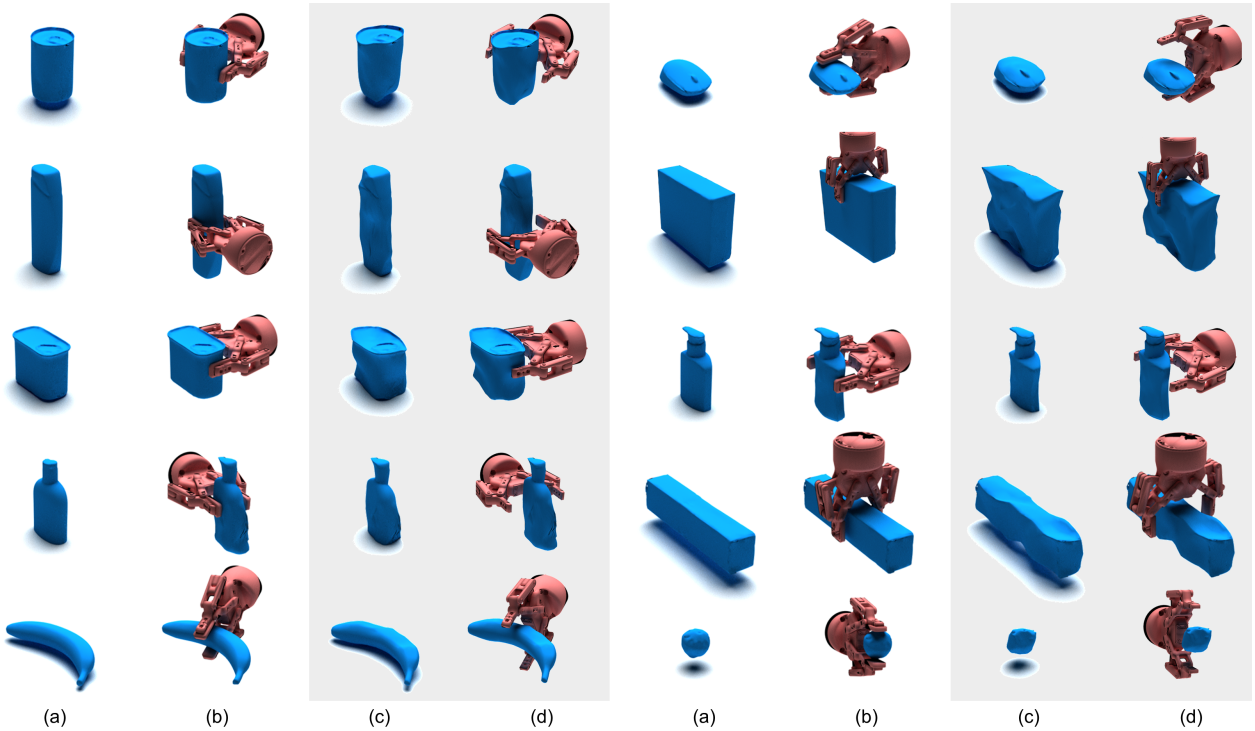

Figure 4: Visualization of the effects of AdvGrasp. (a) The original object, (b) the grasp configuration on the original object, (c) the adversarial object generated by AdvGrasp, and (d) the failed grasp on the adversarial object due to modifications introduced by AdvGrasp.

Object	2-Finger Grasp				3-Finger Grasp			
	Origin	ALC	AGS	AdvGrasp	Origin	ALC	AGS	AdvGrasp
CRACKER BOX	29	**18**	23	28	24	**29**	32	36
TOMATO SOUP CAN	33	14	**14**	24	49	31	28	**25**
MUSTARD BOTTLE	26	**23**	23	27	37	**21**	24	26
POTTED MEAT CAN	37	31	32	**28**	51	38	35	**20**
BANANA	56	33	32	**19**	62	42	48	**32**
BOWL	36	33	**27**	32	68	57	60	**24**
CHIPS CAN	33	**21**	34	35	63	**34**	41	37
STRAWBERRY	41	**28**	30	31	72	44	**35**	36
ORANGE	14	**7**	9	8	67	36	38	**33**
KNIFE	33	20	29	**10**	7	17	23	**13**
RACQUETBALL	21	48	21	**17**	53	**24**	35	27
TOY AIRPLANE	27	35	36	**16**	44	32	44	**21**
WASH SOUP	32	**15**	20	16	54	45	32	**24**
DABAO SOD	34	10	**5**	20	31	33	**21**	24
BAOKE MARKER	37	26	29	**21**	15	**27**	34	40
CAMEL	37	32	**25**	40	23	21	**19**	27
LARGE ELEPHANT	36	30	**19**	23	33	**18**	26	30
DARLIE BOX	50	31	19	**17**	31	**23**	33	25
MOUSE	25	21	6	**1**	35	29	**28**	28
SHAMPOO	19	22	**7**	19	41	**15**	17	20

Table 2: Comparison of attack performance in resisting external forces for two- and three-finger grasping, measured by MaxED.

ger can apply a maximum force of 50 N, three-finger grippers can often counteract gravity for objects with greater mass.

After applying our method, the effectiveness of the original grasp configurations significantly decreases, as reflected by increased MinGF and decreased MaxLM. For most objects, ALC proves more effective in making counteracting gravity more difficult, as it introduces gravity-related factors into the optimization. In contrast, AGS is less effective in this regard.

Attack Performance on Resisting External Forces. We further evaluate the ability of these grasps to resist external

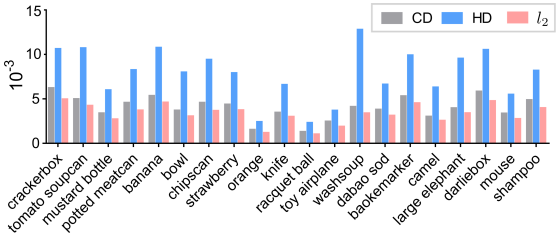

Figure 5: The deformation introduced by AdvGrasp, measured by Chamfer distance (CD), Hausdorff distance (HD), and l_2-norm (l_2) computed on the mesh vertices between the adversarial objects and the original objects.

forces under attack. As shown in Tab. 2, three-finger grasps generally tolerate higher disturbances compared to two-finger grasps, indicating greater robustness to external perturbations. However, after applying our attacks, particularly the unified AdvGrasp, the stability of these grasps is significantly compromised, making them more vulnerable to disruption. These results validate the effectiveness of our methods.

Visualization. Fig. 4 provides a visual comparison between the original objects with their corresponding grasps and the adversarial objects with their respective counterparts. It shows that the overall shape of the objects is preserved. At the same time, localized deformations occur near the grasp contact points, including changes in surface normals and slight shifts in the center of mass. These perturbations disrupt the grasp's wrench equilibrium, validating the effectiveness of our method. This demonstrates how AdvGrasp gener-

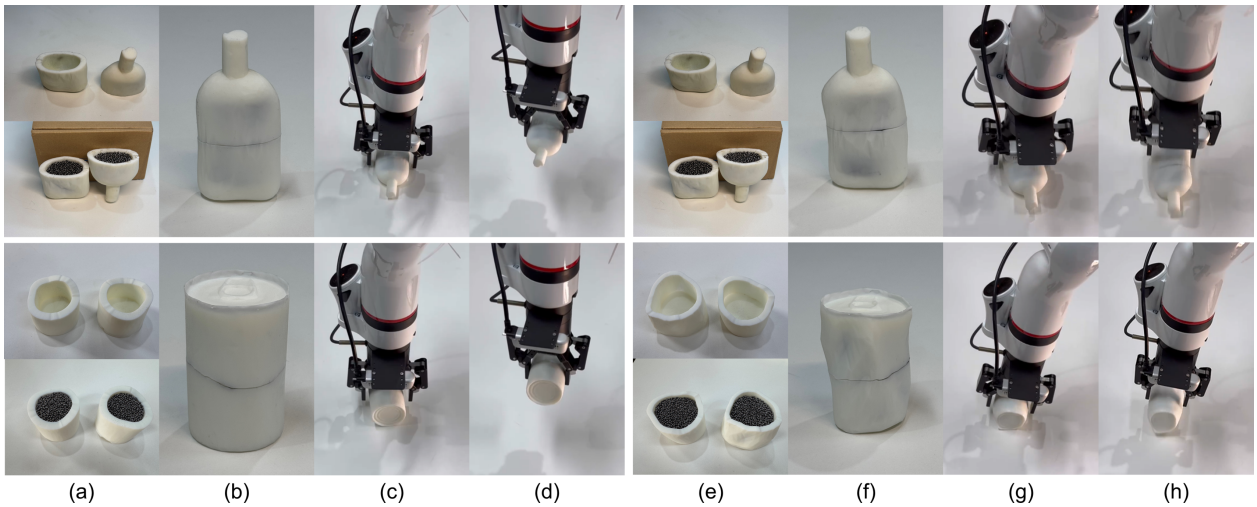

Figure 6: Visualization of the adversarial attack on robotic grasping in a physical scenario: (a) assembly process with lead balls, (b) assembled 3D-printed object, (c) gripper preparing to grasp, and (d) grasping result. The corresponding adversarial process generated by AdvGrasp: (e) adversarial assembly process, (f) assembled adversarial object, (g) gripper preparing to grasp the adversarial object, and (h) adversarial grasping result, which fails due to instability. The detailed process is available in the video provided in the supplementary materials.

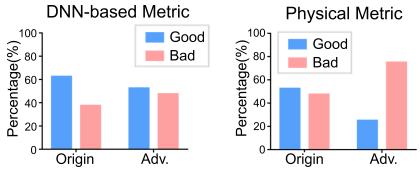

Figure 7: Proportion of original and adversarial grasps generated by AdvGrasp classified as good or bad, evaluated using both the DNN-based metric and the physical metric.

ates imperceptible yet impactful adversarial perturbations to effectively compromise robotic grasping performance.

Analysis on Shape Deformation. We quantitatively analyze the impact of AdvGrasp on the object's geometry. As illustrated in Fig. 5, the perturbations remain minimal, aligning to some extent with the imperceptibility requirements of adversarial attacks. These results underscore the effectiveness of AdvGrasp in achieving adversarial objectives while preserving geometric plausibility.

6.3 Additional Analyses

Physical Results. To validate the effectiveness of our method in real-world scenarios, we select the DABAO SOD and TOMATO SOUP CAN for physical experiments. For each object, we choose a specific grasp configuration and generate adversarial versions using AdvGrasp. We 3D-print these objects and modify them by hollowing them out, splitting them into two halves, and inserting lead balls to increase their mass to 1 kg, addressing the lightweight nature of 3D-printed materials. Grasping experiments are conducted under controlled conditions using a Robotiq 2-finger 2F-85 gripper mounted on a robotic arm. When a maximum grasping force of 30 N is applied to the DABAO SOD and its adversarial version, the original model is successfully lifted, while the adversarial model exhibits slippage and subsequently drops. Similarly, when a 20 N grasping force is applied to the TOMATO SOUP CAN and its adversarial version, the original model is successfully lifted, while the adversarial model fails and is dropped. These experiments highlight the impact of physical properties on grasp quality and confirm the effectiveness of AdvGrasp in generating adversarial objects that compromise grasp stability in real-world conditions.

DNN-based Metric vs. Physical Metric. To validate the importance of physical metrics in grasp evaluation, we assess all grasps in the AdvGrasp-20 benchmark, as well as adversarial grasps generated by AdvGrasp, using both the DNN-based PointNetGPD [Liang *et al.*, 2019] and physical metric. Specifically, we apply the 2-class classification approach from PointNetGPD to determine whether a grasp is good or bad. For the physical metric, a grasp is considered bad if the actual grasp position deviates from the expected position by more than 0.03 m or if the angle deviation exceeds 10 degrees during simulation experiments. As shown in Fig. 7, the physics-based metric identifies most adversarial grasps generated by AdvGrasp as bad. However, PointNetGPD, which relies solely on local geometric properties, fails to accurately assess these cases. These findings underscore the indispensable role of physical metrics in grasp evaluation.

7 Conclusion

In this paper, we have proposed a framework for adversarial attacks on robotic grasping, targeting lift capability and grasp stability. Through shape deformation to increase gravitational torque and reduce stability margin in the wrench space, our method systematically degrades grasp performance. Extensive experiments in simulated and real-world environments validate the effectiveness of our approach. Future work will consider dynamic settings and focus on defense design.

Acknowledgements

This work was supported in part by the Dreams Foundation of Jianghuai Advance Technology Center (2023-ZM01G003), the National Natural Science Foundation of China (62472117), the Guangdong Basic and Applied Basic Research Foundation (2025A1515010157), and the Science and Technology Projects in Guangzhou (2025A03J0137).

References

[Alharthi and Brandão, 2024] Naif Wasel Alharthi and Martim Brandão. Physical and digital adversarial attacks on grasp quality networks. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1907–1902. IEEE, 2024.

[Amaya-Mejía *et al.*, 2022] Lina María Amaya-Mejía, Nicolás Duque-Suárez, Daniel Jaramillo-Ramírez, and Carol Martinez. Vision-based safety system for barrierless human-robot collaboration. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7331–7336. IEEE, 2022.

[Bicchi and Kumar, 2000] A. Bicchi and V. Kumar. Robotic grasping and contact: a review. In *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065)*, volume 1, pages 348–353 vol.1, 2000.

[Carlini and Wagner, 2017] Nicholas Carlini and David Wagner. Towards evaluating the robustness of neural networks. In *IEEE Symposium on Security and Privacy*, pages 39–57, 2017.

[Chen *et al.*, 2015] Kai Chen, Dongcai Lu, Yingfeng Chen, Keke Tang, Ningyang Wang, and Xiaoping Chen. The intelligent techniques in robot kejia–the champion of robocup@ home 2014. In *RoboCup 2014: Robot World Cup XVIII 18*, pages 130–141, 2015.

[Coumans and Bai, 2016] Erwin Coumans and Yunfei Bai. Pybullet, a python module for physics simulation for games, robotics and machine learning. http://pybullet.org, 2016. Accessed: 2024.

[Cutkosky and others, 1989] Mark R Cutkosky et al. On grasp choice, grasp models, and the design of hands for manufacturing tasks. *IEEE Transactions on robotics and automation*, 5(3):269–279, 1989.

[Dieber *et al.*, 2017] Bernhard Dieber, Benjamin Breiling, Sebastian Taurer, Severin Kacianka, Stefan Rass, and Peter Schartner. Security for the robot operating system. *Robotics and Autonomous Systems*, 98:192–203, 2017.

[Dong *et al.*, 2018] Yinpeng Dong, Fangzhou Liao, Tianyu Pang, Hang Su, Jun Zhu, Xiaolin Hu, and Jianguo Li. Boosting adversarial attacks with momentum. In *CVPR*, pages 9185–9193, 2018.

[Fang *et al.*, 2020] Hao-Shu Fang, Chenxi Wang, Minghao Gou, and Cewu Lu. Graspnet-1billion: A large-scale benchmark for general object grasping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11444–11453, 2020.

[Fang *et al.*, 2023] Hao-Shu Fang, Minghao Gou, Chenxi Wang, and Cewu Lu. Robust grasping across diverse sensor qualities: The graspnet-1billion dataset. *The International Journal of Robotics Research*, 2023.

[Ferrari *et al.*, 1992] Carlo Ferrari, John F Canny, et al. Planning optimal grasps. In *ICRA*, volume 3, page 6, 1992.

[Goodfellow *et al.*, 2015] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. In *ICLR*, 2015.

[Guerrero-Higueras *et al.*, 2018] Ángel Manuel Guerrero-Higueras, Noemí DeCastro-García, and Vicente Matellán. Detection of cyber-attacks to indoor real time localization systems for autonomous robots. *Robotics and Autonomous Systems*, 99:75–83, 2018.

[Ju *et al.*, 2023] Tao Ju, Scott Schaefer, and Joe Warren. Mean value coordinates for closed triangular meshes. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pages 223–228. Association for Computing Machinery, 2023.

[Larkins *et al.*, 2012] Robert L Larkins, Michael J Cree, and Adrian A Dorrington. Analysis of binning of normals for spherical harmonic cross-correlation. In *Three-Dimensional Image Processing (3DIP) and Applications II*, volume 8290, pages 195–206. SPIE, 2012.

[Li *et al.*, 2022] Puhao Li, Tengyu Liu, Yuyang Li, Yixin Zhu, Yaodong Yang, and Siyuan Huang. Gendexgrasp: Generalizable dexterous grasping. *arXiv preprint arXiv:2210.00722*, 2022.

[Liang *et al.*, 2019] Hongzhuo Liang, Xiaojian Ma, Shuang Li, Michael Görner, Song Tang, Bin Fang, Fuchun Sun, and Jianwei Zhang. Pointnetgpd: Detecting grasp configurations from point sets. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 3629–3635. IEEE, 2019.

[Lin *et al.*, 2022] Nan Lin, Yuxuan Li, Keke Tang, Yujun Zhu, Xiayu Zhang, Ruolin Wang, Jianmin Ji, Xiaoping Chen, and Xinming Zhang. Manipulation planning from demonstration via goal-conditioned prior action primitive decomposition and alignment. *IEEE Robotics and Automation Letters*, 7(2):1387–1394, 2022.

[Mahler *et al.*, 2017] Jeffrey Mahler, Jacky Liang, Sherdil Niyaz, Michael Laskey, Richard Doan, Xinyu Liu, Juan Aparicio, and Ken Goldberg. Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics. In *Proceedings of Robotics: Science and Systems*, Cambridge, Massachusetts, July 2017.

[Mahler *et al.*, 2018] Jeffrey Mahler, Matthew Matl, Xinyu Liu, Albert Li, David Gealy, and Ken Goldberg. Dex-net 3.0: Computing robust vacuum suction grasp targets in point clouds using a new analytic model and deep learning. In *2018 IEEE International Conference on robotics and automation (ICRA)*, pages 5620–5627. IEEE, 2018.

[Mason *et al.*, 1989] Matthew T Mason, Joey K Parker, and J Kenneth Salisbury. *Robot Hands and the Mechanics of*

Manipulation. The American Society of Mechanical Engineers (ASME), 1989.

[Matheus and Dollar, 2010] Kayla Matheus and Aaron M Dollar. Benchmarking grasping and manipulation: Properties of the objects of daily living. In *IROS*, pages 5020–5027, 2010.

[Mazzeo and Staffa, 2020] Giovanni Mazzeo and Mariacarla Staffa. Tros: Protecting humanoids ros from privileged attackers. *International Journal of Social Robotics*, 12(3):827–841, 2020.

[Meng and Weitschat, 2021] Xuming Meng and Roman Weitschat. Dynamic projection of human motion for safe and efficient human-robot collaboration. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3765–3771, 2021.

[Moosavi-Dezfooli *et al.*, 2016] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, and Pascal Frossard. Deepfool: a simple and accurate method to fool deep neural networks. In *CVPR*, pages 2574–2582, 2016.

[Morris *et al.*, 2020] John Morris, Eli Lifland, Jin Yong Yoo, Jake Grigsby, Di Jin, and Yanjun Qi. Textattack: A framework for adversarial attacks, data augmentation, and adversarial training in nlp. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 119–126, 2020.

[Siciliano *et al.*, 2008] Bruno Siciliano, Oussama Khatib, and Torsten Kröger. *Springer handbook of robotics*, volume 200. Springer, 2008.

[Szegedy *et al.*, 2014] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. In *ICLR*, 2014.

[Tang *et al.*, 2022] Keke Tang, Yawen Shi, Tianrui Lou, Weilong Peng, Xu He, Peican Zhu, Zhaoquan Gu, and Zhihong Tian. Rethinking perturbation directions for imperceptible adversarial attacks on point clouds. *IEEE Internet of Things Journal*, 10(6):5158–5169, 2022.

[Tang *et al.*, 2023] Keke Tang, Jianpeng Wu, Weilong Peng, Yawen Shi, Peng Song, Zhaoquan Gu, Zhihong Tian, and Wenping Wang. Deep manifold attack on point clouds via parameter plane stretching. In *AAAI*, volume 37, pages 2420–2428, 2023.

[Tang *et al.*, 2024a] Keke Tang, Xu He, Weilong Peng, Jianpeng Wu, Yawen Shi, Daizong Liu, Pan Zhou, Wenping Wang, and Zhihong Tian. Manifold constraints for imperceptible adversarial attacks on point clouds. In *AAAI*, volume 38, pages 5127–5135, 2024.

[Tang *et al.*, 2024b] Keke Tang, Lujie Huang, Weilong Peng, Daizong Liu, Xiaofei Wang, Yang Ma, Ligang Liu, and Zhihong Tian. Flat: Flux-aware imperceptible adversarial attacks on 3d point clouds. In *ECCV*, pages 198–215, 2024.

[Tang *et al.*, 2024c] Keke Tang, Tianrui Lou, Weilong Peng, Nenglun Chen, Yawen Shi, and Wenping Wang. Effective single-step adversarial training with energy-based models. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2024.

[Tang *et al.*, 2024d] Keke Tang, Zhensu Wang, Weilong Peng, Lujie Huang, Le Wang, Peican Zhu, Wenping Wang, and Zhihong Tian. Symattack: Symmetry-aware imperceptible adversarial attacks on 3d point clouds. In *MM*, pages 3131–3140, 2024.

[Tang *et al.*, 2025a] Keke Tang, Ziyong Du, Weilong Peng, Xiaofei Wang, Daizong Liu, Ligang Liu, and Zhihong Tian. Imperceptible 3d point cloud attacks on lattice-based barycentric coordinates. In *AAAI*, volume 39, pages 20814–20822, 2025.

[Tang *et al.*, 2025b] Keke Tang, Weiyao Ke, Weilong Peng, Xiaofei Wang, Ziyong Du, Zhize Wu, Peizan Zhu, and Zhihong Tian. Imperceptible adversarial attacks on point clouds guided by point-to-surface field. In *ICASSP*, pages 1–5, 2025.

[Wang *et al.*, 2019] David Wang, David Tseng, Pusong Li, Yiding Jiang, Menglong Guo, Michael Danielczuk, Jeffrey Mahler, Jeffrey Ichnowski, and Ken Goldberg. Adversarial grasp objects. In *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*, pages 241–248. IEEE, 2019.

[Wang *et al.*, 2025] Zhensu Wang, Weilong Peng, Le Wang, Zhizhe Wu, Peican Zhu, and Keke Tang. Eia: Edge-aware imperceptible adversarial attacks on 3d point clouds. In *MMM*, pages 348–361, 2025.

[Xiang *et al.*, 2019] Chong Xiang, Charles R. Qi, and Bo Li. Generating 3d adversarial point clouds. In *CVPR*, pages 9136–9144, 2019.

[Yaacoub *et al.*, 2022] Jean-Paul A Yaacoub, Hassan N Noura, Ola Salman, and Ali Chehab. Robotics cyber security: Vulnerabilities, attacks, countermeasures, and recommendations. *International Journal of Information Security*, 21(1):115–158, 2022.

[Zheng *et al.*, 2021] Baolin Zheng, Peipei Jiang, Qian Wang, Qi Li, Chao Shen, Cong Wang, Yunjie Ge, Qingyang Teng, and Shenyi Zhang. Black-box adversarial attacks on commercial speech platforms with minimal information. In *Proceedings of the 2021 ACM SIGSAC conference on computer and communications security*, pages 86–107, 2021.