文章编号:1671-4598(2025)03-0250-09 DOI:10.16526/j.cnki.11-4762/tp.2025.03.032 中图分类号:TP309 文献标识码:A

针对激光雷达感知算法的黑盒攻击模型与仿真

邵玉亮[1,3],马明晓[2,3]
(1.中国科学技术大学 生物医学工程学院,合肥 230026;
2.中国科学技术大学 人工智能与数据科学学院,合肥 230026;
3.中国科学技术大学 苏州高等研究院,江苏 苏州 215123)

摘要:为了验证自动驾驶中基于激光雷达感知算法面对攻击的脆弱性,提出了一种使用基于点云上采样算法 PC^2-PU 并配合随机下采样来控制攻击点的数量的黑盒攻击模型,这种攻击能够减少因为数据集本身的缺陷而导致的最终结果出现偏差的情况出现,所提出的攻击模型包括攻击点的数量、位置和高度 3 个主要因素,通过实验来验证这些因素对攻击结果的影响程度,实验中使用了 4 种不同类型的感知模型来证明我们提出的攻击的有效性;结果表明,即使只有 20 个攻击点时,在部分区间位置内的攻击成功率超过了 90%,随着攻击点数的增加,在不同位置的攻击成功率均在不断提高;在百度 Apollo 平台中测试了该攻击对决策层产生的影响。

关键词:激光雷达传感器;自动驾驶安全;感知攻击;点云;驾驶决策

Black-box Attack Model and Simulation for LiDAR Perception Algorithms

SHAO Yuliang[1,3], MA Mingxiao[2,3]
(1. School of Biomedical Engineering, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230026, China;
2. School of Artificial Intelligence and Data Science, University of Science and Technology of China, Hefei 230026, China;
3. Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou 215123, China)

Abstract: To verify the vulnerability of LiDAR-based perception algorithms in autonomous driving when facing attacks, a black-box attack model is proposed. This model uses the point cloud upsampling algorithm PC^2-PU in combination with random downsampling to control the number of attack points. This attack can reduce the occurrence of biases in the results caused by inherent flaws in the dataset. The proposed attack model includes three main factors: the number, location, and height of attack points. Experiments are conducted to verify the impact of these factors on attack outcomes. Four different types of perception models are used in the experiments to demonstrate the effectiveness of the proposed attack. The results show that even with only 20 attack points, the attack success rate exceeds 90% in some interval positions. As the number of attack points increases, the attack success rate continues to improve at different positions. The impact of this attack on the decision-making layer is also tested on the Baidu Apollo platform.

Keywords: LiDAR sensor; self-driving security; perception attack; point cloud; driving decision-making

0 引言

随着深度学习的快速发展,视觉感知算法的性能和可靠性也有了很大程度的提高,并被应用于许多的现实场景之中,如自动驾驶汽车、无人机和智能机器人等[1-3]。激光雷达作为一种重要的感知手段不仅能够扫描周围环境来提供三维信息,还能够在能见度较低的条件下正常工作,弥补了传统摄像头在这些方面的感知缺陷。通过将收集的点云送入感知算法模型之中就能够实现对周围物体的识别,自动驾驶汽车根据这些分类后的

收稿日期:2024-12-26; **修回日期**:2025-02-07。
基金项目:中国科学技术大学苏州高等研究院引进人才科研启动专项(KY2260080021)。
作者简介:邵玉亮(1999-),男,硕士研究生。
通讯作者:马明晓(1988-),男,博士,副研究员。
引用格式:邵玉亮,马明晓.针对激光雷达感知算法的黑盒攻击模型与仿真[J].计算机测量与控制,2025,33(3):250-258.

物体做出决策层的判断，并采取不同的决策来应对不同的交通环境。激光雷达传感器采集的点云成为了智能感知不可或缺的一部分，但是攻击者可以通过精心设计的装置使得点云中包含恶意的虚假点，这种攻击严重威胁了感知系统的安全性和可靠性，为了确保激光雷达感知可靠性，切实保障乘客以及行人的安全，已经有许多针对激光雷达感知的安全性的研究[4-6]。

在对自动驾驶汽车中基于激光雷达感知算法攻击的研究中，攻击者可以通过精心设置的攻击装置在不同地点和不同时机发动攻击，使得被攻击者采集的原始激光雷达数据出现偏差。具体来说，攻击者可以通过这种方法向目标系统注入虚假的点云数据，注入的攻击点云不受攻击设备位置的限制，这些攻击点可以比攻击设备所在位置更远或更近。这种攻击能力显著扩大了这类攻击的范围，促进了更多的研究来探讨激光雷达传感器在面对不同攻击场景下的脆弱性[7-8]。

文献［4］中提出了一种名为 ADV-LiDAR 的攻击就是针对基于激光雷达的感知系统的，但是这种基于优化的白盒攻击方法在现实世界部署是非常困难的，这是因为对于现实中的车辆来说，具体的感知模型是绝对保密的，汽车厂家不可能向外界公开量产车型所使用的感知模型信息，所以这种依赖详细模型参数的白盒攻击在真实的交通环境中是很难实现的。文献［6］提出了一种新的对抗点云优化方法，该方法考虑了激光雷达的工作原理，攻击设备的能力以及在优化过程中注入的点的距离误差。又提出了一种控制信号设计方法，将点云的坐标转换为控制信号来精确生成注入点云的所需形状。但是该攻击在部分感知模型上攻击成功率较低。文献［9］又提出了一种具有现实可行性的攻击，这种攻击是在目标受害车辆顶部放置各种物体（如沙发和椅子等）来逃避检测算法的正确识别，攻击所使用的攻击点通过 Moller-Trumbore 算法来获得。但是上述类型的攻击的局限性在于该攻击过于依赖具体目标。文献［10］提出了一种同时攻击激光雷达和摄像头两种传感器的攻击模型，但是该攻击方法需要同时攻击两个信道，并且还要保证攻击的一致性，提高了攻击成功的难度。攻击中也只使用了一个多面体网格作为攻击源，进一步降低了攻击适用范围。但是上述类型攻击的局限性在于该攻击过于依赖具体目标，而且攻击的目标是改变已有物体的正确识别，并不能在点云图中删除目标的点云，其仍然会被感知模型所识别到。文献［5］中所使用的攻击点是在 KITTI 数据集中直接提取或者通过构造渲染器来模拟激光雷达工作原理来渲染汽车网格得到，这使得攻击结果与攻击点的选取有较强的依赖性。随后文献［11］允许攻击者在特定情况下自行放置攻击点，减少了对攻击点的位置和分布的限制，并认为这种简化的攻击更有可能代表有噪声的攻击激光，能够一定程度上反映噪声对攻击的影响，更符合现实攻击中噪声带来的扰动，但是并没有对攻击位置对攻击结果的影响进行详细实验。

使用 Moller-Trumbore 相交算法得到点云集和、直接使用 KITTI 数据集中车辆的点云或通过渲染器生成点云作为攻击轨迹的局限在于这些点的来源与选取的目标物体息息相关，并且对点的分布和质量要求很高，因为这些点云都是通过模拟激光雷达工作原理获得的，但是现实中的攻击点难以完全满足这种理想分布，这种构造攻击点的方式加大了在现实中复现这些攻击点的难度。本文在实验中首先在 KITTI 数据集中提取目标车辆的点集来制作初始数据集，通过使用一种基于深度学习的点云上采样算法 PC^2-PU 来重新生成更多的点[12]，再配合随机下采样的方式控制攻击轨迹中点的数量在 100 以内，这足以满足现实场景中攻击的要求。与未经处理的数据相比，处理之后的点云中点的数量和分布已经发生了很大变化，这种黑盒攻击方式更能真实的反映现实中基于激光雷达的感知算法面对攻击时的脆弱性。之后的实验中又在百度的 Apollo 仿真平台定义了道路阻塞和强制变道两个场景来研究攻击对自动驾驶汽车决策层的影响。

1　基于激光雷达的感知算法

1.1　基于多视图的感知算法

为了减少处理 3D 点云的计算开销和时间开销，可以将点云投影到二维视图中使用二维感知算法进行检测，从而减少了处理数据的时间，提高了算法的运行效率。文献［13］将激光雷达点云投影到鸟瞰图中并使用 RepVGG[14] 作为主干网络增强特征提取能力，使得模型的识别准确率进一步提高。这种将复杂的三维环境信息简化为二维的变换能够使得感知算法满足系统对实时性的要求。为了克服鸟瞰图难以检测距离较远或者体积较小的物体的缺陷，文献［15］提出的 Dynamic Voxelization 是一种融合了鸟瞰图和透视图的端到端的多视图融合算法，该算法首先对原始 LiDAR 点云进行预处理，并嵌入到高维特征空间，之后通过动态体素化将点云进行划分，得到分类之后的鸟瞰图和透视图两种不同视角，在这两种视角下进行特征提取得到局部信息并使用卷积塔对体素级别的特征进行上下文信息编码。随后进行点级别的特征融合使得每个点都能同时利用鸟瞰图和透视图视角的特征，最后采用最大池化（Max Pooling）进行局部特征聚合实现目标检测。该算法能够更好的利用周围环境的点云信息，提高模型预测的精准度。

1.2　基于体素的感知算法

基于体素的点云感知算法将稀疏不规则的点云转换

为密集相邻的体素表示，然后通过卷积神经网络提取用于目标检测的点云特征。相较于直接处理原始点云，这种方式能够在一定程度上提高计算效率，还能够避免多视图算法损失过多的信息的问题。考虑到点云的稀疏性，文献［16］提出了一种稀 Transforme 模块只对非空体素进行特征提取，通过轻量级网络动态筛选关键体素（如前景目标附近的体素），减少无关背景的计算，从而减少计算复杂度。为了进一步减少计算量，PointPillars 将点云编码成一种特殊的体素[17]。首先进行点云离散化使得点云在 x-y 平面上被划分为均匀的网格，垂直方向没有限制且保留单个体素，名为 pillar。之后对这些对每个 pillar 应用简化版的 PointNet 生成对应的特征向量。保持这些 pillar 的相对位置不变创建伪图像，即可使用二维卷积神经网络来实现三维目标检测。

1.3 基于点的检测算法

为了克服基于多视图和体素的检测方法会缺失一部分点云的缺陷，可以直接将原始点云作为模型的输入，直接学习点云的三维表示进行目标检测，而无需将点云转换为体素或其他形式。文献［18］使用原始点云作为输入，通过反向残差瓶颈和可分离的多层感知机（MLP）减少计算量的同时提高检测精度，并使用 AdamW 替代 Adam 提高训练稳定性。Points-RCNN 使用两阶段检测框架处理输入点云[19]，第一阶段使用 PointNet＋＋框架[20]对整个场景进行前景点的分割来预测目标中心偏移量和 3D 框参数，第二阶段使用 ROI Pooling 对这些区域进行局部特征提取，来进一步优化目标检测框和提高置信度预测。

1.4 基于点和体素的检测算法

一方面，基于体素的检测方法会丢失一部分点，并且算法的性能和开销很大程度上受体素参数的设置影响。另一方面，基于点的检测方法可以很更好的利用点云的局部特征，能提供细粒度的邻域信息。所以可以将基于体素和基于点的两种算法结合起来在保证减少计算开销的同时，又尽可能多的利用点云信息。PV-RCNN 是一种结合了点和体素的两阶段检测方法[21]。第一阶段将点云划分为规则的空间体素网格，使用稀疏卷积提取多尺度体素特征生成初始 3D 候选框。第二阶段通过最远点采样发选取出关键点，通过双线性插值，将体素特征赋给关键点。之后从关键点特征中提取更精细的局部特征来实现候选框的精细化回归与分类。文献［22］在 PV-RCNN 的基础上直接使用稀疏卷积在体素上提取关键点特征，避免插值操作来减少计算量，同时在覆盖关键区域的情况下减少选取的关键点数量减少至 512 个来进一步加快推理速度。

为了更好的验证所提出的黑盒攻击模型的有效性，对于上面的 4 种不同类型的感知算法都分别选取了一个感知模型进行攻击检测，即上述介绍的 Dynamic Voxelization、PointPillars、PV-RCNN 和 Points-RCNN 四种感知模型。

2 攻击原理、模型和目标

2.1 攻击原理

激光雷达传感器通过发射激光脉冲照射周围物体并接受返回的信号来测定与目标物体的距离。利用光速恒定的特性，计算激光脉冲从发射到返回的时间差来精确测量距离。激光脉冲发射到物体表面后，被反射回接收器，从而形成完整的测量数据。这种测量方式产生的点云为自动驾驶提供了三维空间的环境信息，使激光雷达成为现代自动驾驶技术的重要组成部分之一。现代激光雷达设备的发射装置通常包括多束激光，可以实现 360 度的无缝扫描，生成大范围的点云数据。某些高端激光雷达设备可以产生每秒数百万个点的深度数据，以更高的精细度捕捉环境中的细节。激光雷达的高分辨率使其能够清晰辨识出物体的形状、大小、位置等重要信息，为自动驾驶系统提供了丰富且详细的环境信息。图 1 展示了激光雷达攻击的基本工作原理。

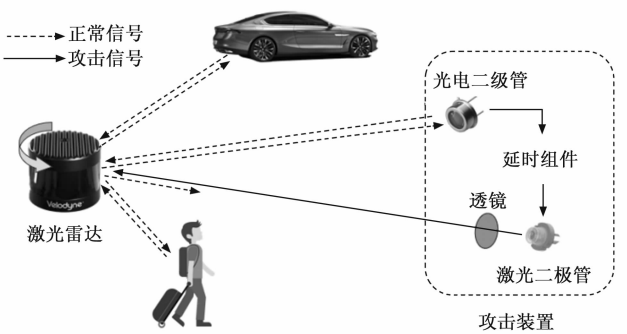


图 1 激光雷达攻击原理示意图

激光雷达发出激光脉冲后，正常照射到行人或者车辆后会返回回波脉冲形成点云。感知算法可以从点云中提取特征，识别出物体的类型、位置、朝向等信息，为车辆路径规划和决策提供支持。但是如果照射到攻击装置后，攻击装置中的光电二极管接收到脉冲信号后传递给延时组件，然后延时组件控制激光二极管在随后的 LiDAR 检测周期中发动攻击来注入虚假点，这样就造成了激光雷达采集的数据中包含了原本不应该存在的攻击点。

这种攻击装置可以被放置在路边或者桥梁之上，对过往的车辆发动攻击会干扰车辆的正常行驶和威胁车内乘客和路上行人的安全。由于发动这种攻击是不需要提前侵入受害车辆的数据通道和数据计算过程，使得这种攻击更加隐蔽。一旦发生攻击之后，很难被车辆自身的检测系统所检测到，只能依赖于车辆本身感知算法的鲁棒性。

2.2　攻击模型

1）攻击点的数量：文献［5］中攻击者已经可以通过精心设计的攻击装置最多可以注入 200 个攻击点，文献［11］又进一步放宽了这种攻击假设，允许攻击者通过随机采样分布来放置点，这使得攻击变得更加容易实现。因此在实验中将攻击者能够实现的攻击点数的最大值定为 100，这在实际攻击中是完全可以实现的。

2）攻击位置：攻击者能够通过改变攻击设备的延迟间隔来控制攻击点距离受害车辆的位置，文献［4］中将攻击距离设置在车辆近前方的位置，文献［5］中则明确了攻击距离，并将攻击位置设置在车辆前方 5～8 m 处，随后提出被检测为车辆的点云移动到车辆前方位置仍然可以被检测为车辆。为了更好的探究不同距离对攻击成功率的影响，实验中对攻击位置进行了更加细粒度的划分，所设置的攻击位置区间为车辆前方 7.5～67.5 m 处，这足以满足绝大部分交通场景。

3）攻击高度：文献［4，5，11］的工作中并没有考虑攻击点距离地面不同高度对攻击成功率的影响，尽管这些研究认为攻击高度可以通过修改延迟间隔来控制。为了探究攻击点高度与攻击成功率的关系，在实验中将攻击点距离地面高度的最大值设置为 2 m 来探究其对攻击结果的影响，因为正常情况下道路上的车辆和行人的高度都不会超过这个临界值。

2.3　攻击目标

基于激光雷达攻击可以大致分为两类：第一种攻击类型的目标是为了欺骗感知算法使其无法正确识别周围物体[5,6,9,23]。例如在目标物体周围构造一些攻击点从而使得感知模型无法对其进行正确分类，攻击目标通常是为了迷惑感知算法使其无法正确识别道路中的车辆或行人。这种攻击的局限性在于即使无法正确识别车辆，但是这些车辆的点云并不会消失，感知算法仍然会将其识别为静止的或者可以移动的障碍物。而且这种识别是持续可追踪的，只要目标车辆仍在激光雷达感知范围之内，就能够被车辆所识别并采取相应的措施避免碰撞，所以这类攻击能够造成的后果有限。与之相对的，另一种攻击类型则致力于注入虚假点使得被攻击目标车辆检测到前方不存在的物体[4,24,25]，这种攻击往往更加隐蔽，而且在精心策划的时间和地点发动攻击能够严重干扰被攻击车辆的决策层判断，对车辆和行人的安全威胁更大，造成的后果也更加严重。实验中所选取的攻击目标是感知算法将攻击点错误的识别为不存在的物体，具体来说就是使得受害车辆错误地将攻击点识别为实际环境中不存在的真实车辆。正常情况下受害车辆前方并没有物体，但是发动攻击之后，感知模型错误的将攻击点识别为不存在的车辆。

攻击设备可以被策略性地放置在路边来对过往车辆发起攻击[4]。当受害车辆受到攻击时，它可能会将攻击点检测为车辆，从而影响其决策层的判断。这种误判可能导致严重的交通事故。在单行道上成功发起攻击可能导致当前道路被堵塞，阻止所有车辆前进。在多车道道路上发起攻击可能导致车辆检测到直接前方的虚假障碍物后重新进行路线规划，更改行驶轨迹。如果虚假车辆就在受害车辆前方不远处，还可能引起紧急制动，会使得后面跟随的车辆难以及时刹车，导致车辆间的追尾碰撞，影响交通安全。

2.4　评价指标

置信度是一个介于 0 和 1 之间的数值，用于评估模型对预测框内存在目标对象的确信程度以及预测框对目标对象位置的准确性。在 KITTI 的训练集中，包含了不同物体的标注信息，其中包括物体的三维信息框。感知模型在训练过程中会给出不同的预测框，通过标注的真实框和当前位置的对象存在概率就可以得到置信度。在训练模型的时候设置一个合理的置信度阈值，只有超过置信度分数阈值才会被检测为目标对象，低于置信度分数阈值的对象则会被丢弃，这样就能保证检测结果的准确性。在实验中，当使用感知算法对攻击场景进行检测，如果感知算法认为攻击点为车辆时，就会给出检测框，此时代表攻击点成功干扰了算法的正常识别。通过检查攻击点位置的三维检测框的生成情况来判断是否攻击成功，并使用攻击成功率作为评价指标来探讨所提出的攻击模型对不同类型感知算法的影响。

3　实验结果

3.1　数据处理

KITTI 数据集采集于德国道路上的真实交通环境，采集的信息包含了彩色图像、灰度图像、激光雷达点云、高精度 GPS 和 IMU 数据等多模态数据，以及包含目标物体 3D 轨迹的标注信息。所有的数据都经过了校准、同步和时间戳处理，能够更好地用于物体检测、跟踪、立体视觉和 SLAM 等任务的研究。

由于感知算法模型会使用 KITTI 中的训练集进行训练，所以使用验证集来提取对应的目标车辆点集，避免数据本身影响到最终的实验结果。由于在验证集中没有标注数据，因此并不知道目标车辆在空间中的中心位置坐标，以及目标物体的尺寸信息。为了解决这一问题，引入了一个额外的空间坐标系来辅助确定物体的位置信息和尺寸。通过改变自定义坐标系原点找到物体中心位置，记录下来的坐标系原点位置即可确定物体位置，实验中坐标系的刻度为 10 厘米，所以通过坐标轴的刻度信息就可以确定目标车辆的长宽高三维信息，之后利用这些信息就可以过滤掉其余点云数据，只保留目标车辆点云来完成数据提取。图 2 展示了通过自定义坐

标轴确定目标车辆位置来进行点云提取的过程。

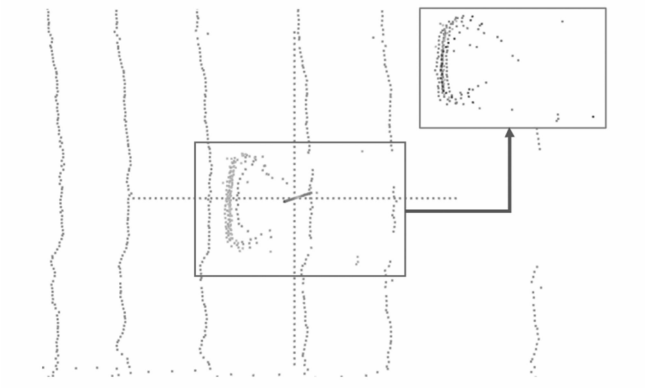


图 2 提取目标车辆点云

PC²-PU 利用了点云中点的相关性来提高上采样的有效性和鲁棒性，具体为通过引入一个 PaCM 模块和 PoCM 模块来确保全局空间的一致性，提升整体上采样效果。点云上采样通常会把点云分成独立区域，而每一个单一的区域只能提供部分结构信息，PaCM 模块是为了解决区域边界上生成的点云分布质量下降的问题而提出的，其通过拓展点的感知范围来提高生成质量。另一个模块 PoCM 揭示了每一个区域内部点的相对位置关系来保持局部空间一致性并恢复细粒度信息，使得网络生成的点能够更加接近真实物体的表面。这两个模块相互补充，能够更好的抑制伪点的生成，提高了网络抗噪声的能力，具有更好的鲁棒性。通过 PC²-PU 算法生成的点云中每一个目标车辆的点的个数都维持在 4 096 个，但是在真实的物理世界中直接注入这么多攻击点是不现实的。在激光雷达扫描周围环境得到的点云图中，距离激光雷达越近，点的密度越高，如此高密度的点云要被控制在一辆车的体积内显然是难以实现的。所以使用了随机下采样的方法来控制点云的数量到 100 个点以内，从而得到最终的攻击轨迹数据集。随机下采样只需要指定采样点的个数就能够实现采样。这种采样方式简单高效，计算复杂度低，适合实时需求。

3.2 攻击位置坐标变换

为了测试这些构造完成的攻击轨迹是否能够攻击成功，本文需要将这些攻击轨迹注入到激光雷达点云中，之后再送入目标感知算法进行检测。与文献［4-5］中的攻击相比，除了能改变攻击所在的平面位置，还能改变攻击轨迹的高度。即利用全局变换矩阵 $\boldsymbol{H}(\theta,\tau)$ 将每一个随机下采样之后的点云集合 V 移动到期望位置 V'：

$$V' = \boldsymbol{H}(\theta,\tau)\cdot V$$

$$\begin{bmatrix} V'_x \\ V'_y \\ V'_z \\ V'_i \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & s_h \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} V_x \\ V_y \\ V_z \\ V_i \end{bmatrix}$$

$$\alpha = \arctan(V_y/V_x)$$

式中，θ 和 τ 分别代表方位角和距离，而向量（V_x，V_y，V_z，V_i）表示 V 中每一个点的 xyz-i 特征向量，对应于立体三维坐标位置和激光返回的强度。通过 s_h 系数可以控制高度变化。将攻击轨迹注入到 KITTI 中的原始激光雷达点云之中进行检测，由于 KITTI 数据集中的点云场景都是来自真实物理世界，所以这种攻击策略更能够代表现实中的攻击。

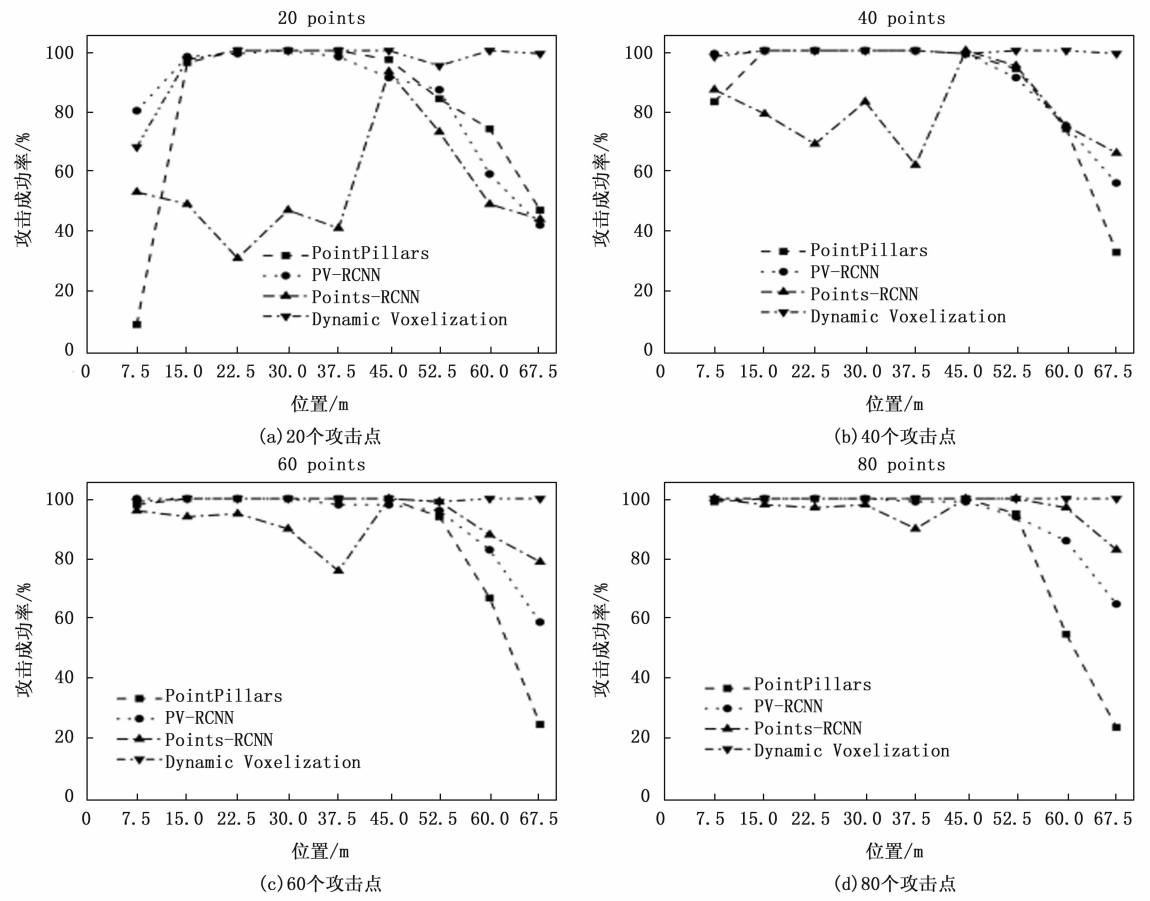
3.3 攻击位置对结果的影响

攻击点距离受害车辆前方不同距离所能够造成的后果有所不同。若攻击位置在车辆前方不远处，不仅会导致受害车辆紧急刹车，危害车内乘客安全，而且还可能造成车辆间的追尾事故，使得攻击的危害被进一步放大。若当攻击位置距离受害车辆较远时，受害车辆将攻击轨迹检测为虚假车辆后会变换车道或更改行驶路线，干扰其他车道上正常行驶的汽车。为了探究攻击位置与攻击成功率之间的关系，实验中将攻击点的数量的初始值设置为 20 并递增至最大值 80，攻击距离从受害车辆前方 7.5 m 处递增至最远距离 67.5 m，实验结果如图 3 所示。

随着攻击轨迹点数的不断增加，整体而言攻击成功率明显提高。即使攻击点的数量为 20 的时候，在 15～45 m 之间该攻击对 PointPillars 和 Dynamic Voxelization 两种感知模型的攻击成功率均值在 90％以上。当攻击点增加到 40 个时，PV-RCNN 模型在上述位置区间的攻击成功率也超过了 90％。当攻击点的数量增加到 60 的时候，在 7.5～52.5 m 之间对 PointPillars、Dynamic Voxelization 和 PV-RCNN 模型的攻击成功率均在 90％以上，在部分位置的攻击成功率甚至达到了 100％。当攻击点数继续增加达到 80 的时候，Points-RCNN 在 7.5～52.5 m 之间的攻击成功率均值也在 90％以上。面对 4 种不同类型的点云感知模型，只要能够在合适的位置区间发动黑盒攻击，即使注入的攻击点的数量较少，也能够干扰感知模型对周围物体的正确识别，使得所提出的攻击模型能够有效的威胁自动驾驶车辆安全。

可以看出基于多视图 Dynamic Voxelization 算法和基于体素的 PointPillars 算法在受到攻击后他们的错误识别率明显高于另外两种算法，这是因为这两种算法本身会对点云进行预处理，这种处理虽然能够使得模型更快收敛，但是也会丢失更多的点云信息，所以其鲁棒性要弱于基于体素和点的 PV-RCNN 算法。而基于点的算法 Points-RCNN 直接处理输入的点云而不用进行信息变换，减少了信息的丢失，所以其鲁棒性也是 4 种不同类型算法中最高的。

虽然攻击成功的迷惑感知算法使其错误感的将攻击轨迹错误的识别为原本不存在的车辆，但是并不是所有



(a)20个攻击点　(b)40个攻击点　(c)60个攻击点　(d)80个攻击点

图 3　不同位置对攻击成功率的影响

的攻击点都会被识别为虚假车辆。如图 4 所示，一部分攻击点（上方圆圈）会被识别为虚假车辆，而另一部分攻击点（下方圆圈）则逃逸出了感知算法的识别，也就是他们并没有被计算在虚假车辆点云之中，被判定为正常点云。这意味着真正造成有效攻击所需要的攻击点的数量比实际注入的点的总数要少，也就是说攻击只需要更少的点就能够成功影响基于激光雷达感知算法的正确识别，这进一步减少了实现该攻击所需要的资源，提高了该攻击在现实中攻击成功的概率。

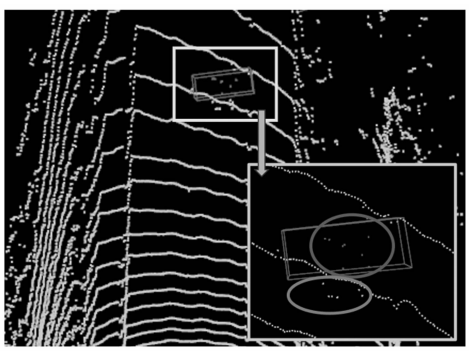


图 4　只有部分攻击点被错误识别

正常情况下，感知算法会将攻击轨迹识别为原本不存在的车辆，但是在某些时候，当我们成功的向点云中注入攻击点之后，检测算法并没有直接将这些攻击点识别为车辆，而是将攻击周围正常的点识别为车辆，尽管这些点并不来自于现实中的一辆汽车，如图 5 所示，感知算法忽略了我们所注入的攻击点（下方圆圈），而是直接将攻击点周围的采集来自地面的正常点（上方圆圈）识别为车辆。这种攻击更加能够逃避防御机制的追捕，因为攻击的成功源自于真实世界的点云，而那些攻击点被排除在外，并没有被识别为车辆，只是单纯的影响了检测算法对正常点云的识别。

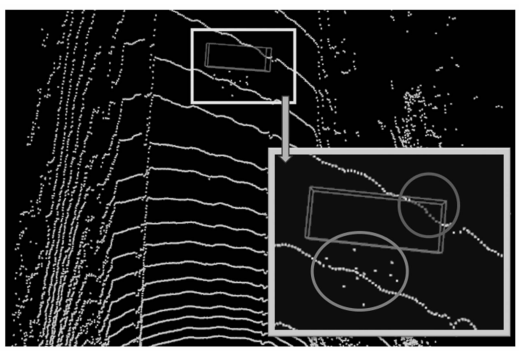


图 5　攻击点干扰了正常点的识别

3.4　攻击高度对结果的影响

在正常的交通场景中，小轿车和人的高度都基本不

会超过 2 m，考虑到这一现实因素，实验中最大高度都不会超过这个限制。攻击轨迹的水平位置设置在距离受害车辆 30 m 处，攻击高度初始值设置为攻击轨迹最低点距离地面的高度，其初始值为 0，这模拟了车辆静止在地面上时车辆底部距离地面的高度。随后攻击高度逐 0.25 m 递增，直至最大高度 2 m。攻击轨迹点的数量分别为 20，40，60 和 80，实验结果如图 6 所示。

即使攻击点的数量增多，不同检测算法的攻击成功率随高度变化的整体趋势基本相同。对于 PointPillars 和 PV-RCNN 两种模型来说，当攻击高度分别低于 1.25 m 和 0.75 m 时，攻击成功率并没有明显的改变。当超过上述高度时，攻击成功率随着高度的增加开始急剧下降，并趋向于 0。对于 Dynamic Voxelization 模型来说，攻击成功率并没有显著改变，始终维持在同一水平。对于 Points-RCNN 模型来说，攻击成功率会随着高度的改变而发生不同程度的变化，但是变化幅度始终维持在一定范围内，并没有快速的增加或减少。因此只要将攻击高度控制在一个合理的区间范围内变化，就能够获得较高的攻击成功率。

4 攻击对自动驾驶决策层的影响

为了更好的了解攻击成功后对受害车辆的影响，实验中使用百度 Apollo 平台来研究攻击对决策层的影响，虽然该模块并能完全代表现真实汽车受到攻击后的决策过程，但是这种仿真方式能够在一定程度上反映攻击的有效性，文献［3］也是用该平台进行仿真。在 Apollo 中定义了两个攻击场景，分别是道路阻塞和强制变道，当攻击成功发动后，受害车辆将攻击轨迹识别为不存在的车辆，这种错误识别使得受害车辆的决策判断发生了改变，这种决策的改变还可能影响到道路上其他正常行驶的车辆。

4.1 道路阻塞

在道路阻塞的攻击场景中，当受害车辆被攻击后检测到前方虚假车辆，将逐渐开始减速，最终停止距离其 12.5 m 处。由于当前道路为单车道，车辆将无法继续前进，持续静止在当前位置，阻塞当前道路其他车辆的正常通行，如图 7 所示。如果在现实世界发生这种攻击，会造成严重的交通阻塞，尤其是在早晚高峰时期，严重影响车辆的正常通行，造成大面积交通拥堵。

4.2 强制变道

在强制变道的攻击场景中，受害车辆正常行驶在多车道上，当其收到攻击后检测到前方虚假车辆，受害车辆将会提前减速并改变当前行驶车道，转入其他车道才能继续行驶。如图 8 所示，受害车辆将会沿着灰色轨迹变换车道。如果精确控制发动攻击的时间和地点，使得

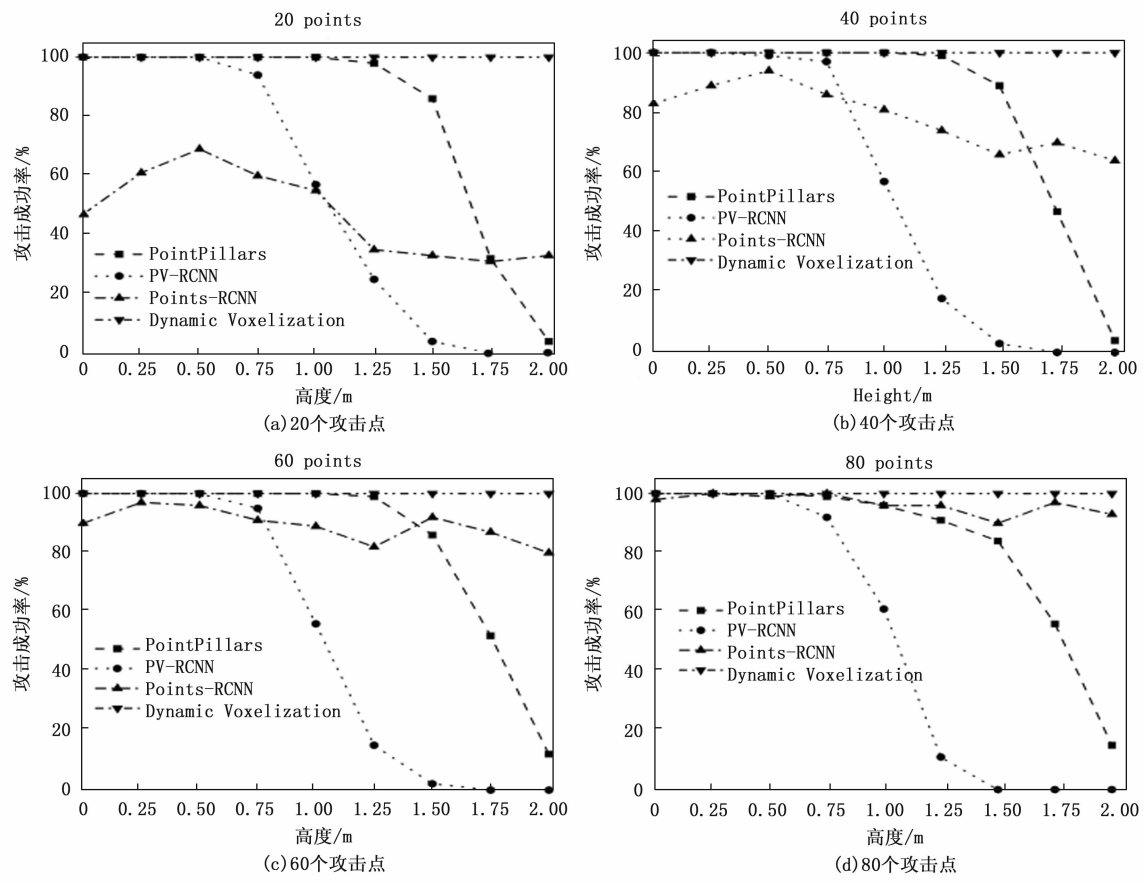


图 6 不同高度对攻击成功率的影响

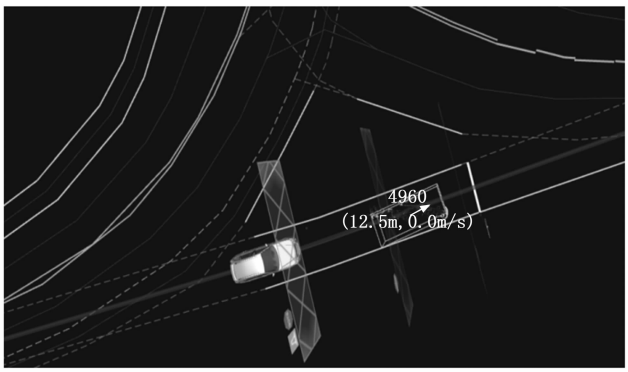


图 7　道路阻塞

受害车辆在某一时刻突然检测到前方车辆，为了避免追尾，受害车辆可能会紧急转向，就会导致其与其他车道上的车辆发生碰撞，不仅会造成个人经济损失，而且还可能威胁到车内乘客的生命安全。

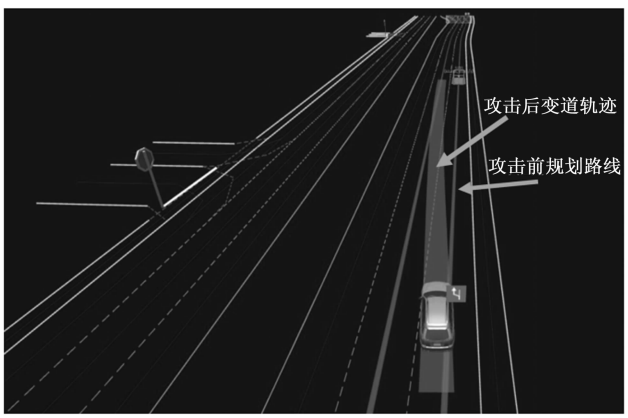


图 8　强制变道

5　结束语

现在的自动驾驶汽车普遍采用了激光雷达传感器对周围环境进行感知，为了研究基于激光雷达的感知模型面对攻击的脆弱性，提出了一种黑盒攻击能够使得感知模型错误地将攻击轨迹识别为不存在的车辆。这种使用深度学习算法的数据处理减轻了数据集对攻击结果的影响。随后选择了 4 种不同类型的感知模型进行检测实验，结果显示即使在攻击点数量相对较少的情况下，某些模型在特定位置的攻击成功率仍然超过 90%。此外还研究了不同垂直高度位置对攻击成功率的影响。结果表明随着攻击点数量的增加，不同算法的攻击成功率变化的整体变化趋势保持一致，只要保证攻击高度在合适区间之内，就能够造成有效攻击。最后使用 Apollo 平台研究了攻击成功后对决策层的影响。模拟的场景包括道路堵塞和强制变道两种情况，突出了攻击可能对道路安全造成的危害。未来的工作将集中在探索可能的防御策略来抵抗此类攻击，确保自动驾驶感知系统的安全性和可靠性。

参考文献：

[1] WANG Z, ZHAN J, DUAN C, et al. A review of vehicle detection techniques for intelligent vehicles [J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 34 (8): 3811 - 3831.

[2] WANG Y, MAO Q, ZHU H, et al. Multi-modal 3d object detection in autonomous driving: a survey [J]. International Journal of Computer Vision, 2023, 131 (8): 2122 - 2152.

[3] CUI C, MA Y, CAO X, et al. A survey on multimodal large language models for autonomous driving [C] //Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2024: 958 - 979.

[4] CAO Y, XIAO C, CYR B, et al. Adversarial sensor attack on lidar-based perception in autonomous driving [C] //Proceedings of the 2019 ACM SIGSAC conference on computer and communications security, 2019: 2267 - 2281.

[5] SUN J, CAO Y, CHEN Q A, et al. Towards robust {LiDAR-based} perception in autonomous driving: General black-box adversarial sensor attack and countermeasures [C] //29th USENIX Security Symposium (USENIX Security 20). 2020: 877 - 894.

[6] JIN Z, JI X, CHENG Y, et al. Pla-lidar: Physical laser attacks against lidar-based 3d object detection in autonomous vehicle [C] //2023 IEEE Symposium on Security and Privacy (SP), IEEE, 2023: 1822 - 1839.

[7] ZHU Y, MIAO C, ZHENG T, et al. Can we use arbitrary objects to attack lidar perception in autonomous driving? [C] //Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security, 2021: 1945 - 1960.

[8] LI Y, WEN C, JUEFEI-XU F, et al. Fooling lidar perception via adversarial trajectory perturbation [C] //Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 7898 - 7907.

[9] 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, 2020: 13716 - 13725.

[10] ABDELFATTAH M, YUAN K, WANG Z J, et al. Adversarial attacks on camera-lidar models for 3d car detection [C] //2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2021: 2189 - 2194.

[11] HALLYBURTON R S, LIU Y, CAO Y, et al. Security analysis of {Camera-LiDAR} fusion against {Black-Box} attacks on autonomous vehicles [C] //31st USENIX Security Symposium (USENIX Security 22), 2022: 1903 - 1920.

[12] LONG C, ZHANG W X, LI R, et al. Pc2-pu: Patch correlation and point correlation for effective point cloud upsampling [C] //Proceedings of the 30th ACM International Conference on Multimedia, 2022: 2191 - 2201.

[13] SHAO Y, SUN Z, TAN A, et al. Efficient three-dimensional point cloud object detection based on improved Complex-YOLO [J]. Frontiers in Neurorobotics, 2023, 17: 1092564.

[14] DING X, ZHANG X, MA N, et al. Repvgg: Making vgg-style convnets great again [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13733 - 13742.

[15] ZHOU Y, SUN P, ZHANG Y, et al. End-to-end multi-view fusion for 3d object detection in lidar point clouds [C] //Conference on Robot Learning, PMLR, 2020: 923 - 932.

[16] CHEN Y, LIU J, ZHANG X, et al. Voxelnext: Fully sparse voxelnet for 3d object detection and tracking [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 21674 - 21683.

[17] 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, 2019: 12697 - 12705.

[18] QIAN G, LI Y, PENG H, et al. Pointnext: Revisiting pointnet++ with improved training and scaling strategies [J]. Advances in Neural Information Processing Systems, 2022, 35: 23192 - 23204.

[19] SHI S, WANG X, LI H. Pointrcnn: 3d object proposal generation and detection from point cloud [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 770 - 779.

[20] QI C R, YI L, SU H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space [J]. Advances in Neural Information Processing Systems, 2017, 30.

[21] SHI S, GUO C, 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, 2020: 10529 - 10538.

[22] SHI S, JIANG L, DENG 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.

[23] CAO Y, WANG N, XIAO C, et al. Invisible for both camera and lidar: Security of multi-sensor fusion based perception in autonomous driving under physical-world attacks [C] //2021 IEEE Symposium on-sec-urity and Privacy (SP), IEEE, 2021: 176 - 194.

[24] CAO Y, BHUPATHIRAJU S H, NAGHAVI P, et al. You can't see me: Physical removal attacks on {lidar-based} autonomous vehicles driving frameworks [C] //32nd USENIX Security Symposium (USENIX Security 23), 2023: 2993 - 3010.

[25] YANG K, TSAI T, YU H, 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, 2021: 349 - 362.

(上接第 249 页)

[7] 韩倩倩，栾　茹，岳云涛，等. 基于 LoRa 的文物建筑无线火灾报警系统研究 [J]. 消防科学与技术，2020，39 (1)：86 - 88.

[8] 刘　颖，鹿玉红，刘　敏，等. 基于 LoRa 无线技术的地震应急救援坍塌信息获取 [J]. 计算机仿真，2020，37 (3)：224 - 228.

[9] 申　翔，吴培仁. 基于北斗卫星导航系统的海上应急搜救系统 [J]. 指挥控制与仿真，2018，40 (6)：49 - 55.

[10] 何　浩，李　阳，赵　晨，等. 搜救卫星系统技术发展与应用探讨 [J]. 卫星应用，2018，(7)：26 - 31.

[11] 刘　磊，孙超山. 低功耗远距离无线通信技术及其军事应用分析 [J]. 通信技术，2018，51 (2)：331 - 336.

[12] 万　云，蒋　阳. 基于 LoRa 技术的温室群远程监控系统的设计 [J]. 计算机工程与设计，2021，42 (2)：595 - 601.

[13] 吴　进，赵新亮，赵　隽. LoRa 物联网技术的调制解调 [J]. 计算机工程与设计，2019，40 (3)：617 - 622.

[14] 宋振雷，吴冬燕，张卫星，等. 基于 LoRa 的智慧工厂环境监测系统的设计 [J]. 电子制作，2021，(9)：79 - 81.

[15] 冯陆崴，朱丰源，田晓华. 基于模拟扩频信号处理的超低功耗 LoRa 唤醒机制 [J]. 信号处理，2023，39 (6)：1079 - 1088.

[16] 万雪芬，崔　剑，杨　义，等. 基于智能手机的 LoRa 无线传输效能测试研究 [J]. 现代电子技术，2018，41 (21)：7 - 11.

[17] 鲍胜文，方拥军，赵　飞，等. 基于 LoRa 技术的无线通信管理系统研究与实现 [J]. 电子世界，2019，(24)：120 - 122.

[18] 桂立君，李继春. 应用新一代物联网技术构建智能集装箱运输标准体系 [J]. 珠江水运，2024，(4)：120 - 122

[19] 李　英，沈金荣. 基于 LoRa 通信技术的智能水表及远程管理平台的研发 [J]. 电子技术与软件工程，2020，(6)：18 - 20.

[20] 吴雅琴，师兰兰. 基于 LoRa 的火灾救援现场人员定位算法研究 [J]. 计算机应用与软件，2020，37 (6)：70 - 75.