

doi: 10.19562/j.chinasae.qcgc.2024.ep.005

# 基于有向图的城市交叉口场景相似性评价方法\*

李江坤<sup>1</sup>, 纵瑞雪<sup>1</sup>, 邓伟文<sup>1</sup>, 王莹<sup>2</sup>, 丁娟<sup>3</sup>(1. 北京航空航天大学交通科学与工程学院, 北京 100191; 2. 吉林大学计算机科学与工程学院, 长春 130025;  
3. 浙江天行健智能科技有限公司, 嘉兴 314000)

**[摘要]** 准确的场景相似性评估对于优化测试场景集极为重要。然而, 现有的基于轨迹的评价方法未能充分捕捉城市交叉口中车辆间的复杂动态交互特征, 从而影响评价结果的准确性。针对此问题, 本研究提出了一种基于有向图的城市交叉口场景相似性评价方法, 该方法通过比较两个场景中车辆的全局交互拓扑关系在时空上的匹配程度来量化场景之间的相似性。首先, 使用有向图来表征各个交叉口中车辆间的交互拓扑结构。然后, 通过比较不同交叉口场景的有向图结构匹配程度来估计它们之间的交互相似性。最后, 采用动态时间扭曲算法在时间维度上对齐场景, 实现两个不同长度测试场景序列的有效比较。3对典型评价案例的定性分析结果显示, 该方法能够细粒度地区分不同相似度水平的场景。进一步地, 为定量验证方法的有效性, 开展了场景相似性与自动驾驶系统性能之间的方差分析实验。实验结果表明, 在不同相似度水平场景簇的测试条件下, 系统的安全性和效率均表现出显著的差异, 从而证实了该方法的有效性。最终, 该方法应用于Apollo的测试场景集的优化, 结果表明方法能够有效指导测试场景集中同质化场景的剔除, 从而提高测试效率。

**关键词:** 自动驾驶; 城市交叉口场景; 相似性评价; 有向图; 动态时间扭曲

## Directed Graph-Based Method for Evaluating Similarity in Urban Intersection Scenarios

Li Jiangkun<sup>1</sup>, Zong Ruixue<sup>1</sup>, Deng Weiwen<sup>1</sup>, Wang Ying<sup>2</sup> & Ding Juan<sup>3</sup>1. School of Transportation Science and Engineering, Beihang University, Beijing 100191;  
2. College of Computer Science and Technology, Jilin University, Changchun 130025;  
3. Zhejiang Tianxingjian Intelligent Technology Co., Ltd., Jiaxing 314000

**[Abstract]** Accurate evaluation of scenario similarity is extremely important for optimizing test scenarios. However, existing trajectory-based evaluation methods fail to adequately capture the complex dynamic interaction characteristics between vehicles at intersections, which affects the accuracy of the evaluation results. To address this problem, in this study a directed graph-based similarity evaluation method for urban intersection scenes is proposed, which quantifies the similarity between scenes by comparing the degree of spatial and temporal matching of the global interaction topologies of vehicles in two scenarios. Firstly, a directed graph is used to characterize the interaction topology between vehicles at each urban intersection. Then, the interaction similarity between different intersection scenarios is estimated by comparing the degree of matching of their directed graph structures. Finally, a dynamic time warping algorithm is used to align the scenarios in the time dimension to effectively compare two test scenario sequences of different lengths. The results of the qualitative analysis of three pairs of typical evaluation cases demonstrate that the method is capable of distinguishing scenes with different similarity levels at a fine-grained level. Furthermore, to quantitatively validate the effectiveness of the method, an ANOVA experiment is conducted

\* 浙江省“领雁”研发攻关计划项目(2023C01238)、浙江省“尖兵”研发攻关计划项目(2023C01133)和嘉兴市重点研发计划项目(2023BZ10002)资助。

原稿收到日期为2024年05月11日, 修改稿收到日期为2024年08月18日。

通信作者: 王莹, 副教授, 博士, E-mail: wangying\_jlu@163.com。

to compare scenario similarity with the performance of the autopilot system. The experimental results reveal that the safety and efficiency of the system exhibit significant differences under test conditions with different levels of scenario similarity, thus proving the method's effectiveness. Ultimately, this method is applied to optimize Apollo. Ultimately, this method rformance of the autopilot system. The experimental results reveal that the safety and efficiency of the system exhibit significantd

**Keywords: autonomous driving; urban intersection scenarios; similarity evaluation; directed graph; dynamic time warping**

## 前言

随着自动驾驶技术的快速发展,确保这些系统的安全和可靠性成为了至关重要的挑战<sup>[1-3]</sup>。特别是在城市交叉口,道路网络中最复杂的部分之一,它被视为测试和验证自动驾驶车辆的最有挑战性的区域<sup>[4-7]</sup>。在路口中,车辆面临多变的交通环境和复杂的车辆互动,对自动驾驶系统的反应速度和适应性提出了极高要求。因此,对城市交叉口场景进行充分的测试验证至关重要。

传统的测试方法,如封闭场地测试和实际道路测试,往往难以在安全可控的环境中充分模拟现实世界的复杂性<sup>[8-9]</sup>。因此,基于场景的测试方法应运而生,允许研究人员在虚拟环境中进行广泛的测试以全面验证和优化自动驾驶系统<sup>[10-11]</sup>。

在仿真测试中,基于机理或数据驱动的场景模型能够迅速生成复杂且多样的海量虚拟测试场景,从而有效地支持自动驾驶系统的全面测试。然而,在大规模自动生成测试场景的过程中,通常会伴随着大量相似场景的产生。这些同质化的场景不仅无法提供新的测试见解,而且还会额外消耗大量计算资源<sup>[12-13]</sup>。若能在测试前对场景集进行精细化的相似性评估,便可有效剔除这些同质化场景,从而显著减少测试场景集的规模。因此,进行场景相似性评估对于提高测试效率具有重要意义。

目前,城市交叉口场景相似性评价的研究仍相对有限。学者们通常通过计算场景中所有车辆之间的轨迹距离差异来间接量化两个场景的相似性。例如,Ries等<sup>[14]</sup>基于轨迹相似性进行场景聚类。Zhang等<sup>[15]</sup>基于生成轨迹与实际轨迹之间的距离来评估重建场景的保真度。然而,基于轨迹的方法无法有效捕捉交叉口中车辆间的复杂动态交互特征,从而影响评价结果的准确性。此外,这些评价结果显著受到道路结构的影响,导致在不同道路结构下

的场景评价结果缺乏一致性。例如,即使不同道路结构上的左转与直行场景在交互模式上可能类似,但由于道路结构的不同,这些场景的行驶轨迹可能会有显著差异,导致看似相似的场景产生完全不同的评估结果。

为了弥补现有研究的不足,本研究提出一种基于有向图的城市交叉口场景相似性评价方法。该方法通过比较两个场景全局交互拓扑关系的匹配程度来量化场景之间的相似性。方法主要包括3个部分。首先,文中采用有向图结构化地表达城市交叉口内车辆间的复杂交互关系。其次,通过对图的分解、聚类和加权求和,评估两个场景中车辆全局交互状态的匹配程度。最后,文中采用基于动态时间扭曲(DTW)的方法,通过时间维度的非线性扭曲,实现不同长度时间序列的有效匹配,以提高评估的准确性和适用性。

## 1 场景相似性评价方法

如图1所示,该方法针对一对包含若干帧的测试场景片段进行评估,主要通过比较两个场景片段中交通车辆间交互状态的匹配程度来评估它们的相似性。

首先,文中基于城市交叉口内所有交通车辆的交互关系构建场景交互拓扑图,将城市复杂交叉口场景转化为图的形式。其中,节点代表车辆,边的方向和权重则分别表示车辆间的主从交互关系和交互风险。接下来,基于闭环检测将该拓扑图分解成多个闭环子图,每个子图均代表一种特定的交互模式。然后,利用图结构相似性检测算法对子图进行聚类,归类那些拓扑结构相同但边权重不同的子图,这些子图反映了类似的交互模式。在同一类中,两个同构子图的邻接矩阵之间的Frobenius范数差异用以衡量子图间的相似度。最后,通过基于数量占比的加权求和方法,计算出任意时刻两个场景之间的相

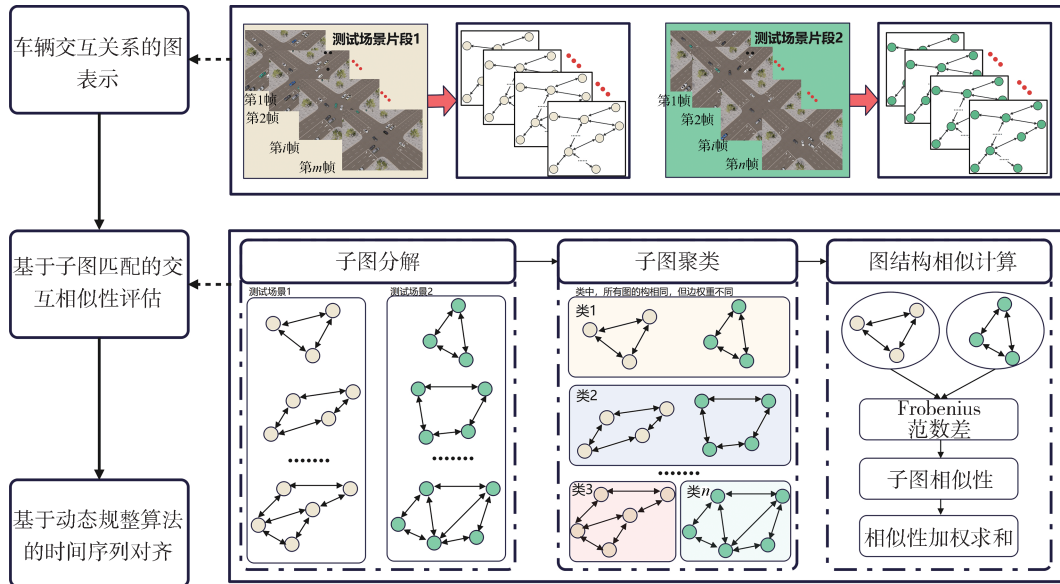


图1 城市交叉口场景相似性评估方法框架

似度。所有时间上的相似性累加平均后,得到两个场景片段的最终相似性值。如果测试场景片段的序列长度不一致,文中将使用动态时间扭曲(DTW)算法通过时间维度的非线性扭曲对齐两个测试场景。

### 1.1 车辆交互关系的图表示

在城市交叉口中,车辆之间的交互关系复杂多样,每辆车的行为都可能直接或间接影响其他车辆<sup>[16-17]</sup>。为了准确描述这些动态交互,需要一种能够全面捕捉并表示这些关系的方法。图提供了一种结构化的方式来表示复杂系统中的多个实体及其相互关系。在图中,实体间的交互关系可以通过边的方向、类型和权重灵活地表达。

基于以上分析,本研究提出了一种基于有向图的新方法来表示城市交叉口车辆交互关系。这种方法被称为场景交互拓扑图,它可以形式化地定义为一个三元组:

$$G = (V, E, W) \quad (1)$$

式中: $V$ 是顶点集合, $V = \{v_1, v_2, \dots, v_n\}$ ,每个顶点 $v_i$ 代表交叉口内的一个交通车; $E$ 是有向边集合, $E \subseteq \{(v_i, v_j) | v_i, v_j \in V \text{ and } i \neq j\}$ ,每条边 $(v_i, v_j)$ 表示车辆 $v_i$ 对车辆 $v_j$ 施加的交互作用; $W$ 是一个映射函数 $W: E \rightarrow \mathbb{R}^+$ ,为图中的每条边赋予一个正的权重值,表示交互风险。

本研究将城市交叉口中的交互作用分为两类:跟驰交互和冲突交互。如图2所示,跟驰交互特指同一车道上前车对后车的影响,这是一种单向的交互作用;而冲突交互发生在两条相交车道上,涉及到

尚未通过交叉点的两辆车之间的双向交互。在本研究中,相交车道指的是路口内交叉的行驶方向车道,例如由南向西的左转车道与由北向南的直行车道为一对相交车道。为了简化模型,冲突交互被定义为发生在交叉车道上且距离冲突点最近的两辆车之间。

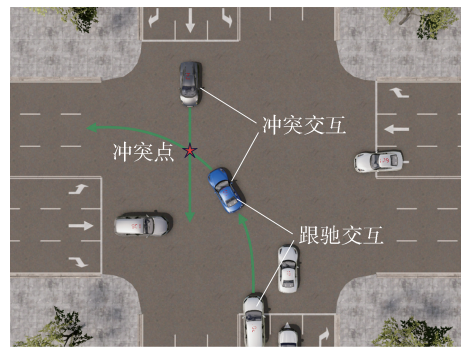


图2 两种交互作用示意图

跟驰交互的交互风险基于碰撞时间TTC(time-to-collision)量化,计算公式如下:

$$TTC_{i,j} = \frac{d_{ij}}{\Delta v_{ij}} \quad (2)$$

式中: $d_{ij}$ 是车辆 $v_i$ 和车辆 $v_j$ 之间沿车道中心线的距离; $\Delta v_{ij}$ 是它们的速度差。进一步地,文中使用高斯函数对其归一化,计算公式如下:

$$\omega_{i,j}^f = \begin{cases} e^{-\frac{(TTC_{i,j})^2}{2c^2}}, & TTC_{i,j} \leq 0 \\ 0, & \text{其他} \end{cases} \quad (3)$$

式中 $c$ 为平滑超参数,参照文献[18], $c$ 的取值为2。

需要注意的是,当 $TTC_{ij} > 0$ 时,两车处于相对远离状态,无交互风险,所以采用分段函数的设计将对应的交互权重设置为0。

冲突交互权重是基于车辆相对于冲突点的交互风险来评估的。到达冲突点的距离和时间差是评估两车交互风险的两个关键指标,分别从空间和时间维度进行量化。车辆离冲突点较远时,随着距离减小,交互风险逐渐增加;而当距离低于一定阈值时,由于安全空间减少,交互风险将急剧上升。从分析中得知,交互风险与距离的关系呈指数函数变化。因此,采用负指数函数来构建距离与交互风险的变化函数。同时,到达冲突点的时间差反映了两车在冲突点相遇的可能性;在相同距离下,时间差越小,交互风险越高。根据文献[19]中的评价框架,文中将距离视为影响交互风险的基本条件,时间差是风险的调节因子,并采用高斯函数对调节因子进行归一化。综上所述,冲突交互权重的计算公式如下:

$$\omega_{i,j}^c = e^{-(d_i/p)} \times e^{-\frac{(\Delta T_{i,j})^2}{2c^2}} \quad (4)$$

式中 $p$ 为超参数,本研究中 $p$ 参数经验取值为10。

图3是一个场景交互拓扑图的构造示例。图3(a)是城市交叉口动态交通场景的关键帧截图,图3(b)是对应的场景交互拓扑图。为了提高图的可读性,图中仅标记了2组权重:2车对1车施加的跟驰交互权重 $w_{2,1}^f$ ,4车和9车之间相互施加的冲突交互权重 $w_{4,9}^c$ 和 $w_{9,4}^c$ 。

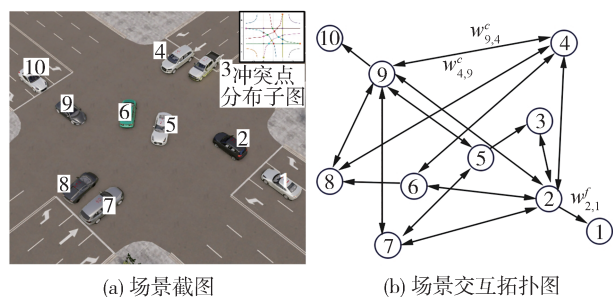


图3 场景交互拓扑图构造示例

## 1.2 基于图的交互相似性评估

本节将交互拓扑图分解成多个子图,每个子图均代表场景中的一个特定的交互模式。通过综合比较这些子图之间的相似性,来全面评估两个场景在交互层面上的整体相似性。

### 1.2.1 基于连通量的子图分解

在图论领域,连通性是有向图的一个关键特征。

定义上,一个有向图被称为连通的,如果图中的任意两个顶点 $u$ 和 $v$ 之间都存在从 $u$ 到 $v$ 的有向路径以及从 $v$ 到 $u$ 的有向路径,称为强连通。这种强连通性与城市交叉口车辆间的交互特征类似,即路口内一组车辆之间的任何一辆车都能直接或间接地影响到组内的其他所有车辆。

因此,本研究使用连通子图来描述路口交通车辆之间的交互模式。例如,如图4所示交通车辆5、7、9之间通过彼此的冲突交互直接相互影响,构成一个连通子图。

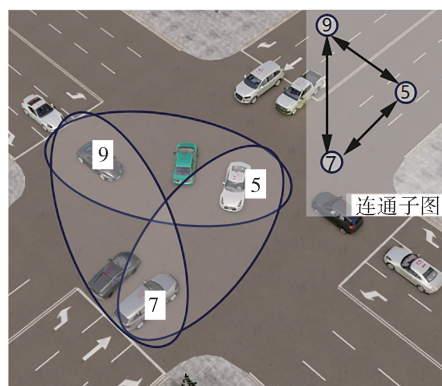


图4 连通子图示意图

为了有效识别这些子图,文中采用了Tarjan闭环检测算法<sup>[20]</sup>来遍历整个交互拓扑图 $G$ 。筛选阈值设置为子图顶点数量大于等于3。最终,场景交互拓扑图 $G$ 被表示为一组连通子图的集合 $\{CG_1, CG_2, \dots, CG_n\}$ ,其中每个连通子图 $CG_i$ 代表路口内车辆间的一种特定的交互模式。

此外,文中还考虑了由跟驰交互导致的车辆间的间接影响。如图5所示,两个或更多的连通子图通过单向的跟驰交互箭头连接起来,形成一个复合子图。这种结构表示了车辆间通过跟驰交互产生的间接影响关系。需要注意的是,车辆11位于车辆10左侧的左转车道上,但由于其距离交叉口较远,因此未在截图中显示出来。

最终,场景交互拓扑图 $G$ 被表示为由连通子图和复合子图构成的集合 $\{CG_1, CG_2, \dots, CG_n, HCG_1, \dots, HCG_m\}$ 。

### 1.2.2 基于同构检测的子图聚类

子图聚类的目标是将结构上完全相同的同构子图分类到同一组中。每个聚类中的子图揭示了类似的交互模式,而主要的差异则体现在边的权重上。通过专注于结构相同的子图进行相似性评估,这种

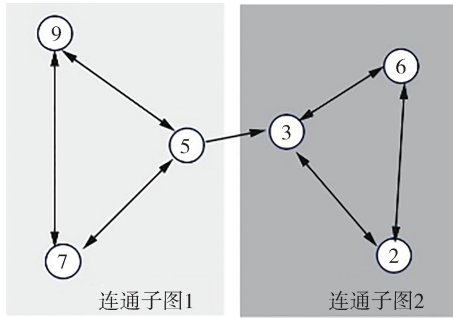


图5 复合子图示例

方法有效地降低了整体图相似性评估的计算复杂性。

图同构检测是图论中的一个核心问题,其目的是确定两个图在结构上是否完全相同。这通常意味着两个图可以通过某种顶点重标记映射成彼此。具体来说,如果存在一种顶点的一一映射  $f: V(G_1) \rightarrow V(G_2)$ ,使得一个图的顶点和边的连接方式可以完全转换成另一个图,则称这两个图是同构的。本研究使用 Ullmann's 算法<sup>[21]</sup>确定两个图是否同构。

### 1.2.3 同构子图的相似性评估

当两个同构子图之间的边权重越接近时,可以认为这两个子图在交互状态上越相似。因此,文中使用两个同构子图的邻接矩阵之间的 Frobenius 范数的差作为评估指标。

然而, Frobenius 范数差的计算高度依赖于两个图的顶点排序。不同的顶点排序会导致邻接矩阵的不同表示,从而影响到最终的相似性评估结果。例如,即使是结构和权重相同的图,在顶点排序不同时,相似性评估结果也会显示出显著差异。

为了解决这个问题,本文提出了一种新的度量方法:计算所有可能顶点排序组合(顶点间前后相对顺序不变)下的 Frobenius 范数的最小值,作为两个同构子图的最终相似性评价价值,计算公式如下:

$$FD_{cc,cc} = \min_{\pi \in S_n} \left( \frac{\|P_{\pi} M_i P_{\pi}^T - P_{\pi} M_j P_{\pi}^T\|_F}{\|M_i\|_F + \|M_j\|_F} \right) \quad (5)$$

式中:  $\|X\|_F$  表示矩阵  $X$  的 Frobenius 范数;  $\pi$  表示顶点的一个排列;  $S_n$  是所有可能的顶点循环排列的集合;  $P_{\pi}$  是根据排列  $\pi$  构造的置换矩阵,用于重新排序邻接矩阵的行和列。

### 1.2.4 相似性综合加权计算

首先,针对每类同构子图,计算所有可能的组合

的相似度,从而得出两个场景在每一类中的相似度,公式如下:

$$S_{\text{class}} = \frac{1}{NP} \sum_{i_{\text{pair}}=1}^{NP} (FD_{i_{\text{pair}}}) \quad (6)$$

式中:  $NP$  是该类中同构子图组合的总数;  $i_{\text{pair}}$  是该类中的第  $i$  对同构子图。

最后,为了获得两个场景之间的总体相似性,文中将每一类的相似度进行加权求和,公式如下:

$$TS = \sum_{k=1}^K (\varphi_k \times S_{\text{class}}) \quad (7)$$

式中  $\varphi_k$  是第  $k$  类的权重,根据每一类同构子图数量占全部子图数量的比例计算。这种权重分配方法确保了在总体相似性的计算中,各个类的贡献得到适当的体现。

## 1.3 基于 DTW 的时间序列对齐

在评估两个场景序列的整体相似性时,直接对齐和累加计算常因不同场景的时间序列长度不一致而变得困难。为应对这一问题,本研究引入了动态时间扭曲(DTW)算法<sup>[22]</sup>,以实现非等长交通场景片段的时序数据对齐。DTW 算法通过构建成本矩阵来测量两个时间序列之间的最小距离,并允许时间序列在时间轴上伸缩,以适应不同的时序长度。

设  $X$  和  $Y$  是两个待比较的测试场景片段序列,其中  $X$  的帧数为  $n$ ,  $Y$  的帧数为  $m$ :

$$X = (x_1, x_2, \dots, x_n) \quad (8)$$

$$Y = (y_1, y_2, \dots, y_m) \quad (9)$$

式中  $x_i$  和  $y_j$  分别为两个交通片段第  $i$  帧的场景交互拓扑图  $G_X^{(i)}$  和  $G_Y^{(j)}$ 。

定义成本矩阵  $D$  大小为  $n \times m$ ,其中  $D(i, j)$  表示序列  $X$  的前  $i$  个元素和序列  $Y$  的前  $j$  个元素之间的最小距离。成本矩阵  $D$  的递推关系如下:

$$D(i, j) = d(x_i, y_j) + \min(D(i-1, j), D(i, j-1), D(i-1, j-1)) \quad (10)$$

式中  $d(x_i, y_j)$  是点  $x_i$  和  $y_j$  之间的距离,基于式(11)计算。

$$d(x_i, y_j) = TS_{G_X^{(i)}, G_Y^{(j)}} \quad (11)$$

最终,通过应用 DTW 算法找到两个场景序列间最佳对齐方式的最小距离  $D(n, m)$ 。这种方法的应用确保即使在帧率不一致或场景持续时间有差异的情况下,也能有效比较两个场景的相似度。

## 2 案例研究

本节的场景案例由 PanoSim 内置场景模型虚拟生成。

### 2.1 评价案例分析

本节选取的3对评价案例分别代表了3种典型的情形:图6展示了相同道路结构下的相似场景,图7展示了相同道路结构下的不相似场景,而图8展示了不同道路结构下的相似场景。图6~图8均采用相同的布局,其中子图a和子图b分别是两个测试场景序列的关键帧截图,子图c则显示了使用动态时间扭曲(DTW)算法计算得出的成本距离矩阵,该矩阵以红色热力图的形式呈现。热力图的横纵坐标对应两个测试场景序列中的帧索引号,帧间隔为300 ms。根据

式(11),较深的红色代表更高的距离成本,即相似度较低,矩阵的每个单元格都展示了两个序列中对对应点之间的距离。此外,图中的黑色折线表示最优匹配路径。

对于第1个评价案例,从图6可以看出,测试场景1和测试场景2展现出较高的交互相似性。首先,在图6(a)和图6(b)中,可以观察到两组场景在路口的交通密度以及车辆的分布和排列在视觉上极为相似。其次,两组场景中的车辆在接近和通过路口时表现出了类似的行为模式。例如,在 $t=18.6\text{ s}$ (图6(a))和 $t=10.5\text{ s}$ (图6(b))的截图中,不仅路口区域内车辆布局显示出高度相似性,而且车辆也展示了相同的交互模式,如等待并让行左转方向的车辆。最后,图6(c)的热力图反映了与这些观察结果相匹配的特征,尤其是对角线附近的浅色区域和最优路径紧贴对角线的位置,这些都指示了两个场景序列之间的高度相似性。

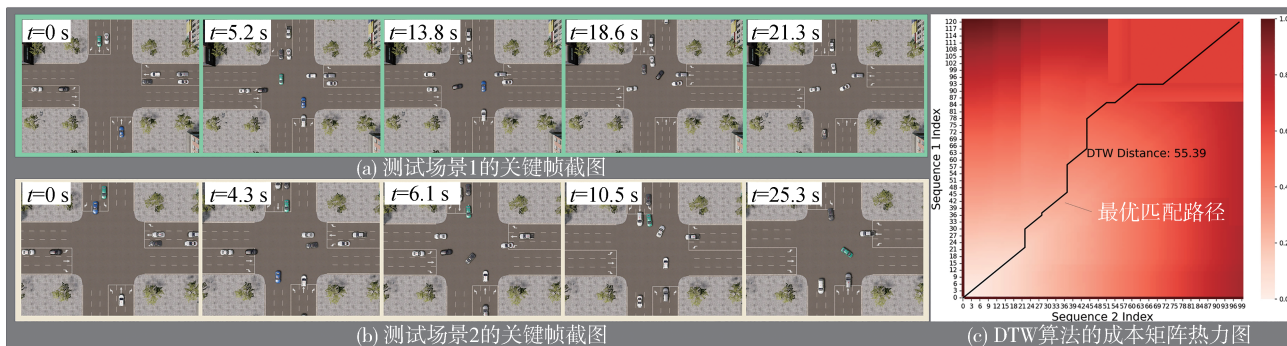


图6 评价案例1:同一道路结构上的相似性场景

对于第2个评价案例,文中详细分析了测试场景1(图7(a))和测试场景2(图7(b))之间的不相似性。具体来说,这两个场景在车辆布局上存在显著差异,这直接影响了它们的交互拓扑结构,从而凸显了两个场景间的不同。例如,在图7(a)中,车辆分布较为分散;而图7(b)中的车辆主要集中在道路的右侧。在图7(b)的集中布局中,车辆主要朝同一方向移动,导致交互相对单一,拓扑结构较为

简单。相比之下,图7(a)中的车辆由于其分散布局,方向更为多样化,导致交互和拓扑结构更为复杂。图7(c)中的DTW成本矩阵和最优路径提供了两个时间序列匹配程度的直观证据。成本矩阵中的深色区域和曲折的最优路径揭示了在时间序列对齐过程中所面临的挑战,这反映了两个场景在时间维度上的匹配难度较大,从而得到了较高的DTW距离。

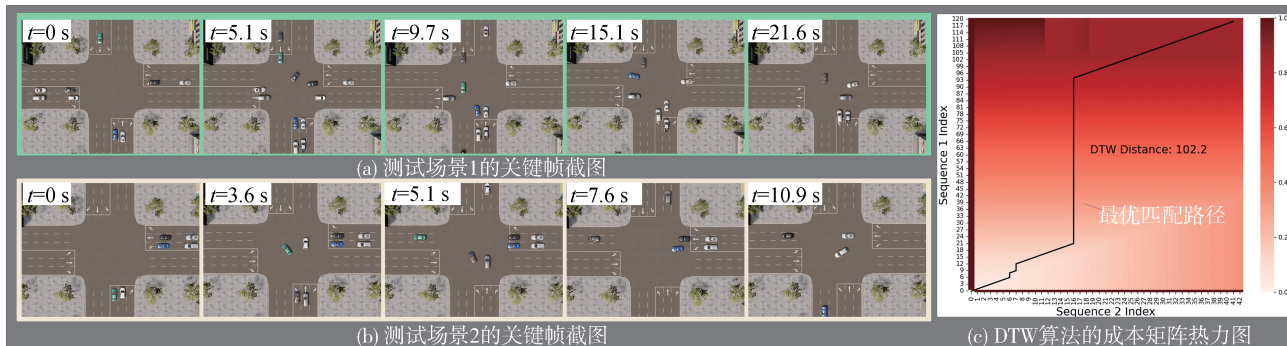


图7 评价案例2:同一道路结构上的不相似场景

针对第3个案例,从图8中可以看出,尽管测试场景1和测试场景2的道路结构不同,它们在路口的车辆布局和交互策略上却表现出了显著的一致性。具体来说,两个场景中路口内的车辆分布较为分散,且来自不同方向的车辆在路口进行了相互等待和让

行,都展示了有序的交通流动模式。图8(c)的热力图中对角线附近的浅色区域以及其上的最优路径清晰地展示了两个场景序列之间的紧密匹配。这一结果强调了所提方法的鲁棒性,能够有效地分析和对比不同道路结构的场景相似性。

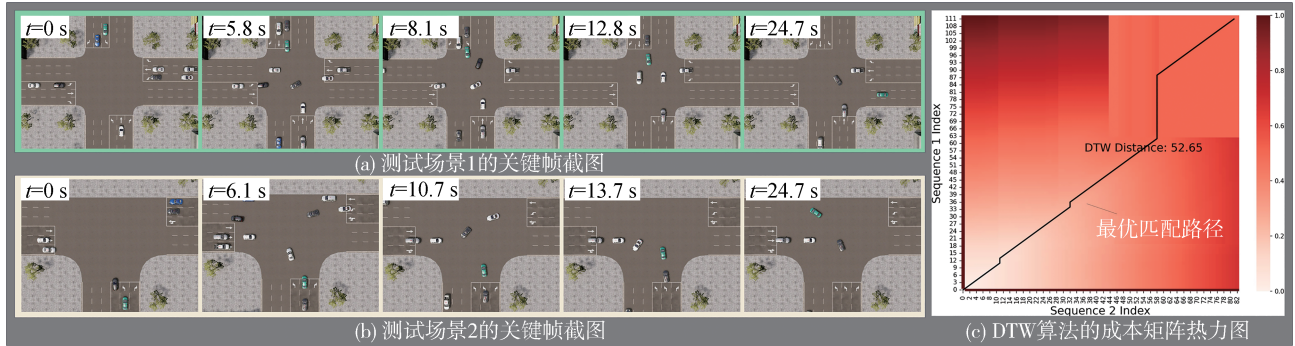


图8 评价案例3:不同道路结构上的相似场景

### 2.2 评价过程展示

在本节中,为了展示文中方法的评价过程,特别选取了案例1中测试场景1的  $t=13.8$  s 帧和测试场景2的  $t=6.1$  s 帧作为评价对象。图9清晰地展示了这两帧场景的图表示形式。交互拓扑图中节点的编号与车辆编号一致,双向箭头表示冲突交互,单向箭头表示跟驰交互。

通过子图分解和聚类,文中共识别出9种不同的子图类别。实验结果如表1所示,详细列出了每

种子图类别的形状、图内相似性及其在总体中的数量占比。表中的相似度值为1意味着该子类中的子图完全来自于一个测试场景,显示出两个场景在该子类图代表的交互模式下完全不同。根据式(7),计算得出这两帧场景的总相似度为0.583。

表1 评价结果汇总

聚类结果	相似性	数量占比
	0.469	0.234
	0.800	0.063
	0.415	0.063
	0.085	0.127
	0.240	0.106
	1	0.234
	0.826	0.042
	1	0.040
	1	0.021

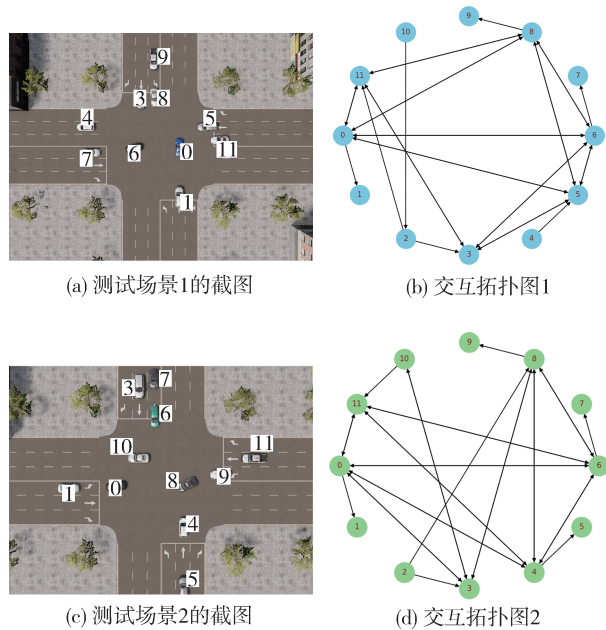


图9 场景图表示

### 3 方法验证及应用

本节旨在通过分析不同相似度场景簇中自动驾驶系统性能的差异来验证所提方法的有效性。具体来说,如果在不同的场景簇中,自动驾驶系统的性能表现出明显的差异性,则表明所提方法具有较高的可靠性。最后,通过测试场景数据集的优化应用证明了所提方法的工程应用价值。

#### 3.1 联合仿真平台

如图 10 所示,本研究搭建了 PanoSim-Apollo 联合仿真框架。在这一框架中,场景仿真平台 PanoSim 提供高逼真的城市复杂交叉口场景仿真服务,高级自动驾驶系统 Apollo 当做测试对象。PanoSim (版本 5.0) 和 Apollo (版本 7.0) 运行于不同的操作系统上,分别是 Windows 11 和 Ubuntu 18.04。这两个系统通过 UDP 协议进行双向数据通信。在仿真过程中,PanoSim 将场景数据通过以太网发送至 Apollo 的决策规划模块。数据处理后,Apollo 将计算得到的车辆控制指令(油门、转向和制动)反馈至 PanoSim 的高精度车辆动力学模块以更新仿真状态。

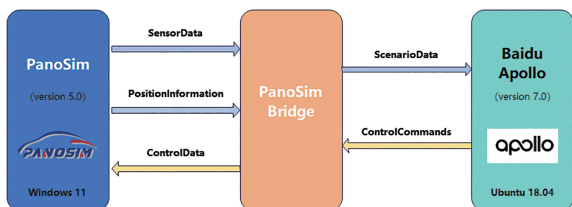


图 10 PanoSim-Apollo 联合仿真框架

联合仿真效果如图 11 所示。左上角为 PanoSim 的高逼真场景渲染效果,其中红车为 Apollo 主车。主图为 Dreamview 平台的 2D 渲染图,其中蓝色车辆为 Apollo 车辆,紫色方框为交通车,右上角区域展示的是 Apollo 的速度状态与控制信息。

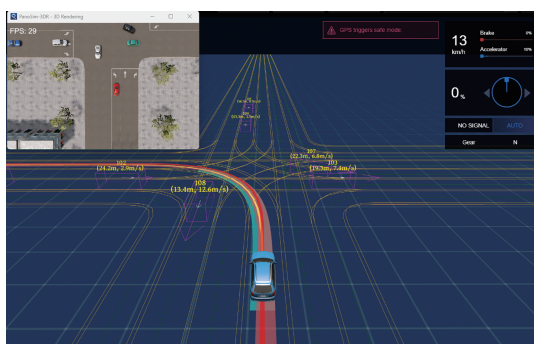


图 11 PanoSim-Apollo 联合仿真效果

#### 3.2 数据收集

为了全面评估自动驾驶系统 Apollo 在各种条件下的行驶安全性和效率,文中使用了 PanoSim 平台的内置场景模型生成了一系列多样的城市交叉口场景。数据收集过程包括两个主要步骤:(1)离线生成城市交叉口场景(主车不在环)并评估场景之间的相似性;(2)使用这些生成的场景对 Apollo 系统进行测试,并详细记录每次测试的行驶风险与效率。采用场景离线生成方式的一个显著优势在于它允许在场景生成阶段进行初步筛选,有效剔除那些同质化或不符合要求的场景,从而避免在自动驾驶系统的全量测试中浪费计算资源。

具体仿真设置展示在图 12 中。在测试中,执行左转的 Apollo 车辆将面对来自北、东、西 3 个方向的交通,涉及 6 条车道。红色曲线标示了 Apollo 车辆的预定行驶路径。每条车道上的最大交通车数量设定为两辆。在仿真初始化阶段,这些交通车将随机放置于各自的车道上,且每辆车的前后间距和初始速度将在  $[3, 10]$  m 和  $[0, 5]$  m/s 的范围内进行随机设定。仿真进行到所有车辆安全驶出交叉口或时间超过设定的最大仿真时长 60 s 时停止。

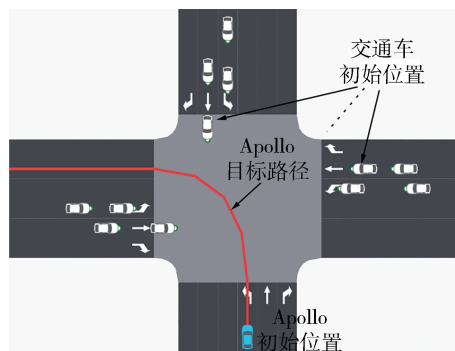


图 12 仿真设置

评估自动驾驶的性能指标如下。

(1)风险指标:车辆在交叉口区域内的平均冲突风险。它通过计算自车与其他所有冲突车辆的风险的和来获得,其中车辆对的冲突风险基于式(4)进行计算,具体采用以下公式计算:

$$R = \sum_{t=1}^T \left( \sum_{j=1}^M \omega_{ego,j}(t) \right) \quad (12)$$

式中: $\omega_{ego,j}(t)$ 表示在  $t$  时刻自车与冲突车  $j$  的冲突风险; $M$  为自车的冲突车总数量; $T$  为仿真总步长。

(2)效率指标:车辆在交叉口区域内的平均车速,单位 m/s。

$$E = \frac{1}{T} \sum_{t=1}^T v_{ego}(t) \quad (13)$$

式中  $v_{ego}(t)$  表示在  $t$  时刻自车的行驶速度。

在本次研究中,总共进行了 6 000 次仿真实验。在数据处理阶段,剔除了由于超时和不合理的碰撞导致的异常数据。超时数据指的是那些未能在规定时间内完成的实验,而不合理的碰撞则指仿真中出现的与现实行驶条件不符的碰撞事件。经过这一筛选过程,最终有效的实验数据总数为 4 365 条。表 2 汇总了在测试过程中得到的自动驾驶系统性能的统计结果。

表 2 验证集的统计结果

指标	均值	标注差	最小值	中位数	最大值
$R$	0.461	0.057	0.285	0.452	0.675
$E$	1.880	0.231	1.160	1.931	2.355

### 3.3 验证过程

首先,根据场景相似性评价结果,采用层次聚类方法对场景进行分类。层次聚类方法主要分为两种类型:凝聚型(agglomerative)和分裂型(divisive)。本实验选择凝聚型层次聚类方法。这种方法从每个场景开始,将每个场景视为一个独立的簇。然后,逐步合并最相似的簇,直到所有场景都被归入一个簇或满足某个停止条件为止。场景之间的距离等于 1 减去场景相似性。

接着,基于聚类的结果,运用单因素方差分析(ANOVA)来分别评估不同场景组别中风险和效率指标分布的差异。单因素方差分析的计算步骤如下。

(1) 计算不同场景簇的组间平方和(between-group sum of squares, SSB):

$$SSB = \sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2 \quad (14)$$

式中: $k$ 为不同水平独立变量的数量,具体指聚类获得场景组数; $\bar{y}_i$ 是第*i*组的场景风险或效率平均值; $n_i$ 是第*i*组的场景数量大小。

(2) 计算同一场景簇的组内平方和(within-group sum of squares, SSW):

$$SSW = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \quad (15)$$

(3) 组间均方  $MS_{\text{between}}$  和组内均方  $MS_{\text{within}}$  分别表示如下:

$$MS_{\text{between}} = \frac{SSB}{k-1} \quad (16)$$

$$MS_{\text{within}} = \frac{SSW}{N-k} \quad (17)$$

(4) 计算  $F$  统计量:

$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}} \quad (18)$$

式中, $F$ 值将用于通过查  $F$  分布表来判断各组均值之间是否存在显著差异。

通过单因素方差分析得知至少有一组与其他组存在显著差异,但它并不具体指出哪些组之间的差异是显著的。为此,文中进一步采用多重比较测试来识别具体哪些组之间的差异达到显著水平,来实现精细化评估。

本实验采用实际应用中流行的最小显著差异(LSD)测试方法。LSD测试的核心是基于  $t$  分布进行两两组间的均值比较。在进行 LSD 测试时,首先计算每对组合的均值差异,然后将这些差异与根据实验误差估计的临界值(LSD值)进行比较。

$$LSD = t_{\alpha/2, df} \times SE_{\text{diff}} \quad (19)$$

式中: $t_{\alpha/2, df}$ 是在给定自由度  $df$  和显著性水平  $\alpha$  的情况下,  $t$  分布的临界值; $SE_{\text{diff}}$ 是两组均值差异的标准误差。

$$SE_{\text{diff}} = \sqrt{\frac{MS_{\text{within}}}{k_1} + \frac{MS_{\text{within}}}{k_2}} \quad (20)$$

如果某一组合的均值差异超过了这个临界值  $LSD$ ,则认为这两个组之间的差异是统计学上显著的。

### 3.4 实验结果

表 3 显示了不同聚类距离阈值条件下获得的聚类数量。聚类阈值的增大显著减少了聚类的数量。当聚类阈值为 0.2 时,得到的聚类数量最多,达到 587 个;而聚类阈值提升到 0.8 时,聚类数量显著减少到只有 5 个。这表明阈值的增加使得聚类条件更为严格,只有距离非常近的场景样本才会被归为同一聚类。

表 4 展示了不同距离阈值下,因变量  $R$  和  $E$  的方差分析结果及 LSD 检测的显著比例。通过  $F$  值和  $p$  值可以看出所有条件下因变量的变化都是统计显著的。从表中对比数据可以看出,随着聚类阈值的减小,单因素方差分析的显著性增加, LSD 检测的显著比例也随之增高,表明较低的聚类阈值能更敏感地捕捉到因变量的变化。这可能是由于较低的阈值条件下,场景间的细微差异更容易被区分开来,从而在统计测试中表现出更高的敏感性。

图 13 直观地展示了距离阈值 0.8 条件下, 5 组

表3 不同聚类距离阈值条件的聚类数量

距离阈值	0.2	0.4	0.8
聚类数量	587	112	5

表4 单因素方差分析结果

距离阈值	因变量	F值	p值	是否显著	LSD检测显著比例/%
0.8	R	3.476	0.004	是	30.121
	E	3.506	0.003	是	43.947
0.4	R	5.440	0.000	是	41.732
	E	6.470	0.000	是	80.261
0.2	R	9.891	0.000	是	71.098
	E	12.666	0.000	是	92.113

测试场景中的行驶风险和效率的分布特征,包括它们的均值和方差。从图表中可以观察到,在行驶风险和效率方面,不同测试场景之间存在显著的差异。这一结果进一步说明了所提出评估方法的有效性,该方法能够以细粒度准确地评估不同场景间的相似性。

此外,从表4的数据分析可见,效率统计结果的显著性显著高于风险统计结果的显著性。这一现象可归因于城市交叉口环境中 Apollo 系统在决策信心

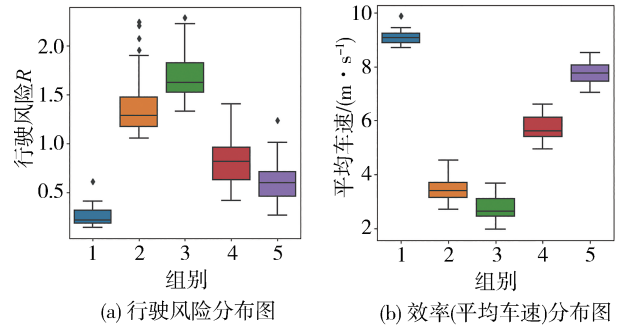


图13 距离阈值0.8对应5组场景的测试结果

不足时通常采取的保守让行策略,导致风险评估结果存在相对较大的波动。决策信心主要受周围交互态势的影响,由于交叉口的安全交互空间较小且交互频繁,即使在相似场景中,由于车辆相遇的细微差异,也可能导致决策信心的发生变化。图14中的两个测试案例直观地验证了这一点。从图中可以看出,在相似的场景下,图14(a)中的 Apollo 采取主动抢行策略,而在图14(b)中,则选择了保守让行策略。尽管风险评估结果存在波动,但在整体趋势的统计结果仍然是显著的。

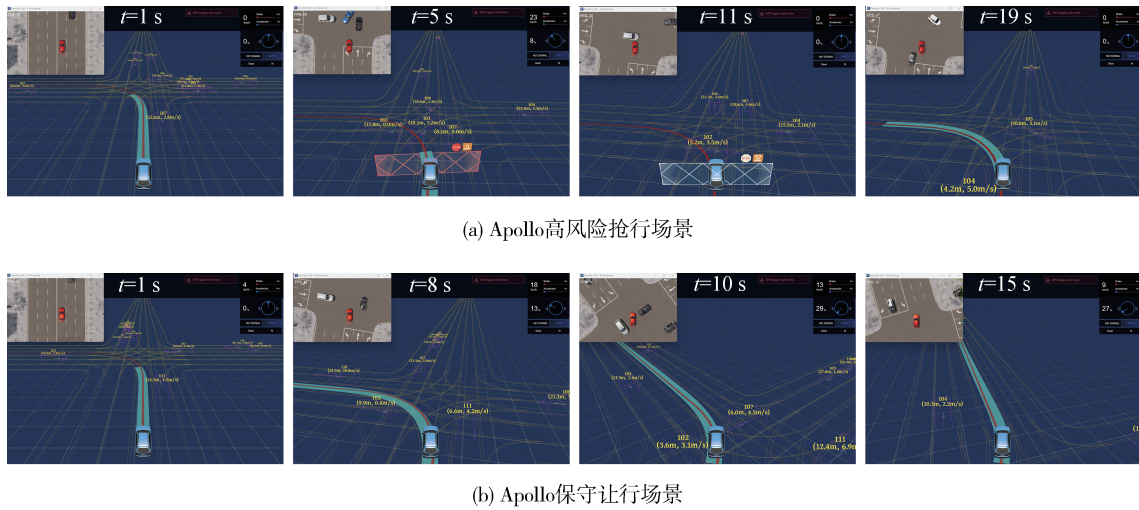


图14 测试案例关键帧截图

相反,效率的计算结果主要取决于场景中的交互密度和车辆的平均交互水平,短暂的让行对整体的速度影响较小,因此效率统计结果较为显著。

综上所述,单因素方差分析的结果显示,在不同相似度的场景簇中,自动驾驶性能的表现存在显著差异,这验证了文中所提出方法的有效性。

### 3.5 测试场景集优化应用

在前文中已验证了所提方法的有效性,本节将展示该方法在测试场景集优化中的实际应用价值。具体实验步骤包括首先对生成的场景集进行优化,即剔除相似场景,然后评估优化前后自动驾驶系统风险和效率分布的一致性。

基于距离阈值0.8的场景聚类结果,实施了4组数据集优化实验,在每组实验中,从5类场景中随机剔除了10%、20%、30%和40%的场景。分布度量基于欧式距离计算:

$$D = \sqrt{\sum_{i=1}^n (f_o(x_i) - f_r(x_i))^2} \quad (21)$$

式中: $f_o(x_i)$ 是原始分布在第*i*个点的概率密度值; $f_r(x_i)$ 是优化后的分布在第*i*个点的概率密度值。

在图15中,展示了不同剔除比例下(10%、20%、30%和40%)风险分布的对比结果。如图15(a)所示,在10%的剔除比例下,优化后的风险分布与原始分布非常接近,分布度量 $D = 0.21$ ,说明此时场景剔除对系统的风险分布影响较小。当剔除比例增加

到20%时,优化后的分布与原始分布之间的差距开始加大,分布度量 $D = 0.91$ 。此时,虽然仍保持了一定的分布一致性,但已经可以观察到一些明显的差异。在剔除比例达到30%和40%时,风险分布出现了明显的变化。优化后的分布与原始分布之间的差异显著增加,表明剔除过多场景可能会影响系统的风险评估精度。图16中的效率分布变化趋势与风险分布的趋势一致,这里不再赘述。

综上所述,文中的方法可以在保持一定测试精度的同时,有效减少测试场景的数量并提升测试效率。然而,需要注意的是,剔除比例的大小直接影响到自动驾驶系统风险和效率评估的准确性。因此,在实际应用中,选择合适的优化比例至关重要。

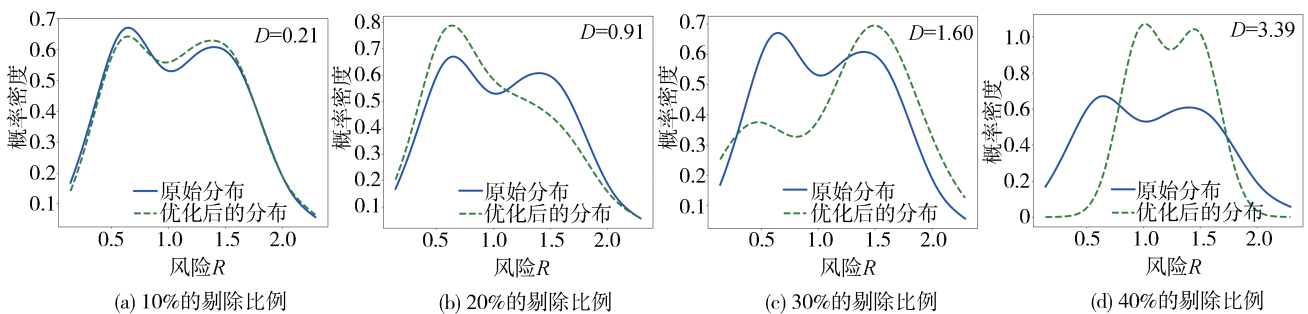


图15 风险分布对比结果

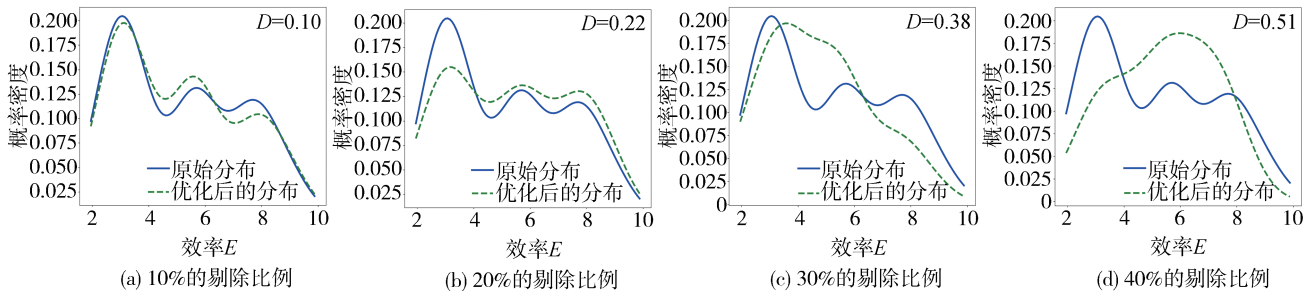


图16 效率分布对比结果

## 4 结论

本文中提出了一种基于有向图的城市交叉口场景相似性评价方法,该方法能够有效捕捉交叉口中车辆间的复杂动态交互特征。具体而言,通过对场景全局交互拓扑关系的时空匹配程度进行量化,实现了场景相似性的准确评价。定性分析和方差分析的实验结果均证明了该方法的有效性。此外,该方

法已被成功地应用于优化Apollo系统中无保护左转的测试场景集,从而进一步验证了该方法在自动驾驶测试评价中的实际应用价值。然而,当前方法的计算量较大,未来的工作将致力于优化这一问题,以进一步提高其实用性。此外,本研究未深入探讨如信号灯相位、道路结构、不同类型交通参与者之间的行为差异以及同类交通参与者不同相遇角度的交互差异,在接下来的研究中将对模型进行进一步的完善和扩展。

## 参考文献

- [1] SONG Y, CHITTURI M V, NOYCE D A. Intersection two-vehicle crash scenario specification for automated vehicle safety evaluation using sequence analysis and Bayesian networks[J]. *Accident Analysis & Prevention*, 2022, 176: 106814.
- [2] WANG X, PENG Y, XU T, et al. Autonomous driving testing scenario generation based on in-depth vehicle-to-powered two-wheeler crash data in China[J]. *Accident Analysis & Prevention*, 2022, 176: 106812.
- [3] 陈君毅, 刘镇源, 杨雪珠, 等. 感知系统触发条件穿透率评估方法研究[J]. *汽车工程*, 2024, 46(4): 577-587, 604.  
CHEN J Y, LIU Z Y, YANG X Z, et al. Evaluation method for the penetration rate of perception system triggering conditions[J]. *Automotive Engineering*, 2024, 46(4): 577-587, 604.
- [4] SCANLON J M, KUSANO K D, DANIEL T, et al. Waymo simulated driving behavior in reconstructed fatal crashes within an autonomous vehicle operating domain[J]. *Accident Analysis & Prevention*, 2021, 163: 106454.
- [5] BERKHAHN V, KLEIBER M, LANGNER J, et al. Traffic dynamics at intersections subject to random misperception [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(5): 4501-4511.
- [6] BANGERT L G, LUBASH T, SCANLON J M, et al. Determination of functional scenarios for intersection collisions [J]. *Accident Analysis & Prevention*, 2023, 193: 107326.
- [7] LU X, ZHAO H, LI C, et al. A game-theoretic approach on conflict resolution of autonomous vehicles at unsignalized intersections [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2023, 24(11): 12535-12548.
- [8] 朱冰, 张培兴, 赵健, 等. 基于场景的自动驾驶汽车虚拟测试研究进展[J]. *中国公路学报*, 2019, 32(6): 1-19.  
ZHU B, ZHANG P X, ZHAO J, et al. Review of scenario-based virtual validation methods for automated vehicles[J]. *China Journal of Highway and Transport*, 2019, 32(6): 1-19.
- [9] 张培兴, 秦孔建, 朱冰, 等. 自动驾驶仿真多逻辑场景综合评价方法[J]. *汽车工程*, 2024, 46(3): 375-382.  
ZHANG P X, QIN K J, ZHU B, et al. Comprehensive evaluation method for automated vehicle in multiple virtual logical scenarios [J]. *Automotive Engineering*, 2024, 46(3): 375-382.
- [10] 邓伟文, 李江坤, 任秉韬, 等. 面向自动驾驶的仿真场景自动生成方法综述[J]. *中国公路学报*, 2022, 35(1): 316-333.  
DENG W W, LI J K, REN B T, et al. A survey on automatic generation methods on simulation scenario for autonomous driving [J]. *China Journal of Highway and Transport*, 2021: 1-21.
- [11] 李江坤, 邓伟文, 任秉韬, 等. 基于场景动力学和强化学习的自动驾驶边缘测试场景生成方法[J]. *汽车工程*, 2022, 44(7): 976-986.  
LI J K, DENG W W, REN B T, et al. Automatic driving edge scene generation method based on scene dynamics and reinforcement learning [J]. *Automotive Engineering*, 2022, 44(7): 976-986.
- [12] ZHU B, ZHANG P X, ZHAO J, et al. Hazardous scenario enhanced generation for automated vehicle testing based on optimization searching method [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(7): 7321-7331.
- [13] ZHU Y, WANG J, GUO X, et al. Functional testing scenario library generation framework for connected and automated vehicles [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2023, 24(9): 9712-9724.
- [14] RIES L, RIGOLL P, BRAUN T, et al. Trajectory-based clustering of real-world urban driving sequences with multiple traffic objects [C]. 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), 2021: 1251-1258.
- [15] ZHANG Q, GAO Y, ZHANG Y, et al. TrajGen: generating realistic and diverse trajectories with reactive and feasible agent behaviors for autonomous driving [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(12): 24474-24487.
- [16] ARVIN R, KHATTAK A J, KAMRANI M, et al. Safety evaluation of connected and automated vehicles in mixed traffic with conventional vehicles at intersections [J]. *Journal of Intelligent Transportation Systems*, Taylor & Francis, 2020, 25(2): 170-187.
- [17] YAN X, ZOU Z, FENG S, et al. Learning naturalistic driving environment with statistical realism [J]. *Nature Communications*, 2023, 14(1): 2037.
- [18] LI J, DENG W, WANG Y, et al. Modeling of complex traffic interaction under uncontrolled intersection scenario [J]. *IEEE Transactions on Intelligent Vehicles*, 2024.
- [19] YU R, ZHENG Y, QU X. Dynamic driving environment complexity quantification method and its verification [J]. *Transportation Research Part C: Emerging Technologies*, 2021, 127: 103051.
- [20] TARJAN R. Depth-first search and linear graph algorithms [J]. *SIAM Journal on Computing*, Society for Industrial and Applied Mathematics, 1972, 1(2): 146-160.
- [21] ULLMANN J R. An algorithm for subgraph isomorphism [J]. *Journal of the ACM*, 1976, 23(1): 31-42.
- [22] Dynamic time warping [G]//MÜLLER M. *Information Retrieval for Music and Motion*. Berlin, Heidelberg: Springer, 2007: 69-84.