

基于激光雷达的点云目标检测安全性综述

项圣¹, 廖丹芯¹, 陈晋音^{1,2}, 郑海斌^{1,2}

(1. 浙江工业大学信息工程学院, 浙江 杭州 310023; 2. 浙江工业大学网络安全研究院, 浙江 杭州 310023)

摘要: 激光雷达技术在自动驾驶、智能交通和环境感知等领域有着广泛应用。在自动驾驶系统中, 利用机载摄像头和激光雷达等对周围环境进行感知, 点云目标检测技术在环境感知方面发挥着重要的作用。近年来的研究指出, 基于激光雷达的点云目标检测系统容易遭受欺骗攻击。然而, 目前对基于激光雷达的点云目标检测的安全性问题的研究相对零散, 设计的安全威胁场景、攻击方法和防御策略相互独立。基于现有研究工作, 首次对基于激光雷达的点云目标检测的安全性进行了综述, 主要关注该技术在实际应用中可能面临的安全挑战和解决方案。首先, 全面回顾了基于激光雷达的目标检测和攻防理论; 其次, 从数字域和物理域双重视角分析了攻击方法, 并对现有的防御策略进行了详尽的分类和总结; 接着, 介绍了该领域目前常用的数据集及评价标准; 最后, 结合面临的挑战和机遇, 对基于激光雷达的目标检测安全的未来研究方向进行了总结和展望, 为基于激光雷达的点云目标检测的安全研究的发展和应用提供参考。

关键词: 激光雷达; 目标检测; 攻击方法; 防御方法

中图分类号: TP183

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2025004

Review on the safety of LIDAR-based point cloud object detection

XIANG Sheng¹, LIAO Danxin¹, CHEN Jinyin^{1,2}, ZHENG Haibin^{1,2}

1. College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China

2. Institute of Cyberspace Security, Zhejiang University of Technology, Hangzhou 310023, China

Abstract: LiDAR technology has been widely applied in fields such as autonomous driving, intelligent traffic, and environmental perception. In autonomous driving systems, the surrounding environment was perceived using on-board cameras and LiDAR, with point cloud object detection playing a crucial role. Recent studies have indicated that point cloud object detection based on LiDAR was susceptible to deception attacks. However, research on the

收稿日期: 2024-04-07; 修回日期: 2024-08-22

通信作者: 陈晋音, chenjinyin@zjut.edu.cn

基金项目: 浙江省自然科学基金 (LDQ23F020001); 国家自然科学基金 (62072406, 62406286); 国家重点研发计划 (2018AAA0100801)

Foundation Items: The Natural Science Foundation of Zhejiang Province (LDQ23F020001), The National Natural Science Foundation of China (62072406, 62406286), The National Key R&D Program of China (2018AAA0100801)

引用格式: 项圣, 廖丹芯, 陈晋音, 等. 基于激光雷达的点云目标检测安全性综述[J]. 网络与信息安全学报, 2025, 11(1): 20-39.

Citation Format: XIANG S, LIAO D X, CHEN J Y, et al. Review on the safety of LIDAR-based point cloud object detection[J]. Chinese Journal of Network and Information Security, 2025, 11(1): 20-39.

security issues of LiDAR-based object detection has been relatively scattered, with security threat scenarios, attack methods, and defense strategies being addressed independently. Based on existing research, the first comprehensive review of the security of LiDAR-based point cloud object detection was presented, focusing on the security challenges and solutions that the technology may face in practical applications. The theory of object detection and the principles of attack and defense based on LiDAR were comprehensively reviewed. Attack methods were analyzed from both the digital and physical domains, and existing defense strategies were classified and summarized in detail. Then, the commonly used datasets and evaluation criteria in this field were introduced. And considering the challenges and opportunities, a summary and outlook on future research directions in the security of LiDAR-based detection were provided, offering valuable insights for the development and application of secure research in LiDAR-based point cloud object detection.

Keywords: LiDAR, object detection, attack method, defense method

0 引言

随着自动驾驶技术的日益成熟，作为核心传感器的激光雷达得到了广泛应用。激光雷达能够精准地进行3D目标检测，不仅能预测目标在3D空间中的类别，如机动车、行人及骑自行车者等，更重要的是能够在真实世界3D坐标系中精确定位与识别目标，并获取目标与本车的实际距离信息。通过获取现实世界坐标中的几何数据，3D目标检测能够直接测量本车与关键目标间的距离，从而有效辅助行驶路径的规划和碰撞风险的规避。

近年来，深度神经网络（deep neural network, DNN）在3D目标检测中大放异彩。然而，研究显示，DNN面临着严重的安全隐患——对抗性攻击，微小的输入扰动会显著改变输出预测。研究者围绕对抗性攻击问题进行了深入研究^[1-7]。在自动驾驶方面，一些研究人员研究了物理上可实现的图像感知系统对抗性攻击^[8,9]，以及针对点云的对抗性攻击^[10,11]。这些攻击在实际交通场景中可能会引发交通事故。例如，在图1所示的基于点云的目标检测模型受攻击场景中，攻击者改变了受害者车辆感知的点云（如通过添加点、移动点或者删除点等操作），导致行人未被正确检测到。因此，对面向激光雷达的安全问题进行系统研究是非常有必要的。

为了深入探讨面向激光雷达的安全问题，本文收集了相关文献并进行了大量的研究。具体来说，本文综述了大约110篇相关论文，这些论文来自知名期刊、会议以及公共网络平台。

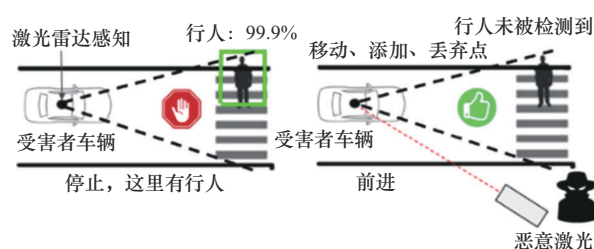


图1 基于点云的目标检测模型受攻击场景
Figure1 Attack scenario of object detection model
based on point cloud

1 激光雷达点云目标检测及攻防原理

本节首先介绍激光雷达的工作原理、基于激光雷达的点云目标检测原理，然后阐述基于激光雷达的点云目标检测对抗攻防原理。

1.1 激光雷达的工作原理

应用于自动驾驶的激光雷达是一种采用非接触激光测距技术的扫描式传感器，其工作原理与一般的雷达系统类似，通过发射激光光束来探测目标，并通过收集反射回来的光束形成点云和获取数据。这些数据经过光电处理后可以生成精确的三维立体图像。在自动驾驶中，激光雷达主要用于实时感知周围环境，包括道路、车辆、行人、交通标志等的位置、形状和运动状态。通过对这些信息的分析和处理，自动驾驶系统可以实现路径规划、避障、跟车等功能，从而提高行车的安全性和舒适性。

(1) 激光雷达的测距方式

激光雷达的测距方式包括三角测距法、飞行时间（time of flight, TOF）测距法和相位测距法，其中相位测距法属于连续波激光测距，飞行

时间测距法和三角测距法属于脉冲激光测距^[12]。

1) 三角测距法。三角测距法的原理是激光器发射的激光在照射到物体后, 反射光由线性电荷耦合元件 (charge-coupled device, CCD) 接收, 由于激光器和探测器之间隔着一段距离, 依照光学路径, 不同距离的物体在 CCD 上的成像位置不同, 按照三角公式进行计算, 就能推导出被测物体的距离^[13]。

2) 飞行时间测距法。在自动驾驶领域, TOF 测距法的基本原理是通过连续发射光脉冲 (一般为不可见光) 到被测物体上, 然后接收从物体反射回来的光脉冲, 通过探测光脉冲的飞行 (往返) 时间来计算被测物体与相机之间的距离。由于 TOF 测距法具有探测精度较高和探测距离较远的优势, 因此大多数激光雷达采用这种方法测距。

3) 相位测距法。相位测距法也是一种基于飞行时间原理的测距方法, 但它通过求解发射波和接收波的相位差来反推 TOF^[14]。值得注意的是, 这个相位并非光的原始相位, 而是被调制的光强的相位。具体来说, 相位测距法对发射光波的光强进行调制, 然后通过测量相位差来间接测量时间。这种方法相较于直接测量往返时间的方法在处理难度上降低了许多, 因此在实际工程应用中具有较高的实用价值。

(2) 激光雷达的扫描方式

根据扫描方式的不同, 激光雷达可分为旋转式激光雷达和面阵式激光雷达两种。

旋转式激光雷达的激光束通常排列成一个垂直的列, 在雷达装置上绕垂直轴旋转。每束激光扫描一个平面, 并在旋转过程中连续扫描周围的环境^[15]。随着雷达的旋转, 通过不同角度和距离的激光束的反射和接收, 可以创建出一个全方位的三维立体图形。这些点云数据随着时间和旋转角度的变化不断积累, 最终构成了整个环境的三维模型。在自动驾驶、环境测绘和工业应用中, 旋转式激光雷达能够提供可靠的高精度数据, 但受制于旋转结构, 可能存在机械磨损和响应速度限制。

相对于旋转式激光雷达, 面阵式激光雷达采用电机带动反射镜运动, 将激光束反射到空间的特定区域。这个反射镜的运动范围较广, 能够覆

盖雷达所需扫描的区域。激光束通过反射镜的移动, 在空间内构建了更广泛的扫描范围。这种方式使得面阵式激光雷达在扫描过程中更加迅速、灵活, 并避免了机械旋转结构的某些限制。这种雷达在需要快速获取数据且对实时性要求较高的场景中表现出色。

每种激光雷达技术有其独特的优势, 适用于不同应用场景。旋转式激光雷达适用于需要高精度和全方位数据的场景, 如制图、建模和长距离探测。而面阵式激光雷达适用于需要快速获取高密度数据的场景, 如移动机器人、实时环境监测和快速响应的自动化系统。这些激光雷达技术不仅提高了环境感知的精度和速度, 也在自动化、智能交通和科学研究等领域推动了许多创新。

1.2 基于激光雷达的点云目标检测原理

基于激光雷达的点云目标检测原理是通过处理和分析激光雷达采集的点云数据, 实现对目标物体的检测和识别。其中, 点云数据是由激光雷达测量发射激光束的飞行时间并计算得出的距离值所组成的。这些距离值被记录下来, 形成一个三维点云数据集。通过对这个数据集进行处理和分析, 可以提取出目标物体的特征信息, 如形状、大小、位置等。在自动驾驶领域, 激光雷达可以通过实时扫描周围环境来检测道路障碍物、行人和其他车辆等目标物体, 从而实现智能驾驶功能。具体检测原理如下。

给定一个深度学习模型 M , 输入数据 D 和相应的地面真实标签 y^* , 目标检测模型的目标是在输入数据 D 时输出的标签与地面真实标签相同, 即 $M(D_{\text{adv}}) = y^*$ 。

基于激光雷达的目标检测方法可以分为基于点的方法、基于网格的方法、基于点云-体素的方法和基于范围的方法^[16]。

1.3 基于激光雷达的点云目标检测对抗攻击原理

基于激光雷达的目标检测技术在自动驾驶领域被广泛用于检测和识别道路上的障碍物和目标物体。攻击者攻击自动驾驶系统通常是为了干扰或破坏系统中目标检测模型的正常运行, 使目标检测功能失效, 从而产生不安全的驾驶或导航行为。具体攻击原理如下。

给定一个深度学习模型 M ，输入数据 D 和相应的地面真实标签 y^* ，对抗性攻击的目标是找到一个对抗性输入 D_{adv} ，使得模型输出的标签与地面真实标签不一致，即 $M(D_{adv}) \neq y^*$ ，对抗性输入 D_{adv} 通常是通过求解一个优化问题而得到的。在物理可实现的攻击中，对抗性输入 D_{adv} 通常是对原始输入 D 进行物理可实现的修改来实现的。要攻击物理世界中的激光雷达感知模型并修改输入的激光雷达点云，最简单的方法是在原始的点云 D 中插入一些点 D' ，得到对抗性的输入 $D_{adv} = D \cup D'$ 。插入点的物理可实现的有效方法之一是在优化得到的特定位置放置一个具有反射面的物体，让它被激光雷达扫描，实现在点云数据中插入点。

1.4 基于激光雷达的点云目标检测防御原理

点云目标检测的防御旨在确保激光雷达系统对周围环境的感知不受干扰或恶意攻击的影响，以维护自动驾驶、机器人和其他应用的安全性和可靠性。

在进行点云防御时，主要目标是构建一个对抗性示例具有鲁棒性的目标检测模型。这意味着该模型能够准确地分类对抗性点云和干净点云，从而使其在未受干扰的干净点云上的性能损失较小。

形式上，给定一个目标检测模型 M 和输入 $\tilde{D}$ ，这可能是一个原始输入 D 或对抗输入 D_{adv} ，防御方法的目标是增强数据训练一个鲁棒的 M' ，这样 $M'(\tilde{D}) = M(D)$ （也就是说，不管输入的是干净样本还是对抗样本，防御方法得到的新模型 M' 都能准确识别，使得识别结果与输入干净样本得到的结果一致），或将 $\tilde{D}$ 进行转换，使得 $M'(T(\tilde{D})) = M(D)$ 。 $T(\cdot)$ 表示对输入的数据进行数据优化，如去噪等。

2 基于激光雷达的点云目标检测攻击

基于激光雷达的点云目标检测攻击主要分为数字域和物理域两大类。数字域攻击主要是利用激光雷达点云数据的漏洞进行攻击，在数字域对点云进行移动、添加、丢弃等操作，使目标检测模型产生错误的检测结果。物理域攻击则主要利用激光雷达自身的工作原理进行攻击。例如，攻击者在物理世界采用制造强光束或者反射面等手段，使得激光雷达接收到错误的回波信号，从而影响检测结果。

2.1 基于激光雷达点云目标检测的数字域攻击

在自动驾驶和机器人领域，激光雷达点云目标检测是一项重要的任务。然而，由于数字域攻击的威胁，该任务的安全性面临着严峻的挑战。本节将介绍当前基于激光雷达点云目标检测的数字域攻击现状，主要包括基于点移动的攻击、基于点添加的攻击、基于点丢弃的攻击和基于点的可迁移性攻击4种类型。

2.1.1 基于点移动的攻击

基于点移动的攻击是一种常见的数字域攻击方法。这种攻击通过微小的扰动误导激光雷达点云目标检测系统^[5,10,11,17-23]，通常通过精细调整点云中点的坐标或反射强度，使目标检测系统将物体错误分类。Xiang等^[10]首次针对点云分类模型实施了C&W攻击。该攻击以L2范数为距离度量，将点向表面微小移动，实现了高效的隐蔽性攻击，攻击成功率高达99%。文献[10]中提出的一个原始点云和4种类型的C&W对抗攻击的示例，如图2所示。

随后的研究工作对这类攻击方法进行了拓展和改进。Liu等^[11]将FGSM和JSMA这两种在图像领域广泛应用的攻击方法应用于3D点云目标检测场景中。Yang等^[17]不仅提出了3种新的点云

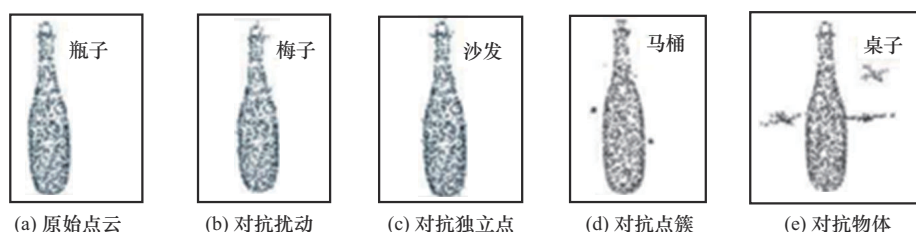


图2 文献[10]中提出的一个原始点云和4种类型的C&W对抗攻击的示例

Figure 2 Example of an original point cloud and four types of C&W counterattacks presented in the reference [10]

攻击方法,还探讨了这些攻击方法在不同点云网络间的迁移性,并通过引入动量增强的点式梯度(point gradient, PG)显著提高了攻击效果。尽管点附着(point attachment, PA)和点分离(point detachment, PD)方法的成功率相对较低,但它们提供了一种在物理世界中可行的攻击方式。

为了生成更具鲁棒性的对抗性扰动, Lee 等^[18]提出了一种基于形状感知的对抗性三维点云生成(generating shape-aware adversarial 3D point clouds, ShapeAdv)框架。该框架不但关注生成能够欺骗目标检测系统的对抗性点云,而且重视保持这些扰动在几何形状上的合理性。通过引入辅助数据来指导形状变形, ShapeAdv 能生成与同一类别的真实物体在形状上相近的对抗性样本。

该方法能够实现对 3D 点云空间形状变化的真实模拟,使得生成的对抗性点云在结构和视觉上与真实物体更接近,从而减少被人工痕迹检测出来的可能性。图 3 所示为文献[18]提出的 ShapeAdv-Latent l_2 方法和 ShapeAdv-Chamfer 方法示例,可以看到,即使在添加了扰动的情况下,点云仍然保持了物体的基本形状特征,这使得攻击更加隐蔽,同时为研究如何提高点云处理系统对此类攻击的鲁棒性提供了有价值的案例。

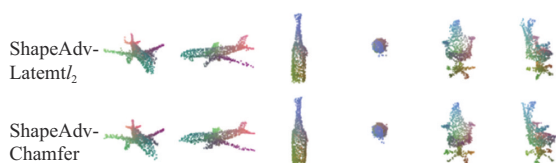


图 3 ShapeAdv-Latent l_2 方法和 ShapeAdv-Chamfer 方法示例

Figure 3 Examples of the ShapeAdv-Latent l_2 and ShapeAdv-Chamfer methods

Tsai 等^[5]的研究工作聚焦于确保对抗性点云在物理实现时的稳定性和可行性。他们采用 K-最近邻(K-nearest neighbors, KNN)算法,通过限制点云中相邻点之间的最小距离,确保在物理世界中实现对抗性点云的连续性和结构完整性。这种方法的创新之处在于将点云的空间关系纳入考虑,从而在保持对抗性效果的同时,避免过度的几何变形。Tsai 等通过 3D 打印验证了所提攻击方法在现实世界中的可行性,还展示了实体对象在保持对抗性属性的同时可以成功地被制

造出来。图 4 所示为文献[5]中攻击算法生成的物理攻击对象示例,它揭示了对抗性攻击在实际物理环境中的潜在影响,对于评估和改进点云处理系统的防御策略具有重要的实践意义。

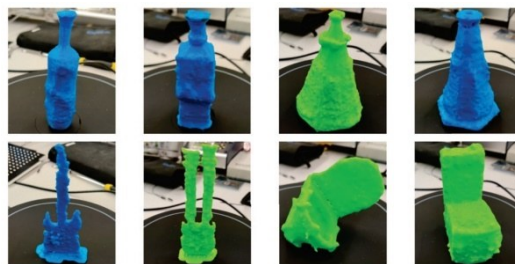


图 4 文献[5]中攻击算法生成的物理攻击对象示例

Figure 4 Examples of physical attack objects generated by attack algorithms in the reference[5]

Wen 等^[19]对人类感知二维图像和三维形状的不同机制进行了分析,并提出了一种新的几何感知对抗攻击(geometry-aware adversarial attack, GeoA3)方法。该方法旨在生成有效的对抗性扰动,以及提高这些扰动的平滑性和公平性,使得扰动在视觉上更加自然,不易被人眼察觉。在错误分类损失和几何感知正则化器的助力下, GeoA3 方法能够同时优化对抗性扰动的隐蔽性和有效性。设计错误分类损失是为了引导生成的对抗性样本朝向模型分类决策边界的弱点,从而增加其被错误分类的概率。几何感知正则化器的引入则是基于人类视觉系统对物体表面光滑性和一致性的敏感性。这种正则化器让生成的对抗性点云具有更接近自然物体的表面属性,减少扰动导致的不自然凹凸或颜色失真,从而使对抗性样本在保持攻击效果的同时,更符合人类观察者对物体表面的期望。图 5 展示的结果清楚地证明了 GeoA3 方法在生成对抗性点云方面的高效性。这些点云与正常点云在视觉上几乎无法区分,却能够有效地误导目标检测系统。这种攻击方法揭示了现有点云处理系统在面对精心设计的对抗性样本时的脆弱性,同时为防御策略的设计带来了新的挑战。防御系统需要能够检测和抵御这种既微小又平滑的扰动,以保护点云处理系统免受此类隐蔽攻击的影响。Wen 等的工作不仅推动了对抗性攻击技术的发展,也为点云安全性研究提供了新的视角和方法论。

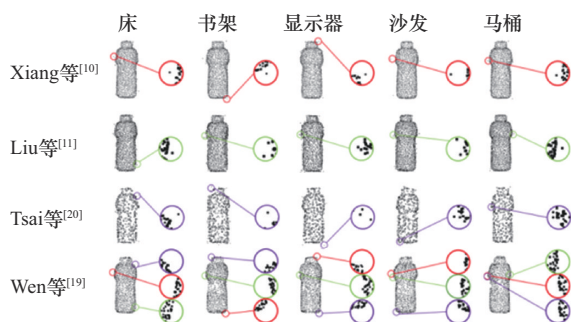


图5 文献[19]中,针对 PointNet模型,对瓶子进行目标攻击的实验结果

Figure 5 Experimental results of targeting bottles against the PointNet model in the reference[19]

这些基于点移动的攻击方法通过有意移动目标物体部分点的坐标,使其被误识别为其他类别的对象,从而干扰系统的正常运作。

2.1.2 基于点添加的攻击

基于点添加的攻击是另一种常见的数字域攻击方法,旨在通过向原始点云中添加额外的点来欺骗目标检测系统^[10, 20-28]。这些额外的点可以是虚假的障碍物、不存在的物体或者与真实场景不符的物体。

Xiang等在文献[10]中除了提出基于点移动的攻击方法,还提出了以下3种基于点添加的攻击方法。

(1) 添加独立点

使用两种不同的距离度量来添加对抗性的独立点。①倒角距离:原始点云与对抗点云之间的倒角距离(倒角距离为每个对抗点找到最近的原始点,将所有最近点对之间的距离求和)。②豪斯多夫距离:原始点云与对抗点云之间的豪斯多夫距离(豪斯多夫距离与倒角距离相似,除了它将找到的所有最近点对之间取最大平方欧几里得距离)。

(2) 添加对抗性点簇

结合以下3种不同的距离度量来添加对抗性点簇。①倒角距离:利用原始点云和对抗点簇之间的倒角距离将簇推向点云表面。②添加的集群数量:使用此度量可以控制点簇数量,文中只添加了1~3个集群。③最小化最远的距离:最小化每个聚类中最远的两个点之间的距离,以约束聚类中添加的点在小区内。

(3) 添加具有一定形状的对抗对象

结合3种不同的距离度量添加对抗对象。①倒

角距离:利用原始点云和对抗对象之间的倒角距离将对抗对象推向点云表面。②添加对象的数量。③ L_2 范数(对抗点云与原始点云之间的欧几里得距离),在一个真实世界的物体和一个对抗的物体之间的 L_2 范数被用来衡量两者之间的相似性,以生成类似于真实世界的形状的对抗对象。

由于搜索空间巨大,直接向无界的3D空间添加点是不现实的,因此上述3种攻击方法都采用临界点的位置作为对抗点(或集群或对象)的初始位置。

为了实现扰动的不可感知性,Liu等^[20]提出了一种名为分布式攻击的PGD攻击,该攻击以三角形网格(将原始点云表面近似为三角形网格)与对抗点云之间的豪斯多夫距离作为距离度量,将对抗点推向三角形网格(即目标表面)。这种方法优化了人类的不可感知性及形状攻击,使其更接近现实。图6展示了PGD攻击生成的两个对抗性点云。

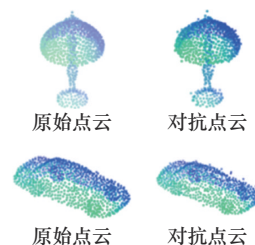


图6 PGD攻击生成的两个对抗性点云

Figure 6 Two adversarial point clouds generated by PGD attacks

为了最小化对抗攻击的扰动点数量, Kim等^[22]提出了最小对抗性攻击。在这个攻击方法中,改变点的数量被保持在最小值,并使用了两种不同的距离度量(倒角距离和豪斯多夫距离)以保持对抗物体和原来物体之间的相似性。图7展示了最小对抗性攻击生成的对抗点云。

为了充分限制对点云的修改,以保持其外观,Arya等^[28]通过在点云表面添加有限数量的点来创建有效的对抗性攻击。该方法引入了一种有效的步长调度算法以解决局部极大值问题,并有效地探索了点云面。结果表明,当添加的点的数量较少时,该方法对统计离群值移除(statistical outlier removal, SOR)防御方法有效。

通过添加这些点,攻击者能够让系统产生错误的检测结果,致使系统无法准确地感知周围环境。

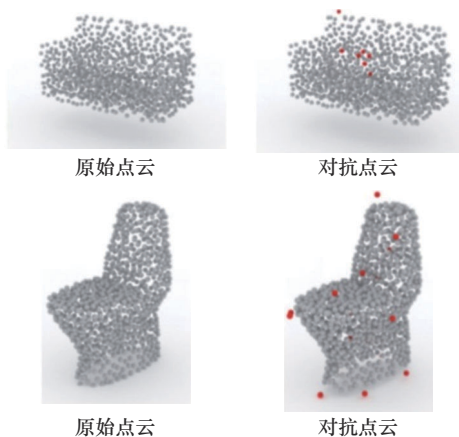


图7 最小对抗攻击生成的对抗点云

Figure 7 Adversarial point cloud generated by minimal adversarial attacks

2.1.3 基于点丢弃的攻击

基于点丢弃的攻击是针对激光雷达点云目标检测系统的另一种数字域攻击方法。攻击者可以选择性地删除原始点云中的一些点，使系统无法正确地感知特定对象或区域^[29,30]。

Zheng 等^[29]提出了一种构建 3D 点云的显著性图的有效方法，以衡量每个点对模型预测损失的贡献。该方法通过连续的点移动过程来近似点的删除，并结合球面坐标系下的损失梯度评分，从而建立一个可靠的显著性映射。利用显著性图，他们进一步标准化了点删除过程。图 8 展示了文献[29]的点云对抗攻击实验结果，从点删除后的物体的点云可以直观地看出，哪些部分的点对模型而言是关键的，哪些点相对不那么重要。这种视觉化不但有助于理解模型对不同部分点云的依赖程度，而且对优化 3D 模型的处理流程具有实际意义。

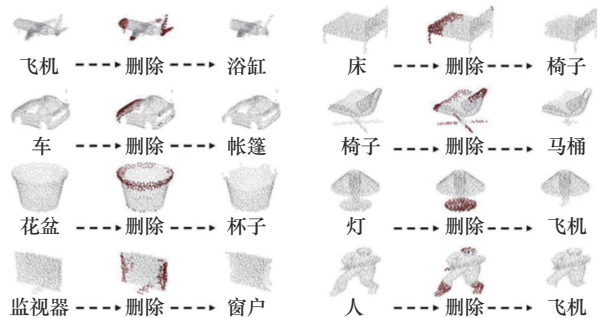


图8 文献[29]的点云对抗攻击实验结果

Figure 8 Experimental results of point cloud adversarial attack in reference[29]

为了研究目标检测模型在对抗环境中的鲁棒性，Wicker 等^[30]证明了在三维深度学习管道中由潜在空间平移引起的临界点集，对于点云和体积表示，暴露了对抗性遮挡攻击的弱点。

这些基于点丢弃的攻击通过删除道路上的障碍物点，使系统无法检测到它们，从而增加发生事故的风险。

2.1.4 基于点的可迁移性攻击

基于点的可迁移性攻击指对抗样本从一个模型迁移到另一个模型的能力。具体来说，如果一个模型生成的对抗样本可以被另一个模型错误分类，那么这个对抗样本就具有可迁移性。

Hamdi 等^[31]提出了 AdvPC 用来在三维点云上生成可迁移的对抗性扰动。这种方法可以生成具有高度一致性和可重复性的对抗性样本，并且可以在多个数据集上进行有效的迁移学习。他们还将对抗性扰动从一个源点云迁移到另一个目标点云以测试其鲁棒性。AdvPC 攻击示例如图 9 所示。

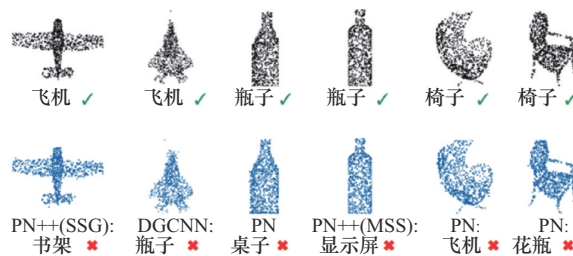


图9 AdvPC 攻击示例

Figure 9 Examples of AdvPC attack

鉴于现有方法要么因优化/迭代方案而极其缓慢，要么因不灵活而无法支持特定类别的针对性攻击，Zhou 等^[32]提出了一种新的标签导向生成对抗网络 (label-guided generative adversarial network, LG-GAN)，可以实现实时灵活的目标点云攻击。将原始点云和目标攻击标签输入 LG-GAN，LG-GAN 可以学习如何使点云变形，从而误导识别网络进入特定的标签。LG-GAN 不但能够有效地攻击多种基于点云的深度网络，而且攻击效果超过现有的一些先进方法。基于扰动的可迁移性攻击方法提高了攻击的通用性。

基于激光雷达点云目标检测的数字域攻击方法对比如表 1 所示

表1 基于激光雷达点云目标检测的数字域攻击方法对比

Table 1 Comparison of digital domain attack methods based on LiDAR point cloud target detection

攻击策略	文献	年份	方法名称	操作方法	白盒/黑盒	优化/梯度	数据集
目标攻击	[10]	2019	Perturbation Independent points Clusters Objects	移动+添加	白盒	优化	ModelNet40 ^[24]
	[27]	2020	AdvPC	添加	白盒	优化	ModelNet40 ^[24]
	[18]	2020	ShapeAdv	移动	白盒	优化	ModelNet40 ^[24]
	[32]	2020	LG-GAN	可迁移性	白盒	/	ModelNet40 ^[24]
	[19]	2020	GeoA3	移动	白盒	优化	ModelNet40 ^[24]
	[5]	2020	KNN	移动	白盒	优化	ModelNet40 ^[24]
	[21]	2020	JGBA	移动	白盒	优化	ModelNet40 ^[24]
	[23]	2022	ITA	移动	黑盒	优化	ModelNet40 ^[24]
非目标攻击	[11]	2019	Extended FGSM	移动	白盒	梯度	ModelNet40 ^[24]
	[29]	2019	Drop100 Drop200	丢弃	白盒	梯度	3DMNIST, ModelNet40 ^[24]
	[17]	2019	FGSM MPG Point-attachment Point-detachment	移动	白盒	梯度	ModelNet40 ^[24] , MNIST
	[20]	2020	Distributional attack Perturbation resampling Adversarial sticks Adversarial sinks	移动+添加	白盒	梯度	ModelNet40 ^[24]
	[22]	2021	Minimal	移动+添加	白盒	优化	ModelNet40 ^[24] ScanObjectNN ^[25]
	[28]	2023	VSA	添加		优化	ModelNet40 ^[24] ScanObjectNN ^[25]
	[30]	2019	Wicker	丢弃	白盒	优化	ModelNet10 ^[24] ModelNet40 ^[24] KITTI ^[26]

2.2 基于激光雷达点云目标检测的物理域攻击

物理域攻击是一种新型的攻击方式，主要通过操纵硬件设备来影响其输出，进而对基于该硬件设备的系统造成危害。这种攻击方式具有隐蔽性强、难以检测和防御的特点，在信息安全领域受到了广泛关注。

在激光雷达点云目标检测中，物理域攻击主要表现为对激光雷达发射的激光信号进行干扰或篡改，从而影响激光雷达对目标的检测和识别。例如，攻击者可以向激光雷达发射器发送强烈的光信号，使其饱和并产生虚假的目标回波；或者在激光信号传输路径上放置障碍物，使激光信号发生散射和衰减，从而降低激光雷达探测目标的精度。

2.2.1 基于激光的攻击

在基于激光的攻击方法^[33-37]中，攻击者试图操纵激光雷达传感器的信号探测系统。文献[33]中基于激光的攻击方法的原理如图10所示，攻击者使用光电二极管接收来自激光雷达的激光脉冲，并通过激活延迟元件触发攻击激光，模拟真实的回波脉冲。

Cao等^[33]首次发现，在实际应用中的自动驾驶系统使用基于激光雷达的感知深度学习模型，通过策略性地注入少量的欺骗性激光雷达点，使感知系统出现误检测，检测到假车辆（现实中不存在的车辆）。这种欺骗攻击可能会导致紧急刹车等严重后果，对乘客造成潜在危险。该攻击评估是在一个特定的模型（百度阿波罗2.5）上进

行的, 且假定了白盒访问, 这在实际应用场景中可能是不现实的。

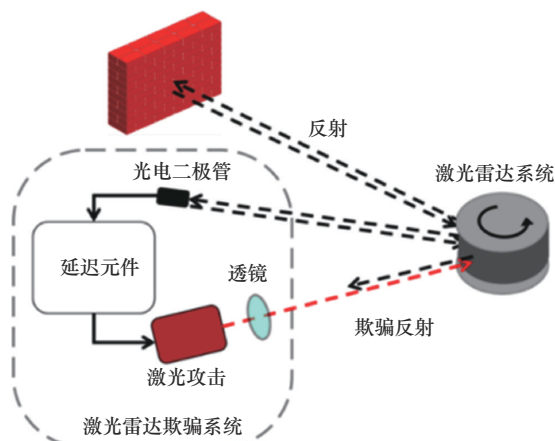


图 10 文献[33]中基于激光的攻击方法的原理

Figure 10 Principle of laser-based attack method in the reference[33]

针对现有的攻击存在有效性和通用性的限制, Sun 等^[34]进一步拓展了针对基于激光雷达感知系统的欺骗攻击, 通过分析流行的 3D 数据集和多个深度学习模型, 对模型本身的脆弱性有了深入了解, 在此基础上提出了黑盒欺骗攻击, 发现在激光雷达点云中被忽略的遮挡模式(遮挡模式通常需要考虑遮挡区域或者传感器感知范围之外的隐藏目标, 以保证被动安全)使自动驾驶汽车容易受到欺骗攻击。对于自动驾驶汽车而言, 正确处理遮挡区域内的目标是一个挑战, 因为这些目标可能在关键时刻出现, 对行车安全构成威胁。然而, Sun 等指出, 如果恶意攻击者使用这些遮挡模式来引入误导性的点云数据, 就有可能使自动驾驶系统做出错误的判断, 从而容易受到欺骗攻击。这种攻击方式的发现表明, 即使是先进的激光雷达感知系统也存在潜在的安全漏洞。这种攻击在多个数据集上取得了成功。

Petit 等^[35]演示了激光雷达欺骗攻击, 通过发射和激光雷达相同频率的激光束来干扰激光雷达接收反射回来的激光。传感器系统很难识别这种攻击, 因为这种攻击不需要对传感器进行任何物理接触或篡改, 也不会干扰数字传感器测量的处理和传输。这种类型的攻击可以欺骗受害者的传感器, 提供看似合法但实际上错误的数据。Hau 等^[36]演示了如何使用相同的物理设备来设计一

个新的、更危险的攻击方法, 即对象删除攻击。对象删除攻击的目的是迫使 3D 目标探测器失效。利用激光雷达的默认设置, 记录每个方向的单一返回信号来干扰 3D 对象感兴趣区域(region of interest, RoI)中的点云。该攻击方法有效降低了常用的三维目标检测模型的性能。

Cao 等^[37]提出了一种基于激光雷达传感器数据固有的自动转换和过滤过程的攻击方法, 实现了以人眼不可见的方式来隐藏物体, 并欺骗自动驾驶车辆的障碍物检测器, 这种攻击称为物理移除攻击(physical removal attacks, PRA)。物理移除攻击的成功实施意味着即使真实世界中的物体并未被移除或改变, 自动驾驶系统的感知设备也可以被有效地欺骗。

基于激光的攻击方法尽管在隐藏目标和欺骗自动驾驶系统方面具有创新性, 但仍存在一些局限性。首先, 这些攻击方法需要事先知道受害者所使用的激光雷达的激光发射频率, 并需要考虑干扰器和被攻击的激光雷达同步问题; 其次, 这些攻击方法只能生成较少数量的随机欺骗点, 难以生成特定的攻击目标。这使得它们在物理世界中难以实现, 特别是当受害者车辆正在移动时。

2.2.2 基于物体的攻击

对于基于目标的攻击方法, 文献[38-45]提出使用具有对抗性形状的物体来攻击激光雷达感知系统。

Cao 等^[38]为了生成物理上可实现的攻击, 提出了一种基于激光雷达渲染器生成对抗性点云的对抗性网格。然而, 他们只考虑了学习特定框架的对抗性网格, 因此生成的 3D 对抗性物体不具有通用性, 难以在其他 3D 场景中重用。此外, 他们仅在一个非常小(几百帧)的内部数据集上评估了攻击方法。如图 11 所示, 第 1 行显示基于激光雷达的检测系统检测到一个正常的框, 第 2 行显示生成的大小相似的对抗对象没有被检测到。Tu 等^[39]通过学习生成了一个通用的且物理可实现的对抗性物体, 展示了在任何目标车辆的车顶上放置一个对抗性物体的攻击方法(如图 12 所示), 该攻击方法的攻击成功率为 80%。Yang 等^[40]在数字域上针对 PointRCNN^[45]的两个阶段分别进行了白盒攻击, 并利用遗传算法优化扰动

点来实现黑盒攻击，然后在物理域将对抗点云打印出来放在路边，使激光雷达将对抗点云误识别为小车（如图13所示）。

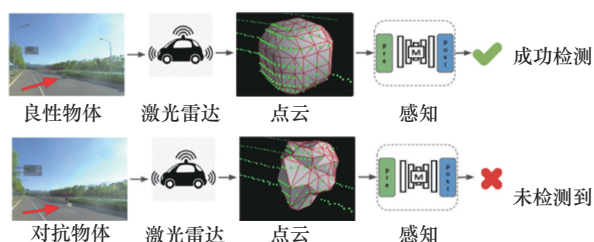


图11 文献[38]对抗性物体攻击示例

Figure 11 Examples of adversarial object attack in reference [38]

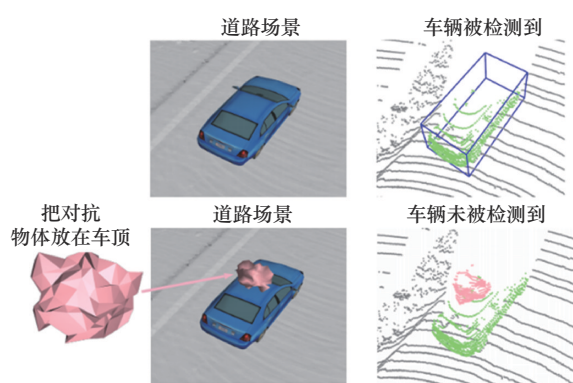


图12 文献[39]对抗性物体攻击示例

Figure 12 Examples of adversarial object attack in reference [39]

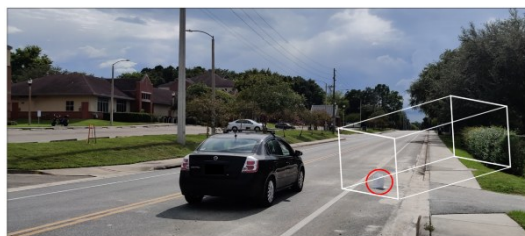


图13 文献[40]对抗性物体攻击示例

Figure 13 Examples of adversarial object attack in reference [40]

对于基于物体的攻击，不仅限于单一模态的，还包括将激光与其他传感器结合进行攻击。Abdelfattah等^[41]的研究表明，基于物体的攻击可以跨越不同感知模态，并成功地整合了激光雷达与其他类型传感器的攻击方法。他们专注于研究级联多模态深度学习网络，这些网络通常用于提高自动驾驶系统对周围环境的理解能力，通过融合来自不同传感器的数据来增强系统的检测性能。实验结果显示攻击成功率达到了近73%。

Tu等^[42]通过在车辆上插入对抗性物体来展

示多传感器检测的脆弱性。他们专注于研究物理上可实现的与输入无关的攻击，这些攻击在实践中是可行的。研究结果表明，一个单一的通用对抗性物体可以在最先进的多模态探测器中成功地隐藏不同的车辆，如图14所示，将对抗物体放置在车顶，多传感器系统未检测到汽车。



图14 文献[42]对抗性物体攻击示例

Figure 14 Examples of adversarial object attack in reference [42]

然而，这些物体被设计成特定的形状和大小，这使得在实践中使用它们进行攻击的灵活性受到限制。此外，这些对抗性物体不同寻常的形状可能引起人类感知的怀疑，容易被防御者和受害者发现。此外，在实践中，高精度地生成这种特定形状的对象是一项挑战。

2.2.3 基于位置的攻击

基于位置的攻击方法致力于寻找目标物体的关键对抗位置，在这些对抗位置放置任何具有反光的物体即可成功发起攻击。

Zhu等^[46]提出了一种创新的攻击框架，攻击者可以在此基础上识别物理空间中的关键对抗位置，通过在这些位置周围放置具有反射表面的任意物体，攻击者能够轻松地欺骗激光雷达感知系统。图15展示了基于位置的对抗攻击。

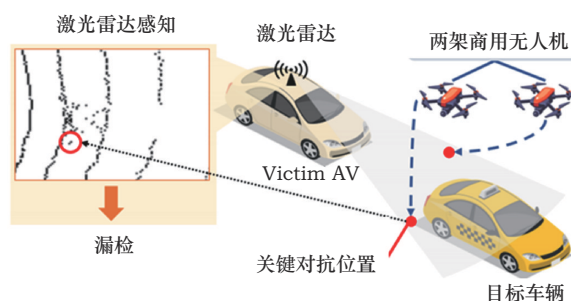


图15 基于位置的对抗攻击

Figure 15 Location-based adversarial attack

基于位置的攻击相对其他的攻击具有更高的灵活性。在这种攻击中,攻击者只需识别物理空间中的一些对抗性位置,然后在这些位置放置任意具有反射激光功能的物体即可,而无须生成特定的物体形状。这种方法的优点是不受特定形状的制约,提高了攻击的灵活性和可实施性。然而,这些位置对场景的要求比较高。攻击者需要准确地确定那些能够有效干扰激光雷达系统的对抗性位置,并在这些位置放置反射激光的物体。因此,攻击者需要对目标系统和场景有深入的了解,这可能增加了攻击的策划和执行难度。

2.2.4 其他类型的攻击

Li 等^[47]通过欺骗自动驾驶汽车的轨迹使得安全关键物体无法被检测到或以不正确的位置检测到。

Zhang 等^[48]提出了一种新的后门攻击策略,攻击者可以通过“毒害”少量的点云样本来实现

攻击目标。这种攻击策略是可物理实现的,允许攻击者使用一些公共对象作为触发器,轻松地执行攻击。这种后门攻击的特殊性在于,攻击者能够以相对较小的代价成功实施攻击,而且攻击的物理实施方式相对容易。

基于激光雷达点云目标检测的物理域攻击研究现状如表 2 所示。

3 基于激光雷达点云的防御方法

针对三维点云的对抗性攻击的防御方法一般分为 4 类:输入变换^[10,11,40,49-55]、数据优化^[11,17,23,56-57]、传感器级别的优化^[58-67]和深度模型的优化^[68-99]。下面分别讨论这些类别下的防御方法。

3.1 输入变换

输入变换是一种预处理方法,用于在深度模型处理输入点云之前引入一些变换。这种变换可

表 2 基于激光雷达点云目标检测的物理域攻击研究现状

Table 2 Research status of physical domain attack based on LiDAR point cloud target detection

物理攻击方式	文献	年份	优化、梯度	黑盒/白盒	数据集	方法简介
发射激光	[35]	2015	—	黑盒	—	发射激光束干扰激光雷达的接收
	[33]	2019	优化	白盒	real-world LiDAR sensor data	向目标物体注入少量对抗点云
	[34]	2020	优化	白盒+黑盒	KITT ^[26]	使用遮挡点云形成对抗
	[36]	2021	—	黑盒	KITTI ^[26]	简单随机点选择策略进行对象删除攻击
	[37]	2023	—	白盒	KITTI ^[26]	利用激光雷达固有的自动转换和过滤过程实现物理移除攻击
放置物体	[38]	2019	梯度	白盒+黑盒	—	优化扰动并打印对抗物体放置在路边(对激光雷达有效)
	[39]	2020	优化	白盒+黑盒	KITTI ^[26]	优化扰动并打印对抗物体放置在车顶(对激光雷达有效)
	[40]	2021	优化	白盒+黑盒	KITTI ^[26]	优化扰动并打印对抗物体放置在路边(对激光雷达有效)
	[41]	2021	优化	白盒	KITTI ^[26]	优化扰动并打印对抗物体放置在车顶(对相机+激光雷达有效)
	[42]	2021	优化	白盒+黑盒	KITTI ^[26] XENITH ^[42]	优化扰动并打印对抗物体放置在车顶(对相机+激光雷达有效)
	[43]	2021	优化	白盒	KITTI ^[26]	优化扰动并打印对抗物体放置在路边(对激光雷达有效)
	[44]	2021	梯度	白盒	KITTI ^[26]	优化扰动并打印对抗物体放置在车顶(对相机+激光雷达有效)
	[48]	2022	优化	白盒	KITTI ^[26]	注入后门攻击
寻找对抗位置	[46]	2021	优化	黑盒	KITTI ^[26]	寻找能够实现对攻击的关键位置,在此处放置反光板实现对抗
扰动轨迹	[47]	2021	优化	白盒	nuScenes ^[31]	通过扰动轨迹实现对抗

以设计为降低模型对对抗性攻击的敏感性, 或者使攻击者更难以生成对抗性点云。

在点云数据处理中, 一种常用的预处理技术是移除异常值。为了减小输入数据中扰动点的负面影响, 研究者提出了多种解决方法。Liu等^[11]设计了一种基于模型梯度的方法来识别并去除那些对模型输出影响显著的点。该方法通过评估每个点的显著性, 优先移除那些对模型变化影响最大的点。该方法适用于模型对输入数据敏感度高的场景, 因为它能够直接移除对模型输出影响最大的点。这种方法的优点是可以有针对性地减少输入数据的噪声。与此相似, Zhou等^[49]采用了一种基于统计的方法, 通过计算每个点与其 k 个近邻之间的欧几里得距离来检测异常值, 随后移除那些远离邻近群体的离群点。该方法侧重于利用局部邻域信息来识别异常值。这种方法的优点是不需要模型的梯度信息, 而是仅依赖于几何距离的度量。这使得它在不知道模型具体信息的情况下也能够有效地使用。然而, 这种方法的性能依赖于 k 的选择及数据的内在分布特性。Xiang等^[10]考虑到扰动点通常少于原始点, 提出了一种简单的随机抽样方法, 即从输入点云中随机抽取并去除一部分点, 以减小扰动点对整体数据的影响。该方法提供了一种简单而有效的策略, 尤其是当扰动点相对较少时。这种方法的优点是计算成本低、易于实施, 但它可能会无差别地移除重要的原始点和扰动点, 因此需要在抽样率上进行权衡, 以平衡去噪和保留重要信息之间的关系。

另一种点云输入变换技术是采用基于密度估计的方法。Yang等^[40]利用从激光雷达传感器获得的点云数据, 通过点的密集程度来区分恶意边界框和正常边界框。研究者首先确定了最靠近激光雷达传感器且位于边界框内的特定点 P_n , 随后计算了以 P_n 为中心、指定半径 r 球体中点的数量。通过恰当地设置半径 r , 该方法能够有效地将恶意对象与正常车辆区分开来。在密度特征提取之后, 使用支持向量机(support vector machine, SVM)对候选对象进行分类, 进一步识别出真实的车辆和恶意对象。该方法基于一个关键假设, 即对抗性边界框和良性边界框在点云中

的密度分布是不同的。这种方法的优点是能够利用点云数据的三维结构信息, 从而提供一种直观的方式来区分不同性质的物体。通过选择恰当的半径 r 来定义点的密度, 这种方法可以适应不同的场景和对象大小。

还有一种输入变换方法是 DUP-Net (denoiser and upsampler network for 3D adversarial point clouds defense)^[50]。DUP-Net 的处理过程分为两个步骤: 首先, 将 SOR 作为去噪器, 用于移除异常值; 然后, 将输出传递给上采样网络, 以产生更密集的点云。该防御方法使用密度更大的点云跟踪均匀分布的点云下表面, 以恢复那些由于对抗性扰动通常会丢失原始点云的临界点。通过结合去噪和上采样技术, 该方法不仅能够清除恶意添加的扰动点, 还能修复由此产生的数据缺失。该方法能够保持原始点云的特性, 这对许多依赖于精确几何形状的应用程序来说是至关重要的。

Liu等^[53]开发了一种方法, 即在原始点云中创建一些随机的点子集, 然后根据这些随机子集的标签中的多数投票来预测原始点云的标签。

输入变换作为一种预处理手段, 通过去除异常值、计算密度和使用去噪上采样技术等方法来提高模型对点云数据中潜在对抗性扰动的鲁棒性。这种方法的优点包括能够降低攻击敏感性、提升输出准确性、保持原始点云特性、实施简便和适应性强。然而, 它也存在缺点, 如可能错误地删除重要点、性能依赖于参数选择和数据分布, 并且不能完全消除所有对抗性攻击的风险。

3.2 数据优化

针对训练的数据优化, 是指优化训练数据以提高深度模型抵御对抗性攻击的鲁棒性, 如数据增强、对抗训练等。

Liu等^[11]首次描述了点云的对抗性训练。为了进一步提升对抗性训练的效果, Liang等^[57]采用了自适应攻击的方法。该方法引入了一个微扰注入模块, 能够巧妙地将各种不同类型的攻击注入深度模型中, 从而生成用于对抗训练的扰动特征。自适应攻击的方法使模型能在多种攻击模式下进行训练, 从而在面对未知攻击时也能有较好的防御能力, 增强了模型在真实环境下的实用

性。而微扰注入模块能将多种攻击类型引入模型训练中,使模型不局限于特定的攻击模式,增加了模型对攻击的泛化能力。Sun 等^[58]将自监督学习应用于三维点云的对抗性训练中。自监督学习是一种无监督学习方法,通过利用未标记的数据来学习数据的内在结构和分布。在对抗性训练中,自监督学习用来提升模型对点云的特征和语义信息的理解,从而提升其抵御对抗性攻击的鲁棒性。数据优化技术在提高深度模型抵御对抗性攻击的鲁棒性方面扮演着重要角色。这些技术通过增加数据的多样性和复杂性,能够显著提升模型的泛化能力和防御能力,减少过度拟合的风险,并增强模型对数据内在结构和语义信息的理解能力。然而,数据优化会增加计算资源的消耗、实施复杂度及潜在的数据偏差问题,且无法完全保证模型抵御所有类型的对抗性攻击。因此,在采用数据优化策略时,需要权衡其带来的正面与负面影响,以确保模型在实际应用中的稳定性和可靠性。

3.3 传感器级别的优化

传感器级别的优化是通过设计多传感器融合(multi-sensor fusion, MSF)的结构,提高自动驾驶系统的检测精度和鲁棒性。例如,一些高级自动驾驶系统,如Waymo、百度Apollo和Pony.ai,通常采用多传感器融合的设计。许多研究工作致力于改进传感器融合模块的性能,以提高自动驾驶系统的安全性和可靠性。

Liang 等^[59]提出了一种端到端可学习的架构,采用连续卷积来融合不同分辨率级别的图像和激光雷达特征图。利用连续卷积层来整合不同来源的数据,能够更高效地处理和融合图像和激光雷达特征,有助于提升数据处理速度和模型的实时性能。该方法通过在不同分辨率级别上融合图像和激光雷达特征,能有效提高三维目标检测的准确性和鲁棒性,尤其是在复杂的环境条件下。他们设计的连续融合层编码了离散状态的图像特征和连续的几何信息,从而实现了一种新颖、可靠、高效、基于多个传感器的端到端可学习的三维目标检测器。Chen 等^[60]提出了多视图三维网络(multi-view 3D network, MV3D),这是一个感知融合框架,以激光雷达点云和 RGB

图像作为输入,用于预测三维边界框的方向。该网络能够综合多种传感器的信息,以提高目标检测的准确性和鲁棒性。MV3D通过融合激光雷达点云和 RGB 图像的数据,有效提高了目标检测的准确性。这种多传感器信息的整合,增强了模型在复杂环境下的鲁棒性,尤其是在光照变化或遮挡等挑战性条件下。

传感器级别的优化使系统能够综合多种传感器数据,从而全面理解周围环境,并提升目标检测的准确性和鲁棒性。然而,这种优化同样会带来增加系统复杂性和计算资源消耗的缺点,特别是在实时处理和模型维护方面需要更多的技术和资源投入。尽管存在这些挑战,传感器级别的优化仍然是提升自动驾驶系统性能和安全性的关键途径,对于推动自动驾驶技术的发展和具有应用具有重要意义。

3.4 深度模型的优化

深度模型优化指的是对深度模型本身的架构进行优化,以提高抵御对抗性攻击的鲁棒性。这可以在训练过程中通过对原始深度神经网络结构的调整来实现。在目标检测领域,基于激光雷达的目标检测模型主要分为基于点、基于网格、基于点云-体素和基于范围 4 种形式。

基于点的 3D 目标检测方法^[68-73]采用深度学习技术处理点云数据,并提出了不同架构用于实现基于原始点云的 3D 目标检测。这些方法主要包括点云的采样和特征的学习两个基本组成部分。通过采样点云数据并学习特征,可以实现对 3D 目标的准确检测。这种方法的优点是能够保留点云数据的完整性,从而准确地捕捉到目标的几何特征;缺点是计算成本较高,因为需要对每个点进行特征提取和处理。

基于网格的三维目标检测方法^[74-90]首先使用光栅化技术将点云数据转化为离散的网格表示形式,如体素、柱状图和鸟瞰视图(bird's eye view, BEV)特征图等;然后采用传统的 2D 卷积神经网络或先进的 3D 稀疏神经网络技术从网格结构中提取关键特征;最后,由 BEV 网格单元完成 3D 对象的精确检测任务。这种方法的优点是可以利用成熟的卷积网络架构提高计算效率,不过,在转换过程中可能会损失一些原始点

云的细节信息。

基于点云-体素的目标检测方法^[91-99]采用混合架构, 结合点云和体素进行三维目标检测。这些方法可分为单阶段和两阶段检测框架。基于单阶段点云-体素的三维目标检测器致力于将点云和体素的特征与骨干网络中的点云到体素及体素到点云变换进行有机结合。而基于两阶段点云-体素的三维对象检测器针对不同的检测阶段采用了不同的数据表示形式。具体来说, 在第一阶段, 采用基于体素的检测框架来产生一组3D对象候选; 在第二阶段, 首先从输入点云中采样关键点, 然后通过点算子从关键点中进一步优化3D候选, 最终实现准确的检测。这种方法的优点是能够同时利用点云的精确性和体素化的计算效率; 缺点是设计和实现较复杂, 需要平衡两种数据形式的转换和特征融合。

基于范围图像的目标检测模型^[100-107]采用了一种密集而高效的二维数据结构, 其中每个像素编码了对应于场景中物体的三维距离信息, 而非传统的RGB颜色值。这种基于范围的检测方法主要通过两个途径来解决目标检测的挑战, 一是开发专为处理范围图像特性而设计的新型模型架构和处理算子, 二是精心选择用于检测的合适视角。范围图像保持了与RGB图像类似的二维结构, 因此基于范围的3D目标检测算法能够借鉴并适应现有的二维目标检测模型, 以实现范围图像的处理。这种方法的优点是能够直接从二维图像中提取深度信息, 简化了数据处理流程。

上述深度模型优化方法各有优势, 可以根据实际应用场景灵活选择, 以有效提升目标检测模型的鲁棒性。

4 基于激光雷达点云的数据集及评价指标

近年来, 基于激光雷达点云的攻击以及防御技术不断涌现, 为了方便不同方法之间的性能比较, 本节总结了目前常用的数据集和评价指标, 便于研究者进行后续研究。

4.1 数据集

在基于激光雷达点云的攻击及防御方面, 常用的数据集有KITTI、nuScenes和Waymo。

KITTI是由德国卡尔斯鲁厄理工学院和丰田美国技术研究院联合创建的自动驾驶场景下的计算机视觉算法评测数据集。该数据集包含市区、乡村和高速公路等场景采集的真实图像数据, 每张图像中最多达15辆车和30个行人, 同时还存在各种程度的遮挡与截断情况。整个数据集由389对立体图像和光流图、39.2 km视觉测距序列和超过20万张带有3D标注的物体图像组成, 以10 Hz的频率采样及同步。对于3D物体检测, 标签细分为车辆、货车、卡车、行人、坐着的行人、骑自行车的行人、有轨电车及Misc(如拖车、代步车等小众类型目标统一归为“Misc”类别)组成。

nuScenes是一个公开的大型数据集, 用于自动驾驶, 由Motional(前身是nuTonomy)团队开发。该数据集由1 000个场景组成, 每个场景的长度为20 s, 包含各种各样的情景。在每个场景中, 有40个关键帧(key frame), 即每秒有2个关键帧, 其他的帧为扫描帧(sweep frame)。这些关键帧经过手工标注, 每一帧都有若干个标注, 标注的形式为边界框。标注的信息包括大小、范围、类别、可见程度等。这样的标注为数据集提供了详细而准确的物体信息。

Waymo是由Google母公司Alphabet旗下的自动驾驶公司Waymo团队开源的自动驾驶数据集, 以帮助研究界在机器感知与自动驾驶技术方面实现突破。Waymo数据集包含3 000段驾驶记录, 总时长达16.7 h、平均每个记录约为20 s。该数据集共有约60万帧, 其中包含大约2 500万3D边界框和2 200万2D边界框。此外, Waymo数据集展示了多样化的自动驾驶场景, 为研究提供了丰富的实际驾驶数据。

4.2 评价指标

基于激光雷达的目标检测的对抗攻击与防御效果需要用指标来衡量。精确率、召回率、交并比、攻击成功率等指标被广泛用于评估基于激光雷达的目标检测模型的性能, 为模型优化和算法改进提供有力的量化依据。

1) 精确率(Precision)。精确率是指所有预测正确值TP(N_{TP})占所有预测值($N_{TotalDetections}$, 包括TP和FP)的分数, 即所有预测为正样本的

样本中有多少是真正的正样本, 计算式为

$$\text{Precision} = \frac{N_{\text{TP}}}{N_{\text{TotalDetections}}} \quad (1)$$

2) 召回率 (Recall)。召回率是指所有预测正确值 TP (N_{TP}) 占有真值 ($N_{\text{GroundTruth}}$) 的分数。

$$\text{Recall} = \frac{N_{\text{TP}}}{N_{\text{GroundTruth}}} \quad (2)$$

3) 交并比 (intersection-over-union, IoU)。IoU 是基于激光雷达的目标检测中常用的一个指标, 是测量在特定数据集中检测相应物体准确度的一个标准。IoU 表示产生的候选框与原标记框的交叠率 (或者说重叠度), 也就是它们的交集与并集的比值。相关度越高, 该值越大。最理想的情况是完全重叠, 即比值为 1。

4) Precision-Recall (P-R) 曲线。P-R 曲线是以召回率为横轴、精确度为纵轴绘制的曲线。通过比较不同阈值下的精确度和召回率, 可以确定最佳的阈值来平衡精确度和召回率, 从而选择最优的分类模型。

5) 平均精度 (average precision)。平均精度是由 P-R 曲线形成的面积。

6) 平均精度均值 (mAP)。mAP 就是对所有类的平均精度 (AP) 进行平均。

7) 攻击成功率 (ASR)。ASR 定义为所有示例中攻击成功的例子 (激光雷达帧) 的百分比。

$$\text{ASR} = \frac{\text{攻击成功的样本数}}{\text{总的样本数}} \quad (3)$$

5 未来研究方向

基于激光雷达点云的目标检测安全研究工作将面临更多的机遇和挑战。例如, 如何高效地对激光雷达点云目标检测系统进行漏洞挖掘和鲁棒加固是安全领域相关研究人员的关注点之一。本节根据目前国内外的研究现状, 基于上述对激光雷达点云攻防相关研究工作的介绍, 探讨其未来的研究发展方向。

(1) 对抗攻击物理域实现

现有的激光雷达点云目标检测对抗攻击方法侧重于数字域的实施, 对于物理域的可行性考虑相对较少。尽管一些方法涉及物理域的实现, 但它们可能缺乏持久的对抗效果, 或者在实际应用中难以实现。因此, 未来的研究可着

眼于设计易于在物理域实施且能够动态变化的持续性对抗攻击技术, 以更全面地挖掘实际系统安全漏洞, 为实际应用中可能面临的各种挑战做好准备。这也有助于提高自动驾驶技术在真实世界中的可靠性和安全性。具体可以采用以下方法提高对抗攻击物理域的实现: ①研究物理世界中对抗样本的生成方法, 考虑环境因素 (如光照、距离、角度) 对激光雷达的影响; ②开发新型材料或技术, 以在物理环境中生成与数字域相对应的对抗性模式, 如利用特殊涂层或 3D 打印材料制造具有扰动效果的物体; ③设立实验场景, 进行模拟攻击测试, 收集不同环境变量下的攻击数据, 用于训练和评估具有更高鲁棒性的目标检测系统。

(2) 高效防御

随着激光雷达技术的不断进步, 攻击者有可能采用更高效且更具隐蔽性的手段来对目标检测系统进行攻击。为此, 防御者需要致力于研发更加精准和高效的防御措施, 以应对日益复杂的攻击。因此, 未来的研究可着力于提升激光雷达系统的精度和鲁棒性, 积极开发新的防御算法和技术, 以确保系统能够更有效地应对各种潜在的攻击。具体来说, 研究人员可从以下几个方面入手: ①优化传感器硬件设计, 以提高信号质量和减少噪声干扰; ②改进数据处理算法, 如滤波、配准等, 以提高目标检测系统的准确性和稳定性; ③利用机器学习算法进行异常检测, 自动识别和过滤掉潜在的对抗性输入, 确保数据的准确性和一致性。

(3) 可解释性

当前, 针对激光雷达系统的对抗性攻击主要依赖于优化生成方法。然而, 这些方法存在一个关键问题, 即可解释性不足。这意味着人们很难理解模型是如何做出决策的。这对于安全性和可靠性来说, 是一个重大隐患。增强模型的可解释性, 有助于理解模型的决策过程, 从而更好地设计和改进漏洞挖掘及防御机制。为此, 未来的研究可以致力于提高激光雷达系统的可解释性, 使其决策过程更加透明。具体可从以下几个方面入手。①利用可视化工具, 如类激活映射 (class activation maps, CAM) [108]、梯度加权类

激活映射 (gradient-weighted class activation maps, Grad-CAM) [109], 展示模型在进行决策时所重视的输入区域。这有助于人们理解模型为何以特定方式响应输入。②开发标准化的可解释性评估框架和指标, 用来衡量不同模型的解释能力, 为设计既准确又具有高度可解释性的系统提供指导。③应用可解释性技术, 如局部解释模型-无关解释 (local interpretable model-agnostic explanations, LIME) [110], 量化各个输入特征对模型预测的贡献度。这有助于揭示哪些特征是决策过程中的关键因素。

6 结束语

随着激光雷达在自动驾驶领域的广泛应用, 深度学习模型也在自动驾驶领域发挥着越来越重要的作用。然而, 面对对抗性输入时, 这些模型往往容易出现错误的识别或决策, 给人们的生命和财产安全带来潜在威胁。因此, 在部署激光雷达的目标检测模型之前, 必须对其进行安全性测试, 以确保模型具有足够的鲁棒性和安全性, 从而保障自动驾驶系统的安全稳定运行。利用生成的对抗扰动对目标模型进行测试, 以提前发现潜在的漏洞, 降低模型出错的风险。对发现的漏洞进行修复加固, 进一步提升模型的安全性。如何全面且高效地进行模型的漏洞挖掘和加固, 以增强激光雷达点云目标检测模型的鲁棒性, 成为当前激光点云目标检测研究领域关注的焦点。

目前, 面向激光雷达点云的安全测试已在自动驾驶领域得到应用, 但对于攻击和防御尚缺少全面、完整的综述。本文总结了基于激光雷达的目标检测的安全问题, 包括激光雷达点云目标检测和攻防原理、攻击方法和防御方法, 以及此领域常用的数据集和评价指标。此外, 对该领域未来的研究方向展开了讨论, 旨在为推动激光雷达安全的进一步发展和应用提供帮助。

参考文献:

- [1] CHANG K, HUANG P, YU H G, et al. Audio adversarial examples generation with recurrent neural networks[C]//Proceedings of the 25th Asia and South Pacific Design Automation Conference (ASP-DAC). Piscataway: IEEE Press, 2020: 488-493.
- [2] GIRALDO J, CÁRDENAS A A, KANTARCIOGLU M, et al. Adversarial classification under differential privacy[C]//Proceedings of 27th Annual Network and Distributed System Security Symposium. Virginia: the Internet Society, 2020. 2020:1-18.
- [3] JIN K, ZHANG T W, SHEN C, et al. A unified framework for analyzing and detecting malicious examples of DNN models[J]. arXiv preprint arXiv:2006.148718.9, 2020.
- [4] PAPERNOT N, MCDANIEL P D, GOODFELLOW I J, et al. Practical black-box attacks against deep learning systems using adversarial examples[J]. arXiv preprint arXiv: 1602.026971.2, 2016.
- [5] TSAI T, YANG K, HO T-Y, et al. Robust adversarial objects against deep learning models[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Hong Kong:AAAI Press, 2020: 954-962.
- [6] ZHANG G M, YAN C, JI X Y, et al. Dolphinattack: inaudible voice commands[C]//Proceedings of the 2017 ACM SIGSAC conference on computer and communications security. New York: ACM Press, 2017: 1-15.
- [7] ZHANG Z H, LIU M X, ZHANG C, et al. Argot: generating adversarial readable chinese texts[C]//Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence. 2021: 2533-2539.
- [8] JIA Y H, LU Y T, SHEN J J, et al. Fooling detection alone is not enough: adversarial attack against multiple object tracking[C]//Proceedings of International Conference on Learning Representations (ICLR'20). 2020: 1-11.
- [9] YANG K C, TSAI T, YU H G, et al. Beyond digital domain: fooling deep learning based recognition system in physical world[C]//Proceedings of the the 34th AAAI Conference on Artificial Intelligence. Hong Kong:AAAI Press, 2020: 1088-1095.
- [10] XIANG C, QI C R, LI B, et al. Generating 3D adversarial point clouds[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 9136-9144.
- [11] LIU D, YU R, SU H. Extending adversarial attacks and defenses to deep 3D point cloud classifiers[C]// Proceedings of 2019 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE Press, 2019: 2279-2283.
- [12] 冉晟晔. 基于旋转扫描的激光雷达的研究与设计[D]. 广州: 广州大学, 2022.
RAN S Y. Research and design of LiDAR based on rotary scanning[D]. Guangzhou: Guangzhou University, 2022.
- [13] 周俞辰. 基于激光三角测距法的激光雷达原理综述[J]. 电子技术与软件工程, 2016(19):94-95.
ZHOU Y C. A review of LiDAR principles based on laser triangulation ranging method [J]. Electronic Technology and Software Engineering, 2016(19): 94-95.
- [14] YANG C, HUO Y, CHEN Q, et al. Time-of-flight measurement in self-triggering pulsed laser ranging[J]. Optical Engineering, 2005, 44(3).
- [15] 汪双. 基于旋转激光雷达的融合定位建图方法[D]. 绵阳: 西南科技大学, 2024.
WANG S. Fusion positioning and mapping method based on rotary LiDAR[D]. Mianyang: Southwest University of Science and Technology, 2024.

- [16] MAO J G, SHI S S, WANG X G, et al. 3D object detection for autonomous driving: a comprehensive survey[J]. *International Journal of Computer Vision* 2023, 131(8): 1909-1963.
- [17] YANG J C, ZHANG Q, FANG R Y, et al. Adversarial attack and defense on point sets[J]. *arXiv preprint arXiv: 1902.10899*, 2019.
- [18] LEE K, CHEN Z Y, YAN X C, et al. ShapeAdv: generating shape-aware adversarial 3D point clouds[J]. *arXiv preprint arXiv: 11626 (2020)*, 2005.
- [19] WEN Y X, LIN J H, CHEN K, et al. Geometry-aware generation of adversarial point clouds[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2022, 44(6): 2984-2999.
- [20] LIU D, YU R, SU H. Adversarial shape perturbations on 3D point clouds[C]//*Proceedings of Computer Vision-ECCV 2020 Workshops*. Berlin: Springer, 2020: 88-104.
- [21] MA C C, MENG W L, WU B Y, et al. Efficient joint gradient based attack against sor defense for 3D point cloud classification[C]// *Proceedings of the 28th ACM International Conference on Multimedia*. New York: ACM Press, 2020: 1819-1827.
- [22] KIM J, HUA B-S, NGUYEN D T, et al. Minimal adversarial examples for deep learning on 3D point clouds[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Piscataway:IEEE Press, 2021: 7778-7786.
- [23] LIU D Z, HU W. Imperceptible transfer attack and defense on 3D point cloud classification[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2022, 45(4): 4727-4746.
- [24] WU Z R, SONG S, KHOSLA A, et al. 3D shapeNets: a deep representation for volumetric shapes[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. Piscataway:IEEE Press, 2015: 1912-1920.
- [25] UY M A, PHAM Q, HUA B, et al. Revisiting point cloud classification: a new benchmark dataset and classification model on real-world data[C]//*Proceedings of the IEEE/CVF international conference on computer vision*. Piscataway: IEEE Press, 2019: 1588-1597.
- [26] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? the kitti vision benchmark suite[C]//*Proceedings of 2012 IEEE conference on computer vision and pattern recognition*. Piscataway:IEEE Press, 2012: 3354-3361.
- [27] HAMDI A, ROJAS S, THABET A K, et al. AdvPC: transferable adversarial perturbations on 3D point clouds[C]//*Proceedings of Computer Vision - ECCV 2020: 16th European Conference*. Berlin:Springer International Publishing, 2020: 241-257.
- [28] ARYA A, NADERI H, KASAEI S. Adversarial attack by limited point cloud surface modifications[C]//*Proceedings of the 6th International Conference on Pattern Recognition and Image Analysis (IPRIA)*. Piscataway:IEEE Press, 2023: 1-8.
- [29] ZHENG T H, CHEN C Y, YUAN J S, et al. Pointcloud saliency maps[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Piscataway: IEEE Press, 2019: 1598-1606.
- [30] WICKER M, KWIATKOWSKA M. Robustness of 3D deep learning in an adversarial setting[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway:IEEE Press, 2019: 11767-11775.
- [31] HAMDI C, BANKITI V, LANG A H, et al. nuScenes: a multimodal dataset for autonomous driving[C]//*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. Piscataway:IEEE Press, 2020: 11618-11628.
- [32] ZHOU H, CHEN D D, LIAO J, et al. LG-GAN: label guided adversarial network for flexible targeted attack of point cloud based deep networks[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway:IEEE Press, 2020: 10353-10362.
- [33] CAO Y L, XIAO C W, CYR B, et al. Adversarial sensor attack on LiDAR-based perception in autonomous driving[C]// *Proceedings of 2019 ACM SIGSAC conference on computer and communications security*. New York: ACM Press, 2019: 2267-2281.
- [34] SUN J C, CAO Y L, CHEN Q A, et al. Towards robust LiDAR-based perception in autonomous driving: general black-box adversarial sensor attack and countermeasures[C]//*Proceedings of the 29th USENIX Security Symposium (USENIX Security 2020)*. Berkeley: USENIX Association, 2020: 877-894.
- [35] PETIT J, STOTTELAAR B, FEIRI M, et al. Remote attacks on automated vehicles sensors: experiments on camera and LiDAR[C]//*Proceedings of Black Hat Europe 11*. 2015 : 995.
- [36] HAU Z Y, Co K T, Demetriou S, et al. Object removal attacks on LiDAR-based 3D object detectors[EB]. *arXiv preprint arXiv:2102.03722*, 2021.
- [37] CAO Y L, BHUPATHIRAJU H S, NAGHAVI P, et al. You can't see me: physical removal attacks on LiDAR-based autonomous vehicles driving frameworks[C]//*Proceedings of the 32nd USENIX Security Symposium (USENIX Security 2023)*. Berkeley: USENIX Association, 2023: 2993-3010.
- [38] CAO Y L, XIAO C W, YANG D W, et al. Adversarial objects against LiDAR-based autonomous driving systems[J]. *arXiv preprint arXiv:1907.05418*, 2019.
- [39] TU J, REN M, MANIVASAGAM S, et al. Physically realizable adversarial examples for LiDAR object detection[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway:IEEE Press, 2020: 13713-13722.
- [40] YANG K, TSAI T, YU H G, et al. Robust roadside physical adversarial attack against deep learning in LiDAR perception modules[C]//*Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security*. New York: ACM Press, 2021: 349-362.
- [41] ABDELFAHATTAH M, YUAN K, WANG Z J, et al. Towards universal physical attacks on cascaded camera-LiDAR 3D object detection models[C]//*Proceedings of 2021 IEEE International Conference on Image Processing (ICIP)*. Piscataway: IEEE Press, 2021: 3592-3596.
- [42] TU J, LI H, YAN X, et al. Exploring adversarial robustness of multi-sensor perception systems in self driving[J]. *arXiv preprint arXiv:2101.06784*, 2021.
- [43] CAO Y L, WANG N F, XIAO C W, et al. Invisible for both

- camera and LiDAR: security of multi-sensor fusion based perception in autonomous driving under physical-world attacks[C]//Proceedings of 2021 IEEE Symposium on Security and Privacy (SP). Piscataway:IEEE Press, 2021: 176-194.
- [44] ABDELFAH M, YUAN K, WANG Z J, et al. Adversarial attacks on camera-LiDAR models for 3D car detection[C]//Proceedings of 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Piscataway:IEEE Press, 2021: 2189-2194.
- [45] SHI S S, WANG X G, LI H S. PointRCNN: 3D object proposal generation and detection from point cloud[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE Press, 2019: 770-779.
- [46] ZHU Y, MIAO C L, ZHENG T H, et al. Can we use arbitrary objects to attack LiDAR perception in autonomous driving? [C]//Proceedings of 2021 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2021: 1945-1960.
- [47] LI Y M, WEN C C, XU F J, et al. Fooling LiDAR perception via adversarial trajectory perturbation[C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Piscataway:IEEE Press, 2021: 7898-7907.
- [48] ZHANG Y, ZHU Y, LIU Z H, et al. Towards backdoor attacks against LiDAR object detection in autonomous driving[C]//Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems. New York: ACM Press, 2022: 533-547.
- [49] ZHOU H, CHEN K J, ZHANG W M, et al. Deflecting 3D adversarial point clouds through outlier-guided removal[J]. arXiv preprint arXiv:1812.11017, 2018.
- [50] ZHOU H, CHEN K J, ZHANG W M, et al. Dup-Net: denoiser and upsampler network for 3D adversarial point clouds defense[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 1961-1970.
- [51] PENG S Y, NIEMEYER M, MESCHEDER L, et al. Convolutional occupancy networks[C]// Proceedings of Computer Vision - ECCV 2020: 16th European Conference. Berlin: Springer, 2020: 523-540.
- [52] DONG X Y, CHEN D D, ZHOU H, et al. Self-robust 3D point recognition via gather-vector guidance[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Piscataway: IEEE Press, 2020: 11513-11521.
- [53] LIU H B, JIA J Y, GONG N Z. PointGuard: provably robust 3D point cloud classification[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 6186-6195.
- [54] SUN J C, KOENIG K, CAO Y L, et al. On adversarial robustness of 3D point cloud classification under adaptive attacks[J]. arXiv preprint arXiv:2011.11922, 2020.
- [55] ZHANG K, ZHOU H, ZHANG J, et al. Ada3Diff: defending against 3D adversarial point clouds via adaptive diffusion[C]//Proceedings of the 31st ACM International Conference on Multimedia. New York: ACM Press, 2023: 8849-8859.
- [56] NADERI H, ETEMADI A, NOORBAKHSH K, et al. LPF-defense: 3D adversarial defense based on frequency analysis [J]. Plos one, 2023, 18(2): e0271388.
- [57] LIANG Q, LI Q, NIE W Z, et al. PAGN: perturbation adaptation generation network for point cloud adversarial defense[J]. Multimedia Systems, 2022, 28(3): 851-859.
- [58] SUN J C, CAO Y L, CHOY C B, et al. Adversarially robust 3D point cloud recognition using self-supervisions[C]//Proceedings of Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021 (NeurIPS 2021). 2021: 15498-15512.
- [59] LIANG M, YANG B, WANG S L, et al. Deep continuous fusion for multi-sensor 3D object detection[C]//Proceedings of the 15th European Conference On Computer Vision (ECCV). 2018: 663-678.
- [60] CHEN X Z, MA H M, WAN J, et al. Multi-view 3D object detection network for autonomous driving[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE Press, 2017: 6526-6534.
- [61] XU D F, ANGUELOV D, JAIN A. PointFusion: deep sensor fusion for 3D bounding box estimation[C]//Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE Press, 2018: 244-253.
- [62] LIANG M, YANG B, CHEN Y, et al. Multi-task multi-sensor fusion for 3D object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE Press, 2019: 7345-7353.
- [63] DU X X, ANG M H, RUS D. Car detection for autonomous vehicle: LiDAR and vision fusion approach through deep learning framework[C]//Proceedings of 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Piscataway:IEEE Press, 2017: 749-754.
- [64] KU J, MOZIFIAN M, LEE J, et al. Joint 3D proposal generation and object detection from view aggregation[C]//Proceedings of 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Piscataway: IEEE Press, 2018: 1-8.
- [65] ARNOLD E, AL-JARRAH O Y, DIANATI M, et al. A survey on 3D object detection methods for autonomous driving applications[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(10): 3782-3795.
- [66] FENG D, HAASE-SCHÜTZ C, ROSENBAUM L, et al. Deep multi-modal object detection and semantic segmentation for autonomous driving: datasets, methods, and challenges[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 22(3): 1341-1360.
- [67] SNOEK C G M, WORRING M, SMEULDERS A W M. Early versus late fusion in semantic video analysis[C]//Proceedings of the 13th Annual ACM international Conference on Multimedia. New York: ACM Press, 2005: 399-402.
- [68] YANG Z, SUN Y N, LIU S, et al. IPOD: intensive point-based object detector for point cloud[J]. arXiv preprint arXiv:1812.05276, 2018.
- [69] YANG Z, SUN Y N, LIU S, et al. 3DSSD: point-based 3D single stage object detector[C]//Proceedings of the IEEE/CVF

- Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 11037-11045.
- [70] NGIAM J, CAINE B, HAN W, et al. StarNet: targeted computation for object detection in point clouds[J]. arXiv preprint arXiv:1908.11069, 2019.
- [71] SHI W G, RAJKUMAR R. Point-GNN: graph neural network for 3D object detection in a point cloud[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 1708-1716.
- [72] PAN X, XIA Z, SONG S, et al. 3D object detection with pointformer[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 7463-7472.
- [73] YANG Z, SUN Y N, LIU S, et al. STD: sparse-to-dense 3D object detector for point cloud[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 1951-1960.
- [74] ENGELCKE M, RAO D, WANG Z D, et al. Vote3Deep: fast object detection in 3D point clouds using efficient convolutional neural networks[C]//Proceedings of 2017 IEEE International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2017: 1355-1361.
- [75] WANG D Z, POSNER I. Voting for voting in online point cloud object detection[C]//Proceedings of Robotics: Science and Systems XI. 2015: 10-15.
- [76] LI B. 3D fully convolutional network for vehicle detection in point cloud[C]//2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2017). Piscataway: IEEE Press, 2017: 1513-1518.
- [77] BELTRÁN J, GUINDEL C, MORENO F M, et al. BirdNet: a 3D object detection framework from LiDAR information [C]// Proceedings of the 21st International Conference on Intelligent Transportation Systems (ITSC 2018). Piscataway: IEEE Press, 2018: 3517-3523.
- [78] YANG B, LUO W J, URTASUN R. PIXOR: real-time 3D object detection from point clouds[C]//Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7652-7660.
- [79] YANG B, LIANG M, URTASUN R. HDNET: exploiting Hd maps for 3D object detection[C]//Proceedings of the 2nd Annual Conference on Robot Learning (CoRL 2018). 2018: 146-155.
- [80] ZHOU Y, TUZEL O. VoxelNet: end-to-end learning for point cloud based 3D object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 4490-4499.
- [81] YAN Y, MAO Y X, LI B. SECOND: sparsely embedded convolutional detection[J]. Sensors, 2018, 18(10): 3337.
- [82] ZHOU Y, SUN P, ZHANG Y, et al. End-to-end multi-view fusion for 3D object detection in LiDAR point clouds[C]// Proceedings of Conference on Robot Learning (CoRL 2019). 2020: 923-932.
- [83] LANG A H, VORA S, CAESAR H, et al. PointPillars: Fast encoders for object detection from point clouds[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 12697-12705.
- [84] WANG Y, FATHI A, KUNDU A, et al. Pillar-based object detection for autonomous driving[C]//Proceedings of Computer Vision-ECCV 2020: 16th European Conference. Berlin: Springer, 2020: 18-34.
- [85] SHI S S, WANG Z, SHI J P, et al. From points to parts: 3D object detection from point cloud with part-aware and part-aggregation network[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(8): 2647-2664.
- [86] DENG J J, SHI S S, LI P W, et al. Voxel R-CNN: towards high performance voxel-based 3D object detection[C]//Proceedings of the AAAI conference on artificial intelligence. 2021: 1201-1209.
- [87] YIN T W, ZHOU X Y, KRAHENBUHL P. Center-based 3D object detection and tracking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 11784-11793.
- [88] MAO J G, XUE Y J, NIU M Z, et al. Voxel transformer for 3D object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2021: 3144-3153.
- [89] FAN L, PANG Z Q, ZHANG T Y, et al. Embracing single stride 3D object detector with sparse transformer[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2022: 8448-8458.
- [90] SUN P, TAN M X, WANG W Y, et al. SWformer: sparse Window transformer for 3D object detection in point clouds[C]// Proceedings of European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 426-442.
- [91] LIU Z J, TANG H T, LIN Y J, et al. Point-voxel CNN for efficient 3D deep learning[C]//Proceedings of Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019. 2019: 963-973.
- [92] TANG H H, LIU Z J, ZHAO S Y, et al. Searching efficient 3D architectures with sparse point-voxel convolution[C]//European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 685-702.
- [93] HE C H, ZENG H, HUANG J Q, et al. Structure aware single-stage 3D object detection from point cloud[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 11873-11882.
- [94] MIAO Z W, CHEN J K, PAN H Y, et al. PVGNet: a bottom-up one-stage 3D object detector with integrated multi-level features[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 3279-3288.
- [95] CHEN Y L, LIU S, SHEN X Y, et al. Fast point R-CNN[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 9774-9783.
- [96] SHI S S, GUO C X, JIANG L, et al. PV-RCNN: point-voxel feature set abstraction for 3D object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern

- Recognition. Piscataway:IEEE Press, 2020: 10526-10535.
- [97] SHI S S, JIANG L, DENG J J, et al. PV-RCNN++: point-voxel feature set abstraction with local vector representation for 3D object detection[J]. International Journal of Computer Vision, 2023, 131(2): 531-551.
- [98] MAO J G, NIU M Z, BAI H Y, et al. Pyramid R-CNN: towards better performance and adaptability for 3D object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway:IEEE Press, 2021: 2723-2732.
- [99] LI Z C, WANG F, WANG N. LiDAR R-CNN: an efficient and universal 3D object detector[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE Press, 2021: 7546-7555.
- [100] BEWLEY A, SUN P, MENSINK T, et al. Range conditioned dilated convolutions for scale invariant 3D object detection[J]. arXiv preprint arXiv:2005.09927, 2020.
- [101] MEYER G P, LADDHA A, KEE E, et al. LaserNet: an efficient probabilistic 3D object detector for autonomous driving [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 12677-12686.
- [102] RAPOPORT-LAVIE M, RAVIV D. It's all around you: range-guided cylindrical network for 3D object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway:IEEE Press, 2021: 2992-3001.
- [103] MEYER G P, CHARLAND J, PANDEY S, et al. LaserFlow: efficient and probabilistic object detection and motion forecasting[J]. IEEE Robotics and Automation Letters, 2020, 6 (2): 526-533.
- [104] LIANG Z D, ZHANG Z H, ZHANG M, et al. RangeIoU-Det: range image based real-time 3D object detector optimized by intersection over union[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE Press, 2021: 7140-7149.
- [105] CHAI Y N, SUN P, NGIAM J, et al. To the point: efficient 3D object detection in the range image with graph convolution kernels[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, Piscataway: IEEE Press, 2021: 16000-16009.
- [106] SUN P, WANG W Y, CHAI Y N, et al. RSN: range sparse net for efficient, accurate LiDAR 3D object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 5725-5734.
- [107] SHENG H L, CAI S J, LIU Y, et al. Improving 3D object detection with channel-wise transformer[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway:IEEE Press, 2021: 2743-2752.
- [108] ZHOU B, KHOSLA A, LAPEDRIZA A, et al. Learning deep features for discriminative localization[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE Press, 2016: 2921-2929.
- [109] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 618-626.
- [110] RIBEIRO M T, SINGH S, GUESTRIN C. Why should I trust you? explaining the predictions of any classifier[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2016: 1135-1144.

[作者简介]



项圣（1993-），男，浙江台州人，浙江工业大学讲师，主要研究方向为机器视觉、人工智能安全。



廖丹芯（2000-），女，广西北海人，浙江工业大学硕士生，主要研究方向为深度学习 and 人工智能安全。



陈晋音（1982-），女，浙江象山人，浙江工业大学教授，主要研究方向为人工智能安全、图数据挖掘和进化计算。



郑海斌（1995-），男，浙江台州人，浙江工业大学讲师，主要研究方向为深度学习、人工智能安全和图像识别。