第 9 卷 第 5 期
2020 年 9 月
集 成 技 术
JOURNAL OF INTEGRATION TECHNOLOGY
Vol. 9 No. 5
Sep. 2020

引文格式：
李慧云, 邵翠萍, 陈贝章, 等. 基于矩阵补全的无人车感知系统的攻击防御技术 [J]. 集成技术, 2020, 9(5): 3-14.
Li HY, Shao CP, Chen BZ, et al. Attack defense technology of unmanned vehicle perception system based on matrix completion [J]. Journal of Integration Technology, 2020, 9(5): 3-14.

基于矩阵补全的无人车感知系统的攻击防御技术

李慧云[1,2,3] 邵翠萍[1,2,3] 陈贝章[1,2,3] 胡延步[1,3,4] 杨赵南[1,2,3]

[1]（中国科学院深圳先进技术研究院 深圳 518055）
[2]（中国科学院人机智能协同系统重点实验室 深圳 518055）
[3]（粤港澳人机智能协同系统联合实验室 深圳 518055）
[4]（西安电子科技大学 西安 710071）

摘 要 环境感知系统是无人驾驶技术中至关重要的一环，是整个无人车安全和稳定的前提。目前无人驾驶领域内对于环境感知技术的研究主要集中在理想环境下的环境信息获取、语义信息高精度识别以及多传感器的信息融合等，而未形成系统全面的攻击检测和防御体系。该研究利用感知系统中多传感器感知信号在时域和空间域上的相关性，建立了多传感器之间的信息交叉数学模型，可有效检测到被攻击的传感器，并基于矩阵补全方法对失真数据进行恢复。实验结果显示，该方法能够较好地检测被攻击传感器，并恢复因攻击而缺失的目标信息。

关键词 无人车；攻击防御；多传感器；数据恢复
中图分类号 U 471.15 **文献标志码** A doi: 10.12146/j.issn.2095-3135.20200509003

Attack Defense Technology of Unmanned Vehicle Perception System Based on Matrix Completion

LI Huiyun[1,2,3] SHAO Cuiping[1,2,3] CHEN Beizhang[1,2,3] HU Yanbu[1,3,4] YANG Zhaonan[1,2,3]

[1](*Shenzhen Institutes of Advanced Technology*, *Chinese Academy of Sciences*, *Shenzhen* 518055, *China*)
[2](*CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems*, *Shenzhen Institutes of Advanced Technology*, *Shenzhen* 518055, *China*)
[3](*Guangdong-Hong Kong-Macao Joint Laboratory of Human-Machine Intelligence-Synergy Systems*, *Shenzhen* 518055, *China*)
[4](*Xidian University*, *Xi'an* 710071, *China*)

Abstract Environmental perception system is an important part of unmanned driving technology, which

收稿日期：2020-05-09 **修回日期**：2020-07-16
基金项目：深圳市无人驾驶感知决策与执行技术工程实验室项目（Y7D004）；深圳电动汽车动力平台与安全技术重点实验室项目
作者简介：李慧云，博士，研究员，博士研究生导师，研究方向为安全芯片、智能系统等；邵翠萍（通讯作者），博士，工程师，研究方向为硬件安全性、数据容错等，E-mail：cp.shao@siat.ac.cn；陈贝章，硕士研究生，研究方向为激光雷达感知与定位；胡延步，硕士研究生，研究方向为 AI 硬件加速计算；杨赵南，硕士研究生，研究方向为计算机视觉。

is the premise of the safety and stability of the unmanned vehicle system. At present, the environmental perception technology is mainly focused on the environmental information acquisition under ideal environment, the high-precision recognition of semantic information, and the multi-sensor fusion etc. A comprehensive attack detection and defense system for the unmanned driving systems is still not available. In this paper, the correlation of multi-sensor perception signals in the perception system in both temporal and spatial domains are considered to establish a mathematical model, which is used to connect the information among different sensors, and to detect the attacked sensors. More important, it can be also used to recover the distorted data based on the matrix completion method. The experimental results showed that, the proposed method can detect the attacked sensor effectively and recover the missed information in the attached perception systems.

Keywords unmanned vehicle; attack defend; multi-sensor; data recovery

1 引 言

信息和通信技术的日新月异带动了智能驾驶技术的发展，以提高交通安全和效率为目标，汽车智能化、网联化已成为汽车产业发展的必然趋势，不同国家和组织争先推出相关政策和新技术[1]。无人驾驶技术是当今社会和前沿科学技术发展的重要方向之一，对于社会的多个领域诸如城市建设、交通出行、经济发展和国防力量有着不可估量的重要意义。

无人驾驶汽车通过车载传感系统感知汽车行驶过程中的道路环境状况，同时对获取的信息进行分析处理，自动规划行车路线并对车辆进行导航，从而到达预定目的地[2]。其中环境感知技术的功能如同人类的眼睛和耳朵一样，主要由激光雷达、视觉摄像头、毫米波雷达、全球定位系统(GPS)等设备组成。该技术主要用来获取无人驾驶汽车周围详细的环境信息，并为规划与决策模块提供丰富的数据，既包括障碍物的位置、形状、类别及速度信息，也包括对一些特殊场景的语义理解(如施工区域、交通信号灯及交通路牌等)[3]。无人驾驶汽车规划与决策环节的安全性以环境感知技术的安全为前提，一旦无人车的感知系统受到攻击，将会导致传感器获取的信息失真及错误的识别结果，进而规划不正确的驾驶策略，极有可能引发车祸，造成严重的生命与财产损失。此类攻击手段廉价、高效且隐蔽、不需要直接访问正在使用的系统，因此对无人车的安全性造成巨大的威胁[4]。例如，2019 年腾讯科恩实验室的研究人员以特斯拉 Model S 为对象，针对其搭载的“Autopilot”进行了安全性研究，找到了使用物理攻击欺骗特斯拉自动驾驶系统的方法。该方法通过在道路特定位置贴上几个贴纸，使得处在自动驾驶模式的汽车并入反向车道[5]。因此，具有一定防御性的环境感知系统对整个无人驾驶汽车的安全性和稳定性起到了先决的作用，研究无人车感知系统对抗主动攻击的防御问题是保证其安全行驶的关键。

本文针对无人系统感知设备研究实时攻击防御与高精度数据恢复方法。通过研究不同传感器之间信息的交叠关系和语义相关性，建立信息交叉数学模型和虚假信息干扰数学模型。同时，根据信息交叉模型判断致错传感器，并采用矩阵补全和矩阵分解等方法对失真信息进行高精度重构恢复。随着无人车产业的发展，无人车攻击等安全性研究将愈发重要，因而本研究在科学研究和

工程应用上都极具现实意义。

2　无人车感知系统攻击

图 1 是整个车载感知系统的架构。该系统的 3 种传感器(激光雷达、相机和毫米波雷达)数据需进行时间同步，将所有的时间误差控制在毫秒级。结合传感器数据，感知系统以帧为基础进行检测、分割、分类等计算，最后利用多帧信息进行多目标跟踪，将相关感知结果输出。由于无人车行驶环境的复杂性，感知系统是多种传感设备间的数据补充与冗余备份的功能模块。如果没有数据恢复的防御机制，任何一个传感设备遭受外部攻击都将对无人车安全行驶造成巨大的威胁。

2.1　摄相机攻击

摄像机作为无人车中必备的一种器件，具有目标检测功能，如行人、车辆、红绿灯，车道线、交通标识检测等[6-8]。但在应对道路结构复杂、人车混杂的交通环境时，相机感知技术还存在很多不足，如存在目标检测困难、易受近距离攻击等问题[9]。而对视觉传感器的近距离攻击主要是通过添加有害信息使视觉检测系统出错，进而导致错误的驾驶策略[10-11]。

添加有害信息的方法一般是通过透镜印刷的方式[12-13]。透镜图像的特点是从不同角度观察同一个交通标志时，得到的结果不同，具体攻击原理如图 2 所示。此外，当标志牌上印刷两种不同的交通标志时，摄像机的角度和人眼角度观察到的标志信息不同，如图 3 所示。

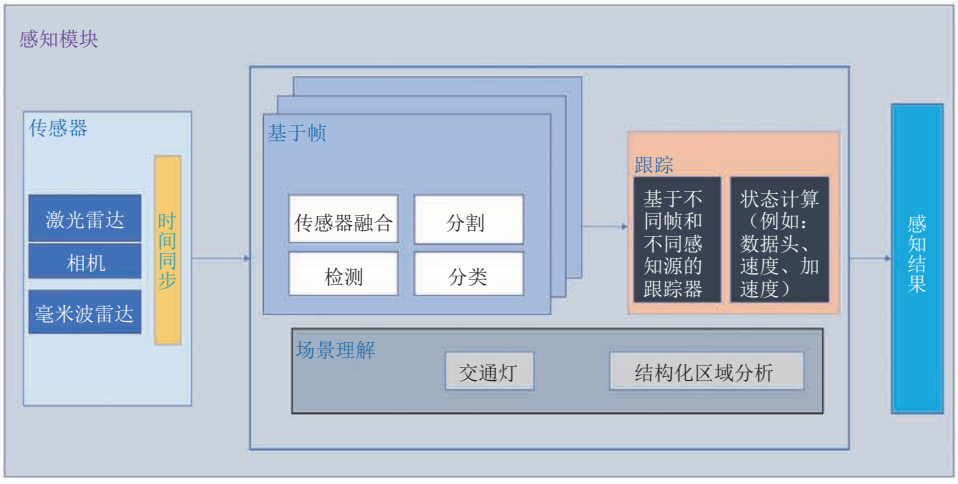


图 1　无人车车载感知系统的架构

Fig. 1　Architecture of perception system of unmanned vehicle

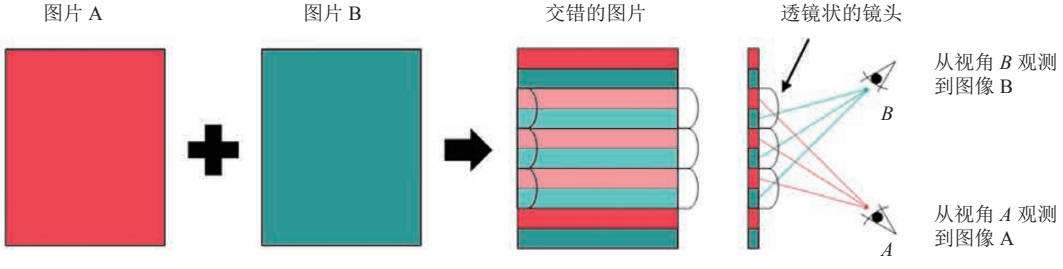


图 2　透镜图像的生成过程及其成像特点

Fig. 2　The process of lens image generation and its imaging characteristics

图 3 攻击示例

Fig. 3 Example of attacks

2.2 激光雷达攻击

针对激光雷达的攻击主要分为对传感器底层感知原理的攻击与感知算法层的攻击。其中底层原理攻击主要包括 3 种方式：激光距离欺骗攻击[14]、激光角度欺骗攻击[15-16]和激光致盲攻击[17]。

激光雷达的底层感知源自激光时间飞行(Time of Flight)原理，即激光发射器发出激光脉冲波，内部定时器开始计算时间 t_1；激光波碰到物体后部分能量返回，当激光接收器收到返回激光波时，停止内部定时器，此时间段记为 t_2；光速表示为 C。则激光雷达到物体的距离如公式(1)所示：

$$S=C\times\frac{(t_2-t_1)}{2} \tag{1}$$

通过发射和接收激光束，分析激光遇到目标对象后的折返时间，计算出到目标对象的相对距离。然后利用此过程中收集到的目标对象表面大量密集点的三维坐标、反射率和纹理等信息，快速得到被测目标的三维模型以及线、面、体等各种相关数据。进一步建立三维点云图，绘制出环境地图，以达到环境感知的目的。激光跟踪测量雷达系统组成中同时包含测距和测角两个探测分系统，对任何一个探测系统的有效干扰都会影响激光跟踪测量雷达的总体性能[18-21]，即可以通过对其中任何一个探测分系统的干扰实现对激光雷达的干扰。

近年来国内外对激光雷达攻击的研究有转向感知算法层攻击的趋势。Cao 等[14]研究认为一些特殊三维结构的物体也会令激光雷达受到对抗攻击，由此错误地把某些物体当做行人，或者对特殊形状的障碍物视而不见。随后，该团队提出了一种 LiDAR-Adv 方法，可生成逃避激光雷达检测的对抗物体。其中，针对激光雷达所制作的对抗样本如图 4 所示。将对抗样本放在路径中央(如图 5)，配置激光雷达的汽车直到逼近对抗样本时才检测出该目标，以至于躲闪不及。该研究揭示了基于自动驾驶的激光雷达感知系统存在潜在的漏洞。

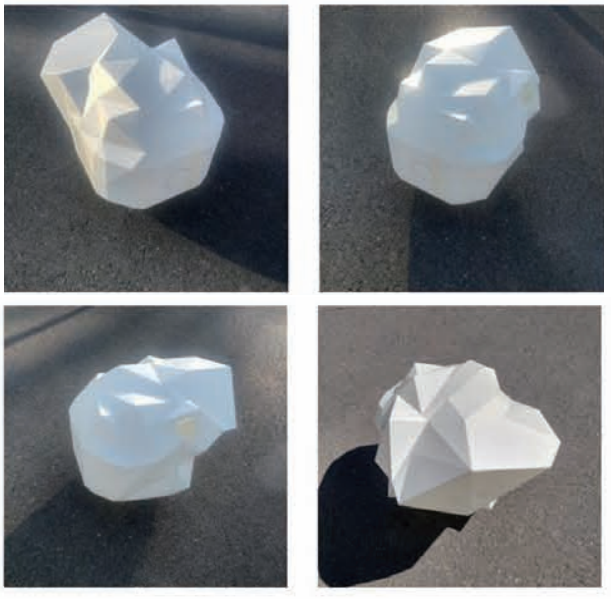


图 4 LiDAR-Adv 生成的激光雷达对抗样本[14]

Fig. 4 LiDAR-Adv generated lidar adversarial samples[14]

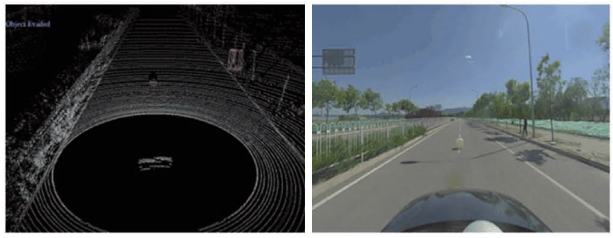


图 5　激光雷达对抗样本（障碍物）的检测失效[14]

Fig. 5　Lidar failed to adversarial samples (obstacles) [14]

3　感知系统防御方法

目前对无人车感知系统攻击的防御方法主要是针对单一传感器采取的一些改善和防御，其可以从一定程度上减小攻击带来的影响，但尚未形成系统全面的攻击检测和防御体系。本文采用矩阵补全和矩阵分解等方法对攻击致错的数据进行高精度恢复，形成一套完整的以检测手段与数据恢复算法为核心的攻击防御方案。

3.1　建立多传感器信息交叉数学模型及攻击检测

虽然无人车在行驶中的环境信息是实时变化的，但是不同的传感器对环境的感知一直存在着空间信息的冗余和交叉，这种相互的交叉和重叠导致传感器两两之间存在相关性。如果能够在空间域上建立传感器之间恒定不变的相关性，那么就能够以此判断某个时刻是否有传感器发生异常或被攻击。

除了空间域上的相关性，每个传感器自身感知的信息在不同的时刻也存在着相关性。本文将传感器自身在不同时间的相关性称为自相关性，而将不同的传感器在空间域上的相关性称为互相关性。利用时间域的自相关性检测被攻击导致的错误，再结合空间域上的互相关性定位被攻击的传感器，这是本文利用感知信号之间的相关性做定位和检测的核心思想。图 6 描述了多传感器信息交叉模型的建立。

在分析多传感数据相关性之前，需先将多传感器（包括摄像头、激光雷达、红外传感器、超声波传感器、GPS 传感器等）在空间或时间上的冗余或互补信息进行提取，以获得被测对象的一致性解释或描述。具体地，多传感器数据提取的方法如下：

（1）收集 N 个不同类型的传感器（有源或无源的）对目标的观测数据。

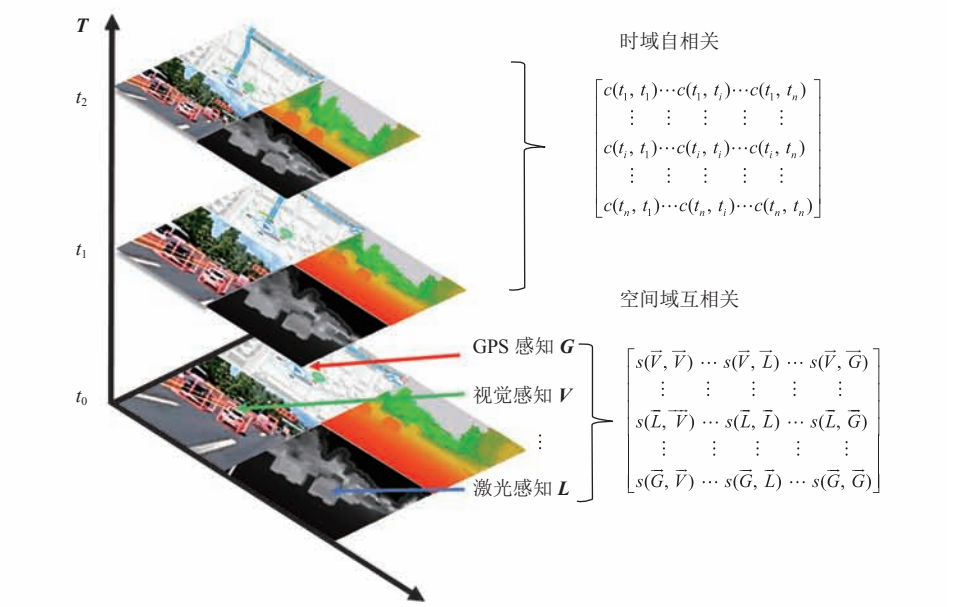


图 6　多传感器之间的信息相关性

Fig. 6　The information correlation among the multi-sensors

(2)对传感器的输出数据(离散的或连续的时间函数数据、输出矢量、成像数据或一个直接的属性说明)进行特征提取的变换，提取代表观测数据的特征矢量 $\boldsymbol{Y}_i$；特征提取的过程可以表征为 $\overrightarrow{Y_1}=f_d\left[f_e\left(\overrightarrow{X_1}\right)\right]$。其中，函数 f_e 表征编码过程，即通过卷积、池化、下采样等方法整合更为底层的特征，如图像纹理、颜色等；函数 f_d 表征解码过程，一般采用上采样来还原到原始图片的分辨率。

(3)对特征矢量 $\boldsymbol{Y}_i$ 进行模式识别处理(如聚类算法、自适应神经网络或其他能将特征矢量 $\boldsymbol{Y}_i$ 变换成目标属性判决的统计模式识别法)完成各传感器关于目标的说明。

以目标检测为例，视觉传感器得到的特征矢量用 $\vec{V}$ 表示；激光雷达得到的特征矢量用 $\vec{L}$ 表示；GPS 得到的特征矢量用 $\vec{G}$ 表示；IMU 得到的特征矢量用 $\vec{I}$ 表示；则所有传感器的特征矢量数据 $\boldsymbol{P}$ 可以表示为 $\boldsymbol{P}=(\vec{V},\vec{L},\vec{G},\vec{I},\cdots)$。这些传感器在做目标检测或定位时，两两之间一定存在着空间上的互相关和时间上的自相关。因此，可以用相关性矩阵 $\boldsymbol{S}$ 表示它们在空间上的互相关性；矩阵 $\boldsymbol{C}$ 表示单个传感器在时间域上的自相关性(以激光雷达为例，$\boldsymbol{C}_L$ 表示激光雷达在不同时刻的自相关性矩阵；$\overrightarrow{L_l}$ 表示不同时刻激光雷达的特征矢量)，具体如公式(2)～(3)所示：

$$\boldsymbol{S}=\begin{bmatrix} s(\vec{V},\vec{V}) & \cdots & s(\vec{V},\vec{L}) & \cdots & s(\vec{V},\vec{G}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ s(\vec{L},\vec{V}) & \cdots & s(\vec{L},\vec{L}) & \cdots & s(\vec{L},\vec{G}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ s(\vec{G},\vec{V}) & \cdots & s(\vec{G},\vec{L}) & \cdots & s(\vec{G},\vec{G}) \end{bmatrix} \tag{2}$$

$$\boldsymbol{C}_L=\begin{bmatrix} c(\overrightarrow{L_1},\overrightarrow{L_1}) & \cdots & c(\overrightarrow{L_1},\overrightarrow{L_l}) & \cdots & c(\overrightarrow{L_1},\overrightarrow{L_n}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c(\overrightarrow{L_l},\overrightarrow{L_1}) & \cdots & c(\overrightarrow{L_l},\overrightarrow{L_l}) & \cdots & c(\overrightarrow{L_l},\overrightarrow{L_n}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c(\overrightarrow{L_n},\overrightarrow{L_1}) & \cdots & c(\overrightarrow{L_n},\overrightarrow{L_l}) & \cdots & c(\overrightarrow{L_n},\overrightarrow{L_n}) \end{bmatrix} \tag{3}$$

通过分析传感器之间的互相关性和自相关性，建立相关性表示的数学模型，研究攻击致错的特征在相关性模型中的传递规律，得到错误传递的函数，可以为进一步定位攻击来源和偏差计算提供理论依据。接着，根据传感器之间信息交叉的特点和不同传感器特征矢量的数据形态选择合适的数据相关性表示形式。同时，建立既能够横向体现随着环境恒定不变的感知数据交叉特性，又能够纵向区分特征矢量中不同物理量在相关性表达中的贡献，且保证错误不在相关性模型中湮没的方法。

3.2 基于矩阵补全的数据恢复方法

接下来重点考虑如何根据错误传递机理去定位被攻击或出现异常的传感器，及如何最大化地修复错误的特征数据，减少错误对感知结果的影响。矩阵补全(Matrix Completion，MC)是一种补全缺失信息的方法[22-23]。在矩阵的元素存在未知或缺失的情况下，矩阵补全可根据已知元素估计出未知元素，从而将矩阵恢复完整。矩阵补全起源于机器学习，即已知部分样本(这些样本来自拥有低秩协方差矩阵的过程)，需要估计那些缺失或未知的数据。目前矩阵补全已广泛应用于机器学习、工程控制、图像和视频处理。本文采用基于矩阵补全的方法对由于无人车被攻击而失真的感知信息进行重建，最终对感知信息进行高精度的恢复。首先，假设矩阵 $\boldsymbol{X}$ 为无人车待恢复的感知信号；$\boldsymbol{M}$ 为感知信号被攻击后的原始信号；且 $\boldsymbol{M}$ 中部分元素因被攻击而失真。然后，通过矩阵补全的方法找到矩阵 $\boldsymbol{X}$，使得 $\boldsymbol{X}$ 中的元素尽量逼近 $\boldsymbol{M}$ 中没有被攻击的部分，而 $\boldsymbol{X}$ 中其他元素作为失真信息的逼近估计。图 7 为采用矩阵补全方法对失真信号进行重建的原理图。标准矩阵补全问题可建模为如下形式的秩最小化约束优化模型：

$$\min_{X\in R^{n1\times n2}} rank(\boldsymbol{X}) \quad \text{s.t. } P_{\Omega}(\boldsymbol{M})=P_{\Omega}(\boldsymbol{X}) \tag{4}$$

其中，$\Omega\in[n1]\times[n2]$ ($n1=\{1, 2, \cdots, n1\}$, $n2=$

$\{1, 2, \cdots, n2\}$）为采样元素的索引集合；$P_{\Omega}(\cdot)$为正交投影算子，表示当$(i, j) \in \Omega$时，$\boldsymbol{M}_{i,j}$为采样元素，则模型可进一步表示为公式(5)：

$$[P_{\Omega}(\boldsymbol{M})]_{i,j}\begin{cases}\boldsymbol{M}_{i,j}, & \text{if}(i,j)\in\Omega\\ 0, & \text{otherwise}\end{cases} \tag{5}$$

当采样数据存在误差时，上述模型可进一步修正为公式(6)：

$$\min_{X\in R^{n1\times n2}} rank(\boldsymbol{X}) \quad \text{s.t.} \ \left\|P_{\Omega}(\boldsymbol{M}-\boldsymbol{X})\right\|_F \leqslant \delta \tag{6}$$

对标准矩阵补全问题中$\boldsymbol{X}$的求解大致可分为4类：基于核范数松弛的矩阵补全模型、基于矩阵分解的矩阵补全模型、基于非凸函数松弛的矩阵补全模型及其他类型的矩阵补全模型。这些矩阵补全模型关注的都是如何基于目标矩阵的先验低秩性从少量采样观察中补全缺失元素。它们的主要区别在于秩函数的松弛方式不同，从而导致模型的凸性各异，模型求解效率和可扩放性也因此不同。在实际使用中，需结合矩阵的规模，以及对数据精度的要求选择合适的松弛补全模型。

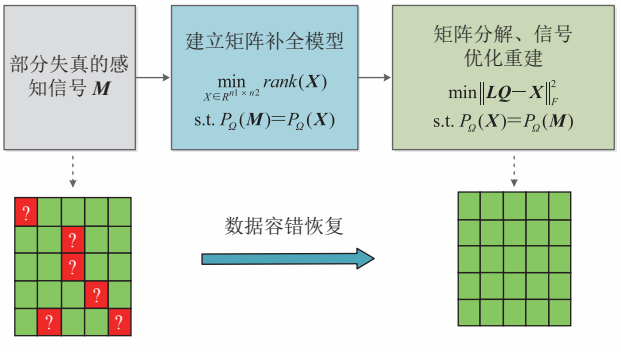


图 7　采用矩阵补全对失真信号的数据容错恢复原理框图

Fig. 7　Principle block diagram of data fault-tolerant recovery for distorted signals with matrix completion

4　实验设计与结果

4.1　实验建立

实验在开源的自动驾驶仿真软件 Carla Simulator 上进行。Carla Simulator 是 Intel Visual Computing Lab 推出的一款开源模拟器，主要用于城市自动驾驶研究。Carla 支持城市自动驾驶系统底层开发、训练和验证。Carla 通过 Server-Client 方式使车辆与虚拟世界进行交互。Client API 采用 Python 编写，Client 向 Server 发送 command 和 meta-command。其中，command 为控制车辆的转向、加速和刹车；meta-command 针对的是 Server 的行为，主要有重启模拟器、改变环境特征和修改传感器组等。Carla 可以加入不同天气和光照等环境特征以及车辆和行人的密度等影响，并且可以根据需求配置传感器（包括彩色相机、深度相机、激光雷达、IMU、GPS 等），从而实时获取动态仿真数据，以供用于自动驾驶仿真测试。视觉相机采用 RGB 彩色相机和深度相机，其图像参数均为 800×600；镜头水平视场角均为 100°；激光雷达为 32 线激光雷达；俯仰角范围为 −26.8°～2°；旋转频率为 1 200 r/min。实验中模拟强光对激光雷达进行攻击，使得点云数据缺失。硬件平台主要为联想移动工作站 TinkPad P51，其配置如表 1，具体过程分为 4 个步骤。

(1) 传感数据采样与特征提取

视觉感知包含目标检测和定位。在自动驾驶感知系统中，只有 GPS 是属于绝对定位的，其他定位方式都需要在结构化的环境下提取相应的环境特征。一旦受到外界的主动攻击，使得特征矢量丢失，很容易出现系统定位或检测失效的问题。例如，如果自动驾驶车辆的视觉定位系统被

表 1　硬件平台配置

Table 1　Hardware platform configuration

项目	参数
CPU	i7-7700HQ 2.8～3.8 GHz
GPU	NvidiaQuadro M2200 4G
OS	Ubuntu 16.04 LTS
CUDA	9.0,with cudnn
SSD	1TB
Software	Python,ROS,Carla Simulator

强光攻击，且没有激光雷达恢复定位的特征，那么将会对自动驾驶的行车带来重大的安全隐患。因此，通过多传感融合来对抗外在攻击的定位方案对于自动驾驶的安全性是必不可少的。

图 8 为激光点云定位框架。首先，通过事先采集的点云信息和激光点云构建激光雷达地图(反射值地图和高度值地图)，并根据激光反射强度与激光高度等物理世界特征量构建地图。然后，通过车辆上的传感器实时匹配自身获取的数据和之前构造的特征地图，从而解算出载体的相对位姿。而点云匹配定位的过程是一个优化问题，只要定义好损失函数，那么求解最小化损失函数就是载体定位的过程。此外，图像对齐是用优化的方法求解航向角 *yaw*，并采用 SSD-HF (SSD-Sum of Squared Difference Histogram Filter) 的优化方法解算 x 和 y。(x, y) 表示载体在激光雷达地图上的平面坐标点，高度信息 z 直接从点云数据中获取。最后，激光定位算法输出位姿信息 $\boldsymbol{X}(x, y, z, yaw)$。当搭载了激光雷达的载体在一次扫描中观测到点云 L_m，则将 L_m 在激光雷达坐标系下的位姿记为 Y_{Lm}；激光雷达载体相对于地图的坐标系为 $\boldsymbol{X}(t)$。根据激光雷达观测原理可写出激光雷达观测公式：

$$Y_{Lm}=\rho\begin{bmatrix}\cos\alpha\cos\beta\\ \cos\alpha\sin\beta\\ \sin\alpha\end{bmatrix}=\boldsymbol{H}(t)\boldsymbol{X}(t)+v \tag{7}$$

其中，ρ 为测量距离；α、β 分别为激光脉冲的测量角；$\boldsymbol{H}(t)$ 表示观测矩阵；v 为测量噪声。若一次扫面的点云数量为 k，则激光雷达的点特征矢量观测公式为：

$$\vec{L}=\begin{bmatrix}\boldsymbol{Y}_{L_1}\\ \boldsymbol{Y}_{L_2}\\ \vdots\\ \boldsymbol{Y}_{L_k}\end{bmatrix}=\begin{bmatrix}\boldsymbol{H}_1(t)\\ \boldsymbol{H}_2(t)\\ \vdots\\ \boldsymbol{H}_k(t)\end{bmatrix}\boldsymbol{X}(t)+\begin{bmatrix}v_1\\ v_2\\ \vdots\\ v_k\end{bmatrix} \tag{8}$$

在定位任务中，激光雷达 Lidar 的特征矢量可以表示为 $\vec{L_l}=[x\ y\ z\ yaw]^{\mathrm{T}}$；在检测任务中，激光雷达 Lidar 的特征向量可以表示为 $\vec{L_d}=[X\ Y\ h\ w\ \theta]^{\mathrm{T}}$；$[X\ Y]^{\mathrm{T}}$ 表示检测目标相对车体坐标系 $[0\ 0]^{\mathrm{T}}$ 的位置；$[h\ w]^{\mathrm{T}}$ 表示检测框的高度和宽度；θ 为检测框内是检测目标的置信度。

惯性测量单元(IMU)的特征矢量为：$\vec{I}=[yaw\ pitch\ roll\ V_x\ V_y\ V_z\ a_x\ a_y\ a_z\ w_x\ w_y\ w_z]^{\mathrm{T}}$，其中 $[yaw\ pitch\ roll]^{\mathrm{T}}$ 表示航向角、俯仰角和横滚角三姿态；$[V_x\ V_y\ V_z]^{\mathrm{T}}$ 表示三轴速度；$[a_x\ a_y\ a_z]^{\mathrm{T}}$ 表示三轴加速度；$[w_x\ w_y\ w_z]^{\mathrm{T}}$ 表示三轴角速度。GPS 特征矢量 $\vec{G}=[x\ y\ z]^{\mathrm{T}}$，$(x, y, z)$ 为车体的三维坐标。自动驾驶车辆通过车载摄像头采集车辆周围的图像。采集到的图片以 RGB 格式输入目标检测系统，检测系统调用深度卷积神经网络算法对 RGB 图像进行特征提取，最终从图片中提取的特征能有效描述目标物体的信息。目标物体检

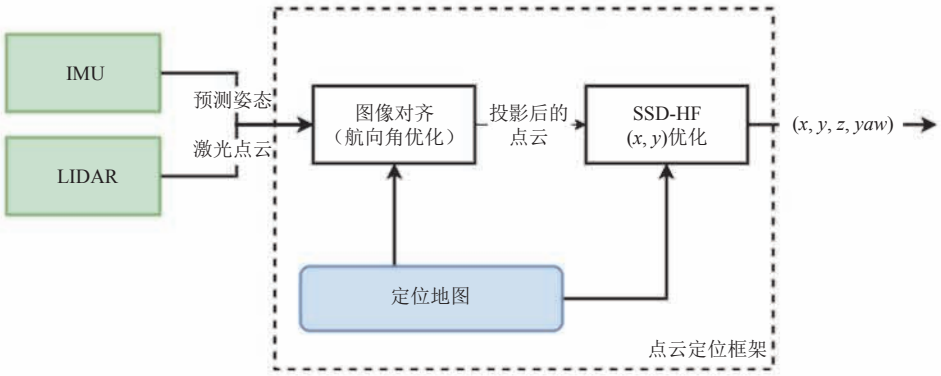


图 8 激光点云定位框架

Fig. 8 Laser point cloud positioning frame

测结果采用向量 $\vec{V}=[X_v\ Y_v\ h\ \omega\ p]^{\mathrm{T}}$ 表示。其中，$[X_v\ Y_v]^{\mathrm{T}}$ 表示目标物体在图像坐标系上的坐标值；$[h\ \omega]^{\mathrm{T}}$ 表示检测框的高度和宽度；v 表示目标物体的类别编号；p 表示目标物体的置信度。

(2) 分析特征数据并建立相关性模型

根据上一步得到的各个传感器感知数据的特征矢量，建立不同传感器之间恒定不变的互相关性和各传感器基于时间不变的自相关性模型。设 Correlation_space() 表示互相关性模型的数学函数，Correlation_time() 表示自相关性模型的数学函数，矩阵 $\boldsymbol{S}_{\text{constant}}$ 表示感知器之间固有的相关性关系矩阵。

(3) 攻击检测和定位

根据相关性模型的数学表示，实时计算传感器互相关性和自相关性，并与 $\boldsymbol{S}_{\text{constant}}$ 表示的感知器之间固有的相关性比对。若完全一致，则表示当前感知系统没有被攻击或出现异常；若两者不一致，则代表传感器出现异常，然后根据自相关性定位被攻击的传感器。攻击检测和定位的实例如图 9 所示，其中 $s(a, b)$ 表示传感器 a 与 b 之间的特征相关性。具体的检测流程如下：

①根据相关性模型，计算没有攻击时传感器之间的互相关性，得到互相关性矩阵 $\boldsymbol{S}_{\text{constant}}$ 和各传感器的自相关性矩阵 $\boldsymbol{C}_1,\boldsymbol{C}_2,\cdots,\boldsymbol{C}_n$；

②采用高重频脉冲激光器作为干扰源对激光雷达进行距离攻击；

③实时采集感知的环境信息，分别对感知数据进行特征提取；

④分别采用时间相关性模型和空间相关性模型对提取的特征进行相关性计算，得到各传感器基于时间的自相关性矩阵 $\boldsymbol{C}_1',\boldsymbol{C}_2',\boldsymbol{C}_3',\cdots,\boldsymbol{C}_n'$，以及传感器之间的互相关性 $\boldsymbol{S}'$；

⑤比较自相关性矩阵 $\boldsymbol{C}_1',\boldsymbol{C}_2',\boldsymbol{C}_3',\cdots,\boldsymbol{C}_n'$ 与 $\boldsymbol{C}_1,\boldsymbol{C}_2,\cdots,\boldsymbol{C}_n$，若前后不一致，则初步定为受攻击传感器；

⑥通过 $\boldsymbol{S}'$ 与 $\boldsymbol{S}$ 比对，进一步确认上一步定位的传感器与其他传感器之间的相关性是否出错，若是，则该传感器被定为受攻击传感器。

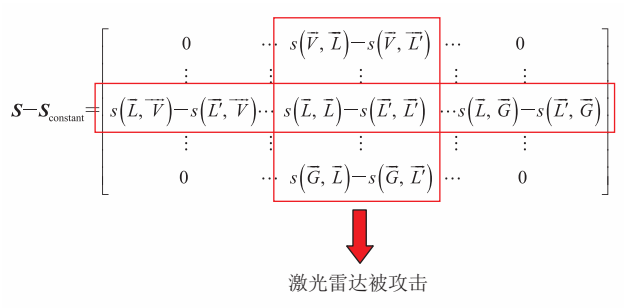


图 9　攻击定位

Fig. 9　Attack localization

(4) 失真数据恢复

当检测到某个传感器被攻击时，联立各个传感器的特征信息，建立原始感知信号矩阵 $\boldsymbol{M}$。其中，$\boldsymbol{M}$ 中未被攻击的传感器感知的特征信息元素集合称为指标集，而在 $\boldsymbol{M}$ 中被攻击的信息称为缺失元素。根据原始感知信号矩阵 $\boldsymbol{M}$，建立矩阵补全数学模型，重构感知信号矩阵 $\boldsymbol{X}$。基于核范数松弛的建模方法涉及复杂的矩阵奇异值分解，所以会导致模型的求解效率和可扩放性 (Scalability，也称为可扩展性) 受限。而基于矩阵分解的建模方法是一类可替代的矩阵补全模型构建方法，其基本思路是将目标矩阵分解为 2 个低秩矩阵 $\boldsymbol{L}$ 和 $\boldsymbol{Q}$ 的乘积，从而避免了复杂的矩阵奇异值分解，加速了算法的执行效率。采用矩阵分解的矩阵补全建模如公式 (9)：

$$\min_{\boldsymbol{L}\in R^{n1\times k},\boldsymbol{Q}\in R^{k\times n2},\boldsymbol{X}\in R^{n1\times n2}}\|\boldsymbol{LQ}-\boldsymbol{X}\|_F^2\quad \text{s.t. } P_{\Omega}(\boldsymbol{X})=P_{\Omega}(\boldsymbol{M})\tag{9}$$

其中，k 为预测的矩阵秩界，该模型采用分块坐标下降算法 (俗称交替最小化算法) 求解，如果能够预先获取合适的 k 值，那么该模型可以在较小的时间复杂度内获得相当精度的解。此外，对任意秩为 r 的矩阵 $\boldsymbol{X}\in Rn1\times n2$，若 $k>r$，则公式 (10) 成立：

$$\|\boldsymbol{X}\|_*=\min_{\boldsymbol{L}\in R^{n1\times k},\boldsymbol{Q}\in R^{n2\times k}}\frac{1}{2}\left\{\|\boldsymbol{L}\|_F^2+\|\boldsymbol{Q}\|_F^2\quad \text{s.t. } \boldsymbol{X}=\boldsymbol{LQ}^{\mathrm{T}}\right\}\tag{10}$$

则公式(10)的解即为全局最优解，即 $\widehat{\boldsymbol{X}}=\widehat{\boldsymbol{L}}\widehat{\boldsymbol{Q}}$，然后通过 $\widehat{\boldsymbol{X}}$ 中对应的值来填充 $\boldsymbol{M}$ 中缺失的部分。

4.2 实验结果

传统点云的补全方法大多依托点云集本身对点云集细节进行补偿，或通过对抗生成网络对 3D 点云模型的部分缺失补偿[24]，不适用于本实验的户外场景。由于针对性攻击造成的包覆目标外侧所有点云缺失的特殊场景，因此本文对比实验采用的是点云查找补全法[25]。图 10 结果显示，本文提出的方法能够较好地恢复被攻击缺失的感知目标信息。其中，图 10(a)表示正常情况下，激光雷达检测到一辆车；图 10(b)表示该车辆遭到攻击后，点云信息完全缺失；图 10(c)表示基于传统补全法恢复的激光点云可视效果图；图 10(d)表示本文方法恢复的激光点云可视效果图。

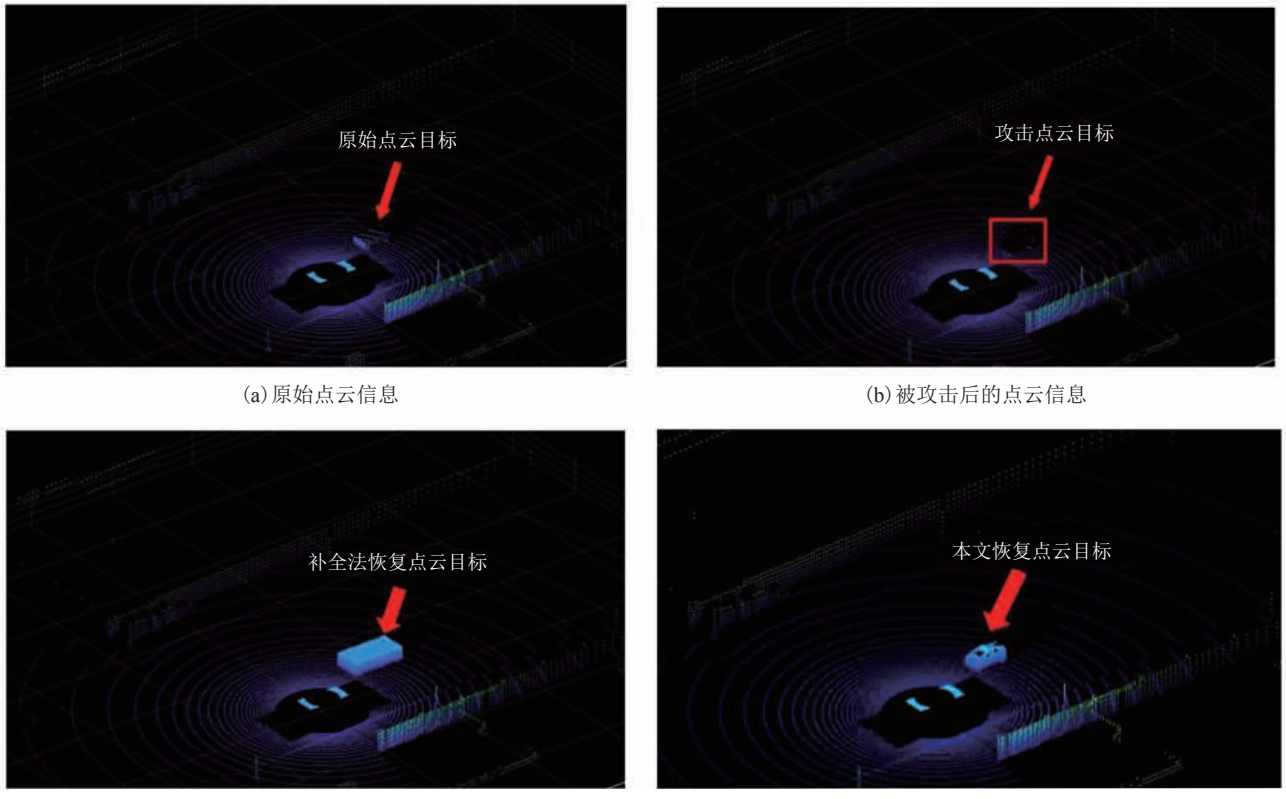
数据恢复的精度通过恢复的点云数据在两个维度上的平均相对误差来评估，可以表示为公式(11)～(13)(n 为实验测试次数)：

$$\delta_x=\frac{|\hat{x}-x|}{x}\times 100\% \tag{11}$$

$$\delta_y=\frac{|\hat{y}-y|}{y}\times 100\% \tag{12}$$

$$\delta=\frac{1}{2n}\sum_{i=1}^{n}\left(\frac{\delta_{x_i}}{x_i}+\frac{\delta_{y_i}}{y_i}\right)\times 100\% \tag{13}$$

表 2 所示为两种恢复方法的实验结果。从表 2 可知，与点云补全方法相比，本文方法所恢复的目标点云信息精度更高、耗时更少。结合图 10 的数据可视图可知，由于雷达激光是呈射线水平均匀散射状发出，故会被目标车辆吸收或遮挡等。因此，造成了依据规则点云恢复的点云



(a)原始点云信息　(b)被攻击后的点云信息

(c)点云补全法恢复的点云信息　(d)本文方法恢复的点云信息

图 10　激光雷达点云数据攻击与恢复

Fig. 10　Point cloud data attack and recovery

表 2　两种恢复方法的实验结果

Table 2　Comparison of two recovery methods

方法	测试序列 1			测试序列 2		
	相对坐标(*X*, *Y*)距离（m）	δ（%）	耗时（s）	相对坐标(*X*, *Y*)距离（m）	δ（%）	耗时（s）
原始目标	(8.90, 4.10)	--	--	(9.31, 4.10)	--	--
点云补全法	(9.84, 5.06)	3.40	4.52	(9.02, 5.62)	4.70	4.85
本文方法	(8.71, 4.15)	0.30	1.46	(9.30, 4.17)	0.20	1.24

注：(*X*, *Y*)表示被攻击目标车辆的相对坐标，*X* 表示沿测试车身纵向距离(车头方向为正)，*Y* 表示测试车辆横向距离(车身左侧为正)；δ 表示恢复误差；“--”表示该项数据不关注

空间大于实际缺失的点云空间，测量位置偏移。而本文采用的方法通过与视觉感知数据的相关性准确定位缺失的点云，有效地约束了失真范围。

5　讨论与分析

目前国内外对于无人车近距离攻击的安全研究尚处于比较早期的阶段，对无人车近距离攻击的防御方法主要集中在单一传感器上，尚未形成系统的攻击检测和防御体系。目前，对于视觉攻击的防御主要有 2 种方法。一种是针对相机感知原理底层的防御方法，有研究[15]通过增加冗余的方式部署激光雷达，针对光学反射伪装背景纹路、亮度相似的障碍物进行防御，但该方法需要探讨两种不同感知设备对同一攻击源产生不同置信度的逻辑判断。而人为地设计冗余传感器置信度判定的方法不适用于多变复杂的道路攻击场景，故此方法属于仅探测，这意味着该类方法在对抗样本上仅能报警，却不能将对抗样本完全识别[16]。另一种是针对感知算法层面的防御方法，通过不断输入新类型的对抗样本并执行对抗训练，从而不断提升网络的鲁棒性。这种基于深度学习的训练方法虽然从一定程度上减少对抗样本对传感器识别的影响，但总是会存在新的对抗样本，同样不适用于复杂多变的道路攻击场景。本研究通过冗余传感之间的相关性，恢复攻击目标的检测，为传统基于规则式的感知算法提供了新的研究思路。

此外，针对激光雷达攻击的防御主要有针对激光感知原理底层的防御方法和针对激光感知算法层面的防御方法。前者的策略包括：(1)通过在光学和光电装置中安装快光电开关、滤光片，防止激光致盲；(2)研究抗激光结构，例如夹层结构，防止敌激光能量对己方装备的破坏[17,26]；(3)通过对技术参数严格保密，如对己方激光信号采取编码技术，加大敌方干扰难度；(4)研制和发展特种耐高温材料的壳体，使其难以被激光武器烧毁和穿透[27]。但是，这些方法只能在一定程度上降低被攻击的风险，没有对被攻击后的数据进行恢复的措施。

针对 LiDAR-Adv[14]生成的激光雷达对抗样本，使用本文算法分析相机与激光的相关性，恢复检测出对抗样本，能够解决自动驾驶激光雷达感知系统的潜在问题。本文利用传感器之间的信息相关性，建立无人车车载传感器信息交叉数学模型与虚假信息干扰数学模型，对被攻击传感器进行实时检测。同时，采用矩阵补全方法对攻击致错的数据进行高精度恢复，形成一套完整的以检测手段与数据恢复算法为核心的攻击防御方案。

6　结　论

本文采用基于矩阵补全的方法对由于无人车被攻击而失真的感知信息进行重建，最终对感知信息进行高精度的恢复。通过矩阵补全的方法找到恢复的感知信号矩阵，尽量逼近原始信号中没

有被攻击的部分。实验结果显示，本文方法能够较好地恢复被攻击缺失的感知目标信息。

由于真实的无人车测试场景需要路测的条件，且需在无人车感知系统基本完善的情况下进行。因此，未来将在实际场景中对不同传感器进行攻击和防御的实验验证，迭代完善攻击检测和数据恢复方法。此外，下一步将优化传感器特征提取和数据融合方式，提高算法执行的效率和精度，使防御技术更为精准高效。

参 考 文 献

[1] Chan CY. Advancements, prospects, and impacts of automated driving systems [J]. International Journal of Transportation Science and Technology, 2017, 6(3): 208-216.

[2] 陈诚, 张永博, 李必军. 激光点云在无人驾驶路径检测中的应用 [J]. 测绘通报, 2016(11): 67-71.

[3] Wang Y, Chao WL, Garg D, et al. Pseudo-lidar from visual depth estimation: bridging the gap in 3D object detection for autonomous driving [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 8445-8453.

[4] 王世峰, 戴祥, 徐宁, 等. 无人驾驶汽车环境感知技术综述 [J]. 长春理工大学学报: 自然科学版, 2017, 40(1): 1-6.

[5] Wu XW, Sahoo D, Hoi SCH. Recent advances in deep learning for object detection [J]. Neurocomputing, 2020, 396: 39-64.

[6] Zhao ZQ, Zheng P, Xu S, et al. Object detection with deep learning: a review [J]. IEEE Transactions on Neural Networks and Learning Systems, 2019: 3212-3232.

[7] Liu L, Ouyang W, Wang X, et al. Deep learning for generic object detection: a survey [J]. International Journal of Computer Vision, 2020, 128(2): 261-318.

[8] Mukhometzianov R, Wang Y. Machine learning techniques for traffic sign detection [Z/OL]. arXiv Preprint, arXiv:1712.04391, 2017.

[9] Sitawarin C, Bhagoji AN, Mosenia A, et al. Darts: deceiving autonomous cars with toxic signs [Z/OL]. arXiv Preprint, arXiv:1802.06430, 2018.

[10] Sitawarin C, Bhagoji AN, Mosenia A, et al. Rogue signs: deceiving traffic sign recognition with malicious ads and logos [Z/OL]. arXiv Preprint, arXiv:1801.02780, 2018.

[11] Liu Y, Chen X, Liu C, et al. Delving into transferable adversarial examples and black-box attacks [C] // International Conference on Learning Representations, 2017.

[12] Brendel W, Rauber J, Bethge M. Decision-based adversarial attacks: reliable attacks against black-box machine learning models [C] // International Conference on Learning Representations, 2018.

[13] 高卫. 激光雷达干扰效果评估方法研究 [J]. 光子学报, 2007, 36(8): 1400-1404.

[14] Cao YL, Xiao CW, Yang DW, et al. Adversarial objects against LiDAR-based autonomous driving systems [Z/OL]. arXiv Preprint, arXiv:1907.05418, 2019.

[15] BarHillel A, Hanukaev D, Levi D. Fusing visual and range imaging for object class recognition [C] // 2011 International Conference on Computer Vision, 2011: 65-72.

[16] Luo Y, Boix X, Roig G. Foveation-based mechanisms alleviate adversarial examples [Z/OL]. arXiv Preprint, arXiv:1511.06292, 2015.

[17] 于国权. 激光角度欺骗干扰半实物仿真系统研究 [D]. 长春: 中国科学院研究生院(长春光学精密机械与物理研究所), 2013.

[18] Cao YL, Xiao CW, 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.

[19] Augere B, Besson B, Fleury D, et al. 1.5 μm lidar anemometer for true air speed, angle of sideslip, and angle of attack measurements on-board Piaggio P180 aircraft [J]. Measurement Science and Technology, 2016, 27(5): 054002.

[20] Westergaard CH. Optical angle of attack detector based on light detection and ranging (LIDAR) for control of an aerodynamic surface: The United States, 8915709 [P]. 2014.

[21] 刘志敬. 光电干扰技术及干扰效果评估研究 [D]. 长春: 长春理工大学, 2012.

[22] 陈蕾, 陈松灿. 矩阵补全模型及其算法研究综述 [J]. 软件学报, 2017, 28(6): 1547-1564.

[23] 肖甫, 沙朝恒, 陈蕾, 等. 基于范数正则化矩阵补全的无线传感网定位算法 [J]. 计算机研究与发展, 2016, 53(1): 216-227.

[24] Huang ZT, Yu YK, Xu JW, et al. PF-Net: point fractal network for 3D point cloud completion [Z/OL]. arXiv Preprint, arXiv: Computer Vision and Pattern Recognition, 2020.

[25] Rusu RB, Cousins S. 3D is here: point cloud library (PCL) [C] // International Conference on Robotics and Automation, 2011: 1-4.

[26] 丁振东, 王娟锋, 朱祺. 实现激光角度欺骗干扰的几个条件 [J]. 电子科技, 2012, 25(10): 122-124.

[27] 宁成达, 田春艳, 蓝志环, 等. 激光致僵武器攻击目标可行性分析 [J]. 科技情报开发与经济, 2009, 19(28): 163-165.