面向可泛化行人过街行为推断的视觉与知识增强型大语言模型

Qingwen Pu[a]、Kun Xie[a,*]、Hong Yang[b]和Guocong Zhai[c]

[a] *Old Dominion University*土木与环境工程系交通信息学实验室，美国弗吉尼亚州*Norfolk 23529* [b] *Old Dominion University*电气与计算机工程系，美国弗吉尼亚州[c] 西南交通大学交通运输与物流学院、综合交通大数据应用技术国家工程实验室、综合交通智能化国家地方联合工程研究中心，中国成都*611756* * 通讯作者；电子邮箱：*kxie@odu.edu*

摘要

现有的行人过街行为推断范式涵盖从统计模型到监督学习方法的各种方法，但其泛化能力有限，在新地点上的表现不佳。大语言模型（LLM）的最新进展推动了从数值模式拟合向语义化、上下文感知行为推理的转变；然而，现有的大语言模型应用缺乏领域特定适应和视觉上下文。本研究提出行人过街大语言模型（PedX-LLM），这是一种视觉与知识增强型框架，旨在将行人过街推断从特定地点的模式识别转变为可泛化的行为推理。PedX-LLM将由LLaVA提取的视觉特征与文本数据及交通领域知识相结合，并通过低秩适应（LoRA）对LLaMA-2-7B基础模型进行微调，以推断过街决策。PedX-LLM的平衡准确率达到82.0%，比表现最佳的统计模型（分层逻辑回归）高7.9个百分点，比表现最佳的监督学习方法（CatBoost）高3.0个百分点。结果表明，视觉增强模块通过捕捉建成环境带来2.9%的性能提升，而整合领域知识又带来额外4.1%的提升。为评估模型在未见环境中的泛化能力，本研究采用基于地点的划分进行跨地点验证。零样本PedX-LLM配置在五个未见测试地点上的平衡准确率达到66.9%，比基线数据驱动方法至少高18个百分点。通过少样本学习向PedX-LLM加入仅五个验证样例后，平衡准确率进一步提升至72.2%。PedX-LLM展现出对未见场景的强泛化能力，证实视觉与知识增强推理使模型能够模拟类人决策逻辑，并克服纯数据驱动方法的局限性。

关键词：大语言模型、行人过街行为、跨地点泛化性、视觉增强推理、领域知识适应

1. 引言

行人安全仍是美国面临的一项重大公共卫生挑战。以弗吉尼亚州为例，2016年至2023年间，行人死亡人数每年增加了41.3%，重伤人数增加了32.2%，呈现出令人担忧的上升趋势。根据弗吉尼亚州交通研究委员会的一份报告（FHWA/VTRC 25-R6），弗吉尼亚州约70%的行人死亡事故发生在路段中部，行人经常选择在指定人行横道以外的地方过街（Xie 等人（2024））。行人过街推断模型提供了一个系统化框架，有助于了解对过街位置选择影响最大的因素，并推断哪些情境下可能出现高风险过街行为（Papadimitriou 等人 2016）。此类模型可为有针对性的干预措施提供依据，并指导基础设施设计改进，以减少路段中部的行人死亡事故。

从统计模型到监督学习方法，现有的行人过街行为推断范式泛化能力有限，在新地点上的表现也不尽如人意（Ham 等2024）。大多数统计模型都施加了线性假设，限制了其捕捉安全因素之间复杂交互作用的能力（Kyriazos and Poga 2024）。例如，这些模型难以表示道路宽度的影响如何随行人年龄而变化，也难以表示照明的影响如何取决于交通状况。这些参数假设往往无法反映真实世界行人决策所具有的非线性、依赖语境的启发式规律。监督学习方法将行为推断视为数值模式识别任务，通过训练数据学习输入特征与过街结果之间的统计关联（Pu 等 2025c）。然而，这些方法专注于拟合特定地点的模式，很少关注其背后的推理过程，因此难以推广到训练环境之外（Elalouf 等 2023）。因此，这些方法主要学习特定地点的模式，而非普遍的行为原则，这限制了它们在未见语境中的适用性。

大语言模型（LLM）推动了范式转变，使研究从数值模式拟合转向语义行为推理（Fahad 等 2025）。大语言模型在海量人类生成文本语料上训练，编码了关于人类认知、推理模式和社会规范的隐性知识（Yao 等 2024）。这一基础使大语言模型能够捕捉依赖语境的非线性决策过程，从而近似人类式推理；而传统的数据驱动方法难以将这类过程形式化（Huang 等 2025）。与需要针对特定地点进行校准的传统方法不同，大语言模型可以通过少样本学习或定向提示适应新语境（Hang 等 2025）。借助通用的行为先验，大语言模型只需极少的额外数据，便能推断行人在未见环境中的决策，为传统方法提供了一种具有泛化能力且数据高效的替代方案（Kim 等 2025）。

然而，将大语言模型应用于行人行为推断面临着根本性挑战。通用预训练模型缺乏交通领域专门知识，无法区分真正的影响因素与虚假相关性（Liu等 2025）。现有的大语言模型应用往往无法对视觉和文本信息进行联合推理，因而难以捕捉建成环境特征（如道路几何形态和交通状况）如何与个体特征相互作用，进而影响行人行为（Blečić等 2024）。此外，使用商业大语言模型应用程序接口（如 GPT-5、Gemini、Claude）会带来隐私问题，因为敏感的视觉和行为数据必须传输到外部服务器（Le

等 2025）。这些局限凸显了构建隐私保护框架的必要性，该框架需将视觉推理与领域知识相结合，以支持准确且具备语境感知能力的行人过街推断。

本研究提出了 PedX-LLM，这是一种视觉与知识增强的大语言模型，如图1所示，它将行人过街预测从数值模式识别转变为具备泛化能力的行为推理。该框架将实地观察结果和基于 LLaVA 的视觉模块与领域知识提示相结合，以支撑行为推理，并采用基于沙普利值的归因方法对决策过程进行明确解码。PedX-LLM 利用低秩适应（LoRA）对 LLaMA-2-7B 模型进行微调，从而实现完全本地化的训练与部署，确保严格的数据隐私。本研究的主要贡献包括：

- 结合卫星影像与文本数据，捕捉建成环境和个体特征对过街决策的综合影响。
- 通过参数高效微调将领域知识融入行为推理，将不透明的模式识别转变为可解释推断。
- 展示了对未见场景的强泛化能力，表明视觉与知识增强的推理能够缓解分布偏移，并支持在异质情境下进行可靠推断。

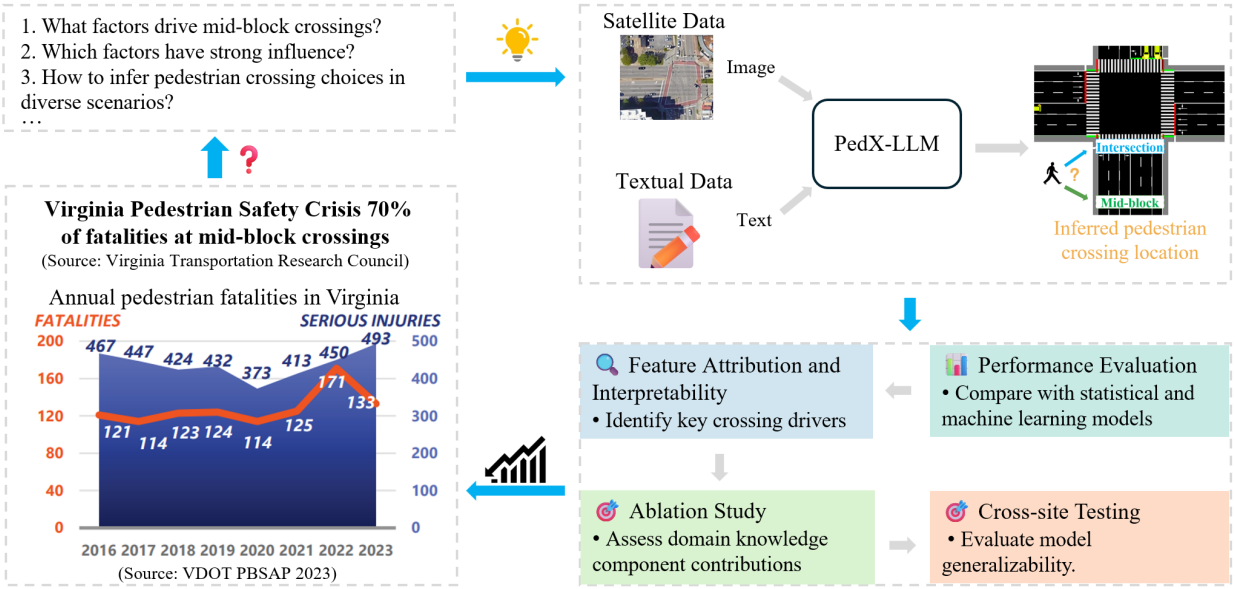


图1：概念框架概述

2. 文献综述

2.1. 行人过街行为推断

选择在路段中部而非指定交叉口过街，体现了便利性与安全性之间的根本权衡。这一选择源于个体特征（如年龄和性别）与环境情境（如照明和交通控制）之间的复杂交互作用

（Holland 和 Hill 2010，Osei 等 2024）。然而，城市环境中的行为变异性使风险评估更为复杂，因为在某一地点促使行人在路段中部过街的因素，往往与另一地点的情况不同（Zhu 等 2025）。因此，从观察数据中推断行为机制的推断框架，对于指导数据驱动的政策决策至关重要，包括基础设施改进和针对性行为干预（Singh 等 2024）。

传统统计框架为量化行人过街地点选择的决定因素提供了基础方法。离散选择方法，尤其是二元逻辑回归和多项逻辑回归，被广泛用于根据可观测变量估计路段中部过街的概率（Ahsan 等 2025）。这些参数方法的主要优势在于可解释性，因为所得系数可直接估计边际效应、支持假设检验，并与既有行为理论相契合（Gu 等 2024）。然而，这些模型依赖线性假设，限制了其推断行人与环境动态关系中固有的复杂非线性交互作用的能力（Hagenaars 等 2024）。此外，统计方法难以处理包含多种人口统计和环境变量的高维特征空间（Koldasbayeva 等 2024）。监督学习方法，包括机器学习算法和深度学习架构，为克服这些统计建模限制提供了一种数据驱动的替代方案。

机器学习算法通过推断行人过街行为中的非线性关系和复杂特征交互，克服了统计框架的参数限制（Pu 等 2025b）。随机森林分类器利用特征重要性排名识别具有显著影响的预测因子（Yuan 等2023）。然而，这种方法无法推断潜在的行为机制，也无法确立人口统计特征与风险感知之间的因果联系（Jacobucci 等 2023）。梯度提升框架，包括 XGBoost 和 CatBoost，能够有效处理非线性交互作用和分类数据（Zhang 和 Jánošík 2024）。然而，这些黑箱模型虽然能够预测结果，却无法解释为何老年行人等特定群体更倾向于选择交叉口（Rafie 等 2025）。此外，支持向量机会生成复杂的决策边界，无法有效地将个体层面的风险容忍度与建成环境影响区分开来（Tanveer 等 2024）。因此，有必要采用能够捕捉高阶特征交互作用的深度学习架构。

深度学习架构具备捕捉行人行为层次模式的先进能力（Huang 等 2024）。多层感知机（MLP）利用全连接网络处理表格化人口统计与地点数据。然而，若缺乏正则化或架构约束，这些模型往往会对特定训练环境过拟合（Przybyła-Kasperek 和 Marfo 2024）。例如，针对特定主干道布局优化的固定参数模型，在应用于几何特征迥异的环境时，性能往往会下降（Pu 等 2025a）。这种失效表明，模型记住的是局部几何特征，而非推断可泛化的行为原则（Shanthini 等 2024）。TabNet 利用序列注意力机制从结构化数据中学习特征重要性（Li 2025）。尽管这种方法减少了手动特征工程，但需要大型数据集才能收敛。此外，它会将文本描述与视觉空间情境分开处理，导致环境推理不完整（Jiang 等 2025）。因此，尽管采用了注意力机制，这些深度学习模型仍存在跨地点泛化能力差、可解释性有限的问题。

尽管该领域已从统计推断发展到深度学习，但仍受制于可解释性与预测准确性之间固有的权衡。此外，现有方法在推断可跨不同城市环境泛化的行为机制方面能力有限（Geiger 等 2025）。这些不足凸显了范式转变的必要性，即转向能够在不同城市环境中实现因果机制可泛化推断的模型。

2.2. 用于行人行为的大语言模型

大语言模型的出现凭借其处理和推理复杂多模态信息的先进能力，为推进行人安全研究提供了重要机遇（Karim *et al.* 2025）。这些模型能够综合大量非结构化数据，例如人口统计属性以及对过街情境的叙述性描述，并将其与结构化道路特征相结合，从而生成全面的行为洞见（Li *et al.* 2025）。在行人行为分析的语境下，大语言模型能够发挥三项关键功能：从文本事故叙述中提取人车冲突模式，识别过街描述中的行为策略，以及从伤害数据库中识别潜在风险因素（Yan *et al.* 2025）。这种集成异构数据源的能力，相较于传统方法，能够更全面地理解行人过街决策机制（Pu *et al.* 2026）。

当前行人安全领域的大语言模型研究主要采用通用提示工程（Abbasi 和 Rahmani，2025）。基于文本的模型从叙述性描述中学习过街模式，却未纳入先验实证知识（Strömel *et al.* 2024）。因此，这些架构能够捕捉统计相关性，却无法理解驱动行人过街决策的因果机制（Yang *et al.* 2025）。当前方法缺乏特定行为理论的集成。在个体层面，模型经常忽视年龄相关风险感知和群体动态的异质性（Deng *et al.* 2025）。同样，风险容忍度方面的性别差异也常常被忽视（Mirza*et al.* 2025）。在建成环境方面，公共交通站点和人行道宽度等物理决定因素很少被显式编码（Rodríguez *et al.* 2009）。因此，该领域缺乏旨在集成有关行人过街行为先验知识的结构化提示工程框架。此外，现有大语言模型研究尚未开展严格的消融分析，以量化个体层面因素与建成环境知识各自的独立贡献。

现有行人安全框架将视觉与文本数据的处理分开。视觉模型利用卫星图像或街景图像提取建成环境特征，例如土地利用模式和人行道布局（Zhou等 2025）。然而，这些视觉组件只是独立的特征提取器，并未作为整合要素融入大语言模型推理逻辑（Torneiro等 2025）。与此同时，文本架构孤立地分析人口统计属性和行为情境，而不考虑空间背景（Peykani等 2025）。这种独立处理未能捕捉建成环境与特定行人特征相互作用的协同机制（Yang等2024）。例如，老年行人将风险规避置于节省时间之上，往往会避开多车道主干道上的路段中部过街（Wilmut 和 Purcell, 2022）。然而，孤立分析道路视觉特征和人口统计特征时，常常忽略此类协同机制（Sharif等 2025）。因此，该领域缺少混合视觉增强架构，无法联合处理建成环境特征与个体属性，从而全面分析行人过街行为。

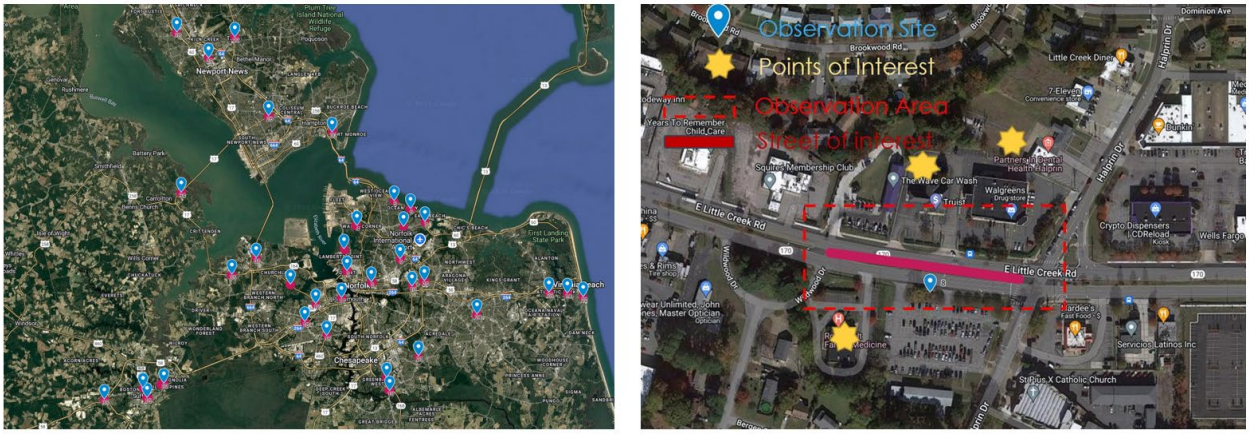
行人行为分析需要使用敏感的个人属性，包括人口统计特征和地理空间坐标（Jie等 2025）。然而，由于训练专有架构需要大量计算资源，目前的研究主要使用商业云端 API 或通用预训练模型（Raiaan等 2024，Choi 和 Chang 2025）。这种对外部服务的普遍依赖，要求将机密观察数据传输至第三方服务器（Ali 和 Ghanem 2025）。此类传输违反了交通机构的监管限制；由于责任风险顾虑，这些机构通常不得共享细粒度行为数据集（Hockstad等 2025）。因此，亟需能够完全在机构控制的基础设施内本地训练专有领域大语言模型的框架。

2.3. 研究空白

文献综述指出，行人过街行为推断存在三个关键空白。首先，传统统计方法和监督学习方法表现出有限的跨地点迁移能力（Efron 2020），学习的是地点特定的数值模式，而非可泛化的行为原则。其次，现有的大语言模型应用缺乏交通领域特定的领域适配，依赖通用提示，未能整合关于人口统计学异质性和建成环境影响的经实证验证的行为理论（Liu *et al.* 2025）。第三，当前框架独立处理视觉与文本数据，未能在统一的推断架构中对建成环境特征和个体属性进行联合推理（Zhang *et al.* 2025）。此外，广泛依赖商业云端 API 也引发了对敏感行为数据隐私问题的担忧（Shanmugarasa *et al.*2025）。这些空白凸显了构建隐私保护框架的必要性，该框架应将视觉增强推理与领域知识相结合，从而实现异质城市环境中的可泛化行人过街推断。

3. 数据准备

如图2（a）所示，本研究在弗吉尼亚州汉普顿锚地地区35处邻近设有人行横道交叉路口的路段中部位置开展了实地观察。本研究依据系统化标准选取地点，以涵盖多种行人过街情境：这些地点存在路段中部过街需求，但人行横道设施缺失或距离较远。选址标准包括位于城市地区或郊区的主干道和集散道路，且附近有显著的行人产生源，例如公共交通站点、住宅区、教育园区、公园、医疗设施、购物中心和停车设施。根据弗吉尼亚州交通部标准 IIM-TE-384.18，所有地点与最近的有标线路段至少相距300英尺。



(a) (b)

图2：(a) 弗吉尼亚州汉普顿锚地地区35个观测点的分布；(b) 一个观测点的布局

这35处选定地点分布于汉普顿路地区的多个辖区，涵盖多种道路布局和土地利用环境。表1总结了这些地点的建成环境特征。这些特征构成了行人进行过街决策时所处的物理框架，也是决定行人在交叉路口与路段中部备选方案之间选择过街地点的关键因素。

表1 研究地点的建成环境特征（N=35 个地点）

变量	类别/范围	描述
车道数 在路段中部	2-10 条车道	平均值：4.5
限速	25-35 英里/小时	25 英里/小时：40%，30 英里/小时：43%，35 英里/小时：17%
是否设有 凸起中央分隔带	存在/不存在	有：37%，无：63%
是否设有 公共交通 车站	存在/不存在	有：77%　，无：23%
人行道总宽度 人行道	10-14英尺	平均值：11.4 英尺
土地利用	5类	教育-居住用地（29%）、商业-居住用地（23%）、办公-居住用地（20%）、教育-办公用地（14%）、和绿地-居住用地（14%）

图2(b)展示了典型的观测点配置，包括观测点、兴趣点、观测区域和目标街道。经过培训的现场调查员会在相邻交叉口人行横道之间预先确定的观测点位进行观察，这些位置可能发生路段中部过街

的情况。他们确保能够清晰观察交叉口和路段中部位置的双向行人流动。调查员对每个地点连续观察六小时，以捕捉白天和夜间条件下的时间变化。由于日落时间存在季节性变化，我们相应调整了观察安排：2022年11月24日至2023年3月31日期间，数据采集时间为下午2:00至晚上8:00（白天87小时，夜间63小时）；其他采集时段则为下午4:00至晚上10:00（白天32小时，夜间24小时）。该方案使我们在35个地点累计观察约210小时，涵盖白天和夜间时段。

对于每位观察到的行人，研究人员手动记录其人口统计特征（年龄组和性别）、步行情境（独自步行或结伴步行）、过街行为（过街地点选择：交叉口或路段中部）、天气状况、交叉口和路段中部位置的照明条件，以及交通控制设施。交通控制设施包括信号配时参数（连续行人过街绿灯开始时间间隔和行人过街绿灯时段）、行人按钮的可用性和功能，以及左转保护相位。数据收集方案经机构审查委员会批准（2022年11月3日），并通过排除个人身份信息确保参与者匿名性。最终数据集包含687条行人观察记录（表2）。其中，259条（37.7%）记录发生在路段中部位置，428条（62.3%）发生在交叉路口。

表2 观察行人的人口统计与行为特征（N=687）

变量	类别	频数	百分比
过街地点	交叉口（有人行横道）	428	62.3%
	路段中部（没有人行横道）	259	37.7%
年龄组	儿童 (<18)	104	15.1%
	成年人 (18-65)	560	81.5%
	老年人 (>65)	15	2.2%
	不确定	8	1.2%
性别	Male	459	66.8%
	女性	210	30.6%
	不确定	18	2.6%
步行情境	独自	435	63.3%
	结伴	252	36.7%
天气状况	晴朗/阳光明媚	325	47.3%
	局部多云	207	30.1%
	多云	79	11.5%

	雨/毛毛雨	76	11.1%
交叉口照明	有	583	84.9%
	无	104	15.1%
街区中段照明	有	630	91.7%
	无	57	8.3%
相邻之间的时间间隔 过街绿灯开始时间（秒）	0 s	60	8.7%
	1 至 85 秒	346	50.4%
	> 85 秒	276	40.2%
	缺失	5	0.7%
行人过街绿灯时段（秒）	0 s	60	8.7%
	1 至 40 秒	373	54.3%
	> 40 秒	249	36.2%
	缺失	5	0.7%
设有行人按钮	Yes	571	83.1%
	No	116	16.9%
影响过街时间的按钮	Yes	492	86.2%
	No	79	13.8%
左转保护相位	Yes	471	68.6%
	No	216	31.4%

观察法能够捕捉不受实验干预影响的真实过街行为。值得注意的是，与交叉路口（84.9%）相比，所选路段中部位置的照明覆盖率更高（91.7%），这反映了研究区域内普遍存在的连续主干道照明。不过，本研究对某些分类（尤其是年龄估计）依赖观察人员的判断。为减少分类偏差，特别是年龄估计方面的偏差，研究实施了严格的研究人员培训方案，以确保统一应用评判标准。

4. 方法论

本研究开发了PedX-LLM框架，用于推断行人过街位置选择。如图3所示，该架构整合了多模态数据收集、结构化提示工程和参数高效微调。传统监督学习方法主要依赖数值特征向量来预测结果，而不推断其潜在机制。相比之下，我们的框架利用大语言模型的上下文推理能力，整合异构数据源，包括实地观察、卫星影像和领域知识，从而通过统一的基于文本的流程，从多模态语境中推导出行为机制。该框架

将二元分类任务（交叉口过街与路段中部过街）转化为自然语言生成问题，使模型能够利用预训练语言知识，同时融入交通领域特定行为原则。

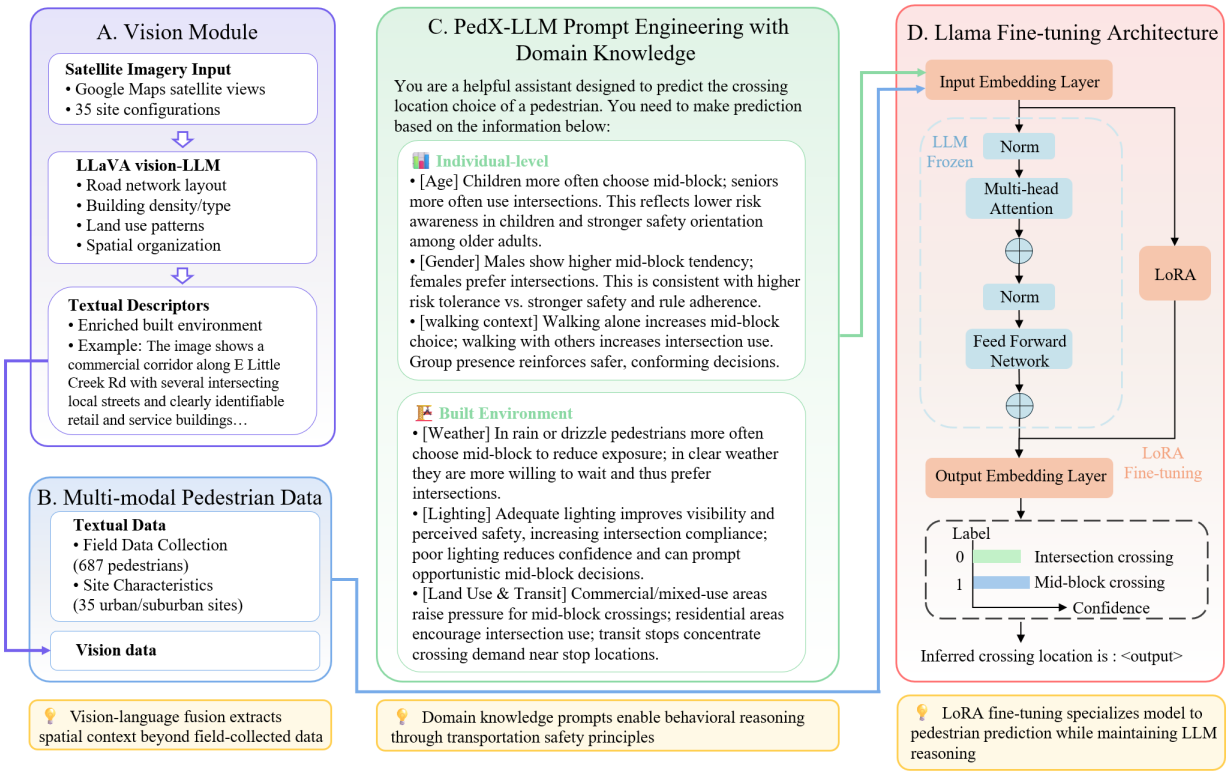


图3：PedX-LLM框架架构。

A. 用于建成环境提取的视觉模块

视觉增强组件（图3中的面板A）利用大型语言与视觉助手（LLaVA）模型（Liu 等 2023）从卫星影像中自动提取建成环境描述符。LLaVA整合了三个组件：(1) 使用预训练 CLIP ViT-L/14（视觉变换器）的视觉编码器，用于处理 336×336 像素卫星影像；其权重被冻结，以保留通过大规模预训练获得的泛化能力；(2) 由带 GELU 激活函数的两层多层感知机构成的投影网络，将 1024 维视觉嵌入映射到 4096 维语言模型嵌入空间；以及 (3) 语言模型（Vicuna-v1.5-7B），根据投影后的视觉特征生成自然语言描述。

通过谷歌地图获取了全部35个地点在缩放级别19下的卫星影像，覆盖约50×50 米范围的区域，其中包括路段中部过街地点及周边城市环境，示例如图2(b)所示。每张图像均通过LLaVA和结构化提示进行处理：“请描述这张卫星影像中可见的城市环境，重点关注：(1) 道路网络布局与组织形式；(2) 可见的建筑密度和类型；(3) 你能够识别出的土地利用模式；(4) 该区域的空间组织。”模型为每个地点生成平均150—200词的文本描述，

刻画宏观尺度的空间特征。示例输出："图中显示，E Little Creek 路沿线是一条商业走廊，分布着数条相交的地方道路，并且可以清晰辨认出零售和服务类建筑。许多商户周围设有停车场，表明该区域采用以汽车为导向的布局。该区域包括便利店、诊所和小型企业，说明在一个发展程度中等的城市环境中，行人活动较为稳定。"

这些通过视觉提取的语境描述提供了定性空间信息，例如建筑规模关系、开发模式和整体城市形态，从而补充定量文本数据。这使下游语言模型能够推理建成环境对行人过街决策的影响。

B. 多模态行人数据

多模态行人数据组件（图3中的面板B）集成了来自卫星影像的视觉数据和来自实地观察的文本数据。如第3节所述，该数据集涵盖了弗吉尼亚州汉普顿路地区35个城市和郊区地点的687次行人过街观察。文本数据组件记录了个体层面属性，包括行人群体特征、步行情境、环境条件以及观察到的过街地点选择，如表2所示。研究还系统记录了建成环境特征，以捕捉道路几何、交通控制设施和土地利用情境。具体变量包括车道数、限速、信号配时参数和分类土地利用定义，如表1所示。这些实地采集的文本数据提供了真实行为观察和定量基础设施测量数据。同时，来自卫星影像的视觉数据（按照第4.A节所述，通过LLaVA视觉大语言模型处理）提取了空间背景信息，包括道路网络布局、建筑密度模式、土地利用配置和空间组织。将这些实地文本指标与通过视觉提取的空间描述信息加以综合，可构建过街场景的整体多模态表征。这种集成弥补了传统方法的局限；传统方法完全依赖观察数据或基础设施清单。

C. 融入领域知识的 PedX-LLM 提示工程

领域知识是指经既有研究实证验证的先验知识。在本框架中，领域知识是指关于行人群体特征（年龄、性别、步行情境）和建成环境因素（照明、土地利用、交通控制）的行为原则，这些原则源自弗吉尼亚州交通部汉普顿路地区行人过街研究（Xie 等 2024）。通过系统化的文本转换，将收集到的多模态数据与特定地点观察相结合，并融入领域知识，从而生成 PedX-LLM 提示。如图3的 C面板所示，领域知识注入分为两个层级类别：个体层面属性和建成环境。

系统提示明确编码了从这些实证研究中得出的行为先验。个体层面的属性将行人群体特征与经验证的行为模式相结合。具体而言，提示指示模型，年龄会显著影响过街选择，并指出老年人通常表现出更强的安全意识和更高的延迟容忍度（这与较低的路段中部过街概率相关）。关于性别，提示编码了相关研究发现：与女性相比，男性参与者更倾向于在路段中部过街。同样，在步行情境方面，系统经过预设，

能够识别出，相较于结伴步行的行人，独自行走的行人在路段中部位置过街的可能性在统计上更高。

关于建成环境，提示中纳入了有据可查的环境影响因素。照明条件被编码为关键因素，并指出充足的视距和照明能够提升感知安全，从而鼓励行人在路段中部过街。对于土地利用和公共交通，模型依据以下原则进行引导：特定的相互作用（如办公-居住用地）会降低过街需求，而公交车站等行人产生源会使过街活动集中，因此需要采取不同的安全考量。

为将这些原则付诸实践，该框架将多模态数据整理为结构化的文本输入，用于训练和推断。如表3所述，数据整合自四个来源：实地观察、建成环境、视觉提取的空间语境和领域知识。每个提示组件都系统地呈现属性信息，并以推断任务说明收尾，引导模型生成二元分类结果（0 = 交叉口过街，1 = 路段中部过街）。这种层级结构使语言模型能够将领域知识与场地特定语境信息综合起来，符合透明的人工智能推理原则；在该原则下，明确整合领域知识有助于提升对推断出的行为机制的可解释性。

表3 PedX-LLM 框架的输入数据结构

提示组件	数据来源	变量名称
领域知识语境	(第4.C节)	
行人群体特征	表2	年龄组
	表2	性别
	表2	步行情境
环境条件	表2	天气状况
	表2	交叉口照明
	表2	街区中段照明
道路几何	表1	街区中段车道数
	表1	限速
	表1	人行道总宽度
	表1	是否设有凸起式中央分隔带
交通控制	表2	相邻之间的时间间隔过街绿灯开始时间
	表2	行人过街绿灯时段
	表2	提供行人按钮
	表2	影响过街时间的按钮
土地利用与公共交通	表1	土地利用
	表1	公共交通站的存在
建成环境（视觉）	（第 4.A 节）	

D. LLM 微调架构与训练配置

PedX-LLM 框架采用 LoRA（Hu等 2022）对 LLaMA-2-7B 模型进行微调。选择这一架构，专门用于应对交通机构面临的计算资源和隐私限制。与需要重新训练所有模型参数的全量微调不同，LoRA 引入了一种“旁路”机制，使模型在保留其海量预训练语言知识的同时，学习新的行人过街行为。

为实现任务特定学习并避免灾难性遗忘，我们冻结 LLaMA 基础模型的预训练权重（W_0）。随后，将可训练的低秩矩阵注入注意力层，以实现领域特定适应。

对于输入 x，前向传播的输出 h 在数学上计算为通用语言知识（固定）与交通领域适配（可训练）之和：

$$h = W_0 x + \frac{\alpha}{r} BAx \tag{1}$$

$W_0 x$ 表示从基础模型保留的冻结通用知识（例如理解英语语法和逻辑）。BAx 表示学习到的领域知识，其中 $B \in \mathrm{R}^{d \times r}$ 和 $A \in \mathrm{R}^{r \times d}$ 是低秩矩阵（秩为 $r = 32$）。$\frac{\alpha}{r}$ 是一个缩放因子，用于控制新领域知识对最终决策的影响。

这种分解确保模型在学习行人行为的细微特点时（通过矩阵 A 和 B），不会覆盖其基本推理能力。此配置产生约 3250 万个可训练参数，仅占 LLaMA-2-7B 总计 70 亿个参数的 0.46%，因此可部署在标准机构工作站上。

为进一步将内存占用从 28 GB 降低至约 7 GB，我们采用 4 位 NormalFloat（nf4）量化，在将模型权重压缩为 4 位精度的同时，通过混合精度计算保持计算准确率。模型采用因果语言建模目标进行优化。分词器词汇表中新增了领域特定标记（<INTERSECTION>、<MIDBLOCK>），损失函数仅针对这些答案标记最小化负对数似然：

$$L = -\sum_{i=1}^{N} I_{\text{answer}}(i) \log P(t_i \mid t_1, \ldots, t_{i-1}; \theta) \tag{2}$$

通过屏蔽提示上下文（$I_{\text{answer}} = 0$），我们确保模型只从答案标记中学习，促使其形成可泛化的推理模式，而非记忆提示结构。详细的超参数设置，包括 AdamW 优化器设置和余弦学习率调度，总结于表4。

表4 模型架构与训练配置

类别	参数	配置 / 值
基础模型	架构	LLaMA-2-7B-hf
	隐藏层大小（d）	4096
量化	精度类型	4位 NormalFloat（nf4）
	计算数据类型	float16
LoRA适配器	秩（r）	32
	Alpha（α）	64
	缩放因子	2.0
	目标模块	q_proj,k_proj, v_proj, o_proj
	可训练参数	≈32.5 百万（0.46%）
	丢弃率	0.1
优化	优化器	AdamW
	学习率	2×10^{-5}
	权重衰减	0.01
	学习率调度	带 50 步预热的余弦调度
	梯度裁剪	0.5
训练	有效批量大小	16
	最大序列长度	512 个词元
	最大训练轮数	250
	早停耐心值	15 个训练轮次
	评估指标	平衡准确率
硬件	GPU	NVIDIA Quadro RTX 5000（32 GB）

5. 结果与讨论

5.1. 模型性能与基线比较

为严格评估所提出框架的有效性，我们将其与一系列代表不同方法范式的成熟基线方法进行了全面比较。这些基线方法分为两类：（1）统计模型，包括逻辑回归和分层逻辑回归，后者作为行为可解释性的标准；（2）监督学习基线，包括随机森林、XGBoost、支持向量机（SVM）、CatBoost（这些模型是表格数据预测准确性的常用基准）、多层感知机（MLP）以及 TabNet，后两者用于评估先进架构处理高维特征交互和文本输入的能力。通过分层随机抽样，将包含687条观测的数据集划分为训练集（70%）、验证集（15%）和测试集（15%），以严格保持各个子集中固有的类别不平衡（路段中部占37.7%，交叉口占62.3%）。所有模型

使用验证集进行训练和超参数调优，并在五次随机划分的留出测试集上计算最终性能指标（平均值和标准差）。

表5展示了五次随机划分所得的性能指标。鉴于类别不平衡（路段中部占37.7%，交叉口占62.3%），本研究将平衡准确率作为主要指标，其定义为敏感度与特异度的算术平均值，以确保两个类别权重相等。其他指标包括：准确率（整体预测正确率）、精确率（TP/(TP+FP)，其中路段中部过街为正类）、F1分数（精确率与召回率的调和平均值）。标准差（Std Dev）用于量化五次随机划分中平衡准确率的变异性，以反映模型稳定性。

表5 行人过街位置推断模型的性能比较

模型	准确率	精确率	F1-分数	平衡准确率	Std Dev
统计模型					
逻辑回归	73.5%	70.2%	72.6%	73.2%	2.3%
分层逻辑回归	74.2%	71.5%	72.7%	74.1%	2.2%
监督学习					
随机森林	77.3%	74.9%	77.0%	77.1%	2.1%
XGBoost	78.9%	76.4%	78.3%	78.7%	1.9%
支持向量机	75.1%	72.6%	74.7%	75.0%	2.7%
CatBoost	79.2%	76.9%	78.8%	79.0%	2.0%
多层感知机	78.0%	75.3%	77.6%	77.8%	2.2%
TabNet	79.7%	77.2%	79.2%	79.4%	1.9%
大语言模型框架					
基线 LLaMA-2-7B	62.3%	59.9%	62.0%	62.1%	3.1%
PedX-LLM（纯文本）	75.2%	72.7%	74.8%	75.0%	1.8%
PedX-LLM（视觉增强）	77.4%	74.2%	76.7%	77.9%	1.7%
PedX-LLM（视觉与知识增强）	**82.1%**	**80.0%**	**81.7%**	**82.0%**	**1.4%**

统计模型和监督学习基线建立了具有竞争力的性能标准。统计方法的平衡准确率分数介于73.2%至74.1%之间，其中分层逻辑回归表现最佳。监督学习基线在所有指标上均展现出更优的性能。CatBoost取得了79.0%的最高平衡准确率，精确率为76.9%，F1分数为78.8%。TabNet取得了79.4%的平衡准确率、79.7%的总体准确率和79.2%的F1分数。

基线 LLaMA-2-7B 的表现不佳，准确率为 62.1%，这反映出模型的预训练数据（通用网络文本）与我们所处的专业交通安全领域背景之间存在显著的领域差距。通过 LoRA 进行监督微调显著缩小了这一差距。PedX-LLM（纯文本）版本将平衡准确率提升至 75.0%（提高了 12.9 个百分点），精确率和 F1分数也分别提升至 72.7% 和 74.8%。然而，这种纯文本方法仍未能超越 CatBoost（79.0%）等强基线方法，这表明单纯依靠模式识别不足以完成这项任务。

集成视觉提取的建成环境语境，进一步优化了模型表现。PedX-LLM（视觉增强）版本无需领域知识，仅利用卫星影像特征，便实现了 77.9% 的平衡准确率，较纯文本版本提高了 2.9%。这些结果证实，自动提取的视觉特征能够捕捉宏观尺度空间环境，与人工采集的场地数据形成互补。

在 PedX-LLM（视觉增强）的基础上，PedX-LLM（视觉与知识增强）融入了交通领域知识，取得了 82.0% 的最高平衡准确率，较视觉增强配置提高了 4.1 个百分点。这表明，领域知识对于提升行为推理能力至关重要，其作用超越了单纯的多模态数据融合。完整框架的表现比最佳统计模型（分层逻辑回归）高 7.9 个百分点，比最强的监督学习基线（CatBoost）高 3.0 个百分点，同时标准差最低，仅为 1.4%，表明其在不同场地配置下具有稳健的泛化能力。

5.2. 领域知识的消融研究

为量化领域知识提示的贡献度（图3中的C面板），本节针对两类知识开展了系统的消融实验：个体层面知识（年龄、性别和步行情境）以及建成环境知识（天气、照明和土地利用与公共交通）。通过分别或组合选择性地整合这些要素，我们分离并分析了它们各自对过街位置推断性能的贡献。表6展示了消融结果，其中类别（Δ）表示各知识类别相较于基线带来的提升，因素贡献度（Δ）则量化了各个因素的贡献（计算方法为类别内权重乘以该类别的总体增益），结果表明，全面整合的框架准确率最高。

表6 领域知识贡献度的消融分析

知识类别	知识因素	因素贡献度（Δ）	平衡准确率	类别(Δ)
PedX-LLM（视觉-增强）	None		77.9%	
+个体层面	Age	+1.19%	80.9%	+3.0%
	性别	+0.80%		
	步行情境	+1.01%		
+建成环境	天气	+1.27%	81.2%	+3.3%
	照明	+1.11%		
	土地利用与公共交通	+0.92%		

PedX-LLM（视觉增强-和知识增强）	Both	82.0%	+4.1%

注：改进幅度（Δ）是相对于基线（PedX-LLM（视觉增强））计算得出的。

基线 PedX-LLM（视觉增强）模型仅使用视觉和文本特征，不包含领域知识提示，其平衡准确率为77.9%。消融结果揭示了两类知识各自的贡献。整合个体层面知识（年龄、性别和步行情境）使性能适度提升了3.0个百分点，准确率达到80.9%。相比之下，纳入建成环境知识（天气、照明和土地利用与公共交通）带来了更显著的单独增益，提升了3.3个百分点，准确率达到81.2%。这表明，单独考察各类别时，环境情境对行为推断的影响比个体人口统计特征更为显著。整合两类知识流的完整 PedX-LLM（视觉与知识增强）框架准确率最高，达到82.0%，综合提升4.1个百分点，表明整合两类知识能够显著提高推断性能。

为了解这些类别层面增益背后的具体驱动因素，本节采用基于置换的特征归因方法，将每个知识类别分解为其构成因素。在个体层面知识类别中，年龄是最主要的因素，贡献了1.19个百分点（占该类别总增益的39.6%）。这反映出不同年龄组在风险感知和过街能力方面存在显著差异，尤其是儿童与老年人的安全倾向截然不同。步行情境的影响次之，贡献了1.01个百分点（33.5%），体现了行人结伴步行时，社会从众和集体决策所带来的安全益处。性别贡献了0.80个百分点（26.9%），反映了已有研究所记录的男性与女性行人在风险容忍度方面的差异。

建成环境知识类别呈现出不同的重要性层级。天气被确定为首要的环境驱动因素，贡献了1.27个百分点（占该类别增益的38.7%）。这一主导作用符合行为原则：恶劣天气条件会提高行人在交叉路口等待时感知到的成本，从而提升将路段中部过街作为节省时间策略的效用。照明贡献了1.11个百分点（33.5%），证实能见度限制会影响行人的感知安全以及对过街机会的察觉，进而从根本上塑造其过街地点选择。土地利用与公共交通贡献了0.92个百分点（27.8%），反映出公共交通站点和商业活动的邻近性会使过街需求集中，并影响路线选择。这些结果验证了整合两类知识能够构建一个稳健的行人过街行为推断框架，其表现显著优于仅依赖其中一种知识的模型。

5.3. 评估跨地点泛化性

将行人行为模型泛化到几何特征和运行特征各异的未见位置，是一项重大挑战。在特定地点训练的模型往往会过拟合局部特征，导致跨地点泛化能力较差。采用基于地点的

划分策略进行系统的跨地点验证，评估了 PedX-LLM 框架的稳健性，并确保训练环境与测试环境完全分离。

为防止数据泄漏，我们将汉普顿路地区 35 个地点的 687 条观测数据划分为三个互不重叠的组。训练集包含 22 个地点（N=455），涵盖了多种城市主干道和郊区集散道路。验证集包含 5 个地点（N=102），用于超参数调优。测试集由另外 5 个地点（N=130）组成，这些地点经过筛选，旨在代表最具挑战性的场景，其特点是存在显著的几何变化。

通过与既有基线方法进行基准测试，我们得以了解该框架的性能。除传统算法（逻辑回归、分层逻辑回归、CatBoost 和 TabNet）外，我们还通过两种不同的推理配置评估了所提出框架的泛化能力。PedX-LLM（零样本）配置将领域知识增强提示直接用于未见测试地点，不提供特定地点的示例。另一种 PedX-LLM（少样本）配置则将从验证数据集中随机选取的五个示例纳入提示上下文，从而通过上下文学习（ICL）实现适应，无需进一步更新参数。表 7 展示了留出测试集上的汇总性能指标。

表 7 留出测试集上的跨地点验证结果（5 个地点，130 个观测值）

模型	准确率	精确率	F1-得分	平衡准确率	Std Dev
逻辑回归	58.5%	59.6%	48.7%	41.2%	4.8%
分层逻辑回归	61.4%	60.1%	60.6%	46.2%	4.5%
CatBoost	62.1%	64.5%	55.3%	48.3%	4.2%
TabNet	69.7%	61.9%	51.1%	43.6%	4.6%
PedX-LLM（零样本）	75.8%	68.5%	67.6%	66.9%	4.1%
PedX-LLM（少样本）	79.4%	73.0%	72.6%	72.2%	3.5%

PedX-LLM框架（零样本）的平衡准确率达到66.9%，较基线方法（CatBoost，48.3%）高出18.6个百分点。传统监督学习模型表现出对多数类的系统性偏向；值得注意的是，TabNet取得了相对较高的总体准确率（69.7%），但平衡准确率较低（43.6%）。相比之下，PedX-LLM在所有指标上均保持了更优的平衡性，其F1分数显著高于基线方法。通过少样本配置引入五个验证示例后，平衡准确率进一步提升至72.2%（+5.3%），同时降低了性能方差（标准差：3.5%），表明其仅需少量校准数据即可实现稳健的适应能力。

为详细评估跨站点泛化性能，表8展示了模型在五个未见测试地点的性能，这些地点涵盖从双车道到十车道的道路设施，以及简单和复杂的过街场景。

表8 模型在未见测试地点上的泛化性能

Site ID	位置	车道	逻辑回归	分层逻辑回归	CatBoost	TabNet	PedX-LLM（零-镜头）	PedX-LLM（少镜头）
1	汉普顿 Blvd	6	38.9%	42.5%	46.1%	41.6%	65.3%	69.1%
9	杜鹃花花园 Rd	2	45.1%	48.0%	51.4%	46.5%	70.7%	74.7%
10	N 军事 Hwy	10	35.2%	39.1%	44.6%	36.9%	61.8%	67.7%
27	汉普顿 Hwy	4	38.5%	41.7%	45.2%	40.0%	62.6%	68.8%
35	桥 Rd	4	47.7%	51.1%	54.0%	49.3%	73.8%	76.5%
Mean			41.2%	44.8%	48.3%	43.6%	66.9%	72.2%
Std Dev			4.8%	4.7%	3.8%	4.9%	5.0%	3.8%

PedX-LLM 框架在所有测试环境中均表现出稳定优势。与监督学习基线 CatBoost 相比，其零样本性能提升幅度从地点 10 的 17.2% 到地点 35 的 19.8% 不等。地点 10 是一条十车道主干道，其中路段中部过街率较低（12%），其极端道路几何特征使其成为最具挑战性的场景。基线模型未能实现泛化，平衡准确率分数介于 35.2% 至 44.6% 之间；而 PedX-LLM 在零样本配置下达到了 61.8%。这一成功源于模型能够运用领域知识，理解车道数过多对过街行为的抑制作用。同样，该模型也有效适应了地点 9 的分布偏移（这是一条双车道集散道路，其中路段中部过街率最高），准确率达到 70.7%，而 CatBoost 为 51.4%。

基线模型在不同地点的性能差异显著（平衡准确率为 35.2%-54.0%），在极端几何配置下表现最差。地点 10（十车道主干道）的基线准确率最低（35.2%-44.6%），原因是训练数据未能充分涵盖罕见的宽路幅配置；而地点 35（四车道集散道路）的基线性能相对较好（47.7%- 54.0%），因为该几何配置更接近典型训练条件。PedX-LLM 通过领域知识增强推理，在所有配置下均保持了稳定的性能。

这些结果证实，PedX-LLM 框架的领域知识增强的语言推理能力使模型能够在不同地点之间实现可靠的泛化。将特定地点的模式

识别转化为可泛化的行为推断，克服了纯数据驱动方法固有的分布偏移局限，为多样化的城市环境提供了可扩展的解决方案。

5.4. 可解释性与演示

为了解释 PedX-LLM 框架如何推断行人过街行为，本节采用了一种基于句子的特征归因方法，该方法以夏普利值理论（Winter 2002）为基础。夏普利值源自合作博弈论，通过计算每个特征在所有可能特征组合中的平均边际贡献，为将模型输出归因于输入特征提供了严格的数学框架。该方法通过效率（归因值之和等于模型输出）和对称性（等价特征获得相同的贡献度）等关键性质，确保归因公平。夏普利值广泛用于监督学习可解释性，在理解复杂模型行为方面尤其有效。在我们的实现中，输入提示被分解为七个不同的组成部分（行人群体特征、交通控制、道路几何、建成环境、土地利用与公共交通、环境条件和领域知识背景）。每个组成部分都被视为一个特征，其夏普利值用于量化该组成部分对模型过街地点推断置信度的贡献。

本研究通过汇总整个包含687条观测的数据集中的绝对夏普利值，量化了全局特征重要性。表9展示了各提示组件的贡献度排名。

表9 过街位置推断中的特征重要性汇总

Rank	提示组件	平均绝对夏普利值	贡献度（%）
1	行人群体特征	0.304	25.8%
2	交通控制	0.257	21.8%
3	领域知识背景	0.150	12.7%
4	道路几何	0.147	12.5%
5	建成环境（视觉）	0.142	12.1%
6	土地利用与公共交通	0.090	7.6%
7	环境条件	0.088	7.5%

行人群体特征（25.8%）和交通控制（21.8%）是主导因素，合计占模型决策权重的近一半。这与行为学文献的观点一致，即个体风险容忍度和延误是主要动机。领域知识背景（12.7%）、道路几何（12.5%）和建成环境（12.1%）构成了第二梯队的重要决定因素。这表明，尽管个体特征驱动决策，模型仍高度依赖物理框架和领域原理来调节这些选择。土地利用与公共交通（7.6%）和环境条件（7.5%）提供了情境背景，进一步优化推断。

为展示模型在预测过街行为时捕捉关键的人口统计特征和建成环境差异的能力，图4通过代表性案例展示了特征归因结果。每个案例由四个部分组成：原始数据部分（左侧），包括文本数据和视觉数据；PedX-LLM 输入部分，展示输入模型的结构化提示；模型推理过程部分（中间），说明模型如何推理不同因素对过街决策的影响；以及带有沙普利值图表的推断结果框（右侧），展示预测的过街地点、模型置信度和特征贡献，其中负值（左侧）有利于交叉口过街，正值（右侧）有利于路段中部过街。



图4：个案的特征归因：（a–b）相同建成环境下不同的行人群体特征（结伴的老年女性与独自一人的成年男性）；（c–d）相同行人群体特征下不同的建成环境（第17和第19号地点）。

案例 1 和 2 展示了地点 10 完全相同的建成环境条件如何产生截然相反的过街预测（92.8% 选择交叉口过街，65.2% 选择路段中部过街），而这一差异完全由行人群体差异驱动。行人群体特征呈现出最显著的反转（−0.239 至 +0.128），反映了从老年女性群体的谨慎转向成年男性独自行走的风险容忍度。领域知识背景的作用方向也发生反转（−0.118 至 +0.104），从适应老年人偏好路口过街转为适应成年人接受路段中部过街。交通控制在两种情况下均维持相同的 50 秒周期，却产生相反的影响（−0.203 对比 +0.148）：老年人认为等待可以接受，而独自行走的成年人则将延误视为选择路段中部过街的诱因。道路几何的影响始终为负（−0.215 对比 −0.128）：十车道主干道对老年人构成难以逾越的障碍，对成年人则只是中等程度的阻碍。环境条件的解读因语境而异（+0.048 对比 +0.057）：路段中部照明优势对风险规避型老年人而言微不足道，却会积极影响成年人的决策。土地利用与公共交通以及建成环境的影响均从负转正（−0.070 至 +0.066；−0.113 至 +0.082）：前者引导老年人选择正规过街设施，而分散式商业布局则强化成年人选择路段中部过街的倾向。这验证了模型会依据人口群体脆弱性调节相同基础设施的影响，而非套用固定权重。

案例 3 和 4 展示了不同的建成环境如何产生截然相反的过街预测（路段中部为 58.6%，交叉口为 62.2%），尽管行人群体特征完全相同（独自步行的成年女性）。行人群体特征始终表现出稳定的正向贡献度（+0.338 对比 +0.262），反映了独自行走的成年人偏好直接路线。交通控制的变化幅度最大，从促进路段中部过街转变为最强的交叉口预测因素（−0.212 变为 −0.301）：地点 17 的绿灯时段较短，尽管信号配时可预测，仍提高了对交叉口过街的偏好；而地点 19 较长的绿灯时间和保护性信号配时则减少了延误，提升了感知到的路口安全性。道路几何从促进因素转变为阻碍因素（+0.196 变为 −0.196）：地点 17 的凸起中央分隔带使安全的两阶段过街成为可能，而地点 19 缺少中央分隔带，加之车速较高且人行道宽度为零，大幅增加了路段中部过街暴露风险。环境条件则从正向转为负向（+0.071 变为−0.067）：地点 17 的晴朗天气有助于保持足够的能见度，且没有额外限制；地点 19 的毛毛雨则增加了感知风险，使偏好转向受控路口。建成环境的影响方向也发生逆转（+0.188 变为 −0.156）：地点 17 的中央安全岛和走廊出入口支持直接在路段中部过街，而地点 19 的标线人行横道和信号引导则强化了在交叉口过街的行为。领域知识背景以及土地利用与公共交通也呈现类似的反转（+0.174 变为 −0.168；+0.113 变为 −0.084）：地点 17 的商业活动和公共交通可达性使行人需求沿走廊分布；相比之下，地点 19 所处的高速环境缺少显著的路段中部吸引点。这验证了道路几何是关键调节因素，而基础设施缺陷会加剧不利环境条件的影响。

将可解释性分析与案例级归因相结合，可以为呈现高路段中部过街率的地点揭示可执行的干预优先事项：（1）优化信号配时，以减少相邻交叉口的行人延误，从而缩小不同穿越地点之间的效用差距，因为交通控制始终是贡献最大的因素之一；（2）安装中央隔离带护栏或围栏等实体阻隔设施，以提高感知到的路段中部过街复杂度，因为跨案例分析表明，道路几何具有显著影响；（3）通过增设响应式按钮和清晰标线的人行横道来完善交叉口基础设施，从而提升合规穿越吸引力，尤其是针对

弱势人群。该模型能够联合考虑几何特征、交通控制参数和人口统计特征，这证实了视觉与知识增强推理能够实现具有可解释性且具备语境感知能力的行为推断，为有针对性的、地点特定的安全干预奠定了基础。

6. 结论

本研究提出了 PedX-LLM，这是一个旨在将行人过街推断从数值模式拟合转向具备泛化能力的上下文感知行为推理的框架。该框架集成了多模态输入，包括卫星影像（用于捕捉建成环境）和文本数据，并对行人行为的领域知识进行编码，从而构建行人过街决策的全面表征。我们采用 LoRA，利用训练数据对该框架进行微调，并在此前未见过的环境中评估其性能。

研究结果表明，PedX-LLM 框架能够有效推断行人过街行为并识别关键决定因素。该框架整合了个体层面属性、建成环境特征和领域知识，平衡准确率达到 82.0%，比 CatBoost 高 3.0 个百分点，比统计模型（分层逻辑回归）高 7.9 个百分点。基于 Shapley 值的归因分析显示，行人群体特征（25.8%）和交通控制（21.8%）是主要行为驱动因素。通过 LLaVA 模块从卫星影像中提取视觉增强的建成环境特征后，性能进一步提升了 2.9%，这表明集成视觉上下文具有重要价值。消融研究量化了领域知识整合的贡献度；结合个体层面知识和环境知识，相较于基线累计提升了 4.1 个百分点。此外，跨地点评估证实了模型具有稳健的泛化能力。在未见地点的零样本配置下，该框架的平衡准确率达到 66.9%，显著优于分层逻辑回归（46.2%）和 CatBoost（48.3%）。少样本适配进一步将性能提升至 72.2%，表明该模型能够有效克服纯数据驱动方法所受的分布偏移限制。

本研究为专用于行人行为分析的大语言模型建立了一种范式，从而为相关文献作出了贡献。首先，研究提出了一种多模态架构，将视觉提取的情境信息与文本行为记录相结合，克服了单模态方法无法捕捉环境协同效应的局限。其次，通过采用 LoRA 进行参数高效微调，所提出的框架能够有效利用本地数据，从而提升性能。第三，该框架通过结构化提示工程，将交通领域知识融入模型的推理过程。这种集成使特定地点的模式识别提升为可泛化的行为推断。最后，该框架基于部署在本地服务器上的开源大语言模型实现，从而确保专有数据得到保护并维护隐私。

从实践角度来看，PedX-LLM 支持在数据稀缺且异质的城市环境中开展行人行为分析，而传统模型在这些环境中表现不佳。其出色的零样本和少样本泛化能力，使交通工程师能够利用极少的本地数据，评估新建或改造地点的行人过街行为。通过联合利用视觉上下文和领域知识，

该框架还有助于开展行人设施的设计评估和基于场景的分析，为基础设施和交通控制决策提供依据。此外，将其部署在本地托管的开源大语言模型上，可为面向机构的决策支持工具提供兼顾隐私保护与可扩展性的基础。

未来的工作将着重于扩展数据集，以涵盖更多样化的城市环境，包括高密度大都市地区和郊区走廊；在交通法规和文化规范各异的不同地理区域验证该框架；并探索集成交通流量和天气状况等实时数据流，从而实现动态行为推断和自适应对策部署。

致谢
本研究部分由Old Dominion University（ODU）土木与环境工程系交通信息学实验室资助。本研究使用的行人过街数据收集自以下项目：“影响行人选择路段中部过街的因素及潜在对策，”该项目由弗吉尼亚州交通研究委员会（VTRC）资助。本文内容仅代表作者观点，作者对文中所述事实和数据的准确性负责。本文内容不代表VTRC或其他相关机构的官方观点或政策。

参考文献
Abbasi, S., Rahmani, A.M., 2025. 面向自动驾驶车辆大语言模型的上下文感知提示工程。Concurrency and Computation: Practice and Experience 37 (27-28), e70392. Ahsan, M.J., Abdel-Aty, M., Anwari, N., Barbour, N., 2025. 利用视频数据和具有均值异质性的混合Logit模型评估驾驶员在路段中部人行横道处的制动行为。Transportation Research Interdisciplinary Perspectives 31, 101380. Ali, A., Ghanem, M.C., 2025. 超越检测：大语言模型与新一代网络安全。SHIFRA 2025, 81-97. Blečić, I., Saiu, V., A. Trunfio, G., Year. 利用多模态大语言模型提升城市步行适宜性评估，Cham，第394-411页。Choi, W.C., Chang, C.I., 2025. 开源与商业大语言模型（llms）的优势与局限：DeepSeek和OpenAI ChatGPT的比较研究。Deng, Z., Nagel, R., Yan, J., Zhang, T., 2025. 大语言模型智能体能否识别经济博弈中的人口统计学异质性？可见于SSRN。Efron, B., 2020. 预测、估计与归因。International Statistical Review 88, S28-S59. Elalouf, A., Birfir, S., Rosenbloom, T., 2023. 开发基于机器学习的模型，以减轻行人在道路交通事故中所受伤害的严重程度。Heliyon 9 (11). Fahad, S.A., Zhengkui, D.W., Chet, N.P., Wong, N., Ng, A.B., See, S., 2025. 多模态大语言模型的进展与应用：集成、挑战与未来方向。人工智能驱动：社交媒体分析与网络安全。Springer，第309-336页。Geiger, A., Ibeling, D., Zur, A., Chaudhary, M., Chauhan, S., Huang, J., Arora, A., Wu, Z., Goodman, N., Potts, C., 2025. 因果抽象：机制可解释性的理论基础。Journal of Machine Learning Research 26 (83), 1-64。

Gu, C., Liu, X., Mao, N., 2024. 隧道—互通立交连接路段中的驾驶员行为机制与冲突风险模式：基于行为适应理论的全面调查。Sustainability 16 (19), 8701. Hagenaars, J.A., Kühnel, S., Andress, H.-J., 2024. 解读与比较Logistic、Probit和Logit回归中的效应。Sage Publications。Ham, J.-S., Huang, J., Jiang, P., Moon, J., Kwon, Y., Saripalli, S., Kim, C., Year. Omnipredict：GPT-4o增强的多模态行人过街意图预测。载于：《自适应基础模型：迈向个性化高效学习的人工智能》论文集。Hang, T., Liu, S., Feng, J., Djigal, H., Huang, J., 2025. 基于提示学习的少样本关系抽取：分类体系、综述、挑战与未来方向。《ACM计算调查》。Hockstad, T., Rahman, M., Jones, S., Chowdhury, M., 2025. 交通运输网络安全与数据隐私的监管差距分析。Transportation Journal 64 (1), e12036. Holland, C., Hill, R., 2010. 成年行人全生命周期中不安全过街决策预测因素的性别差异：一项模拟研究。《事故分析与预防》 42 (4), 1097-1106. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., 2022. LoRA：大语言模型的低秩适应。ICLR 1 (2), 3. Huang, J., Xu, Y., Wang, Q., Wang, Q.C., Liang, X., Wang, F., Zhang, Z., Wei, W., Zhang, B., Huang, L., 2025. 基础模型与智能决策：进展、挑战与展望。《创新》。Huang, X., Zeng, L., Liang, H., Li, D., Yang, X., Zhang, B., 2024. 利用大数据和深度学习技术全面评估城市行人环境的步行适宜性。《科学报告》 14 (1), 26993. Jacobucci, R., Grimm, K.J., Zhang, Z., 2023. 面向社会与行为研究的机器学习。Guilford Publications。Jiang, J.-P., Liu, S.-Y., Cai, H.-R., Zhou, Q., Ye, H.-J., 2025. 表格数据的表示学习：全面综述。arXiv预印本 arXiv:2504.16109。Jie, T., Khan, A., Azam, S., Khalil, U., Ejaz, N., Afzal, M., Said, Y., Kucher, D.E., 2025. 行人安全分析与预测的先进方法：多传感器数据融合与GIS技术的进展。《IEEE应用地球观测与遥感精选主题期刊》。Karim, M.M., Shi, Y., Zhang, S., Wang, B., Nasri, M., Wang, Y., 2025. 大语言模型及其在道路安全与出行能力提升中的应用：全面综述。arXiv预印本 arXiv:2506.06301。Kim, S., Hong, S., Jin, J., Noh, B., Year. 利用多模态大语言模型进行驾驶场景上下文增强的轨迹预测，并结合风险感知决策推理。载于：EPIA人工智能会议论文集，第192-204页。Koldasbayeva, D., Tregubova, P., Gasanov, M., Zaytsev, A., Petrovskaia, A., Burnaev, E., 2024. 环境研究与实践中数据驱动地理空间建模的挑战。《自然通讯》 15 (1), 10700. Kyriazos, T., Poga, M., 2024. 机器学习模型在社会科学中的应用：管理非线性关系。《百科全书》 4 (4), 1790-1805. Le, T.-D., Jouvet, P., Noumeir, R., 2025. 在有限数据集上利用混合专家模型提升Transformer对法语临床笔记进行分类的性能。《IEEE转化工程与健康医学期刊》。Li, J., Li, J., Yang, G., Yang, L., Chi, H., Yang, L., 2025. 大语言模型与多模态大模型在自动驾驶中的应用：全面综述。Li, W., Year. 面向高维表格数据的TabNet：通过特征融合提升可解释性与性能。载于：IET会议论文集 CP915，第168-173页。

Liu, H., Li, C., Wu, Q., Lee, Y.J., 2023. 视觉指令微调. 《神经信息处理系统进展》 36, 34892-34916. Liu, T., Yang, J., Yin, Y., 2025. 面向交通运输系统的基于大语言模型智能体的建模：概念框架. 《交通运输人工智能》 1, 100001. Mirza, I., Jafari, A.A., Ozcinar, C., Anbarjafari, G., 2025. 不同大语言模型的性别偏见分析. Osei, K.K., Obiri-Yeboah, A.A., Adu-Gyamfi, L., Ackaah, W., 2024. 行人道路过街行为与偏好：基于人际行为理论的视角. 《交通伤害预防》 25 (1), 91-100. Papadimitriou, E., Lassarre, S., Yannis, G., 2016. 将人为因素纳入行人过街行为模型. 《交通研究 F 辑：交通心理学与行为》 36, 69-82. Peykani, P., Ramezanlou, F., Tanasescu, C., Ghanidel, S., 2025. 大语言模型：结构化分类体系，以及对挑战、局限、解决方案和未来方向的综述. 《应用科学》 15 (14), 8103. Przybyła-Kasperek, M., Marfo, K.F., 2024. 用于表格分散数据中多样条件属性的多层感知机神经网络. 《公共科学图书馆·综合》 19 (12), e0311041. Pu, Q., Xie, K., Guo, H., 2025a. 对安全关键事件中自动驾驶车辆与人工驾驶车辆交互式跟驰行为的建模：基于tgsim数据集的多智能体状态空间注意力增强框架. 见SSRN 5239734。Pu, Q., Xie, K., Guo, H., Zhu, Y., 2025b. 对车辆与行人近失碰撞场景中碰撞规避行为的建模：曲线碰撞时间与Mamba驱动的深度强化学习. 《事故分析与预防》 214, 107984. Pu, Q., Xie, K., Guo, H., Zhu, Y., 2026. 对交互式碰撞规避行为的建模：多智能体状态空间Transformer增强强化学习框架. 《事故分析与预防》 226, 108334. Pu, Q., Zhu, Y., Wang, J., Yang, H., Xie, K., Cui, S., 2025c. 利用无人机数据分析测量高密度区域交叉口的交通指标. 《交通研究记录》, 03611981241311566. Rafie, Z., Talab, M.S., Koor, B.E.Z., Garavand, A., Salehnasab, C., Ghaderzadeh, M., 2025. 运用XGBoost和可解释人工智能准确预测2型糖尿病. 《BMC公共卫生》 25 (1), 3688. Raiaan, M.a.K., Mukta, M.S.H., Fatema, K., Fahad, N.M., Sakib, S., Mim, M.M.J., Ahmad, J., Ali, M.E., Azam, S., 2024. 大语言模型综述：架构、应用、分类体系、未解决的问题与挑战. 《IEEE Access》 12, 26839-26874. Rodríguez, D.A., Brisson, E.M., Estupiñán, N., 2009. 波哥大快速公交车站周边路段层面的建成环境属性与行人活动之间的关系. 《交通研究D辑：交通与环境》 14 (7), 470-478. Shanmugarasa, Y., Ding, M., Arachchige, C.M., Rakotoarivelo, T., 年份. 大语言模型的隐私悖论（SoK）：进展、隐私风险与缓解措施。载于：第20届ACM亚洲计算机与通信安全会议论文集，第425-441页。Shanthini, S., Devi, M.S., Grace, R.S., 2024. 人工神经网络（ANN）中感知机学习规则与架构的全面综述. 《工程、能源与金融中的深度学习》，115-163。Sharif, H.O., Joseph, J., Dessouky, S., Eid, E., Weissmann, J., 2025. 选定高风险行人路段中部过街点并开展安全评估，以规划行人过街设施。Singh, D., Das, P., Ghosh, I., 2024. 使用机器学习算法预测无信号交叉口的行人过街行为：分析与比较. 《多模态用户界面期刊》 18 (2), 239-256.

Strömel, K.R., Henry, S., Johansson, T., Niess, J., Woźniak, P.W., 年份. 讲述健身故事：利用大语言模型解读健身追踪器数据以促进反思。载于：2024年CHI计算系统中的人为因素会议论文集, 第1-16页。Tanveer, M., Rajani, T., Rastogi, R., Shao, Y.-H., Ganaie, M., 2024. 孪生支持向量机综合综述. 《运筹学年鉴》 339 (3), 1223-1268. Torneiro, A., Monteiro, D., Novais, P., Henriques, P.R., Rodrigues, N.F., 2025. 迈向基于视觉语言模型的通用城市监测：综述、评估与研究议程. arXiv预印本 arXiv:2510.12400. Wilmut, K., Purcell, C., 2022. 为什么老年行人面临更高风险？一项系统综述. 《人为因素》 64 (8), 1269-1291. Winter, E., 2002. 夏普利值. 《博弈论手册：经济学应用》 3, 2025-2054. Xie, K., Yang, H., Ishak, S., Zhai, G., Yan, Z., 2024. 影响行人选择在路段中部过街的因素及潜在对策. Yan, Y., Liao, Y., Xu, G., Yao, R., Fan, H., Sun, J., Wang, X., Sprinkle, J., An, Z., Ma, M., 2025. 大语言模型在交通与运输研究中的应用：方法论、最新进展与未来机遇. arXiv预印本 arXiv:2503.21330. Yang, L., Yang, H., Cui, J., Zhao, Y., Gao, F., 2024. 建成环境因素对老年人步行行为的非线性与协同效应：结合LightGBM和SHAP的分析. 《城市数据、科学与技术汇刊》 3 (1-2), 46-60. Yang, Y., Li, Z., Bizimana, J.P., Zafri, N., Dong, Y., Li, T., 2025. 大语言模型与行人安全：在无信号交叉口预测驾驶员礼让行为中的应用. arXiv预印本 arXiv:2509.19657. Yao, Y., Duan, J., Xu, K., Cai, Y., Sun, Z., Zhang, Y., 2024. 大语言模型（LLM）安全与隐私调查：优点、缺点与丑陋之处. 《高可信计算》 4 (2), 100211. Yuan, X., Liu, S., Feng, W., Dauphin, G., 2023. 基于随机森林的端到端学习算法的特征重要性排序. 《遥感》 15 (21), 5203. Zhang, L., Jánošík, D., 2024. 结合混合机器学习模型提升短期负荷预测：CatBoost和XGBoost方法. 《应用专家系统》 241, 122686. Zhang, X., Ouyang, T., Shang, Y., Liao, Q., Li, Y., 2025. UrbanMLLM：联合学习跨视角图像以实现城市理解。Zhou, Q., Zhang, J., Zhu, Z., 2025. 利用多模态大语言模型和街景图像评估城市视觉吸引力感知. 《建筑》 15 (16), 2970. Zhu, M., Graham, D.J., Zhang, N., Wang, Z., Sze, N., 2025. 天气对路段中部过街行人安全感知的影响：一项基于CAVE的研究. 《事故分析与预防》 215, 107988.