本文已发表于《机器人：科学与系统会议论文集》（RSS）2023。请按以下格式引用本文：P. Antonante、S. Veer、K. Leung、X.Weng、L. Carlone和M. Pavone，“面向任务的自动驾驶车辆感知失效风险估计”，机器人：科学与系统会议论文集（*RSS*），2023。

面向任务的自动驾驶车辆感知失效风险估计

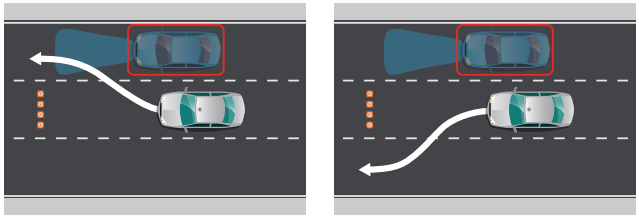
Pasquale Antonante*、Sushant Veer†、Karen Leung†‡、Xinshuo Weng†、Luca Carlone*和Marco Pavone†§
* 麻省理工学院
† NVIDIA Research
‡ 华盛顿大学§ 斯坦福大学

arXiv:2305.01870v1 [cs.RO] 3 May 2023

摘要——安全性与性能是实现自动驾驶的关键因素：一方面，我们希望自动驾驶车辆（AVs）安全；另一方面，其性能（例如，乘坐舒适性或行驶进度）也至关重要，关系到人们是否愿意采用自动驾驶。为了在安全性与性能之间取得微妙平衡，自动驾驶汽车需要规避风险，但不能完全避免风险。为实现兼顾安全与性能的驾驶，本文开发了一种任务感知 风险估计器，用于评估感知失效对自动驾驶车辆运动规划造成的风险。如果感知失效与运动规划的安全性无关，那么无论该感知失效多么严重，我们的任务感知风险估计器都将其视为低风险；反之，如果看似无关紧要的感知失效严重影响了运动规划，估计器就会将其视为高风险。本文提出一种任务感知风险估计器，用于判断是否需要触发安全机动。为估计任务感知风险，我们首先利用感知监测器检测到的感知失效，合成一个关于车辆周围环境的替代性合理模型。随后，将感知失效导致的风险形式化为自动驾驶车辆在感知到的场景与替代性合理场景下的运动规划之间的“相对”风险。我们采用一种名为联结函数的统计工具，对分布之间的尾部依赖进行建模，以估计该风险。联结函数的理论性质使我们能够计算风险的概率近似正确（PAC）估计。我们使用 NuPlan 对任务感知风险估计器进行评估，并将其与已有基线方法进行比较。结果表明，所提出的风险估计器取得了最佳 $F1$-分数（达到最佳基线方法的两倍），并且在召回率与精确率之间实现了良好平衡，即在安全性与性能之间实现了良好平衡。

一、引言

尽管机器人学和自主系统发展迅速，自动驾驶车辆（AVs）的感知模块仍会遇到大量故障模式（例如物体错误分类或误检测、幽灵障碍物、分布外（OOD）物体等），这些问题可能危及乘客、其他驾驶员和行人的安全。因此，开发此类感知故障检测器的问题近期受到越来越多的关注 [4, 44, 35, 10]。然而，这些故障发生得足够频繁，以至于每次检测到此类故障都采取备用安全机动，会严重损害自动驾驶车辆的性能。本文致力于开发一种任务感知型感知监测器，仅当感知故障对自动驾驶车辆的运动规划构成重大风险时才触发，从而在确保安全的同时保持高效驾驶；图 1 展示了一个示例，凸显开发任务感知型感知监测器的重要性：该监测器聚焦于与任务相关的感知故障。



（a）任务相关故障　　(b) 任务无关故障

图 1. **任务感知型感知故障检测示意图**。白色车辆为自车，蓝色车辆为外部（非自车）车辆。在此示例中，自车的感知系统未能检测到该非自车车辆。随后，图 1(a) 展示了与任务相关的障碍物漏检，因为由于误检测，自车的运动规划很可能会与非自车车辆发生碰撞。图 1(b) 展示了与任务无关的车辆漏检，因为无论是否发生感知失效，自车的运动规划都不会导致其与非自车车辆发生碰撞。

我们构想了一种任务感知型感知监测器，它包含三个主要组成部分，如图 2 所示。首先，感知故障检测与识别模块识别感知故障，并隔离相关模块和故障模式。其次，合理场景生成器利用故障识别所提供的感知失效模式知识，构建对自动驾驶车辆周围环境的合理替代模型的概率性（可能为多模态）描述，以支持真实世界场景。[1]最后，任务感知风险估计器评估感知失效导致自动驾驶车辆运动规划风险增加的程度。近期关于感知故障检测的研究相当丰富 [4, 40, 35, 10], , 也有一些关于合理场景生成的研究 [11, 24, 3],, 而关于任务感知风险估计的研究则少得多。事实上，本文的主要研究内容正是任务感知风险估计器的开发。

本文提出的任务感知风险估计器，将感知场景中的自动驾驶车辆运动规划所面临的风险，与生成的合理场景中的风险进行比较。在这两种场景（感知场景

这项工作是在 Pasquale 于 NVIDIA Research 实习期间完成的。

[1]合理的自动驾驶车辆周围环境的概率描述可能具有高度随机性和多模态特征。在合理场景中进行规划并不切实际，而且可能不利于制定良好的规划方案；然而，我们仍然可以利用本文提出的方法，借此估计感知故障给自动驾驶车辆带来的风险。

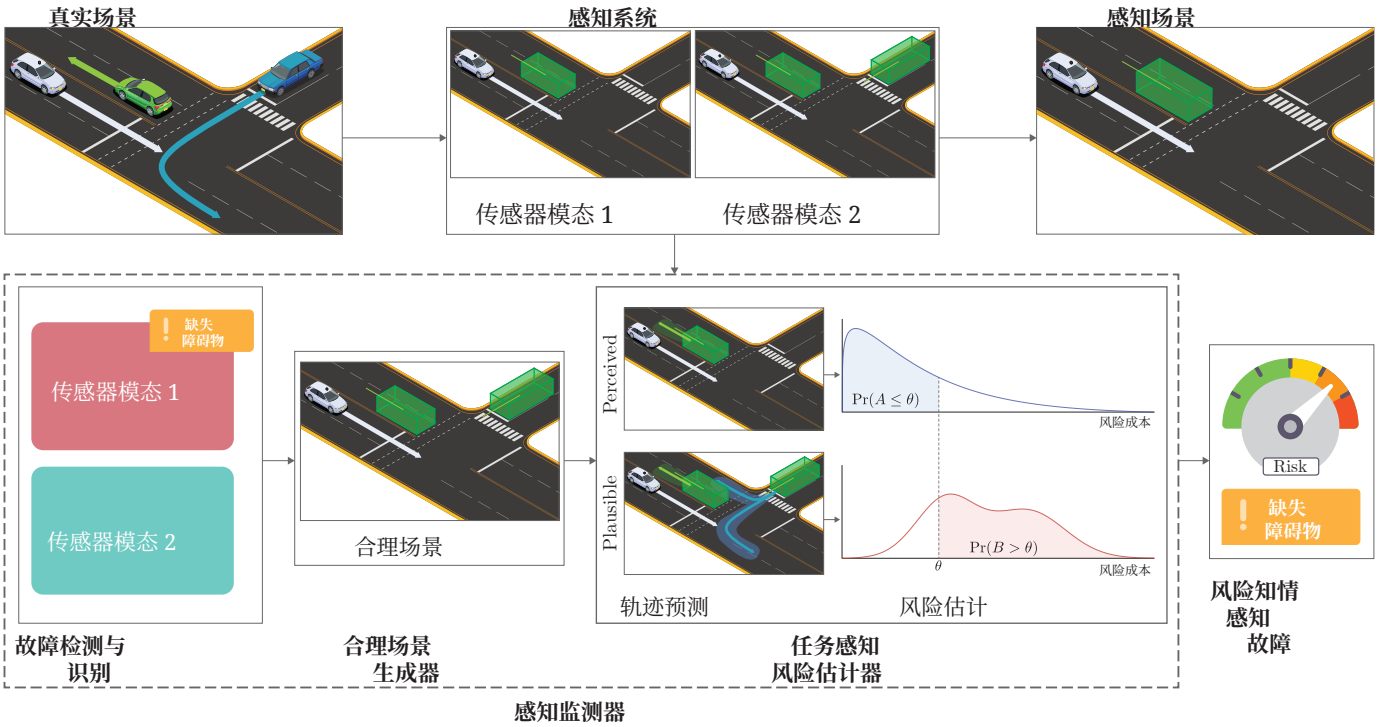


图 2. **任务感知型感知监测器概览。** 该场景包含自车（白色车辆）和两个非自车交通参与者（绿色和蓝色车辆）。顶行展示了感知系统未能检测到障碍物（蓝色车辆）的场景：感知系统使用的两种传感器模态之一无法检测到该障碍物（顶部中央子图），从而导致感知输出中出现障碍物缺失故障（顶部右侧子图）。底行描绘了所提出的任务感知型感知监测器。故障检测与识别模块检测到传感器 1 正在发生故障（例如使用时空信息）。合理场景生成器利用当前故障信息，从感知场景生成合理场景。最后，任务感知风险估计器计算与该故障相关的风险。底行场景中阴影（绿色和蓝色）区域表示由非自车轨迹预测模块计算得到的轨迹不确定性。可能的轨迹为每个场景生成风险代价分布，用于估计与感知失效相关的风险。如果合理场景中的风险显著高于感知场景中的风险，我们将该故障检测为与任务相关。我们的检测器使用一种称为联结函数的统计工具来估计两个代价分布之间的尾部依赖。

和合理场景）中，自动驾驶车辆运动规划所面临的风险均表示为某种风险度量的概率分布，例如，碰撞时间。我们提出相对场景风险（RSR）的概念，用以衡量感知场景风险不高时，合理场景风险较高的概率。为了通过实证估计 RSR，我们采用名为联结函数 [38]，的统计工具来建模分布间的尾部依赖，并给出 RSR 估计值的概率近似正确（PAC）界。最后，我们提出一种基于 RSR PAC 界的检测算法，该算法在面对高风险的任务相关故障时，能够以较高概率触发警报。

贡献声明。 本文的贡献如下：(i) 我们形式化定义了相对场景风险 （RSR），这是我们进行任务感知风险估计的基础；(ii) 我们开发了一种算法，通过利用联结函数在运行时估计 RSR，并为估计结果的正确性提供概率保证；(iii) 最后，我们通过在 NuPlan [18] 中创建的包含 100 个逼真感知失效场景的数据集上，将我们的方法与先前方法进行比较，展示了我们框架的有效性。结果表明，我们的风险估计器取得了最高的 F1 分数，精确率和召回率之间表现出良好的平衡，并能提前预测碰撞，从而留出充足时间采取缓解措施。代码可在 https://github.com/NVlabs/persevere 获取。

II. 相关工作

我们讨论了感知监测方案三个阶段的相关现有研究，即 感知故障检测、合理场景生成和任务相关风险估计。

感知故障检测与识别。自动驾驶车辆依靠车载感知系统来提供态势感知，并为车载决策与控制系统提供信息。感知系统的可靠性对于自动驾驶汽车的安全运行至关重要。尽管理想情况下，感知系统在任何条件下都不应发生故障，但要保证这一点并不容易 [10]；因此，在运行时检测和识别感知系统中的故障日益受到关注。 [3]研究了故障检测与识别问题，作者提出了一种用于在线监控自动驾驶车辆感知系统的系统级框架。除了检测故障外， [3] 中的框架还会在运行时从预先已知的故障列表中识别系统正在经历的故障模式。文献中的其他方法包括利用运动预测中的时空信息评估三维目标检测 [47], 、使用定时质量时序逻辑（TQTL）推理感知算法的理想时空属性 [5, 6], , 或通过设置逻辑约束模型断言来检测异常 [28]。

由于感知系统由多个模块组成，针对特定子模块的故障检测也受到关注。以往的研究主要关注目标检测 [35],、语义分割 [9, 39], 、定位 [25, 19],、分布外（OOD）检测 [42, 21], ，或高精地图的变化 [30]。上述研究均侧重于检测和识别感知系统中的故障，但并未评估这些故障对自动驾驶车辆运动规划的影响。

合理场景生成。尽管关于显式合理场景生成的先前研究有限，但文献中的许多工作会推理合理的替代场景，以检测或避免故障。事实上，故障检测方法常利用不同传感器模态之间的时空不一致性，其中每种模态都隐式提出一个合理场景。You 等人 [47]利用此前多个三维场景的历史信息来预测一个合理场景，再将其与感知场景进行比较，以检测错误。类似地， [33] 关联相机图像和激光雷达点云，以检测缺失的（或虚假的）障碍物。除了障碍物检测，先前的研究还探讨了定位错误。Li 等人 [32] 利用粒子滤波器保留自动驾驶车辆多个可能的位置。此外，关于遮挡条件下规划的文献包含基于模型的方法 [17, 29, 48]和基于学习的方法 [11, 24, 20]。这些工作会扩展场景，将可能缺失（被遮挡）的障碍物纳入其中。每当故障检测未能提供关于合理场景的足够信息时，也可以使用相同技术生成合理场景。

风险评估。文献中有多种风险评估技术 [12, 46]。一种“衡量”风险的方法是监测运动规划的成本恶化情况，如 [15, 14, 16] 中所做的那样。另一种方法是使用关键性指标，例如碰撞时间（TTC），它会根据两个交通参与者的速度和朝向保持不变这一假设，计算二者发生碰撞前的时间。文献中的关键性指标种类繁多，感兴趣的读者可参阅 [46], 了解全面的综述；然而，通常而言，这些指标只评估某个场景是否危险，同时假设从感知中推断出的场景是正确的。近期有一种方法使用神经网络，利用带标签的数据对风险等级进行分类 [1]。然而，基于学习的方法缺乏分布外鲁棒性，通常也较难解释，且缺乏保证。其他关于感知相关风险评估的研究，例如 [7], 提出的目标检测系统风险排序召回率，以及 [8], 开发的基于感知不确定性的风险包络，都没有对其他智能体的未来动作进行推理。Topan 等人[44] 则通过 Hamilton-Jacobi（HJ）可达性分析 [36], 计算“不可避免”碰撞集合（ICS），并在另一智能体即将进入 ICS 时将情况标记为不安全，从而考虑感知故障情况下其他智能体的未来动作 [31, 22, 23]。然而，[44]假设最坏情况场景必然发生，即 自动驾驶车辆和其他道路交通参与者会试图彼此碰撞，并且未考虑多个智能体之间的交互，导致过度-保守性。本文中，我们通过利用轨迹预测网络 [41], 在考虑其他智能体未来动作的同时，估计感知失效情况下自动驾驶车辆（AV）的运动规划所面临的风险。此外，我们的方法还提供 PAC 界，并具备实时运行能力。

III. 任务感知型感知监测器：概述与风险估计问题表述

本节概述任务感知型感知监测器的组成模块，该监测器由三个部分构成：感知失效检测与识别、可信场景生成器以及任务感知风险估计器。此外，本节还对任务感知风险估计问题进行了形式化表述，这是本文的主要研究重点。在本文其余部分中，我们将自动驾驶车辆称为自车，而将任何其他智能体称为非自车智能体。

感知失效检测与识别。我们假设可以使用感知失效检测与识别模块，例如 [3], 中介绍的算法，该模块能够识别感知系统当前经历的故障模式集合。感知系统由一组模块组成，每个模块负责一项特定任务，例如目标检测、定位等。每个模块都可能出现有限种故障模式。感知失效检测与识别模块会计算出故障状态向量 $\boldsymbol{f}$，其中包含有关当前故障的相关信息，即当前故障模式集合及相应的感知诊断信息（例如中间检测结果、原始传感器数据等）。

合理场景生成器。 在评估感知失效对自动驾驶车辆运动规划造成的风险之前，我们需要了解自动驾驶车辆运行所处的实际场景。为此，我们使用所谓的合理场景生成器，构建对周围场景的新估计。设 $\boldsymbol{x}_t \in \mathbb{X} \subseteq \mathbb{R}^n$ 为时间 t 的世界状态估计，该估计由感知模块提供（即“感知场景”）。我们假设世界状态由自车状态 $\boldsymbol{x}^{\mathrm{e}}$、非自车智能体状态 $\boldsymbol{x}^{\mathrm{ne}}$ 以及地图属性 $\boldsymbol{x}^{\mathrm{m}}$ 构成，例如车道线、停止标志、交通信号灯等。给定感知模块提供的感知世界状态 $\boldsymbol{x}_t$ ，以及来自感知失效检测与识别的当前感知失效模式信息 $\boldsymbol{f}$ ，合理场景生成器以概率分布 $\zeta(\hat{\boldsymbol{x}}_t|\boldsymbol{x}_{0:t}, \boldsymbol{f})$ 的形式，返回时间 t 时合理世界状态 $\hat{\boldsymbol{x}}_t \in \mathbb{X}$ 上的替代合理场景；我们要求合理场景生成器支持实际世界状态。

尽管有一些方法可用于场景生成， [3, 47, 11, 24, 20], 我们在此采用的方法基于 [3]中的感知故障识别方法。简而言之， [3] 中的方法会检测中间感知结果之间的不一致性，并推断可能导致这些不一致性的故障集合。[2] 例如，考虑这样一种情况：基于雷达的检测

[2]该方法 [3] 还可以整合更复杂的测试，包括逻辑公式、数学正确性证明和先验界。

模块检测到自车前方有一个障碍物，但基于摄像头的检测模块却漏检了同一障碍物：在这种情况下，感知系统可能会优先采用摄像头检测结果，并将雷达检测结果判定为误检而予以舍弃；因此，感知场景中自车前方将没有障碍物。然而， [3]中的方法可以检测出雷达与摄像头之间的不一致性，并结合时空考量，推断雷达检测结果实际上是正确的。随后，合理场景生成器便可利用故障信息，生成一个包含雷达检测到的障碍物的替代合理场景。一般而言，替代场景可能并非唯一。例如，由于基于雷达的检测模块可能只能检测缺失障碍物的位置和速度，而无法识别其类别（如汽车、行人等），因此，此类检测到的故障可能会产生一个覆盖可能场景的概率分布 $\zeta(\hat{\boldsymbol{x}}_t|\boldsymbol{x}_{0:t},\ \boldsymbol{f})$（例如，对缺失障碍物类别采用均匀分布）。类似地，该分布 $\zeta(\hat{\boldsymbol{x}}_t|\boldsymbol{x}_{0:t},\ \boldsymbol{f})$也可以对雷达位置和速度测量中的不确定性进行建模。假设感知流水线中至少有一个模块计算出了正确结果，那么生成场景的分布就会涵盖实际场景。[3] 本文不假设合理场景生成器的具体实现方式，该实现方式可能取决于感知系统架构；我们关注的重点是风险评估。我们只要求可能场景分布涵盖实际场景，而对于最常见的感知失效模式，已有研究表明这一点是可以实现的 [3, 47, 11, 33, 17]（如第二节所述）。

相对场景风险。我们关注的是，在生成的合理场景 $\zeta(\hat{\boldsymbol{x}}_t|\boldsymbol{x}_{0:t},\ \boldsymbol{f})$ 中，自车的运动规划所面临的风险相较于感知场景 $\boldsymbol{x}_{0:t}$ 增加了多少。设 $\boldsymbol{x}^{\mathrm{e}}_{t:t+T}$ 为规划模块针对时间范围 T 生成的自车运动规划。我们假设存在一个轨迹预测模块，该模块能够提供未来世界状态轨迹的分布 $\psi(\boldsymbol{x}_{t:t+T}|\boldsymbol{x}_{0:t},\ \boldsymbol{x}^{\mathrm{e}}_{t:t+T})$，该分布以世界状态历史和自车运动规划为条件。轨迹预测器ψ 能够推理智能体交互，并提供考虑多种智能体意图的多模态预测；在我们的实验中，我们使用 Trajectron++ [41] ，这是一种满足上述标准的最先进的轨迹预测模型。我们所描述的方法不依赖于规划模块和预测模块的具体选择。设 $c:\ \mathbb{X}\to\mathbb{R}^+$ 为一个代价函数，其值越高，表示对自车而言场景的风险越高。此类函数的示例包括自车与最近的非自车智能体之间的距离，或碰撞时间指标 [46] 的替代指标（使得数值越高意味着时间越短）。未来世界状态 $\boldsymbol{x}_{t:t+T}$ 上的分布ψ 会在未来的每个时间步上诱导出一系列关于预测代价 c 的单变量分布 $\{\phi_{t+\tau}(c_{t+\tau}|\boldsymbol{x}_{0:t},\ \boldsymbol{x}^{\mathrm{e}}_{t:t+T})\}_{\tau=1}^{T}$。在本文的其余部分，我们将使用预测代价分布。

[3]无论感知架构如何，跨帧时空信息都能提供有用的诊断信息，因此这种方法也可以推广到早期融合和中期融合感知系统。

$\phi_{t+\tau}$ 针对特定的 τ 和特定的运动规划$\boldsymbol{x}^{\mathrm{e}}_{t:t+T}$。为使记号简洁，我们省略 ϕ 对 $t+\tau$和 $\boldsymbol{x}^{\mathrm{e}}_{t:t+T}$的显式依赖，将预测代价分布表示为 $\phi(c|\boldsymbol{x}_{0:t})$。类似地，合理场景分布 $\zeta(\hat{\boldsymbol{x}}_t|\boldsymbol{x}_{0:t},\boldsymbol{f})$在合理世界状态$\hat{\boldsymbol{x}}_t$上诱导出代价分布 $\phi(c|\boldsymbol{x}_{0:t},\ \boldsymbol{f})$】【。

现在，我们可以通过以下定义对任务感知风险概念进行形式化。

定义1（相对场景风险（RSR））。设 $\boldsymbol{x}_{0:t}$ 为感知场景的世界状态历史，设$\zeta(\hat{\boldsymbol{x}}_t|\boldsymbol{x}_{0:t},\boldsymbol{f})$ 为由感知故障 $\boldsymbol{f}$所生成的可能场景的分布，这些故障由感知失效检测与识别模块检测得到。设$\phi_A := \phi(c|\boldsymbol{x}_{0:t})$ 为感知场景的成本分布，$\phi_B := \phi(c|\boldsymbol{x}_{0:t},\boldsymbol{f})$ 为可能场景的成本分布。设 $\theta\in\mathbb{R}^+$ 为规划器希望保持在其以下的成本阈值。于是，可能场景与感知场景之间的相对场景风险（*RSR*）定义如下：

$$\hat{\mathcal{R}}:\theta\mapsto\Pr_{A\sim\phi_A,B\sim\phi_B}(B>\theta\mid A\le\theta). \tag{1}$$

对于给定的 θ，RSR 越高，可能场景的成本分布 ϕ_B 就越向较高成本方向偏移，如图 2 所示。因此， $\mathcal{R}$ 的值越大，意味着可能场景相比感知场景具有更高的风险。请注意，一般而言， ϕ_A 和 ϕ_B 并不相互独立，因为底层场景在很大程度上是相同的。

θ 的选择定义了期望的安全阈值：例如，如果成本是智能体之间的距离，那么 θ 可以是自车与最近智能体之间可接受的最小距离。 θ 的选择可能取决于场景（例如，在高速公路上行驶与处于交通拥堵中时，最小距离可能不同）。为克服这种依赖关系，我们取 θ 来捕获感知场景成本分布（ϕ_A）中的大部分概率质量。具体而言，设 Φ_A 为 ϕ_A 的边缘累积分布函数（ CDF ），设 Φ_B 为 ϕ_B 的 CDF。回顾一下，CDF Φ 的广义逆（此处记为 Φ^{-1}）定义为：

$$\Phi^{-1}(p):=\inf\{c\in\mathbb{R}\ :\ \Phi(c)\ge p\}. \tag{2}$$

然后，我们选择 $\theta=\Phi_A^{-1}(p)$。这等价于将 θ 设为感知场景中最常见情形下风险成本的最大值。我们将 p 称为风险厌恶参数，因为它表示自车智能体愿意在其运动规划中接受的风险程度。

现在我们可以定义 p分位数相对场景风险。

定义 2 (p-分位数相对场景风险)。设 $\mathcal{R}$ 为定义 *1* 中的相对场景风险。设 $p\in(0,\ 1)$ 为上述描述的风险厌恶参数。那么， p 分位数相对场景风险（*p-RSR*）定义如下：

$$\mathcal{R}:p\mapsto\hat{\mathcal{R}}\circ\Phi_A^{-1}(p). \tag{3}$$

请注意，上述定义只是用风险厌恶参数 p重新表述定义 1。

问题陈述。给定感知到的世界状态历史 $\boldsymbol{x}_{0:t}$、感知模块故障模式 $\boldsymbol{f}$ 和风险厌恶 p，我们希望估计定义 2 中的 p-分位数相对场景风险 $\mathcal{R}(p)$。值得指出的是，这是一个颇具挑战性的问题，因为分布 ϕ_A 和 ϕ_B并不相互独立，而且我们没有它们或其累积分布函数 Φ_A、 Φ_B的显式解析表示；因此，我们无法通过解析方法计算 $\Phi_A^{-1}(p)$ 或 $\mathcal{R}(p)$。不过，我们可以从这些分布中独立地抽取样本。具体而言，我们从这些分布中抽取样本，并确保这些样本相互独立（任何新样本都不依赖于先前样本）且服从相同分布（底层场景和智能体的行为固定不变）。下一节中，我们将介绍一种利用这些样本估计 $\mathcal{R}(p)$ 的方法。

IV. 任务感知风险估计

在本节中，我们使用联结函数 [26], 这一用于对分布之间的尾部依赖进行建模的统计工具，提出一种算法来估计定义 2 中定义的相对场景风险。

A. Copula 简介

为了对两个单变量分布 ϕ_A 和 ϕ_B之间的依赖关系进行建模，我们采用联结函数 的概念（更详尽的介绍请参见 [26, 第 1]章）。Copula 是用于对多个随机变量之间的依赖关系进行建模的工具；“copula”这一名称旨在强调联结函数将联合分布函数与其单变量边缘分布连接起来的方式。我们将在下列定义中对此进行精确的数学描述。

定义 3 (联结函数 [38])。*A* d维联结函数 C :$[0,1]^d \to [0,1]$ 是在 d维单位立方体 $[0,1]^d$ 上定义的函数，并满足以下条件：

1) $C(u_1,\dots,u_{i-1},0,u_{i+1},\dots,u_d) = 0$ *for any* u_i, $i \in \{1,\dots,d\}$,
2) $C(1,\dots,1,u,1,\dots,1) = u$ *for any* $u \in [0,1]$ *in any position, and*
3) C *is d-non-decreasing.*[4]

这三个性质确保联结函数具有类似联合分布函数的行为。为了直观理解，设每个u_i 都是值域为 $[0,1]$的概率。第一个条件表示，如果与 u_i 相关联的事件发生概率为零，那么无论其他事件发生的概率如何，所有事件同时发生的联合概率都为零。反过来，如果除一个事件外，其他所有事件都必然发生，那么联合事件的概率就等于那个不确定事件发生的概率。最后一个条件要求联结函数在每个分量上都是非递减的。

下面给出的 Sklar 定理 [43], 为应用联结函数提供了理论基础，并说明了联结函数存在（且唯一）的条件。请注意，在本文中，我们只要求联结函数存在，但为求完整，下面也一并列出唯一性条件。

定理1 (Sklar 定理 [43])。 设 $\Phi(x_1,\dots,n_d)$ 为一个联合分布函数，且 Φ_i, $i = 1,\dots,d$ 为其边缘分布。那么，存在一个联结函数 C : $[0,1]^d \to [0,1]$，使得对于所有属于 $[-\infty,+\infty]$ 的 $x_1,\dots,x_d$

$$\Phi(x_1,\dots,x_d) = C(\Phi_1(x_1),\dots,\Phi_d(x_d)). \tag{4}$$

此外，如果边缘分布是连续的，那么 C 是唯一的；否则， C 在值域 $\Phi_1 \times \cdots \times$值域 Φ_d 上唯一确定，其中值域 Φ_i 表示 Φ_i 的值域（像）。

Sklar 定理强调了联结函数在研究多元分布中的重要性。该定理首先表明，所有多元分布都可以用联结函数表示；其次表明，可以利用联结函数从单变量分布函数构造多元分布函数。后一点对我们尤其重要，因为正如我们在上一节中所指出的，我们无法从联合分布 $\Pr(A,B)$ 中进行采样，但可以从边缘分布 $\Pr(A)$ 和 $\Pr(B)$ 中进行采样。

B. 使用 Copula 估计 p-相对场景风险

设 A、 B 是分别从 ϕ_A 和 ϕ_B 中抽取的随机变量，其累积分布函数分别为 Φ_A 和 Φ_B（记号在定义1中介绍）；简要回顾一下， ϕ_A 是感知场景的代价分布， ϕ_B 是合理场景的代价分布。暂且假设我们能够估计描述 ϕ_A 和 ϕ_B 之间关系的联结函数。由于联结函数 $C(A,B)$ 包含了关于 (A,B) 之间依赖结构的信息，因此我们可以利用它来度量这两个分布之间的尾部依赖关系。于是，根据条件概率的定义以及定理1中的公式（4），我们可以将公式（1）中的 p-相对场景风险 表示如下：[5]

$$\begin{aligned}
\mathcal{R}(p) &= \Pr(B > \Phi_A^{-1}(p) \mid A \le \Phi_A^{-1}(p)) \\
&= 1 - \Pr(B \le \Phi_A^{-1}(p) \mid A \le \Phi_A^{-1}(p)) \\
&= 1 - \frac{\Pr(A \le \Phi_A^{-1}(p), B \le \Phi_A^{-1}(p))}{\Pr(A \le \Phi_A^{-1}(p))} \\
&= 1 - \frac{C(\Phi_A \circ \Phi_A^{-1}(p), \Phi_B \circ \Phi_A^{-1}(p))}{\Phi_A \circ \Phi_A^{-1}(p)} \\
&= 1 - \frac{C(p, \Phi_B \circ \Phi_A^{-1}(p))}{p}.
\end{aligned} \tag{5}$$

遗憾的是，我们无法获得两个累积分布函数 Φ_A 和 Φ_B 以及联结函数 C的显式表达式，因此，$\mathcal{R}$ 无法通过解析方法计算。接下来，我们将分别利用来自 ϕ_A 和 ϕ_B的 n 个独立同分布样本，构造累积分布函数 Φ_A 和 Φ_B的经验估计 $\Phi_A^{(n)}$ 和 $\Phi_B^{(n)}$ ，从而对 $\mathcal{R}$ 给出可证明的界。

[4] $[0,1]^d$ B $= \prod_{i=1}^d [x_i, y_i]$ $\subseteq$也就是说, 对于每个超矩形，其 C-体积均为非负数：() $= \int_B dC\ u$ $\sum_{z \in \prod_{i=1}^d \{x_i, y+i\}} -1^{\ N(\boldsymbol{z})} C\ \boldsymbol{z} \ge 0$ $N\ \boldsymbol{z} = \{k\ \ z_k = x_k\}$ ()(), 其中 ()#:

[5]为简洁起见，我们省略概率符号 Pr 下随机变量 A 和 B 所抽取自的分布。

定理 2 (关于 p-相对场景风险的 PAC 界)。 设 $\{A_i\}_{i=1}^n$ 和 $\{B_i\}_{i=1}^n$ 分别是来自累积分布函数 Φ_A 和 Φ_B的 n 独立同分布样本。设

$$\Phi_A^{(n)}(A) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}[A_i \le A], \quad \Phi_B^{(n)}(B) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}[B_i \le B]$$

Φ_A 和 Φ_B的经验估计值，分别记为.....。风险厌恶参数 $p \in (0, 1)$ 如第三节所述，且令 $\alpha \in (0, 1)$。那么，以至少 $1-\alpha$ 的概率:

$$1 - \frac{\min\{p, \bar{v}(p,\alpha,n)\}}{p} \le \mathcal{R}(p) \le 1 - \frac{\max\{p + \underline{v}(p,\alpha,n) - 1, 0\}}{p},$$

其中

$$\underline{v}(p,\alpha,n) = \Phi_B^{(n)} \circ \left[\Phi_A^{(n)} + \epsilon(\alpha,n)\right]^{-1}(p) - \epsilon(\alpha,n)$$

$$\bar{v}(p,\alpha,n) = \Phi_B^{(n)} \circ \left[\Phi_A^{(n)} - \epsilon(\alpha,n)\right]^{-1}(p) + \epsilon(\alpha,n)$$

并且 $\epsilon(\alpha, n) = \sqrt{\ln(2/\alpha)/(2n)}$。

证明：见附录 *A*。 ■

这两个界都是 Sharp 的，即它们都可以取到。具体而言，当 ϕ_A 和 ϕ_B完全正相关时，下界可以取到；这里的完全正相关是指 B 几乎必然是 A的严格递增函数。相反，当 ϕ_A 和 ϕ_B 完全负相关时，上界可以取到；这意味着 B 几乎必然是 A的严格递减函数。

定理 2 中的 PAC界易于计算，并使我们能够在运行时估计 p-相对场景风险 $\mathcal{R}(p)$。特别是，独立同分布样本这一假设并不苛刻：正如我们在第三节中指出的，我们的问题表述允许使用独立同分布样本。此外，我们不假设特定的联结函数，只假设其存在性，而这已由 Sklar 定理证明。

C. 触发安全机动

根据定理 2 中给出的结果，我们可以通过概率来衡量:与感知场景相比，合理场景是否会使自车面临不必要的风险。然而，如果概率较低，触发安全机动来降低风险可能会损害自车的性能。接下来，我们设计一种检测算法（算法 1），用于检测系统是否可能正处于高风险情形，并可据此触发安全机动。

考虑一个表示高风险情形的风险阈值 $\gamma \in (0,1)$。如果定理 2 中 $\mathcal{R}(p)$ 的下界高于γ，则意味着当前场景以至少 $1-\alpha$ 的概率确实对应于一个高风险情形（从 p- 相对场景风险（RSR）的意义上来说）。在这种情况下，算法 1 会检测到一个任务相关故障（第 5 行），从而可用于触发安全机动。这一推理可以轻松扩展到考虑多个关键性等级对应的多个阈值，每个阈值分别关联不同的缓解策略或不同的驾驶场景（例如，高速公路、城市驾驶、接送上下客等）。

算法 1 可总结如下：在识别出故障并生成合理场景后，该算法

算法 1 p-相对场景风险检测算法

需要： 状态 $\boldsymbol{x}_{0:t}$、故障 $\boldsymbol{f}$、代价度量 c、风险厌恶 p、置信水平 $1-\alpha$ 和风险阈值 γ。

确保： 如果是关键场景则为 TRUE，否则为 FALSE。

1: $\{A_i\}_{i=1}^n \sim \phi(c|\boldsymbol{x}_{0:t})$ $\{B_i\}_{i=1}^n \sim \phi(c|\boldsymbol{x}_{0:t}, \boldsymbol{f})$
2: $\Phi_A^{(n)}(A) \leftarrow 1/n \sum_{i=1}^n \mathbb{1}[A_i \le a]$
3: $\Phi_B^{(n)}(B) \leftarrow 1/n \sum_{i=1}^n \mathbb{1}[B_i \le b]$
4: **如果** $\min\{p, \bar{v}(p, \alpha, n)\} < p(1-\gamma)$ **则**
5: **返回** TRUE
6: **否则**
7: **返回** FALSE
8: **结束 if**

对两个场景进行采样（第 1 行），并估计经验累积分布函数（第 2–3 行）。如果定理 2 中的下界高于风险阈值 γ，则返回 TRUE，否则返回 FALSE（第 4 行）。算法步骤如图 3 所示。

该算法有四个参数：风险厌恶度 $p \in (0, 1)$、风险阈值 $\gamma \in (0, 1)$、样本数量$n \in \mathbb{N}$以及置信水平 $1-\alpha \in (0, 1)$。风险厌恶度 p 根据感知场景风险分布的分位数来衡量自车的风险容忍度。数值越高表示风险容忍度越高，即，预期自车会在风险更高的情况下安全行驶。我们的风险度量 p-RSR 表示合理场景比感知场景风险更高的概率。如果该概率较大，即，高于风险阈值 γ，算法会将场景分类为高风险。为了估计 p-RSR，我们使用定理 2 中的 PAC 界，这需要选择一个期望的置信水平 $1-\alpha$和预测成本样本的数量 n。显然， n 应尽可能高，但这受到系统计算预算的限制。如果 p、 γ和 $1-\alpha$ 的值接近 1，算法将是不审慎的，即，会将大多数场景分类为低风险；另一方面，如果这些值较小，算法将过于审慎，即，会将大多数场景分类为高风险。在我们的实验中，我们发现， p、γ和 $1-\alpha$在 $[0.9, 0.99]$范围内的取值能够在审慎性与不审慎性之间实现良好平衡。

V. 实验结果

在本节中，我们将任务感知风险估计器的性能与其他各种基线方法进行比较。我们的实验在一台台式计算机上进行，该计算机配备英特尔 i9-10980XE 4.7 GHz CPU（36 个核心）和英伟达 GeForce RTX 3090 GPU。

A. 数据集

我们使用公开可用的 NuPlan 数据集 [18] 测试了所提出的方法。为了测试风险估计，我们在 NuPlan 场景中实现了一种故障注入机制。我们考虑了 4 类故障，即误检测、障碍物漏检、虚假障碍物和定位错误。每类故障还进一步划分为多个子类。

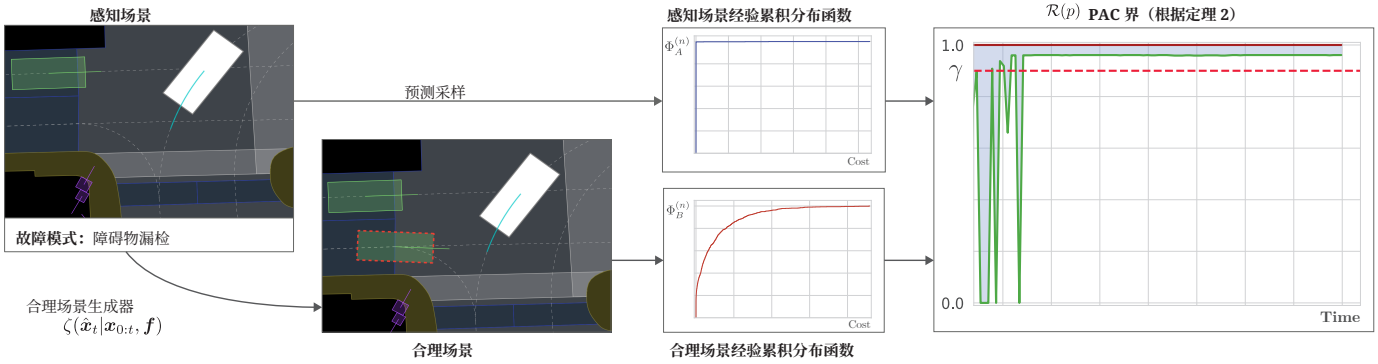


图 3. **算法 1 示意图**。感知场景中存在漏检障碍物故障，合理场景生成器对其进行处理并生成合理场景。我们分别对这两个场景进行采样，并估计成本的经验累积分布函数。感知场景的经验累积分布函数表明其风险较低，因为场景中唯一的车辆处于静止状态，正在让自车优先通行。然而，合理场景的风险较高，因为此时自车正处于与一辆运动车辆发生碰撞的路径上。我们使用这两个累积分布函数计算定理 2 中的 PAC 界。红色实线表示上界，绿色实线表示下界，红色虚线表示风险阈值 γ。只要下界高于 γ，算法就会将该场景标记为高风险。

误检测。误检测表示对车辆周围某个智能体/物体的估计出错。我们考虑：(i) 朝向：自车感知系统错误估计智能体的朝向；(ii) 尺寸：自车感知系统错误估计智能体的尺寸；(iii) 速度：自车感知系统错误估计智能体的速度（方向和/或大小）；最后，(iv) 交通信号灯：自车感知系统错误估计交通信号灯的状态。所有误检测子类型（交通信号灯除外）都受噪声影响，且噪声可能因场景而异。例如，朝向误检测可能使车辆航向角产生偏移，其偏移服从均值为 $\pi/6$ 、标准差为 0.1 的高斯分布。 **虚假障碍物**。自车感知系统错误地检测到一个并不存在的障碍物（即虚假障碍物）。虚假障碍物可以是：(i) 路径内，即位于自车轨迹上；或 (ii) 路径外，即不在自车轨迹上。**障碍物漏检**。自车感知系统未能检测到某个智能体；漏检的障碍物可以是：(i) 路径内，即位于自车轨迹上；或 (ii) 路径外，即不在自车轨迹上。 **定位错误**。自车感知系统未能在地图中定位自身。每种故障模式可以是：(i) 静态，即故障在整个场景持续时间（20 秒）内持续存在；或 (ii) 动态，即故障随时间随机出现/消失。在我们的实验中，动态故障模式出现的概率为 0.25，且消失前至少持续 1 秒。

我们手动设计了 100 个用于评估的现实场景，每个场景至少包含一种自动驾驶车辆中常见的故障模式；各场景中的故障分布详见表 I。这些现实场景包括：行人过马路时检测结果闪烁、对拥有优先通行权的车辆的朝向或速度估计错误、对有来车时的交通信号灯发生误检测等。有关所有场景的详细说明，请参阅网站 https://anonymous.4open.science/w/task-relevant-risk-estimation/。

B. 实施细节

我们使用 Python 实现了所提出方法的所有组件。如第 III 节所述，所提出的

表 I
场景

故障模式	子类型	静态	动态
虚假障碍物	路径内	5	5
	路径外	10	10
障碍物漏检	路径内	5	10
	路径外	10	10
误检测	朝向	10	
	速度	10	
	Size	5	
	交通信号灯	5	
定位错误		5	

方法与规划器无关。在我们的实验中，我们使用了 NuPlan-devkit [37] 提供的智能驾驶员模型（IDM）规划器 [45, 2] 。该规划器旨在朝目标行驶，沿车道前进，同时避免与自车前方的领航智能体发生碰撞。对于非自车预测模块，我们则使用了 Trajectron++ [41]。

为了创建高风险情形（例如碰撞），我们利用 NuPlan 提供的闭环能力。闭环仿真使自车和其他智能体能够偏离专家驾驶员在数据集中原本记录的行为。在我们的仿真中，每辆车也遵循 IDM 策略 [45, 2]。然而，由于 NuPlan 仿真器的限制，行人和自行车仍沿着数据集中原始记录的轨迹行进（开环）。

1）合理场景生成：实验评估的主要目标是关注与任务相关的风险估计。我们使用一种合理场景生成方法，该方法根据感知失效模式，通过向真实标注信息（即，智能体的速度、尺寸、朝向或位置）添加高斯噪声来构造合理场景。具体而言，我们在位置上添加均值为零、标准差为0。2m的高斯噪声，在航向角上添加0。1rad的高斯噪声，并在速度大小和方向上添加0。1m/s的高斯噪声。这种方法源于以下常见场景。考虑一个具有两种传感器模态（例如，摄像头和雷达）以及一个传感器融合算法的感知系统。不失一般性地假设，

传感器融合由于基于摄像头的检测错误而错误检测智能体的速度，而雷达未发生故障。一旦故障检测与识别模块识别出摄像头是错误感知场景的原因，合理场景生成器就可以使用卡尔曼滤波器跟踪雷达检测结果（未发生故障的传感器模态），从而提出车辆的合理速度。由于卡尔曼滤波器会生成不确定性的高斯估计，因此车辆的速度也是高斯分布的。这一逻辑可以扩展到其他故障模式（例如，车辆漏检），本文使用的合理场景生成器正是对其进行模拟。直接根据感知故障监视器和原始传感器数据生成合理场景超出了本文的范围；正如第六节所述，我们将在未来的工作中进一步探索这一主题。

2) 基线方法：我们将我们的方法与两个基线方法进行了对比，一个基于 Hamilton-Jacobi (HJ) 可达性分析 [44]，另一个基于碰撞概率。 **HJ可达性**。HJ 可达性的核心思想是计算一个目标状态集合，在固定的时间范围内，智能体会据此决定寻求碰撞还是避免碰撞。HJ 可达性中有两种智能体反应，即寻求碰撞（min）和避免碰撞（max）。 [44] 中提出的方法旨在弥补感知失效相关信息的缺失：它考虑了一种保守情况，即智能体和自车都处于首选动作是寻求碰撞的情形，也就是*min-min* 策略。对于场景中的每个智能体，HJ 可达性都会计算一个价值函数（在我们的案例中，该函数基于两个包围框之间的有符号距离），其零次水平集表示存在一组会导致碰撞的控制输入。价值函数会预先离线计算；在运行时，我们只需查表，因此这种方法极快。由于场景中有多个智能体，我们会为每个智能体计算价值函数，然后取整个场景中的最小值；如果该值小于零，我们就认为该场景属于高风险。**碰撞概率**。第二个基线使用轨迹预测模块计算与场景中任一智能体发生碰撞的概率。与我们提出的方法类似，该基线使用合理场景估算风险。它通过轨迹预测模块对感知场景和合理场景中的轨迹进行采样；如果合理场景中的碰撞概率高于感知场景，且前者超过阈值 γ，我们就认为该场景属于高风险。我们的方法与碰撞概率基线之间的关键区别在于，后者无法捕捉感知场景与合理场景之间的依赖关系。两个基线——碰撞概率和 HJ 可达性——都使用无法适应不同场景的绝对风险阈值。相比之下， p-RSR 衡量感知误差导致的感知场景与合理场景之间风险分布的变化，从而隐式地将风险阈值适应于具体场景。例如，假设一辆被正确检测到的车辆切入自车所在车道（高风险），但一名远处骑自行车者的速度（低风险）被低估了。两种基线方法均将其视为高风险场景，因为尽管感知错误本身属于低风险，但被正确感知的车辆正在进行高风险操作。由于该失效并未显著改变风险分布，我们的方法正确地将其分类为低风险。

C. 结果

在实验中，我们使用来自每个场景的 $n = 1000$ 个样本，置信度 $1 - \alpha$ 和风险阈值 γ 均设为 0。9。我们测试了风险厌恶的多个取值，即 $p = \{0.9, 0.95, 0.99\}$，结果见表 II。我们测试了碰撞时间 [46] 和动量加权距离代价指标（详见附录 B），以评估风险。如前所述，碰撞时间用于计算在两个交通参与者的速度和朝向保持不变时，两者发生碰撞前的时间。动量加权距离则计算两个交通参与者的包围框之间的距离，同时考虑两者的相对速度和朝向。这两种度量的详细说明见附录 B。如果发生碰撞，我们就将该场景视为高风险。这也使我们能够计算报警至碰撞时间 指标，该指标衡量感知监测器首次发出警报与实际碰撞之间的时间。

表 II 报告了在 100 个场景中取平均值的结果。如上文所述，如果发生碰撞（真实标签），我们便将该场景视为高风险场景，并报告 F1 分数、精确率、召回率、准确率以及 警报-碰撞 指标。所提出的方法在 F1 分数、精确率、召回率和准确率方面均优于两种基线方法，即 即，HJ 可达性和碰撞概率。特别是，当使用风险厌恶为 0.99 的动量加权距离时，我们的方法优于其他所有方法。因此，尽管基线方法在显然具有风险的场景中表现相近（召回率相似），但我们的方法在细微场景中展现出更高的辨识能力，提高了精确率（从而提升了 F1 分数）。除了相关的分类结果外，该方法还能提前平均 4.72 秒预测碰撞，为自动驾驶车辆采取风险缓解措施提供充足时间。HJ 可达性的运行时最快（请注意，其价值函数已预先计算，在运行时该方法只需使用查找表），但它也最为保守；其召回率较高（与我们的方法相当），但精确率、准确率和 F1 分数最低。

表 III 展示了动量塑形距离度量的混淆矩阵。该表显示，我们的方法能够可靠地检测高风险和低风险场景，错误分类极少。值得注意的是，部分假阳性源于这样的情况：与靠近自车的某个智能体相关的故障发生了，但自车并未与其发生碰撞。

其他运行时间考量。 风险厌恶为 0.99 的动量加权距离，运行时平均值为 0.29 秒，中位数为 0.2 秒。所提出方法的瓶颈是轨迹采样，该过程由轨迹预测模块执行；在我们的案例中，该模块为 Trajectron++ [41]。从图5

表 II 结果

算法	参数	F1 分数	准确率	精确率	召回率	报警至碰撞时间 [s] 平均值	报警至碰撞时间 [s] 中位数	运行时 [s] 平均值	运行时 [s] 中位数
动量加权距离（提出的方法）	$p=0.90,\gamma=0.9,\alpha=0.1$	0.70	0.80	0.55	**0.96**	5.28	3.60	0.29	0.2
	$p=0.95,\gamma=0.9,\alpha=0.1$	0.79	0.88	0.68	**0.96**	5.18	3.60		
	$p=0.99,\gamma=0.9,\alpha=0.1$	**0.86**	**0.93**	**0.81**	0.92	4.72	3.03		
碰撞时间（拟议）	$p=0.90,\gamma=0.9,\alpha=0.1$	0.70	0.80	0.55	**0.96**	4.61	2.45	0.26	0.17
	$p=0.95,\gamma=0.9,\alpha=0.1$	0.76	0.86	0.65	0.92	4.16	2.05		
	$p=0.99,\gamma=0.9,\alpha=0.1$	0.79	0.90	0.79	0.79	4.13	2.25		
碰撞概率	$\gamma=0.90$	0.40	0.32	0.26	**0.96**	6.53	4.95	0.25	0.17
	$\gamma=0.95$	0.41	0.33	0.26	**0.96**	6.31	4.95		
	$\gamma=0.99$	0.43	0.41	0.28	0.92	4.33	3.48		
HJ 可达性	$\gamma=0$	0.39	0.28	0.24	**0.96**	**7.10**	**5.75**	**0.01**	**0.01**

		预测值：高风险	预测值：低风险
Actual	高风险	22（真正例）	2 假负例
	低风险	5 假正例	71 真负例

表 III 动量形状距离混淆矩阵

我们可以看到，轨迹预测模块的平均耗时为 0.22 秒，约占总运行时间的 75%。只需使用更快的轨迹预测模块，例如 PredictionNet [27]，，即可轻松克服这一限制；它的速度比 Trajectron++ [27]快两个数量级。此外，所提出的方法很容易并行化，因为可以针对场景中的每个智能体并行计算成本，并将两个场景批量合并为一次预测网络查询。通过采用建议的实现改进，我们预计运行时将显著缩短。

VI. 结论与未来工作

本文提出了一个框架，用于评估感知失效对自动驾驶车辆运动规划所构成的风险。我们首先识别感知失效模式，随后为当前场景合成合理的替代场景，然后评估自动驾驶车辆在合理场景中所面临的风险相较于其感知场景增加了多少，从而实现了这一目标。我们将任务感知风险形式化为 p-分位数相对场景风险，并进一步开发了一种利用独立同分布样本对其进行估计的算法。此外，我们为风险估计提供了 PAC 界，以高概率确保算法的正确性。最后，对该方法的实验评估表明，在精确率、召回率、准确率和 F1 分数方面，我们的检测器优于基线方法。

作为我们未来工作的一部分，我们将开发完整的集成式任务感知型感知监测器，其中包含三个组成模块（感知失效检测与识别模块、合理场景生成器、任务感知风险估计器），并评估其闭环性能。我们还将探索风险规避参数和风险阈值参数的数据驱动校准与在线自适应。最后，我们将开展以提升我们方法的运行时。如上所述，我们实验中的计算瓶颈源于对 Trajectron++ 的查询。我们将使用更快的预测网络，例如 [27],，以加快计算速度。

附录

A. 定理 2 的证明

在证明定理 2 之前，我们先介绍两个有用的结果。首先注意到，式（5）中的风险函数依赖于联结函数以及 A 和 B 的累积分布函数，即 Φ_A 和 Φ_B。因此，为了给出风险函数 $\mathcal{R}$ 的性能界，我们首先需要界定联结函数的取值。为此，我们可以使用著名的 Fr´echet–Hoeffding 联结函数界。

定理 3 (弗´雷歇–霍夫丁定理（[38],定理 2.2.3）。 对于任意联结函数 $C: [0,1]^d \to [0,1]$ 以及任意 $(u_1, \ldots, u_d) \in [0,1]^d$，以下界限成立：

$$W(u_1,\ldots,u_d) \le C(u_1,\ldots,u_d) \le M(u_1,\ldots,u_d), \quad (6)$$

其中

$$W(u_1,\ldots,u_d) = \max\left\{1-d+\sum_{i=1}^{d} u_i, 0\right\},$$
$$M(u_1,\ldots,u_d) = \min\{u_1,\ldots,u_d\}.$$

本文关注的是二维联结函数，因此我们可以将定理 3 应用于联结函数 $C(p,v)$，并得到：

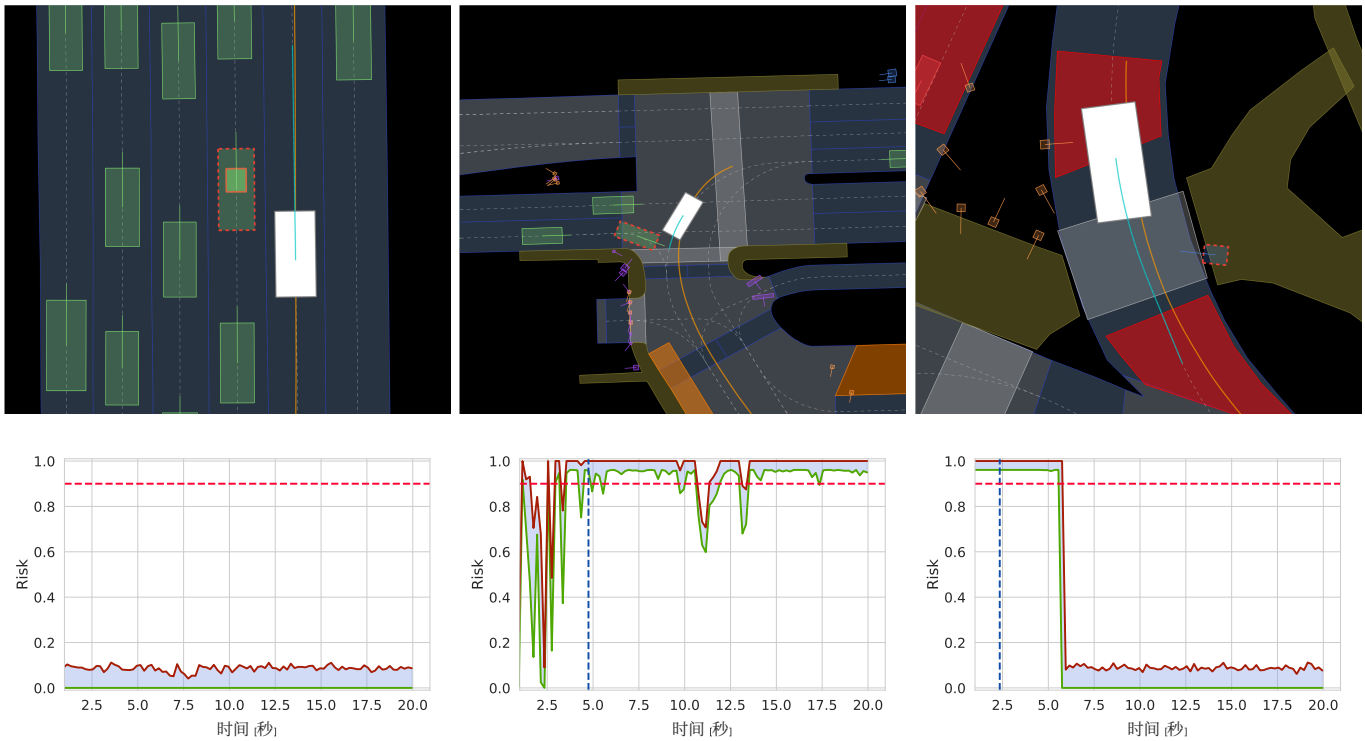
$$\max\{p+v-1,0\} \le C(p,v) \le \min\{p,v\}.$$

特别是，对于风险估计，我们采用 $v = \Phi_B \circ \Phi_A^{-1}(p)$。然而，我们无法获取两个累积分布函数 Φ_A 和 Φ_B 的显式表达式；因此，需要通过经验方法对其进行估计。以下定理给出了经验累积分布函数的估计界。

定理4 （Dvoretzky–Kiefer–Wolfowitz 置信区间 [13, 34]）。 设 Φ 为某个未知分布的累积分布函数，$\Phi^{(n)}$ 为使用来自 Φ的 n 个独立同分布样本计算得到的经验累积分布函数，则其成立的概率至少为 $1-\alpha$，

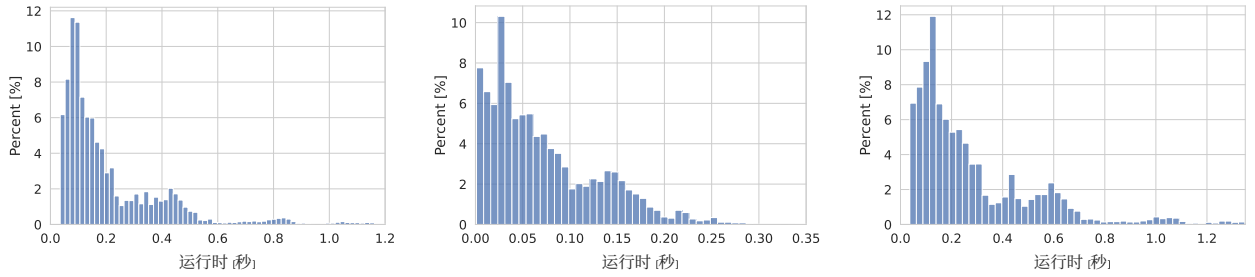
$$\Phi^{(n)}(x) - \epsilon(n,\alpha) \le \Phi(x) \le \Phi^{(n)}(x) + \epsilon(n,\alpha), \quad (7)$$

其中 $\epsilon(n, \alpha) = \sqrt{\ln(2/\alpha)/(2n)}$。



(a) 误检测（尺寸） (b) 障碍物漏检 (c) 行人漏检

图 4. **场景示例及其相应的估计风险**。顶行展示场景，其中自车以白色方框表示，其他车辆以绿色方框表示，行人以蓝色方框表示。红色虚线表示智能体的真实值位置和尺寸，实线则表示自车感知系统感知到的位置和尺寸。底行展示顶行对应场景的估计风险。水平虚线表示风险阈值 γ。红色实线表示风险上界，绿色实线表示风险下界。蓝色竖直虚线表示碰撞时间。值得注意的是，在我们的仿真中，碰撞后自车和非自车智能体的行为不会发生变化，即，仿真会持续运行至场景结束。



(a) 预测运行时间 (b) 风险估计运行时间 （c）总运行时间

图5。 **采用所 提出的方法，使用动量加权距离**。预测运行时间平均为 0.22 秒（中位数运行时间 : 0.14 秒）。风险估计运行时间平均为 0.07 秒（中位数运行时间：0.06 秒）。总运行时间平均为 0.29 秒（中位数运行时间：0.2 秒）。

我们使用定理4在下一个引理中建立 $\Phi_B \circ \Phi_A^{-1}(p)$ 的界。

引理 1. 设 A、 B 是两个随机变量，其累积分布函数分别为 Φ_A 和 Φ_B。设 $\Phi_A^{(n)}$ 和 $\Phi_B^{(n)}$ 分别是使用来自 Φ_A 和 Φ_B 的 n 个独立同分布样本估计得到的经验累积分布函数。那么，以至少 $1-\alpha$ 的概率，有：

$$\underline{v}(p,\alpha,n) \leq \Phi_B \circ \Phi_A^{-1}(p) \leq \bar{v}(p,\alpha,n),$$

其中

$$\bar{v}(p,\alpha,n) = \Phi_B^{(n)} \circ \left[\Phi_A^{(n)} - \epsilon(\alpha,n)\right]^{-1}(p) + \epsilon(\alpha,n)$$

$$\underline{v}(p,\alpha,n) = \Phi_B^{(n)} \circ \left[\Phi_A^{(n)} + \epsilon(\alpha,n)\right]^{-1}(p) - \epsilon(\alpha,n)$$

and $\epsilon(\alpha,n) = \sqrt{\ln(2/\alpha)/(2n)}$.

证明： 我们关注量 $\Phi_B \circ \Phi_A^{-1}(p)$。注意， $\Phi_A^{(n)}$ 和 $\Phi_B^{(n)}$ 是递增函数。根据定理4，我们知道 $\underline{x} \leq \Phi_A^{-1}(p) \leq \bar{x}$ （见图6），其中

$$\bar{x}(p,\alpha,n) = \left[\Phi_A^{(n)} - \epsilon(\alpha,n)\right]^{-1}(p)$$
$$\underline{x}(p,\alpha,n) = \left[\Phi_A^{(n)} + \epsilon(\alpha,n)\right]^{-1}(p) \tag{8}$$

同样地，我们有 $\underline{v} \le \Phi_B \circ \Phi_A^{-1}(p) \le \bar{v}$ 其中

$$\bar{v}(p,\alpha,n) = \Phi_B^{(n)} \circ \bar{x}(p,\alpha,n) + \epsilon(\alpha,n)$$
$$\underline{v}(p,\alpha,n) = \Phi_B^{(n)} \circ \underline{x}(p,\alpha,n) - \epsilon(\alpha,n) \tag{9}$$

将式（8）代入式（9），即可完成证明。 ■

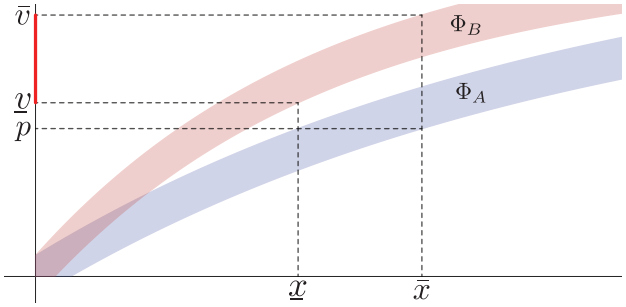


图6. CDF 复合界。阴影区域表示两个累积分布函数的置信区间，即 $\Phi^{(n)}(x - \epsilon(n,\alpha) \le \Phi_A(x) \le \Phi^{(n)}(x + \epsilon(n,\alpha))$。蓝色和橙色区域分别表示和。

现在我们可以证明定理 2 了。

定理 *2* 的证明：由定理 3 可知，

$$\max\{p + v - 1, 0\} \le C(p,v) \le \min\{p, v\},$$

其中 $v = \Phi_B \circ \Phi_A^{-1}(p)$。我们可以使用引理 1 对 v 进行界定，从而得到

$$\max(p + \underline{v} - 1, 0) \le C(p,v) \le \min(p, \bar{v}). \tag{10}$$

由于 $\mathcal{R}(p) = 1 - C(p,\Phi_B \circ \Phi_A^{-1}(p))/p$，我们有 $C(p,\Phi_B \circ \Phi_A^{-1}(p)) = p(1 - \mathcal{R}(p))$。将其代入公式（10），即可完成证明。 ■

B. 成本函数

设 $\boldsymbol{x}_e$、$\boldsymbol{v}_e$ 分别为自车的位置和速度。同样，设 $\boldsymbol{x}_a$、$\boldsymbol{v}_a$ 分别为任意非自车智能体的位置和速度。此外，设 ν 为用于惩罚违反交通规则行为的项，例如在道路错误的一侧行驶、逆着交通流方向行驶，或在红色交通信号灯亮起时通过交叉路口。

碰撞时间成本定义如下：

$$c_{\text{TTC}} = 1 - \max_{a \in \text{Agents}} \min\left\{\frac{TTC(x_e, x_a, v_e, v_a)}{m}, 1\right\} + \nu,$$

其中，m 表示 TTC 函数的最大值。该函数输出自车与另一个智能体发生碰撞前的时间（假设它们以当前行驶方向恒定速度行驶）；如果不会发生碰撞，则输出无穷大。在我们的实验中，我们使用 $m = 3$。该代价函数的取值范围为 $[0, 1],$，数值越大表示碰撞时间越短，因此风险越高。

动量形状距离则定义如下：

$$c_{\text{MSD}} = \max_{a \in \text{Agents}} e^{\epsilon\delta/2} + \nu,$$
$$\delta = \left((\boldsymbol{x}_{a,\parallel} - \boldsymbol{x}_{e,\parallel})(\boldsymbol{v}_{a,\parallel} - \boldsymbol{v}_{e,\parallel})\right)^2 + \left((\boldsymbol{x}_{a,\perp} - \boldsymbol{x}_{e,\perp})(\boldsymbol{v}_{a,\perp} - \boldsymbol{v}_{e,\perp})\right)^2$$

我们使用下标 $\parallel$ 和 $\perp$ 分别表示向量沿着与自车航向角平行的方向和垂直方向的投影。缩放因子 ϵ用于衡量智能体的重要性：在实验中，当智能体为车辆时，我们设置 $\epsilon = 0.5$；当智能体为行人时，设置 $\epsilon = 1$。

参考文献

[1] Nakul Agarwal 和 Yi-Ting Chen。驾驶场景中的风险感知。载于*NeurIPS* 自动驾驶机器学习研讨会论文集，2022 年。[2] Saleh Albeaik、Alexandre Bayen、Maria Teresa Chiri、Xi- aoqian Gong、Amaury Hayat、Nicolas Kardous、Alexan- der Keimer、Sean T McQuade、Benedetto Piccoli 和 Yiling You。智能驾驶员模型（IDM）的局限性与改进。*SIAM*应用动力系统杂志，21(3):1862–1892，2022 年。[3] P. Antonante、H. Nilsen 和 L. Carlone。感知系统监控：确定性、概率性以及基于学习的故障检测与识别。*arXiv* 预印本 *arXiv: 2205.10906*，2022 年。[4] Pasquale Antonante、David I. Spivak 和 Luca Carlone。感知系统的监控与可诊断性。载于*IEEE/RSJ*智能机器人与系统国际会议（*IROS*）论文集，第 168–175 页，2021 年。doi: 10.1109/IROS51168.2021.9636497。

[5] A. Balakrishnan、A. G. Puranic、X. Qin、A. Dokhanchi、J. V. Deshmukh、H. Ben Amor 和 G. Fainekos。为基于视觉的感知系统指定并评估质量指标。载于欧洲设计、自动化与测试会议暨展览论文集（*DATE*），第 1433–1438 页，2019 年。

[6]Anand Balakrishnan、Jyotirmoy Deshmukh、Bardh Hoxha、Tomoya Yamaguchi 和 Georgios Fainekos。Percemon：感知系统的在线监控。载于运行时验证国际会议论文集，第 297–308 页。Springer，2021 年。

[7] Ayoosh Bansal、Jayati Singh、Micaela Verucchi、Marco Caccamo 和 Lui Sha。风险排序召回率：自动驾驶车辆目标检测系统的碰撞安全指标。载于地中海嵌入式计算会议（*MECO*）论文集，第 1–4 页，2021 年。[8]Julian Bernhard、Patrick Hart、Amit Sahu、Christoph Sch¨oller 和 Michell Guzman Cancimance。感知不确定性下自动驾驶车辆的基于风险的安全包络。载于*IEEE* 智能车辆研讨会（*IV*）论文集，第 104–111 页，2022 年。[9]Victor Besnier、Andrei Bursuc、David Picard 和 Alexandre Briot。触发故障：通过局部对抗攻击学习进行分布外检测，应用于

语义分割。载于*IEEE/CVF*国际计算机视觉会议论文集，第 15701–15710 页，2021 年。[10]Daniel Bogdoll、Maximilian Nitsche 和 J Marius Z¨ollner。自动驾驶中的异常检测：综述。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第 4488–4499 页，2022 年。[11] Filippos Christianos、Peter Karkus、Boris Ivanovic、Stefano V Albrecht 和 Marco Pavone。使用双层变分遮挡模型规划被遮挡交通智能体的行为。*arXiv* 预印本 *arXiv:2210.14584*，2022 年。[12] John Dahl、Gabriel Rodrigues de Campos、Claes Olsson 和 Jonas Fredriksson。碰撞规避：威胁评估技术文献综述。*IEEE*智能车辆汇刊，4(1)：101–113，2018 年。[13]Aryeh Dvoretzky、Jack Kiefer 和 Jacob Wolfowitz。样本分布函数和经典多项分布估计量的渐近极小极大特性。《数理统计年刊》，第 642–669 页，1956 年。[14] Alec Farid、Sushant Veer 和 Anirudha Majumdar。面向机器人学习、具有统计保证的任务驱动分布外检测。载于机器人学习会议（*CoRL*）论文集，第 970–980 页，2021 年。[15] Alec Farid、David Snyder、Allen Z Ren 和 Anirudha Majumdar。基于视觉的机器人控制中具有统计保证的故障预测。*arXiv* 预印本 *arXiv:2202.05894*，2022 年。[16] Alec Farid、Sushant Veer、Boris Ivanovic、Karen Leung 和 Marco Pavone。自动驾驶车辆轨迹预测器的任务相关故障检测。*arXiv* 预印本 *arXiv:2207.12380*，2022 年。[17] Enric Galceran、Edwin Olson 和 Ryan M Eustice。自动驾驶决策中遮挡情况下的增强车辆跟踪。载于*IEEE/RSJ*智能机器人与系统国际会议（*IROS*）论文集，第 3559–3565 页，2015 年。[18] K. Tan 等、H. Caesar、J. Kabzan。Nuplan：面向自动驾驶车辆的基于机器学习的闭环规划基准。载于*CVPR ADP3*研讨会论文集，2021 年。[19] O. A. Hafez、G. D. Arana、M. Joerger 和 M. S penko。量化机器人定位安全性：一种用于固定滞后平滑的全新完整性监测方法。*IEEE*机器人与自动化快报，5(2)：3182–3189，2020 年。[20] Yutao Han、Jacopo Banfi 和 Mark Campbell。通过想象未知空间中的情形来规划穿越路径。载于机器人学习会议（*CoRL*）论文集，第 905–914 页，2021 年。[21] Dan Hendrycks、Steven Basart、Mantas Mazeika、Mohammadreza Mostajabi、Jacob Steinhardt 和 Dawn Song。面向真实场景扩展分布外检测。*arXiv* 预印本 *arXiv:1911.11132*，2019 年。[22]Kai-Chieh Hsu、Allen Z Ren、Duy Phuong Nguyen、Anirudha Majumdar 和 Jaime F Fisac。从仿真到实验室再到真实环境：通过屏蔽机制实现安全强化学习，并提供泛化保证。*arXiv* 预印本 *arXiv:2201.08355*，2022.[23]Haimin Hu、Kensuke Nakamura 和 Jaime F Fisac。Sharp：考虑屏蔽机制的鲁棒规划，实现安全高效的人机交互。*IEEE*机器人与自动化快报，7(2):5591–5598，2022.[24] Masha Itkina、Ye-Ji Mun、Katherine Driggs-Campbell 和 Mykel J Kochenderfer。利用人作为传感器进行多智能体变分遮挡推断。载于机器人与自动化国际会议（*ICRA*）论文集，第4585–4591页，2022.[25] Hao Jing、Yang Gao、Sepeedeh Shahbeigi 和 Mehrdad Dianati。自动驾驶车辆GNSS/INS定位系统的完整性监测：最新进展与开放性挑战。*IEEE*智能交通系统汇刊，23(9):14166–14187，2022。doi: 10.1109/TITS.2022.3149373.[26] Harry Joe。使用联结函数进行依赖关系建模。CRC出版社，2014.[27] Alexey Kamenev、Lirui Wang、Ollin Boer Bohan、Ishwar Kulkarni、Bilal Kartal、Artem Molchanov、Stan Birch-field、David Nist´er 和 Nikolai Smolyanskiy。Predictionnet：用于规划、控制和仿真的实时联合概率交通预测。载于机器人与自动化国际会议（*ICRA*）论文集，第8936–8942页，2022.[28] Daniel Kang、Deepti Raghavan、Peter Bailis 和 Matei Zaharia。用于调试机器学习的模型断言。载于神经信息处理系统进展（*NeurIPS*）论文集，2018.[29] Markus Koschi 和 Matthias Althoff。考虑遮挡和交通规则的交通参与者集合预测。*IEEE*智能车辆汇刊，6(2):249–265，2020.[30] John Lambert 和 James Hays。信任，但要验证：用于高清地图变化检测的跨模态融合。载于神经信息处理系统进展（*NeurIPS*）论文集，2021.[31]Karen Leung、Edward Schmerling、Mengxuan Zhang、Mo Chen、John Talbot、J Christian Gerdes 和 Marco Pavone。将基于可达性的安全保障融入人机车辆交互规划框架。国际机器人研究期刊，39(10-11):1326–1345，2020.[32]Franck Li、Philippe Bonnifait 和 Javier Iba˜nez- Guzm´an。利用车道级地图进行地图辅助航位推算与完整性监测。*IEEE*智能车辆汇刊，3(1):81–91，2018.[33] Jinshan Liu 和 Jung-Min Park。“眼见不一定为实”：检测针对自动驾驶车辆的感知错误攻击。*IEEE*可靠与安全计算汇刊，18(5):2209–2223，2021.[34]Pascal Massart。Dvoretzky-Kiefer-Wolfowitz不等式中的紧确常数。概率年刊，第1269–1283页，1990.[35]Dimity Miller、Peyman Moghadam、Mark Cox、Matt Wildie

目标检测器中的假负例机制。*arXiv* 预印本 *arXiv:2203.07662*，2022。[36]Ian M Mitchell、Alexandre M Bayen 和 Claire J Tomlin。连续动态博弈中可达集的时变 Hamilton-Jacobi 公式。*IEEE* 自动控制汇刊，50(7):947–957，2005。[37]Motional。NuPlan-devkit。网址 https://github.com/motional/nuplan-devkit。[38] Roger B Nelsen。*Copula* 简介。Springer Science & Business Media，2007。[39] Quazi Marufur Rahman、Niko S¨underhauf、Peter Corke 和 Feras Dayoub。FSNet：用于语义分割的故障检测框架。载于*IEEE*机器人与自动化快报，第7卷，3030–3037页，2022。doi: 10.1109/LRA.2022.3143219。[40]Manikandasriram Srinivasan Ramanagopal、Cyrus Anderson、Ram Vasudevan 和 Matthew Johnson-Roberson。未能学会：自动识别自动驾驶汽车的感知故障。*IEEE*机器人与自动化快报，3(4):3860–3867，2018。[41]Tim Salzmann、Boris Ivanovic、Punarjay Chakravarty 和 Marco Pavone。Trajectron++：利用异构数据进行满足动态可行性的轨迹预测。载于欧洲计算机视觉会议（*ECCV*）论文集，683–700页，2020。[42]Apoorva Sharma、Navid Azizan 和 Marco Pavone。通过曲率草图高效检测深度神经网络的分布外样本。载于人工智能中的不确定性会议（*UAI*）论文集，1958–1967页，2021。[43]M Sklar。多维分布函数及其边缘分布。巴黎大学统计研究所出版物´，8:229–231，1959。[44]Sever Topan、Karen Leung、Yuxiao Chen、Pritish Tupekar、Edward Schmerling、Jonas Nilsson、Michael Cox 和 Marco Pavone。面向交互动力学的感知区域：用于障碍物检测安全评估。载于*IEEE* 智能车辆研讨会（*IV*）论文集，1201–1210页，2022。doi: 10.1109/IV51971.2022.9827409。[45]Martin Treiber、Ansgar Hennecke 和 Dirk Helbing。实证观测和微观模拟中的拥堵交通状态。《物理评论*E*》，62(2):1805，2000。[46] Lukas Westhofen、Christian Neurohr、Tjark Koopmann、Martin Butz、Barbara Sch¨utt、Fabian Utesch、Birte Neurohr、Christian Gutenkunst 和 Eckard B¨ode。自动驾驶关键性指标：现状综述与适用性分析。《工程计算方法档案》，30(1):1–35，2023。[47] Chengzeng You、Zhongyuan Hau 和 Soteris Demetriou。通过时间一致性检查检测针对自动驾驶车辆感知的激光雷达欺骗攻击。载于移动人工智能安全与隐私研讨会论文集，13–18页，2021。[48]Peter Zechel、Ralph Streiter、Klaus Bogenberger 和 Ulrich G¨ohner。自动驾驶车辆的行人占用预测。载于*IEEE*国际机器人计算会议（*IRC*）论文集，第 230–235 页，2019 年。