ROCO：环岛交通冲突数据集

Depu Meng[1]
中国科学技术大学自动化系，中国合肥 电子邮箱：
mdp@mail.ustc.edu.cn

Owen Sayer
密歇根大学安娜堡分校机械工程系，美国密
歇根州 电子邮箱：osayer@umich.edu

张如生
密歇根大学安娜堡分校土木与环境工程系，密
歇根州安娜堡 邮箱：rushengz@umich.edu

沈胜银
密歇根大学交通研究所，密歇根大学安娜堡分校，
密歇根州安娜堡 邮箱：shengyin@umich.edu

李厚强，哲学博士
中国科学技术大学电子工程与信息科学系，中国合肥 邮箱：
lihq@ustc.edu.cn

刘恒，哲学博士，通讯作者
密歇根大学安娜堡分校土木与环境工程系，密歇根
州安娜堡 电子邮箱：henryliu@umich.edu

arXiv:2303.00563v2 [eess.SY] 2 Mar 2023

[1]这项工作是在 Depu Meng 作为访问学者于密歇根大学安娜堡分校土木与环境工程系工作期间完成的。

摘要

几十年来，交通运输研究界一直将交通冲突作为替代安全度量进行研究。然而，由于交通冲突十分罕见，收集大规模现实世界交通冲突数据极具挑战性。本文介绍并分析了ROCO——一个真实世界环岛交通冲突数据集。数据采集于密歇根州安娜堡 State St. 与 W. Ellsworth Rd. 交叉口的一处双车道环岛。我们将安装在该环岛处的四台鱼眼摄像机所捕获的原始视频数据流作为输入数据源。我们采用基于学习的冲突识别算法从视频中寻找潜在的交通冲突，然后对其进行人工标注，以完成数据集的收集与标注工作。2021年8月至2021年10月期间，共收集到557起交通冲突和17起交通事故。我们提供了利用路侧感知系统从交通冲突场景中提取的轨迹数据，并根据交通冲突严重程度、交通冲突原因及其对交通流的影响进行分类。通过收集的交通冲突数据，我们发现，车辆驶入环岛时未让行环行车辆是导致交通冲突的最大原因。ROCO数据集将公开发布于 https://github.com/michigan-traffic-lab/ROCO。

关键词：交通冲突、安全指标、路侧感知、安全分析

引言

数十年来，提升交通安全一直是交通运输研究中一项至关重要且长期以来的目标。交通事故（如碰撞）是交通运输领域造成死亡、受伤和财产损失的主要原因。为了更好地了解汽车碰撞，需要大规模的事故数据。然而，由于碰撞事故极少发生（*1, 2*）（约每英里 1.9×10^{-6} 次），且数据收集方法十分有限（大多数情况下，事故数据取自警方报告），因此很难在相对较短的时间内获取大量事故数据。此外，警方报告所包含的信息往往缺少必要内容，例如，碰撞发生前车辆的确切速度和行驶操作。

由于事故数据获取存在上述局限，交通冲突等替代安全指标已得到广泛研究和讨论。交通冲突一词尚无统一定义。本文采用 NCHRP 219（*3*）给出的定义：“交通冲突是指涉及两名或两名以上道路使用者的交通事件，其中一名道路使用者采取某种非典型或异常行为，例如改变方向或速度，从而使另一名道路使用者面临碰撞危险，除非其采取规避操作。”图1展示了碰撞事故与交通冲突的对比。交通冲突被视为正常交通行为与交通事故之间缺失的“纽带”（*4*）。许多研究探讨了交通冲突与碰撞事故之间的关系，包括金字塔式安全连续谱（*5*）、病因学一致性（*6*）和因果模型（*7*）。由于交通冲突发生得更加频繁，且通常不会造成损害，因此更容易在较短时间内获取大量交通冲突数据。已有研究探讨了交通冲突的多种应用，包括关键行为分析（*8–10*）、前后对比研究（*11–13*）、识别高风险地点（*14*）以及实时安全预测（*15–17*）。

尽管交通冲突数据比事故数据更容易获取，但交通冲突研究中仍存在若干与数据相关的问题。其中最关键的问题之一是缺乏公开可用的大规模交通冲突数据。目前，交通运输研究人员和工程师需要各自收集交通冲突数据。此外，正如（*18*）中所述，由于每个研究团队使用自己的数据，数据不一致也成为一个问题。不同的冲突识别标准会导致不同的分析结果，因此难以对不同研究的发现进行讨论和概括。大多数交通冲突数据是通过路侧摄像机或网联车辆收集的。一些可能导致交通冲突的因素，例如周围其他车辆的状态、天气条件和光照条件，可能无法在交通冲突数据中观测到。这些未观测因素可能导致交通安全建模与分析不准确。

为解决上述交通冲突数据问题，我们提出了 ROCO，一种新的环岛交通冲突数据集。为收集罕见的交通冲突数据，我们利用预先安装的路侧摄像机和边缘云基础设施 [2]。环岛的四个角落各安装了一台路侧鱼眼摄像机。同时，我们部署了边缘云基础设施，以低延迟、全天候（24/7）的方式，将路侧采集的原始传感器数据流式传输至基于云端的 Mcity Mobility Data Center（*19, 20*），如图 2(a) 所示。收集 ROCO 数据的环岛是位于美国密歇根州安娜堡 State St. 与 W. Ellsworth Rd. 交叉口处的双车道环岛（42.229379, -83.739013）。即使有了收集到的 $24/7$ 视频数据，识别交通冲突仍然过于耗时，难以

[2]Mcity 赞助的研究项目 N029385：人工智能与车联网技术助力安全高效环岛。

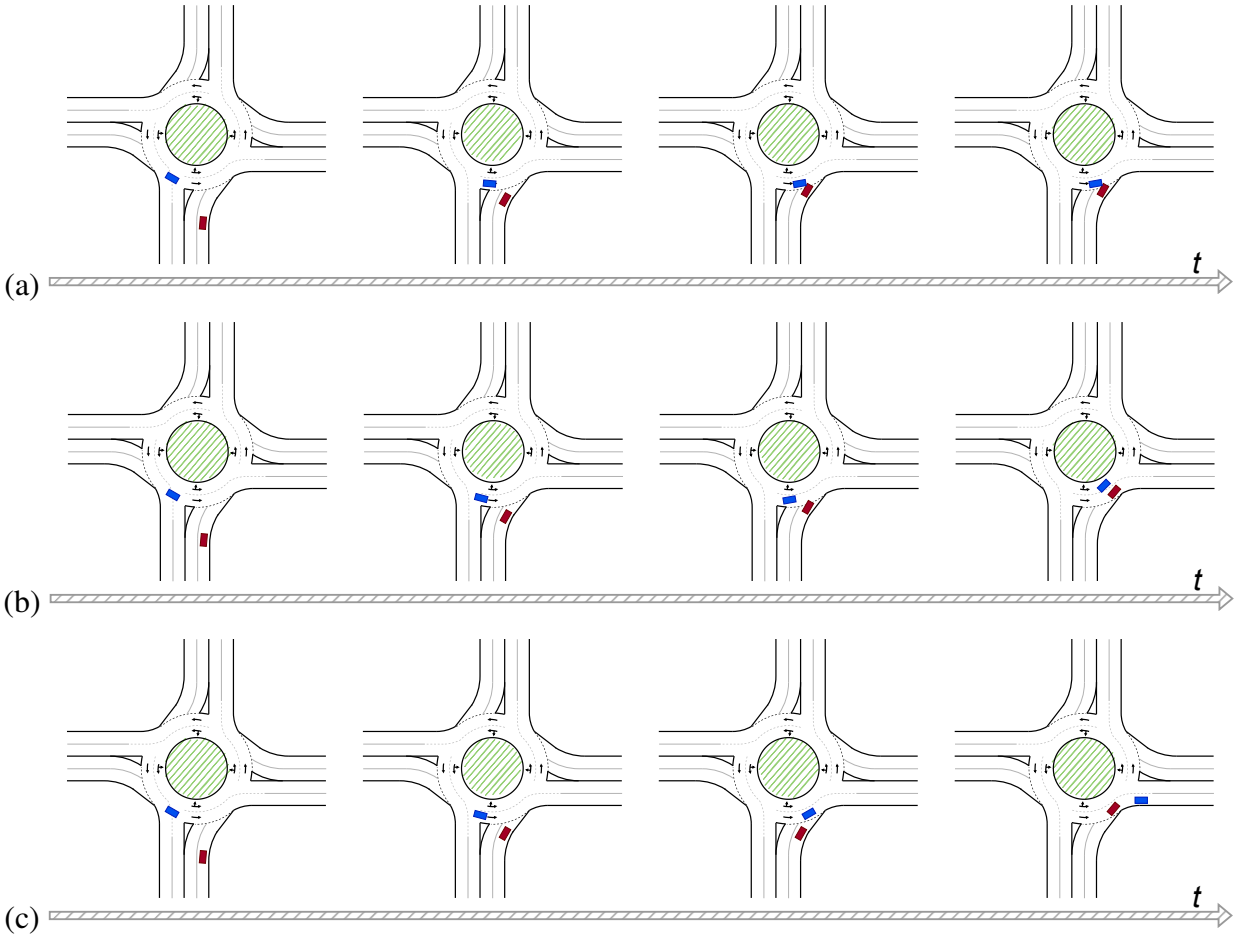


图1 （**a**）碰撞事故、（**b**）交通冲突和（**c**）正常驾驶行为示意图。在（a）中，红色车辆未礼让蓝色车辆便驶入环岛，导致碰撞事故。在（b）中，红色车辆未礼让，迫使蓝色车辆采取规避动作（急剧变道至内侧车道），从而引发交通冲突。在（c）中，红色车辆礼让蓝色车辆，没有发生碰撞或冲突。交通冲突是一个广义概念，包括时空接近或规避动作、险些相撞事件、碰撞险情以及其他安全关键事件（*18*）。交通冲突可作为交通事故的替代安全度量，帮助交通运输研究人员和工程师更好地提升交通安全。

人工从正常交通场景中识别交通冲突。因此，我们开发了一套数据收集与标注流程：首先使用基于深度学习的视频处理算法识别潜在交通冲突，然后人工检查并标注所有被判定为可能包含交通冲突的视频片段。交通冲突识别算法采用两阶段架构，每个阶段都使用经过改进的 SlowFast（*21*）网络。第一阶段模型识别出相当一部分视频片段。第二阶段模型对第一阶段筛选出的视频进行再识别。经过第二阶段处理后，约 23% 的筛选视频包含一个或多个交通冲突。

经过两阶段算法筛选视频片段后，我们开始进行核验与标注。我们不对交通冲突作定量定义，因为

(a) 

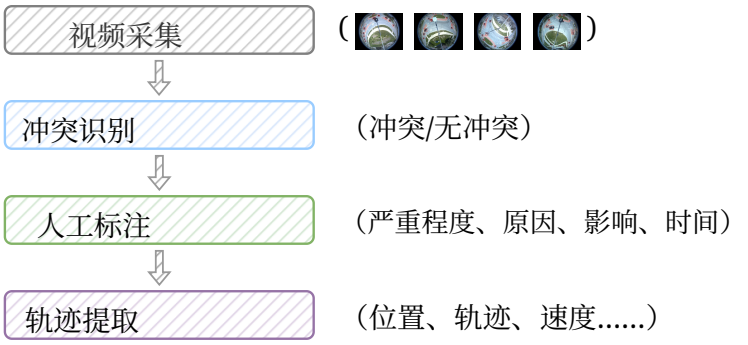
(b) 

图 2 （a）数据采集地点环岛示意图。该环岛为双车道环岛，位于美国密歇根州安娜堡 State St. 与 W. Ellsworth Rd. 的交叉口。（b）交通冲突数据采集流程。我们首先使用基于学习的冲突识别算法找出潜在冲突，然后人工检查并标注这些潜在冲突，并提取轨迹数据。

交通冲突是一个多维问题（*22–24*）。要为交通冲突提出通用且明确的定量条件非常困难。因此，我们采用基于示例的标注方法。我们为环岛中每种典型的交通冲突类型提供若干示例，由标注人员对潜在冲突与现有示例之间的相似程度进行定性评估。我们根据冲突严重程度、交通冲突的原因，以及交通冲突对环岛交通流的影响，设计了交通冲突分类体系。ROCO数据集中共有557个交通冲突样本和17个交通事故样本。此外，我们使用路侧感知系统从每个视频片段中提取轨迹数据。由于环岛的传感器覆盖范围完整，环岛内或附近的所有车辆都会被记录。我们还记录了照明和天气条件，以便未来的研究使用。

本文提供了 ROCO 数据集的一些初步统计结果。一项有趣的发现是，交通冲突的最主要原因是车辆驶入环岛时未让行于环行车辆。我们收集的交通冲突中，超过 71% 是由未让行造成的。此外，尽管卡车等大型车辆在环岛中并不常见，但它们造成了许多未让行事件（所有未让行事件中约有 $1/3$ 是由大型车辆造成的）。

我们相信，借助 ROCO数据集，交通运输研究人员将能够轻松获取标注一致的真实世界交通冲突数据。对交通冲突进行建模，并研究碰撞事故与冲突之间的关系，将为如何提升环岛交通安全提供指导。目前，我们仅收集了一个环岛的数据。不过，未来我们将继续从多个环岛和信号交叉口收集更多交通冲突数据，以提升数据集的多样性和泛化能力。

总而言之，本文的主要贡献如下：1. 我们提出了 ROCO，这是一个包含丰富情境信息的环岛交通冲突数据集，包括周边车辆状态和天气条件。2. 我们设计了交通冲突数据采集流程，其中包括两阶段冲突识别框架，以及人工核验与标注。3. 我们对 ROCO 数据集进行了统计分析，并发现了交通冲突的规律，使交通运输机构能够据此提升环岛安全性和效率。

效率。

相关研究

交通安全分析研究，尤其是交通冲突的自动检测，是一个不断发展的领域。随着研究的推进，相关技术和方法也在不断发展，并建立在前人的工作基础之上。正如交通冲突最初只是理论构想，随后它们也能够得到验证，成为事故预测指标，并用于分析我们周围的道路。

交通冲突

交通冲突的最初思想可追溯至 20 世纪 30 年代，当时赫伯特·W·海因里希提出，所有与安全相关的交通事故都与险些相撞事件有着共同的原因或根源（*25*）。这一观点可以具体应用于交通冲突和碰撞事故：假设碰撞事故是冲突初始状态的一种潜在结果（此时若不采取规避动作，两辆或多辆车辆就会发生碰撞）（*18, 26*）。此后，这一定义得到了扩展。随着人们进一步认识到基于事故的分析存在的问题，交通分析中对交通冲突的应用也日益增多。

出于多种原因，交通冲突被用作事故数据的替代指标。交通冲突的发生频率远高于碰撞事故（*26, 27*），尽管识别起来更加困难；碰撞事故的定义十分明确，即两辆或多辆车辆发生接触，但交通冲突的识别则主观得多（*18*）。不过，交通冲突仍使我们能够在交通分析中发现并利用规模更大、类型更多样的数据集。使用交通冲突而非事故数据的另一个原因是，事故数据“本质上具有滞后性”（*18, 28*）：要获得事故数据，就必须先发生碰撞事故。这可能会大幅延误交通管制的改进和调整，而这些改进和调整本可避免伤亡。此外，事故数据往往不可靠，或难以广泛获取（*18, 28*）。这些数据通常来自警方报告，而报告并不总是包含重要因素，例如碰撞事故的确切位置或准确的车辆速度。警方报告也未涵盖所有碰撞事故，因为许多事故都未被报告（*29*）。

人们认为，交通冲突可以作为准确或合理的碰撞事故先兆（*26*）。如果交通冲突发生时，一名或多名道路使用者采取了规避动作以避免碰撞或事故，那么顺理成章地可以推断，规避动作失败将导致碰撞事故。这一观点已由多篇不同的论文和模型加以阐述。其中包括金字塔式安全连续体模型，该模型根据交通事件的危险程度对所有交通事件进行排序；这些事件分为三类：无干扰通行、交通冲突和碰撞事故（*18, 30*）。此外，其他许多研究（包括（*31*）和（*32*））也分别发现，交通冲突与碰撞事故之间存在很强的相关性。

交通冲突通常通过人工观察并依据研究或研究小组制定的一套标准来确定，或者使用事故预测指标和算法来确定。这类指标包括TTC（碰撞时间）和PET（侵入后时间）。TTC是指在所有条件保持不变的情况下，两辆或多辆车辆发生碰撞前的时间。TTC已得到广泛应用，例如（*26*）和（*31*）中的研究。PET是指一辆车辆离开潜在碰撞点后，到另一辆车辆抵达该点之间的时间（*18*）。通过为TTC或PET设定阈值，便可以自动检测交通冲突。

尽管交通冲突分析有着明显的优势，但交通冲突数据及其收集仍存在问题。首先，也是最重要的一点，目前尚无统一的交通冲突识别标准（*18*）。这导致不同研究之间存在不一致。对于由算法检测出的冲突，必须制定一套规则或标准来指导计算机的判断，但这些规则或标准几乎完全由研究人员自行决定

（*18*）。因此，比较不同研究的结果并验证其有效性变得极具挑战性。此外，交通冲突数据与事故数据一样，通常是在较短时间内收集的。仅收集几天甚至几周的数据，无法反映驾驶员行为和交通流量随季节、事件或天气变化而产生的变化（*33*）。这会导致人们对不同时间段的交通冲突数量作出有偏差的假设（*18*）。了解这些缺陷，是认识到为何需要更标准化的数据的关键。

通过制定具体标准或汇集大量一致的数据（如本研究所做），我们让研究人员能够获取便于相互比较和参考的数据。这有助于验证和交叉比对模型，并整体提升模型性能。此外，我们还可以进一步考察其他常被忽略或被认为不太重要的变量，例如照明、天气、一天中的时段、交叉口或环岛内的车道等。

路侧感知

路侧感知系统是智能交通系统、网联车辆和自动驾驶车辆的重要组成部分（*34–36*）。与车载感知系统相比，路侧感知系统受遮挡影响较小，并且在长期事件跟踪和预测方面表现更好（*37*）。（*37*）提出了一个大规模路侧三维目标检测数据集。（*38*）提出了一个大规模车路协同三维目标检测数据集。（*39*）开发了一种路侧感知单元（RSPU）。（*19,20*）设计并部署了一套用于环岛的路侧感知系统。路侧摄像机也用于交通冲突研究（*40*）。在大多数情况下，先从路侧感知系统中提取目标位置和轨迹，再将其用于交通冲突研究。然而，我们直接使用原始视频作为输入，而非轨迹。根据我们的观察，提取出的轨迹中存在的噪声会对交通冲突识别产生负面影响。在识别交通冲突时，我们使用感知系统进行交通冲突轨迹提取。

交通冲突识别

本节介绍我们如何识别和收集冲突数据。由于交通冲突属于罕见事件，完全依靠人工识别交通冲突并不现实。另一方面，由于交通冲突难以进行定量描述（与碰撞不同，碰撞可以通过道路使用者之间的物理接触来定义），因此仅依靠算法识别交通冲突也不可靠。因此，我们采用交通冲突识别算法来找出潜在的交通冲突，然后对其进行人工核查。

输入数据

我们首先将视频流转换为较短的视频片段，每段时长为30秒，采样率为2.5Hz。视频分辨率设为 960×720。由于鱼眼摄像机造成的畸变，环岛对角处的车辆在视频中显得很小且模糊。因此，我们分别使用四台鱼眼摄像机拍摄的视频，而不将四台摄像机的视频合并起来进行冲突识别。

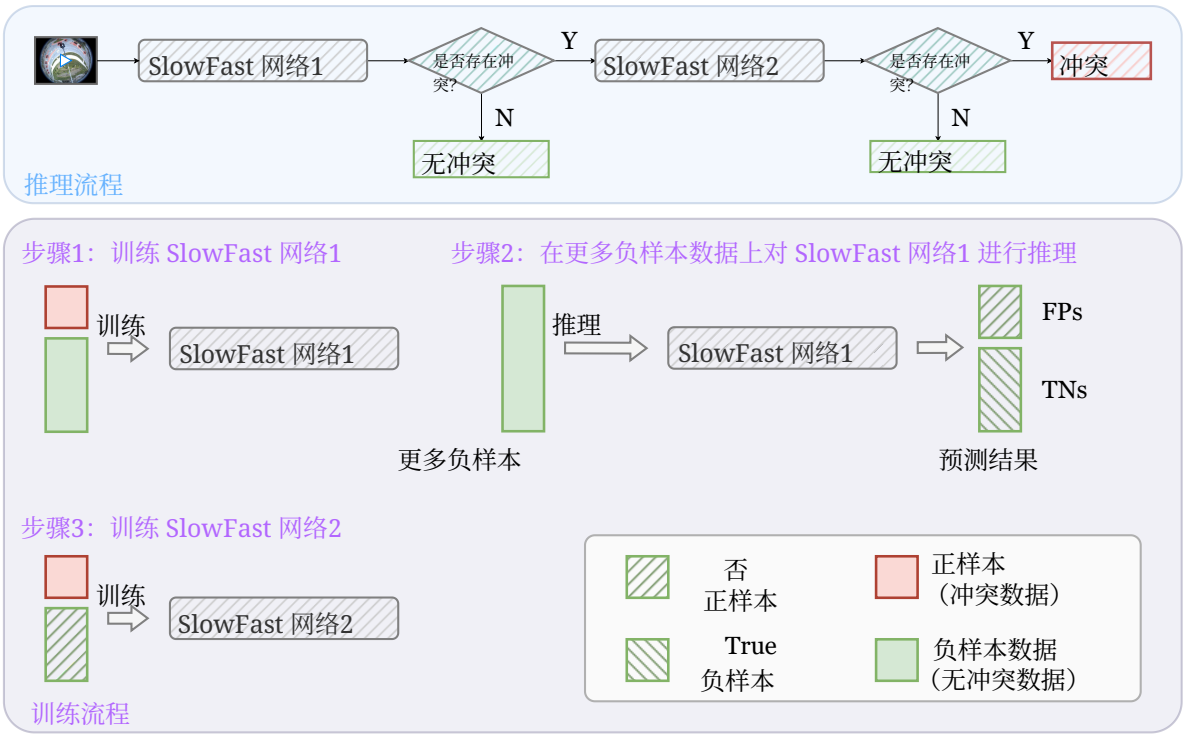


图3 级联冲突识别算法的推理和训练流程示意图。该算法由两个 SlowFast 网络组成。在推理阶段，只有当两个网络都预测存在冲突时，最终预测结果才为阳性。训练分为三个步骤：首先训练 SlowFast 网络1。然后，在更多负样本上对该网络进行推理，以获取 SlowFast 网络1的假阳性预测结果。最后，将这些假阳性作为负样本来训练 SlowFast 网络2。通过这种方式，SlowFast 网络2能够学会区分真实冲突与 SlowFast 网络1产生的假阳性。

冲突识别算法

首先，我们开发了一种基于规则的算法来识别潜在冲突，并对其进行人工检查和标注，得到232个初始交通冲突样本。随后，我们使用这232个交通冲突样本训练一个基于学习的冲突识别级联框架。如图3所示，该级联框架包含两个阶段。两个阶段的模型架构完全相同，均为采用二分类（冲突或非冲突）输出的SlowFast（*21*）网络。SlowFast 网络是一种常用的视频处理网络，以原始视频为输入。SlowFast 网络包含两条通路：一条慢速通路，以较低帧率运行，用于学习空间特征；另一条快速通路，以较高帧率运行，用于学习有用的时间信息。两个网络的架构相同，但由于训练过程不同，参数也有所不同。在推理过程中，只有当SlowFast 网络1和SlowFast 网络2都将样本预测为冲突时，我们才将最终预测结果判定为冲突。

训练时，如图3所示，我们首先将冲突视频设为正样本，并随机抽取负样本数据，以此构建 SlowFast 网络1 的训练集。然后使用该训练集训练 SlowFast 网络1。由于交通冲突在现实世界中非常罕见，SlowFast 网络1 的精确率相当低（7%）。SlowFast 的大多数预测结果

都是假阳性，这意味着大多数被预测为冲突视频的视频实际上并没有冲突。为了减少假阳性的数量，我们实现了另一个网络 SlowFast 网络2，专门用于筛除 SlowFast 网络1 的假阳性。我们在更多负样本上运行 SlowFast 网络1，并收集假阳性样本。然后，我们为 SlowFast 网络2 构建训练集：将冲突视频设为正样本，将 SlowFast 网络1 的假阳性设为负样本。这样，SlowFast 网络2 便经过训练，能够区分真实冲突与 SlowFast 网络1 的假阳性。

我们估算第一阶段模型和两阶段模型的精确率。精确率定义为

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

此处，TP 表示真阳性的数量，FP 表示假阳性的数量，FN 表示假阴性。精确率是所有检测到的冲突中真实冲突所占的比例。对于第一阶段模型，输出视频的精确率约为 7%。当我们将第一阶段模型输出的视频输入第二阶段模型进行再识别后，精确率提高到 23%。借助交通冲突识别模型，标注成本大幅降低。

数据标注

交通冲突标注与分类体系

基于收集到的潜在冲突样本，我们为 ROCO 设计了标注说明，并对潜在冲突样本进行标注。我们为交通冲突定义了三个要素：冲突的严重程度、冲突的原因，以及冲突对交通流的影响。

严重程度。 对于严重程度，我们在冲突数据集中定义了三个等级。等级0表示该数据为负样本，即片段中不存在交通冲突。有些车辆可能违反规则或法律，但并未使其他道路使用者面临碰撞危险。我们不将此类事件视为交通冲突。等级1表示片段中存在一个或多个交通冲突。等级2表示片段中发生了碰撞，即道路使用者之间发生了物理接触。由于我们在整个数据集中找不到统一的严重程度表示方式，因此不再定义更细的严重程度等级，只采用最简单的指标。未来可以在数据集中应用其他安全指标。

原因。 我们将环岛内交通冲突的原因划分为以下常见类别：

- (0) “进入环岛时未让行环行车辆”。根据密歇根州的环岛交通法规，车辆进入环岛时，应让行环岛内的车辆，并等待安全间隙后再进入。未能做到这一点可能引发安全关键事件。图4(a)展示了未让行的示例。

- (0*)“公交车/卡车/拖车等大型车辆进入环岛时未礼让环行车辆”。这是原因 (0) 的特殊情况，在我们的数据集中，其优先级高于原因 (0)。这意味着，如果大型车辆未能礼让并造成交通冲突，则会将其标记为原因 (0*)，而不是原因 (0)。由于它

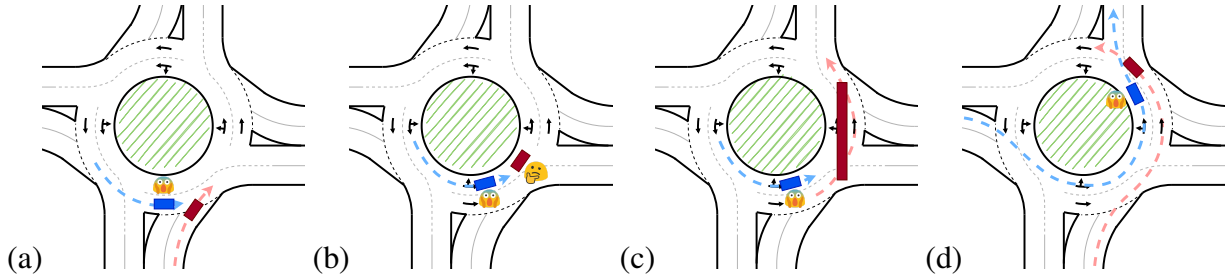


图 4 我们归纳的环岛交通冲突常见原因示意图。(a) 红色车辆驶入环岛时未礼让蓝色车辆。(b) 红色车辆在环岛内突然停车，造成追尾险情。(c) 红色卡车过长，占据环岛的两条车道，迫使蓝色车辆减速或停车。(d) 红色车辆本应从顶部出口驶出，却继续在环岛内绕行，造成碰撞险情。

对大型车辆而言，进入环岛时不干扰环岛交通流更加困难，因此我们将其单独列为一种原因。

- (1)“在环岛内不必要地突然制动/停车”。车辆在环岛内行驶时，除非遇到紧急情况，否则不应突然开始制动或停车。突然制动/停车可能导致追尾碰撞或碰撞险情，如图 4(b) 所示。
- (2)“卡车/公交车/拖车体型过大，妨碍其他道路使用者”。由于有些车辆过长或过宽，难以在环岛内舒适通行，因此可能占据部分两条车道。这会妨碍环岛内的其他车辆。大型车辆与周围车辆之间可能发生安全关键事件，如图 4(c) 所示。
- (3)“不当车道使用”。从环岛错误出口驶出，或在环岛内穿越车道，会扰乱交通流，可能导致侧擦碰撞。图 4(d) 展示了一个示例。
- (4)“其他原因”。许多其他原因可能导致交通冲突，包括交通拥堵，以及警车/救护车/消防车或其他应急车辆。我们将这些情况全部归入“其他原因”类别。

根据我们的观察，上述原因涵盖了该环岛交通冲突的大多数成因。对于其他环岛，未来可能需要相应扩展常见原因集合。借助这些交通冲突原因，交通运输研究人员和政府机构或许能够更好地改进驾驶教育、环岛处的交通标志等，以提高交通安全。

对交通流的影响。为了更好地了解交通冲突如何影响其他车辆，我们对交通冲突对环岛交通流的影响进行了标注。我们将交通冲突造成的干扰严重程度分为四个类别：

- (0) 交通流未受影响。即使是引发交通冲突的车辆，也不会减速或停车。例如，驾驶员猛打方向盘以避免碰撞，但幸运的是周围没有其他车辆，因此没有车辆被迫减速。
- (1) 一辆或两辆车辆因交通冲突而被迫减速。这是一个



图 5 交通冲突的提取轨迹可视化。五张图像展示了交通冲突期间的五个帧。蓝色车辆未礼让粉色车辆便驶入环岛，从而引发交通冲突。

常见场景，因为有些车辆会减速以避免冲突。

- (2) 由于交通冲突，一辆或两辆车辆被迫在环岛内停车。当交通冲突较为严重时，驾驶员可能需要大力制动以避免碰撞。车辆停在环岛内对其他驶来的车辆非常危险。

- (3) 三辆或更多车辆被迫在环岛内减速或停车。这意味着交通冲突严重影响了环岛内的交通。

车辆检测与跟踪系统

确定包含交通冲突的视频片段后，我们采用车辆检测与跟踪系统进行轨迹提取。我们重新使用了（*19, 20*）开发的路侧感知框架，但将车辆检测器替换为性能更强的 YOLOX（*41*）检测器。借助这一性能更强的检测器，我们能够提取更准确、更可靠的轨迹。图 5 展示了交通冲突轨迹提取的示例。

数据收集与标注效率

目前，借助冲突识别算法，我们成功大幅降低了数据收集和标注成本。与人工标注相比，冲突识别算法的推理速度非常快。使用一块 NVIDIA RTX 3080 GPU，对一段 30 秒的视频片段进行推理约需 1 秒。对于人工标注，我们估计，每段视频片段需要一人花费 60 秒来判断其中是否存在交通冲突（识别），并额外花费 20 秒标注交通冲突的严重程度、原因和影响。表 1 展示了在 ROCO 中收集并标注一个交通冲突数据样本的预估成本。在 ROCO 中，我们估计每 1,200 段视频片段中会有一段包含交通冲突，因此，如果完全依赖人工标注，收集并标注一个交通冲突将耗费约 20 小时的人工劳动。幸运的是，我们的交通冲突识别算法能够筛除大多数不包含冲突的视频片段。识别精确率约为 23% 时，经识别算法筛选出的约 4 段视频片段中，就有一段包含交通冲突。在这种情况下，每个样本的预估人工成本降至约 280 秒。如果我们获得完美的交通冲突识别算法，达到 100% 的精确率，人工成本还会进一步降低，但每个样本仍需 80 秒。

方法	GPU 时间	人工识别时间	人工标注时间	总劳动时间	
纯人工标注	0s	72,000s	20s	72,020s	
我们的方法	1，200s	260s	20s	280s	
完美识别	1,200s	60s	20s	80s	

表 1 ROCO 中一个交通冲突数据样本的预估数据收集和标注成本。GPU 时间指我们的交通冲突识别算法的 GPU 推理时间。人工识别时间指一个人判断视频中是否存在冲突所需的时间。人工标注时间指一个人标注交通冲突各项要素（严重程度、原因等）所需的时间。总人工时间指人工劳动所需的总时间。完美识别指一种理想情况，即冲突识别算法达到 100% 精确率。在这种情况下，每个样本所需的人工劳动仅为80 秒。与纯人工标注相比，我们的方法显著降低了总人工成本。

数据集统计与分析

该数据集包含 24/7 段视频，拍摄时间为 2021 年 8 月初至 2021 年 10 月底。每段被标记为存在潜在冲突的视频，都标注了相关信息，例如是否发生冲突/碰撞事故、冲突/碰撞事故的原因，以及事件对交通流的影响。

对每个潜在冲突进行标注和标记后，我们对数据集进行了基本统计，以分析结果。总计发生了 17 起碰撞事故和 557 起交通冲突。在这些冲突和碰撞事故中，最常见的原因是未让行于已进入环岛的车辆，其次是卡车、公交车或拖车未让行，第三常见的原因是不当车道使用。图 6 展示了所有原因及其在冲突和碰撞事故中所占的比例。这些发现颇为有趣，因为这三种原因都相当容易避免。对于未让行的情况，驾驶员或许需要积累更多环岛驾驶经验，以了解并适应何时驶入、何时等待。对于不当车道使用，驾驶员可能需要接受更多关于环岛及车道的教育，同时环岛处也应增设交通标志或使标志更加清晰。

我们还按一天中的小时对冲突和碰撞事故进行了分类，如图 7(a) 所示。可以看出，人们在下午发生碰撞事故的频率远高于上午，尤其是下午 3 点至 6 点之间；这可能是因为下午交通流量增加。中午时分似乎也发生了大量冲突。值得注意的是，午夜至凌晨 6 点之间没有标记出任何潜在冲突。这表明在光线较暗时，摄像头或程序可能难以检测车辆和／或碰撞事故，不过夜间驾驶员较少，这或许也能部分解释这一现象。我们还按星期几对碰撞事故和冲突进行了分类，如图 7(b) 所示。一个有趣的现象是，与工作周的大多数其他日子相比，星期二的冲突数量明显更多。原本可能会预期各天的数量大致相同。这或许可以用未纳入统计的本地事件来解释。

为了更好地了解冲突的不同原因对交通流的影响，我们绘制了几张图表，即图8所示的图表。这些图表展示了具有特定原因的冲突中，对交通流严重程度等级产生特定影响的冲突所占百分比。从图中可以得出一个有趣的结论：不当使用车道最有可能导致交通流影响的严重程度达到3级，这意味着有3辆或更多车辆减速或停车。虽然所有

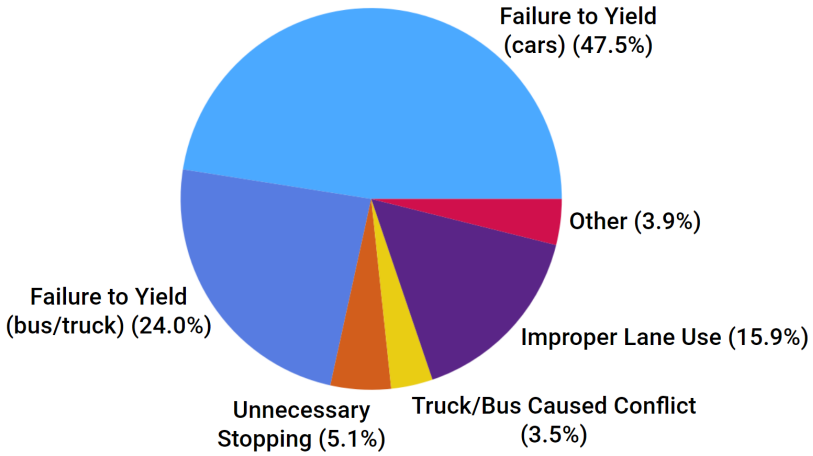


图 6 按原因划分的碰撞事故和冲突。由不同原因导致的冲突和碰撞事故所占比例。未让行显然是最主要的原因，其次是不当车道使用。这表明，驾驶员可能不熟悉环岛，或者需要在更安全或受控的环境中积累更多驾驶经验。

其他原因显然都表明，最可能对交通流产生的影响是严重程度为1级，即只有一两辆车减速。

总体而言，ROCO数据集包含17起碰撞事故和557起交通冲突，其中绝大多数是由于未让行迎面驶来的车辆所致。这些冲突和碰撞事故大多发生在下午3点至6点左右；也大多发生在工作日，其中星期二达到峰值。大多数冲突和碰撞事故对交通流的影响较小，而不当使用车道往往会造成更显著的影响。

局限性与未来工作

ROCO数据集存在一些局限性。首先，由于人工成本较高，目前的数据规模相对较小；其次，我们所采用的交通冲突识别算法可能带来潜在的数据偏差问题。本节还将讨论如何将数据泛化至更多交通场景。

数据集规模

我们当前的数据集仅包含557起交通冲突和17起碰撞事故。如数据收集部分所述，在ROCO数据集中识别一起交通冲突的人工成本约为280秒。由于交通冲突是一种非常专业的事件，我们只能信任具备交通运输知识的人员进行数据标注。我们计划进一步降低数据标注的人工成本，以便收集大规模交通冲突数据。

潜在数据偏差

由于我们使用交通冲突识别算法来识别潜在的交通冲突，因此识别算法可能会引入一些数据偏差。例如，该算法可能更容易识别由“未礼让”导致的交通冲突，而不是由其他原因导致的交通冲突。由于我们没有某一特定时段内的全部交通冲突数据，目前无法估算数据偏差问题。我们将在

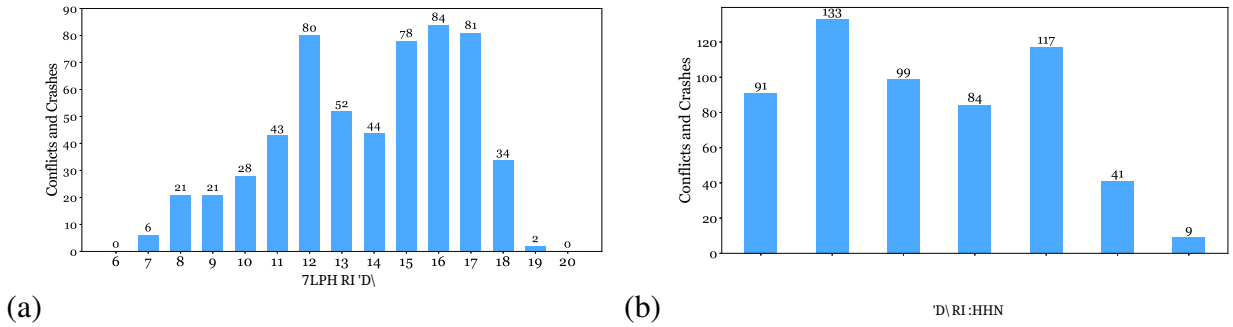


图 7 （a）按一天中的时段划分的冲突和碰撞事故。（b）按星期几排列的冲突和碰撞事故。（a）图中按小时显示了早上 6 点至晚上 8 点之间的所有碰撞事故和冲突。数据显示，随着一天的推进，冲突总体呈上升趋势，直到下午 6 点左右才开始减少。这至少在一定程度上可能是因为标准工作日通常在下午 5 点左右结束。（b）有趣的是，星期二的冲突数量明显多于工作日中的大多数其他日子。工作日每天的数量相当一致似乎才是正常情况。

未来。

推广到其他交通场景

目前，我们仅收集了一个环岛的交通冲突数据。不同环岛或不同交叉口的交通冲突模式和统计特征可能有所不同。我们计划在不同的交叉口和环岛收集更多交通冲突数据，以增加数据集的多样性和覆盖范围。

结论与更广泛的影响

本文介绍了 ROCO，一个环岛交通冲突数据集。通过采用交通冲突识别算法，我们将标注成本降低到了可承受的范围。我们为交通冲突提供了多种标签，包括严重程度、原因以及对交通流的影响。我们基于这些数据进行了基础统计分析。我们发现，在我们的数据集中，进入环岛时未让行是导致交通冲突的首要原因。

与事故数据不同，事故数据必须等到发生一起碰撞事故才能获得一条记录，而交通冲突数据可以在没有实际碰撞事故发生的情况下收集。借助数据收集与标注流程，我们相信能够持续向研究界和政府机构提供越来越多真实的交通冲突数据。ROCO数据集公开后，世界各地的交通运输研究人员可以开展更多分析。根据统计分析结果，可以采取相应措施（例如，使交通标志更加清晰），以预防环岛中的交通冲突及潜在碰撞事故。此外，在自动驾驶汽车开发中，在安全关键场景下测试自动驾驶车辆对于确保其可靠性至关重要。我们的 ROCO 数据集有望用作安全关键场景下的自动驾驶汽车测试数据集。

致谢

作者谨此感谢密歇根大学安娜堡分校的 Mcity 提供经费支持。作者还要感谢密歇根大学

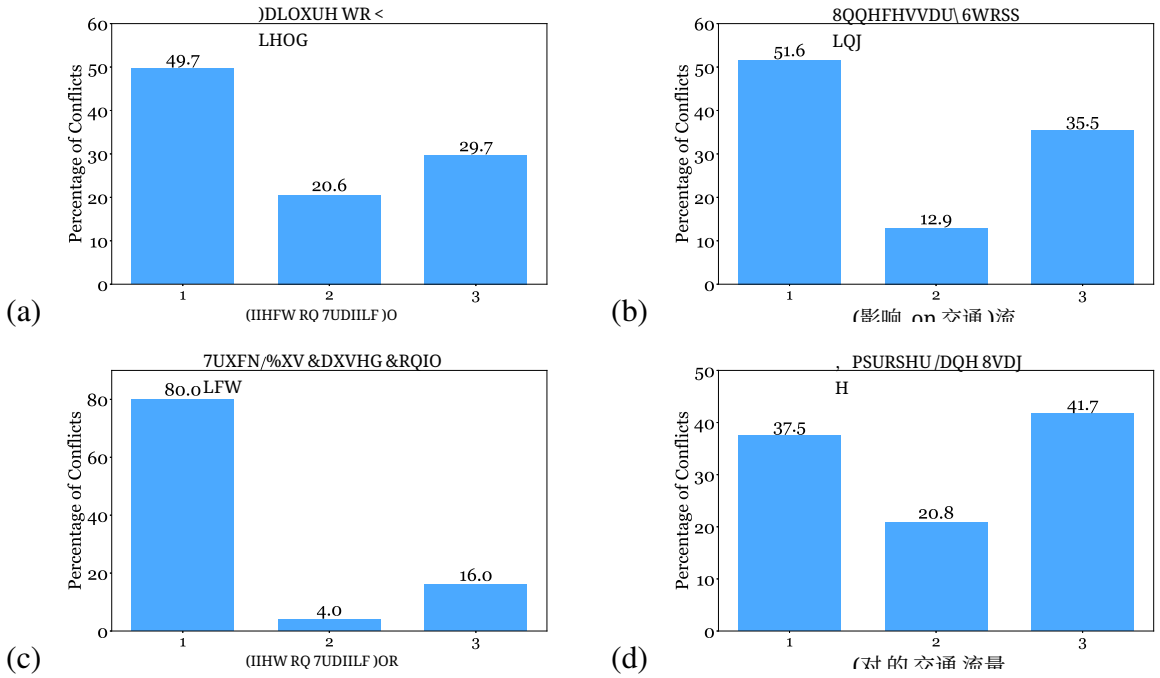


图8 这些图展示了四种不同冲突原因所对应的交通流影响中，各类冲突所占的百分比：(a) 未让行，(b) 不必要停车，(c) 卡车/公交车引发的冲突，(d) 不当使用车道。等级0–3表示对交通流影响的严重程度。因此，0表示没有影响，1表示1–2辆车减速，2表示1–2辆车停车，3表示3辆或更多车辆减速或停车。例如，约49.7%的未礼让冲突严重程度为1，因此导致1–2辆车减速。

安娜堡分校密歇根交通实验室的所有成员。他们在标注交通冲突数据方面提供的帮助，对ROCO数据集和本文至关重要。

作者贡献

作者确认对本文的贡献如下：研究构思与设计：Depu Meng、沈胜银、刘恒；数据收集：Depu Meng、Owen Sayer、张如生、沈胜银；结果分析与解读：Depu Meng、Owen Sayer；论文初稿撰写：Depu Meng、Owen Sayer、张如生、沈胜银、李厚强、刘恒。所有作者均审阅了研究结果，并批准了论文最终版本。

参考文献

1. Liu, H. X. 和 S. Feng, 《自动驾驶车辆的“稀缺性诅咒”》。*arXiv*，2022年。 2. Feng, S.、X. Yan、H. Sun、Y. Feng 和 H. X. Liu，利用自然驾驶与对抗性环境对自动驾驶车辆进行智能驾驶智能测试。《自然·通讯》，第12卷，第1期，2021年，第1–14页。 3. Glauz, W. D. 和 D. J. Migletz，交通冲突分析在交叉口的应用，1980年。 4. Cooper, P, 《加拿大交通冲突研究现状报告》。载于第一届交通冲突研讨会论文集，1977年。 5. Hydén, C.，交通安全评价方法的发展：瑞典交通冲突技术。隆德理工学院技术公报，系，，第70号，1987年。 6. Tarko, A. P.，利用交通冲突和洛马克斯分布估算预期事故数——理论与数值探索。《事故分析与预防》，第113卷，2018年，第63–73页。 7. Davis, G. A.、J. Hourdos、H. Xiong 和 I. Chatterjee，交通冲突与事故因果模型概要。《事故分析与预防》，第43卷，第6期，2011年，第1907–1919页。 8. Madsen, T. K. O. 和 H. Lahrmann，利用交通冲突研究比较信号交叉口的五种自行车设施设计。《交通研究*F*辑：交通心理学与行为》，第46卷，2017年，第438–450页。 9. Beitel, D.、J. Stipancic、K. Manaugh 和 L. Miranda-Moreno，利用骑行者与行人交互及自动化视频冲突分析评估共享空间的安全性。《交通研究*D*辑：交通与环境》，第65卷，2018年，第710–724页。 10. Ma, Y.、X. Qin、O. Grembek 和 Z. Chen，结合冲突概率与严重程度构建无信号控制交叉口安全热力图。《事故分析与预防》，第113卷，2018年，第303–316页。 11. Brow, G. R.，道路使用者安全研究中的交通冲突。《加拿大土木工程期刊》，第21卷，第1期，1994年，第1–15页。 12. Tarrall, M. B. 和 K. K. Dixon，保护式与许可式兼用信号相位下双左转车道的冲突分析。《交通研究记录》，第1635卷，第1期，1998年，第105–112页。 13. Ismail, K.、T. Sayed 和 N. Saunier，行人与车辆冲突的自动化分析：前后对比研究的背景。《交通研究记录》，第2198卷，第1期，2010年，第52–64页。 14. Tiwari, G.、D. Mohan 和 J. Fazio，利用冲突分析预测混合交通流中的致命事故地点。《事故分析与预防》，第30卷，第2期，1998年，第207–215页。 15. Hossain, M.、M. Abdel-Aty、M. A. Quddus、Y. Muromachi 和 S. N. Sadeek，实时事故预测模型：现状、设计路径与普遍性要求。《事故分析与预防》，第124卷，2019年，第66–84页。 16. Machiani, S. G. 和 M. Abbas，安全替代指标直方图（SSH）：针对信号交叉口困境区相关冲突的一种新型实时安全评估方法。《事故分析与预防》，第96卷，2016年，第361–370页。

17. Essa, M. 和 T. Sayed，采用交通冲突模型评估信号交叉口在周期层面的安全性。《交通研究*C*辑：新兴技术》，第89卷，2018年，第289–302页。18. Zheng, L.、T. Sayed 和 F. Mannering，面向道路安全分析的交通冲突建模：分析方法综述与未来方向。《事故研究分析方法》，第29卷，2021年，第100142页。19. Zhang, R.、Z. Zou、S. Shen 和 H. X. Liu，路侧协同感知系统的设计、实现与评估。《交通研究记录》，第0卷，第0期，0，第03611981221092402页。20. Zou, Z.、R. Zhang、S. Shen、G. Pandey、P. Chakravarty、A. Parchami 和 H. X. Liu，利用路侧摄像机进行自动驾驶实时全栈交通场景感知，2022年。21. Feichtenhofer, C.、H. Fan、J. Malik 和 K. He，用于视频识别的 SlowFast 网络。载于*IEEE/CVF*国际计算机视觉大会论文集，2019年，第6202–6211页。22. Vogel, K.，车头时距与碰撞时间作为安全指标的比较。《事故分析与预防》，第35卷，第3期，2003年，第427–433页。23. Guido, G.、F. Saccomanno、A. Vitale、V. Astarita 和 D. Festa，比较基于视频采集数据得出的安全绩效指标。《交通工程学报》，第137卷，第7期，2011年，第481–491页，被引次数：56。24. Laureshyn, A.、T. De Ceunynck、C. Karlsson、Å. Svensson 和 S. Daniels，探寻交通事件的严重程度维度：将扩展Delta-V作为交通冲突指标。《事故分析与预防》，第98卷，2017年，第46–56页。25. Yorio, P. 和 S. Moore，利用来自25,000多家机构的特定场所健康与安全数据，考察影响海因里希安全金字塔存在性的因素：考察影响海因里希安全金字塔存在性的因素。《风险分析》，第38卷，2017年。26. Orsini, F.、G. Gecchele、R. Rossi 和 M. Gastaldi，一种用于实时道路安全分析的基于冲突的方法：与基于事故的模型进行比较评估。《事故分析与预防》，第161卷，2021年，第106382页。27. Glauz, W. D.、K. M. Bauer 和 D. J. Migletz，预期交通冲突率及其在预测事故中的应用。《交通研究记录》，1985年。28. Zheng, L.、K. Ismail 和 X. Meng，道路安全分析中的交通冲突技术：未解问题与一些见解。《加拿大土木工程期刊》，第41卷，2014年。29. Ye, F. 和 D. Lord，事故数据漏报对三种常用交通事故严重程度模型的影响研究。《交通研究记录：交通研究委员会期刊》，第2241卷，2011年，第51–58页。30. Zheng, L.、K. Ismail 和 X. Meng，采用平移伽马-广义帕累托分布模型绘制安全连续谱并估算碰撞事故。《安全科学》，第64卷，2014年，第155–162页。31. El-Basyouny, K. 和 T. Sayed，利用交通冲突构建安全绩效函数。《安全科学》，第51卷，第1期，2013年，第160–164页。

32. Xie, K.、C. Li、K. Ozbay、G. Dobler、H. Yang、A.-T. Chiang 和 M. Ghandehari，基于视频的安全评估综合框架的开发，2016年，第2638–2643页。33. Nofal, F. 和 A. Saeed，利雅得市道路交通事故的季节变化与天气影响。《公共卫生》，第111卷，第1期，1997年，第51–55页。34. Chen, Q.、S. Tang、Q. Yang 和 S. Fu，Cooper：基于三维点云的网联自动驾驶车辆协同感知。载于*2019*年*IEEE*第*39*届国际分布式计算系统会议（*ICDCS*），IEEE，2019年，第514–524页。35. Rauch, A.、F. Klanner、R. Rasshofer 和 K. Dietmayer，面向协同感知系统的基于Car2x的高层融合架构感知。载于*2012*年*IEEE*智能车辆研讨会，IEEE，2012年，第270–275页。36. Wang, T.-H.、S. Manivasagam、M. Liang、B. Yang、W. Zeng 和 R. Urtasun，V2vnet：用于联合感知与预测的车对车通信。载于欧洲计算机视觉会议，Springer，2020年，第605–621页。37. Ye, X.、M. Shu、H. Li、Y. Shi、Y. Li、G. Wang、X. Tan 和 E. Ding，*Rope3D*：面向自动驾驶和单目三维目标检测任务的路侧感知数据集，2022年。38. Yu, H.、Y. Luo、M. Shu、Y. Huo、Z. Yang、Y. Shi、Z. Guo、H. Li、X. Hu、J. Yuan 和 Z. Nie，*DAIR-V2X*：大规模车路协同三维目标检测数据集，2022年。39. Tsukada, M.、T. Oi、M. Kitazawa 和 H. Esaki，面向自动驾驶的联网路侧感知单元。《传感器》，第20卷，第18期，2020年，第5320页。40. Autey, J.、T. Sayed 和 M. H. Zaki，采用自动化交通冲突分析评估右转智能通道的安全性。《事故分析与预防》，第45卷，2012年，第120–130页。41. Ge, Z.、S. Liu、F. Wang、Z. Li 和 J. Sun，*YOLOX*：*2021*年超越*YOLO*系列，2021年。