

街道视觉智能在城市规划领域的分析范式转型*

Paradigm Shift of Street Visual Intelligence in Urban Planning

尹瀚琦 孙玉梅 邬伦 张帆 YIN Hanyu, SUN Yumei, WU Lun, ZHANG Fan

摘要 街道作为城市的基本单元,其视觉特征不仅直观反映了城市空间的品质,还深刻影响着居民的生活福祉。街景图像以人的视角反映街道场景,为微观层面的城市环境分析提供了独特视角,成为城市规划领域不可或缺的数据。系统梳理街景数据在城市规划中的应用历程,重点分析其与人工智能技术发展的紧密关联。从早期的传统手工审计,到基于统计机器学习的特征工程,再到深度学习驱动的自动化分析,街景数据的利用效率和分析精度得到了极大提升。近年来,自监督学习和大语言模型的应用,进一步拓展了街景数据的潜力,使其能够支持更复杂的城市分析任务。通过回顾街景数据在不同技术阶段的应用,揭示了人工智能如何重塑其分析范式,并展望其在未来城市规划建设中的潜力。

Abstract Streets serve as the foundational elements of urban environments. Their visual characteristics directly reflect urban spatial quality and influence residents' well-being. Street view imagery captures street scenes from a human perspective, providing a unique viewpoint for micro-level urban environment analysis and serving as an indispensable resource for urban planning. This study systematically reviews the development of street view image applications in urban planning, especially focusing on their integration with advancing artificial intelligence (AI) techniques. From early manual audits to feature extraction powered by statistical machine learning, and later to automated analysis driven by deep learning, the efficiency and accuracy of street view images utilization have markedly improved. In recent years, the incorporation of self-supervised learning and large language models has markedly enhanced the application potential of street view images, enabling more complex urban analysis compared to earlier approaches. By reviewing the application of street view data across different technological stages, this study illustrates how artificial intelligence has reshaped its analytical paradigms and explores its promising potential for future urban planning.

关键词 街景图像;人工智能;城市规划;分析范式

Key words street view images; artificial intelligence; urban planning; analytical paradigm

文章编号 1673-8985 (2025) 02-0025-07 中图分类号 TU981 文献标志码 A

DOI 10.11982/j. sup. 20250204

作者简介

尹瀚琦

北京大学地球与空间科学学院遥感与地理信息

系统研究所

博士研究生

孙玉梅

石家庄铁路职业技术学院

副教授,硕士生导师

邬伦

北京大学地球与空间科学学院遥感与地理信息

系统研究所

教授,博士生导师

张帆(通信作者)

北京大学地球与空间科学学院遥感与地理信息

系统研究所

研究员,博士生导师

fanzhanggis@pku.edu.cn

0 引言

街道作为城市空间的基本单元,不仅承载着居民日常生活、经济活动和社会文化交流,更是城市形态和功能的重要体现。在我国“新型城镇化”和“智慧城市”建设战略背景下,街道的规划与治理已成为提升城市品质、

*基金项目:国家自然科学基金面上项目“基于街景大数据的城市物质空间多重语义表达”(编号42371468)资助。

促进社会公平和实现高效治理的关键环节。街道的布局、环境和功能状态直接反映出城市的活力与韧性,为城市更新和精细化治理提供了基础数据和实践窗口。

传统城市规划主要依赖于地图数据和文字描述,往往难以捕捉街道环境中建筑风貌、绿化覆盖、交通设施和人流动态等微观细节。而街景图像等视觉信息则以直观、生动的方式还原了城市建成环境的真实状态,为规划者提供了多维度、时空连续的现场数据^{[1]1049}。这类数据不仅弥补了传统方法在细节捕捉上的不足,更为区域功能划分、环境美学评价和安全风险评估等决策提供了坚实的实证依据,助力构建宜居、和谐、可持续的城市空间^{[1]1045}。

伴随着人工智能和大数据技术的迅速发展,街道视觉信息的分析范式经历了从传统手工调查与静态照片记录,到统计机器学习与特征工程,再到深度学习及自监督学习驱动的全自动化分析的重大转型。每一阶段的技术进步不仅提高了数据采集和处理的效率,也不断拓展了对街道空间多维特征的理解和应用。这种从“经验驱动”向“数据驱动”的范式转变,正契合我国当前推动城市更新、优化存量资源和提升城市治理精细化水平的总体要求^[2]。本文将系统梳理这一范式转型的历程,旨在为实际的街道规划与城市高质量发展提供一定的理论支持和参考。

1 街景数据:概念、来源与优势

街景数据是指通过街道级别的影像采集技术记录的城市环境数据,通常以全景图像或视频形式呈现。与传统地图、航拍或遥感影像不同,街景数据以人眼视角捕捉城市空间的细节,直观呈现建筑立面、街道家具、绿化覆盖、道路状况和社会活动^{[1]1044}。由于其高精度、连续性和可感知性,街景数据在城市规划、环境评估、智能交通等领域具有重要应用价值,尤其在城市更新研究中,成为监测街道空间演变、评估环境品质的重要数据来源。

街景数据的获取主要分为两类。第一类

是商业平台提供的大规模影像数据,通常通过装载在街景采集车上的摄像机进行采集,具有数据覆盖广、视角相对固定的特点。以Google街景(Google Street View, GSV)为例,其自2007年开始提供服务,时效性较强,一般每1—3年更新一次,覆盖全球90多个国家,但由于数据限制,在中国大陆地区没有完整的GSV数据。在中国大陆地区,腾讯街景和百度街景是重要的街景数据来源:腾讯自2011年开始提供街景服务(现已停止更新),百度自2013年起提供了中国多个地级市的街景服务(更新频率因地区而异,多个城市近年仍有更新)。第二类是基于志愿者服务的街景数据采集,如Mapillary和KartaView,依靠用户自行使用智能手机或相机拍摄上传,其特点是能够获取商业平台覆盖不到的偏远地区数据,但覆盖范围不均匀、视角不一致,数据质量受用户设备和贡献活跃度影响。总结而言,目前研究中主要使用商业平台的数据,因其精度高、覆盖广且更新稳定,更能满足大规模城市分析和应用需求。

与传统城市环境数据相比,街景数据具备多项优势,使其在街道建筑环境研究中具有独特价值。首先,其高空间分辨率能够清晰展现建筑外立面、绿化景观、公共设施等微观环境要素,是研究街道尺度空间品质的理想数据源。其次,街景数据的时间序列性为其赋予了对比分析的潜力,例如Google街景、百度街景等提供了多期街景图像,使研究者能够追踪不同时间段的街道变迁,为城市更新过程的量化分析奠定基础。街景数据可以整合建筑足迹、三维点云等多种模态数据,将全景视图与三维特征相结合,通过三维重建技术直观展现街道空间,以支持虚拟城市环境重建。这些优势赋予街景数据独特的价值,使其成为城市规划中连接微观观察与宏观决策的重要桥梁。

街景数据的应用价值在于可以利用技术方法量化其展现的城市街道物理特性,提取关键指标,从而为城市规划决策提供支持。例如,通过计算机视觉技术分析街景图像,提取

基础设施破损程度的指标(如路面裂缝、路灯损坏),用于为维护优先级排序提供依据^[3];通过对街景图像中的植被进行分割,提取区域绿化率指标,用于指导城市绿地布局优化^[4];通过街景元素分布、建筑密集度和行人活动的评估,提取感知指标(如安全感),用于提升居民生活品质^[5];通过分析城市形态特征,提取天空开阔度等指标,用于反映热岛效应和通风条件,支持热环境改善规划^[6];通过测量人行道宽度、障碍物分布和交叉口设计,提取可步行性指标,用于为步行友好型街道设计提供数据支持^[7]。这些指标提取后,可用于确定城市更新区域的优先级、制定绿化提升方案或优化公共设施配置,从而将街景数据的分析成果转化为城市规划的实际决策依据,增强其实用性。

随着数据采集技术的进步和人工智能的发展,街景数据在城市研究中的应用范围不断扩展^[8]。从最初的视觉记录到如今的智能分析,街景数据不仅提升了城市空间研究的精度,还推动了规划决策的科学化和数据驱动化。未来,结合计算机视觉与深度学习,街景数据将在城市更新、环境评估、交通优化等领域发挥更大作用,为构建更宜居、更智慧的城市提供坚实的数据支撑。

2 街景数据在城市规划领域的分析范式转型

街景数据作为城市环境分析的重要数据源,其应用方式随着人工智能(AI)技术的发展经历了显著转型。本节旨在从数据来源、分析方法及应用任务等角度,系统梳理街道视觉智能分析范式在城市规划领域的发展演进(见图1)。总的来说,数据来源从早期的人工拍摄逐步转向多平台街景数据;分析方法由传统手工审计及统计分析,逐渐升级为基于人工智能驱动的自动化分析;应用任务则从静态的单维度评估,发展为动态预测与多模态融合分析。这些转变提高了城市环境分析的精细度和自动化水平,为智慧城市建设提供了更加精准、有力的决策支持。笔者将通过梳理不同技术阶段的核心方法及其在城市

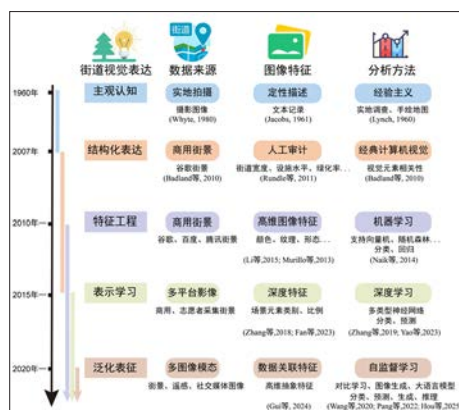


图1 街道视觉分析范式的演变:从经验驱动到人工智能驱动

Fig.1 Evolution of street visual analysis paradigms: from empirically driven to AI-driven

资料来源:笔者自绘。

规划中的典型应用,系统阐述人工智能技术如何重塑街景数据的分析范式。

2.1 主观认知与经验主义阶段

在Google街景出现之前,街景数据的获取主要依赖人工实地调查和摄影。研究者通过走访街道,拍摄照片,并手工记录街道布局、建筑形态、公共空间等物理特征,以分析城市空间和人类行为之间的关系。例如,Whyte^[9]通过摄影与录像的方式观察公共空间中人群的行为,揭示局部区域视觉环境与人类生活的联系。类似地, Jacobs^[10]在《美国大城市的死与生》中强调了街道的“活力”和“多样性”对城市更新的重要性,采用实地观察的方法为街道功能优化提供了理论支持。Lynch^[11]在《城市意象》一书中,通过实地观察和手绘认知地图的方式,总结出城市空间中影响居民感知的5大要素(路径、边界、区域、节点和地标),奠定了城市空间感知研究的基本框架。然而,这一阶段的研究主要属于现象学范畴,依赖经验主义进行定性分析,而非标准化的定量研究。此外,这种实地调研方法仅适用于小范围区域,难以进行大范围的推广应用。

2.2 街道视觉结构化表达阶段

Google街景的推出极大推动了街道尺度

上的城市规划与建成环境研究。其高覆盖率、低成本和可远程访问的特点,有效突破了传统实地调查在人力与空间尺度上的局限,使得大范围、标准化的街道环境分析成为可能。在Google街景服务普及初期,研究者主要依赖人工目视解译和简单图像特征提取进行建成环境审计。例如, Rundle等^[12]通过人工标注纽约社区街景的步行友好性指标(如人行道宽度、过街设施),结合线性回归模型揭示了街道环境与公共健康的显著关联。Badland等^[13]开发了标准化街景审计工具“SPACES”,通过人工识别绿化率、建筑维护等指标,支持墨尔本街道质量评估提供量化依据。尽管这类研究为街景数据在城市规范方面的应用提供了实证依据,但其方法论仍存在显著瓶颈:人工标注耗时数百小时,且依赖传统的线性回归模型,难以捕捉街道环境的复杂空间关联。例如,对百万级街景图像的安全评估需要超过千小时的人力^{[14]793},并且简单的线性回归模型在处理多种数据时可能面临共线性问题,从而影响分析结果。

2.3 特征工程与统计机器学习阶段

随着计算机视觉技术的快速发展,街景图像分析逐渐从依赖人工审阅向自动化、可扩展的处理方式转变,以应对传统方法在效率与精度上的双重瓶颈。在这一转型过程中,研究者开始借助特征工程方法,从图像中自动提取具有判别力的结构化特征,并结合能够有效应对共线性问题的机器学习模型,对街道环境进行更高效和更精准的分析。特征工程是指通过人工设计特征提取算法,从原始图像中提取色彩、纹理、边缘等维度的数值特征,以供后续建模分析使用。例如, Naik等^{[14]795}通过提取街景图像的颜色和纹理特征,结合支持向量机(SVM)算法,构建了街道安全感预测模型(StreetScore)。该模型能够自动分析并评估街道的安全感,尤其是对百万级街道的安全评估,提供了重要的数据支持,推动了城市安全导向的街道更新。Li等^[15]提出了基于颜色阈值分割的绿视率指数(Green View Index),通

过量化街道的绿化水平,为城市街道绿化改造提供了科学依据。

尽管这些方法成功突破了人工标注的局限,实现了部分自动化,但仍面临一些挑战。传统的图像特征,如颜色和纹理,虽然能描述街景的基本视觉元素,但难以有效捕捉诸如街道氛围、空间形态等更为复杂的高层次语义信息。为了弥补这一不足,研究者们进一步尝试使用能够反映街道整体布局和结构的特征。例如,通过提取街景图像的场景特征描述子(GIST),计算图像的宏观语义特征,进一步分析街景图像对应的土地利用类型和街区的空间布局^[16-17]。总的来说,这种指标在一定程度上关注了街景图像的宏观特征,但仍难以全面捕捉其中丰富且复杂的语义信息,限制了其在更复杂、更综合任务中的应用。

2.4 街道视觉表示学习阶段

深度学习技术的引入彻底革新了街景数据的分析方法,不仅显著提升了效率、降低了成本,也为城市研究带来了全新的技术视角与方法体系。其中,表示学习作为深度学习的核心理念,指的是模型能够从原始图像中自动学习多层次、高维度的特征表示,从而提升对复杂场景的理解能力。相较于依赖人工设计的特征工程方法,表示学习通过端到端的方式实现了街景图像的自动化特征提取,极大地拓展了街景数据在城市研究中的应用边界。它不仅提升了街景图像的大规模处理效率,也推动了城市分析从宏观量化向微观细节解析的范式转变。

在实际应用中,基于街景图像的深度学习研究通常围绕两类核心任务展开:分类任务与预测任务。分类任务主要包括物体检测与语义分割,分别用于识别街景图像中的典型元素(如建筑、车辆、行人等)以及实现对图像中各区域的像素级语义标注;而预测任务则通过将图像特征与其他多源城市数据相结合,实现对居民感知、安全感、交通噪声、人流密度等城市指标的定量评估与空间建模。

在深度特征驱动的分类任务中, Faster R-CNN、YOLO等目标检测模型的出现助力了街景中建筑、车辆、行人等物体的自动识别。这些模型能够高效、准确地定位街景中的各类目标并将其分类,极大地降低了传统人工审计所需的大量时间与人力成本。例如,利用模型可以自动检测城市中的破损建筑或道路^[18]、有待提升的基础设施^[19],发现车辆违规停放^[20]等,为城市更新与管理提供了精准的数据支持。此外,PSPNet、DeepLab等语义分割模型被广泛应用于街景图像的精细化解析。这些模型能够实现像素级的图像分割,从而计算绿视率、天空开阔度、建筑物比例、街道围合度等精细指标^[21-23]。这些指标为城市设计中的绿化覆盖率、视线通廊和街道空间品质等规划目标提供了科学依据,推动街景研究从宏观到微观的转变,为提升城市空间品质提供了新的思路。

同时,在深度特征驱动的预测任务中, ResNet、VGG等预训练卷积神经网络被用于从街景图像中提取高层语义特征。这些高维特征捕捉了图像中的抽象场景信息,并可应用于城市规划相关的预测任务。例如,Zhang等^[24]通过提取街景的深度特征,成功预测了居民对街道环境的多种感知评分(安全、美丽、活力、富裕、无聊、压抑),实现了从图像到居民感知的端到端映射。此外,基于这种端到端思路,利用模型提取的高维特征还可以预测交通噪声^[25]^[28]、交通流量^[26]、建筑物类型^[27]、街区犯罪率^[28]、商业活力^[29]等。这种预测能力使研究者能够从街道视觉角度评估城市规划质量、发展水平及优化需求,为城市政策的制定提供数据驱动的支持。同时,对这种深度特征进行可解释分析,有助于识别街景图像中与预测结果高度相关的主要区域。

总体而言,深度学习在街景分析中的应用不仅显著提高了效率和精度,更为城市研究带来了全新的方法论。它使城市空间分析从传统的描述性研究转向量化建模与区域预测,进而为城市规划、设计和管理提供了更加科学、

系统的技术工具。然而,这些深度学习模型主要依赖于监督学习框架,即需要大量高质量的标注数据来进行模型训练。数据标签的准确性和完整性直接影响模型的性能和分析结果,因此,数据标注的质量成为这一分析范式中的重要挑战。

2.5 自监督学习驱动下的场景泛化表征阶段

自监督学习(Self-Supervised Learning, SSL)通过挖掘数据内部结构(如时空关系与跨模态信息),在无需人工标注的条件下,学习具泛化能力的特征表征^[30]。凭借对标注依赖的突破以及对模型迁移性与自主学习能力的增强,SSL正逐渐成为当前街景分析方法演进中的关键方向。其核心框架主要包括对比学习、生成式学习和自回归模型。对比学习旨在通过拉近相似样本表示、拉远不相似样本表示的方式,有效提取数据的内在特征,并在图像及时空数据的特征学习中得到广泛应用。生成式学习利用数据重建或合成任务捕获复杂语义信息,在城市模拟与规划设计领域具有重要应用价值。自回归模型(如典型的GPT)能够自动分析和生成复杂的城市环境信息,如区域功能分类、环境预测和规划方案生成。上述各框架为城市规划中的数据分析和决策支持提供了坚实的方法论基础。

近年来,研究者基于街景数据,并借鉴自监督学习的理论与方法,在城市规划领域开展了多方向的探索与应用研究。在基于街景图像的对比学习方面,研究主要集中于空间和时间维度的特征对比。在空间对比中,通过构建地理邻近的街景图像正负样本对,对比学习模型能够捕捉城市建成环境中区域间的邻近特征,充分揭示当前街景图像与其地理邻近图像之间的关联,从而为区域社会经济指标的预测提供支持^[31]。而在时间对比中,研究则侧重于同一地区不同时期街景图像之间的对比,以使模型区分并学习场景中的不变特征(如建筑轮廓、树木高度、道路位置等)与变化特征(如树叶颜色、车辆、行人、天气等),从而对区域建成环境进行更为深入的分

析^[32]。除了仅依赖街景图像进行时空对比之外,一些研究还尝试实现不同模态数据的对齐,以期为区域综合表征提供更丰富的视角。例如,将地理邻近的遥感图像与街景图像进行融合,可以从多个视角增强对区域特征的理解,从而为区域建成环境分析和灾后损伤评估^[33]等任务提供更为有力的信息支持。同样,将街景图像与建筑物足迹、兴趣点(POI)等多源数据融合,有助于全面表征区域的建筑属性^[34],并在建筑物功能类型分类等应用中展现出较高的有效性^[35]。此外,传感器数据与街景图像的融合研究有助于对城市建成环境质量进行精细化预测,如空气质量^[36]、噪声水平^[25]^[141]等多个维度,从而为城市环境管理提供科学依据。

街景图像相关的重建与生成是计算机视觉与城市规划交叉领域的研究热点,其核心目标是通过数据驱动方法恢复或补全城市空间的语义信息,解决数据缺失问题,并为规划决策提供支持。在街景图像重建方面,当前主要采用自监督学习中的掩码建模技术,即通过学习图像中已知区域的信息来推测未知区域的内容。该方式有助于提升街景数据质量,如提升街景图像清晰度,修复过曝部分,弥补缺失区域等。街景图像生成则旨在构造原本可能不存在的数据。例如,通过路网俯视图生成街景图像,可有效补全非街景采集区域的数据^[37]。从街景图像的视角重建建筑物三维形态,能够补全建筑足迹数据^[38]。值得注意的是,目前大部分公开的街景数据(如Google、百度提供的)均为白天的图像,而真实采集夜间街景图像成本较大。因此,利用图像生成技术将白天图像转为不同时段的夜间图像,对于夜间街道安全性评估、夜间经济活力分析具有重要应用价值^[39]。此外,在城市设计领域,街景图像生成技术可用于更直观地展示设计效果,如模拟建筑拆建或街区风格更新后的景观样貌。该技术不仅能够生成单幅图像,还可以生成视频,从而呈现街区在不同时期的动态变化信息。这对于辅助研究区域交通优化、人群流量分析、灾害逃生情景以及历史街区演变规律等有重要

意义。

大语言模型 (Large Language Models, LLMs) 通过归纳学习海量数据,生成逻辑一致的回答。其强大的语义理解和推理能力,结合人类反馈,形成了中立的“通用世界观”^{[40]1}。基于此,LLMs能够优化城市规划分析,提升数据质量和模型精度。在数据层面,LLMs通过对街景图像、遥感数据、社交媒体文本等多源信息的归纳学习,生成更符合人类认知的调查数据,丰富城市感知任务(如街道美感、安全性、步行友好性评估等)的数据支持^[41]。在模型层面,LLMs可辅助机器学习模型提升分析精度,例如,通过对比LLMs的场景理解结果与街景语义分割模型的预测,精准识别错分类元素,进而优化分类性能^[42]。

此外,LLMs不仅能够提升现有城市规划分析范式的精度与效率,还能引领范式转型,深化数据驱动与智能化决策^[43]。借助人机知识交互、通用世界观的深度理解以及辅助决策能力,LLMs正推动街道视觉智能分析在城市规划中的应用变革^{[40]2}。例如,对LLMs提针对特定任务的提示指令(prompt)时,其能够充分利用先验知识,实现更精准的城市空间分析。例如,当询问街景图像所对应的城市功能区时,LLMs会对用户的prompt针对性地检索知识并进行合理的推理,能够高效判别出具体的建筑^[44]、用地^[45]、街道功能类型^[46]。此外,LLMs具备整合多模态数据(如街图像、文本、传感器数据等)的能力,能够实现基于街景图像的地理定位^[47],量化建成环境质量^{[48]87718}等经典城市规划任务。最后,LLMs还具备评估和决策优势,例如,在基础设施评估中,LLMs通过分析街景图像中的建筑状态、道路状况等信息,辅助判断城市基础设施的维护需求^{[48]87722};在城市景观评价中,LLMs可结合绿化覆盖率、建筑风格协调性及视觉吸引力等因素,提供城市空间美学的定量评估^[49];在交通安全性分析中,LLMs则通过解读街景图像,分析道路通行状况、步行友好性及潜在交通风险^[50],从而为城市道路设计和行人安全规划提供决策支持。

自监督学习 (SSL) 通过大规模街景图像数据学习城市环境的通用表征(如绿化覆盖、建筑密度),在城市规划的多种任务中具有潜力,但其局限性不容忽视。首先,SSL生成的通用表征虽在多种可视任务中表现突出,却难以胜任城市规划中需人类决策支持的下游任务,例如城市交通模式的判别、社会公平的评估、土地利用的优化及环境可持续性的分析。其次,SSL对“异常”数据的处理能力不足,例如极端天气(如暴雪、洪水)或突发事件(如交通事故、施工),尽管这些情况在街景图像中数量较少,但对灾害监测或异常交通分析等特定任务至关重要。而SSL学习的是大量数据统计规律,容易稀释此类图像信息,难以凸显其价值。最后,SSL难以捕捉街景图像中因地理位置差异导致的空间异质性(如北京与深圳街区的气温差异),且缺乏对空间关系的推理能力。以LLMs为例,其虽擅长语言理解和逻辑推理,却难以学习空间拓扑结构,理解城市空间关系,限制了其在城市规划决策中的应用^[51]。未来研究中可以考虑将城市规划领域知识与SSL融合,通过知识增强检索或对特定下游任务微调的方式,增强SSL在城市规划分析方面的适用性。

3 总结与展望

街道是城市的基本组成单元,承载着交通、生活、经济和文化等多重功能,而街景数据作为城市视觉分析的重要数据源,能够直观反映城市环境的物理特征和社会功能。随着人工智能技术的快速发展,街景数据的分析方法经历了从传统手工审计到深度学习驱动的自动化分析的深刻变革。早期研究依赖人工实地调查和摄影,局限于小范围定性分析。Google街景的推出使大范围、标准化分析成为可能,但人工标注耗时且模型能力有限。统计机器学习阶段实现了部分自动化,但仍难以全面捕捉街景的复杂语义信息。表示学习通过自动化、精细化的特征提取,有效捕捉街景图像中的复杂语义信息,显著提

升了其在城市规划领域的应用潜力。然而,这一阶段仍高度依赖大量高质量的人工标注数据。近年来,自监督学习的兴起弥补了这一短板,它利用数据自身结构生成监督信号,实现了在无需标注的情况下获取图像的通用表征。与此同时,大语言模型凭借其通用的“世界观”,能够对图像信息进行更深层次的理解与推理,为多种城市规划领域下游任务提供了更加智能、科学且通用的支持工具。展望未来,街道视觉智能在城市规划领域的突破可以围绕数据、方法和应用3个层面展开。

(1) 数据层面:数据补全与生成及多模态数据融合

数据补全与生成:通过提升图像清晰度、修复过曝图像、补全缺失图像等技术,提升现有数据质量,为模型训练提供高质量元数据;利用生成式技术生成夜间街景图像或自定义编辑街景元素,解决特定领域的数据缺失问题;通过图生视频技术模拟历史街区演变或未来场景变化,为城市规划和设计提供动态可视化支持。

多模态数据融合:将街景图像与遥感图像、POI数据、传感器数据、建筑足迹等多源数据结合,构建城市建成环境的综合表征,为区域功能分析、环境质量评估等下游任务提供更丰富的知识支持。

(2) 方法层面:多模型优势互补与空间关系建模

多模型优势互补:结合深度学习、自监督学习和大语言模型的优势,例如利用对比学习挖掘街景数据的时空特征,或借助大语言模型的通用世界观和推理能力优化区域表征效果,提升街景数据分析的精度和效率。

空间关系建模:为强化SSL的空间理解能力,可采用两种策略:一是空间关系学习,将地理拓扑知识(如道路、街景点、建筑物的九交模型)融入SSL预训练,构建街景间的空间联系,生成空间感知表征;二是空间关系推理,通过知识增强检索从外部地理数据库提取关键空间信息,优化SSL的空间表征

效果。

(3) 应用层面:LLMs驱动的新范式

将LLMs的语义理解与推理能力应用于城市规划领域的多种下游任务。对于可视类任务,LLMs可以实现街景图像到具体任务的端到端智能分析。例如,自动识别建筑功能类型(如住宅、商业)、量化街道美学价值、预测居民对街道环境的感知评分,评估城市基础设施的更新需求等。对于非可视类任务,将LLMs的通用“世界观”和城市规划领域专家知识结合,可在复杂情境下提供决策帮助。例如,为土地使用分区优化、社会公平性提升以及空间隔离现象的缓解等具体议题提供参考建议,助力规划者制定更具针对性与前瞻性的策略。

总之,随着技术的不断进步,街景数据的分析方法将更加智能化、多元化和精细化。通过数据补全与生成、多模态数据融合、多重模型互补,以及LLMs驱动的新范式,街景数据将在新的街道视觉分析范式下为城市规划和城市管理提供更加科学和精准的结果,推动城市的可持续发展。

参考文献 References

[1] 张帆,刘瑜. 街景影像——基于人工智能的方法与应用[J]. 遥感学报, 2021, 25 (5): 1043-1054.
ZHANG Fan, LIU Yu. Street view images: methods and applications based on artificial intelligence[J]. Journal of Remote Sensing, 2021, 25(5): 1043-1054.

[2] 姜鹏,倪砦,郝望. 面向未来的DAD与智慧城市[J]. 上海城市规划, 2016 (3): 52-55.
JIANG Peng, NI Tong, XI Wang. DAD and smart cities for the future[J]. Shanghai Urban Planning Review, 2016(3): 52-55.

[3] 金盛,郭文彤,江杨,等. 街景影像在城市交通研究中的应用:回顾、分析和展望[J]. 交通运输工程与信息学报, 2024, 22 (2): 191-209.

JIN Sheng, GUO Wentong, JIANG Yang, et al. Applications of street view images in urban transportation research: review, analysis, and prospect[J]. Journal of Transportation Engineering and Information, 2024, 22(2): 191-209.

[4] 陈虹,汤军,龚阳春,等. 基于街景与高分遥感影像的超大城市绿地高精度识别与空间特征解析[J]. 地球信息科学学报, 2024, 26 (12): 2818-2830.
CHEN Hong, TANG Jun, GONG Yangchun, et al. High-precision identification and spatial characterization of green spaces in megacities based on street view and high-resolution remote sensing images[J]. Journal of Geo-Information Science, 2024, 26(12): 2818-2830.

[5] 李心雨,闫浩文,王卓,等. 街景图像与机器学习相结合的道路环境安全感知评价与影响因素分析[J]. 地球信息科学学报, 2023, 25 (4): 852-865.
LI Xinyu, YAN Haowen, WANG Zhuo, et al. Evaluation of road environmental safety perception and influencing factors based on the integration of street view images and machine learning[J]. Journal of Geo-Information Science, 2023, 25(4): 852-865.

[6] 冯叶涵,陈亮,贺晓冬. 基于百度街景的SVF计算及其在城市热岛研究中的应用[J]. 地球信息科学学报, 2021, 23 (11): 1998-2012.
FENG Yehan, CHEN Liang, HE Xiaodong. SVF calculation based on Baidu Street View and its application in urban heat island research[J]. Journal of Geo-Information Science, 2021, 23(11): 1998-2012.

[7] 周垠,龙瀛. 街道步行指数的大规模评价——方法改进及其成都应用[J]. 上海城市规划, 2017 (1): 88-93.
ZHOU Yin, LONG Ying. Large-scale evaluation of walkability index: method improvement and its application in Chengdu[J]. Shanghai Urban Planning Review, 2017(1): 88-93.

[8] 张昊,尹力. 大数据在评价有关公共健康的建成环境中的应用:文献综述[J]. 上海城市规划, 2020 (5): 36-40.
ZHANG Hao, YIN Li. Application of big data in evaluating built environment related to public health: a literature review[J]. Shanghai Urban Planning Review, 2020(5): 36-40.

[9] WHYTE W H. The social life of small urban spaces[M]. Washington, DC: Conservation Foundation, 1980.

[10] JACOBS J. Death and life of great American cities[M]. New York: Random House, 1961.

[11] LYNCH K. The image of the city[M]. Cambridge, MA: MIT Press, 1960.

[12] RUNDLE A G, BADER M D M, RICHARDS C A, et al. Using Google Street View to audit neighborhood environments[J]. American Journal of Preventive Medicine, 2011, 40(1): 94-100.

[13] BADLAND H M, OPIT S, WITTEN K, et al. Can virtual streetscape audits reliably replace physical streetscape audits?[J]. Journal of Urban Health, 2010, 87(6): 1007-1016.

[14] NAIK N, PHILIPOOM J, RASKAR R, et al. Streetscore - predicting the perceived safety of one million streetscapes[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Columbus, OH, USA: IEEE, 2014: 793-799.

[15] LI X, ZHANG C, LI W, et al. Assessing street-level urban greenery using Google Street View and a modified green view index[J]. Urban Forestry & Urban Greening, 2015, 14(3): 675-685.

[16] LI X, ZHANG C, LI W. Building block level urban land-use information retrieval based on Google Street View images[J]. GIScience & Remote Sensing, 2017, 54(6): 819-835.

[17] MURILLO A C, SINGH G, KOSECKÁ J, et al. Localization in urban environments using a panoramic gist descriptor[J]. IEEE Transactions on Robotics, 2013, 29(1): 146-160.

[18] CAMPBELL A, BOTH A, SUN Q. Detecting and mapping traffic signs from Google Street View images using deep learning and GIS[J]. Computers, Environment and Urban Systems, 2019, 77: 101350.

[19] DAI Y, LIU L, WANG K, et al. Using computer vision and street view images to assess bus stop amenities[J]. Computers, Environment and Urban Systems, 2025, 117: 102254.

[20] PENG X, SONG R, CAO Q, et al. Real-time illegal parking detection algorithm in urban environments[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(11): 20572-20587.

[21] FAN Z, ZHANG F, LOO B P Y, et al. Urban visual intelligence: uncovering hidden city profiles with street view images[J]. Proceedings of the National Academy of Sciences, 2023, 120(27): e2220417120.

[22] ZHANG F, ZHANG D, LIU Y, et al. Representing place locales using scene elements[J]. Computers, Environment and Urban Systems, 2018, 71: 153-164.

[23] 李海薇,陈崇贤,刘欣宜,等. 人视街景图像和机器学习结合的城市街道适老性水平空间效应研究[J]. 地球信息科学学报, 2024, 26 (6): 1469-1485.
LI Haiwei, CHEN Chongxian, LIU Xinyi, et al. Exploring the spatial effects of urban street age-friendliness using human-perspective street view images and machine learning[J]. Journal of Geo-Information Science, 2024, 26(6): 1469-1485.

[24] ZHANG F, ZHOU B, LIU L, et al. Measuring human perceptions of a large-scale urban region using machine learning[J]. Landscape and Urban

- Planning, 2018, 180: 148-160.
- [25] HUANG J, FEI T, KANG Y, et al. Estimating urban noise along road network from street view imagery[J]. *International Journal of Geographical Information Science*, 2024, 38(1): 128-155.
- [26] ZHANG F, WU L, ZHU D, et al. Social sensing from street-level imagery: a case study in learning spatio-temporal urban mobility patterns[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2019, 153: 48-58.
- [27] KANG J, KÖRNER M, WANG Y, et al. Building instance classification using street view images[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018, 145: 44-59.
- [28] YAO Y, DONG A, LIU Z, et al. Extracting the pickpocketing information implied in the built environment by treating it as the anomalies[J]. *Cities*, 2023, 143: 104575.
- [29] 司睿, 林姚宇, 肖作鹏, 等. 基于街景数据的建成环境与街道活力时空分析——以深圳福田区为例[J]. *地理科学*, 2021, 41 (9): 1536-1545.
- SI Rui, LIN Yaoyu, XIAO Zuopeng, et al. Spatiotemporal analysis of the built environment and street vitality based on street view data: a case study of Futian District, Shenzhen[J]. *Scientia Geographica Sinica*, 2021, 41(9): 1536-1545.
- [30] GUI J, CHEN T, ZHANG J, et al. A survey on self-supervised learning: algorithms, applications, and future trends[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, 46(12): 9052-9071.
- [31] WANG Z, LI H, RAJAGOPAL R. Urban2Vec: incorporating street view imagery and POIs for multi-modal urban neighborhood embedding[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(1): 1013-1020.
- [32] LI Y, HUANG Y, MAI G, et al. Learning street view representations with spatiotemporal contrast[J]. *arXiv*, 2025: 2502.04638.
- [33] LI H, DEUSER F, YIN W, et al. Cross-view geolocalization and disaster mapping with street-view and VHR satellite imagery: a case study of Hurricane IAN[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2025, 220: 841-854.
- [34] LI Y, HUANG W, CONG G, et al. Urban region representation learning with OpenStreetMap building footprints[C]//*Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery, 2023: 1363-1373.
- [35] WANG J, HUANG W, BILJECKI F. Learning visual features from figure-ground maps for urban morphology discovery[J]. *Computers, Environment and Urban Systems*, 2024, 109: 102076.
- [36] QI M, HANKEY S. Using street view imagery to predict street-level particulate air pollution[J]. *Environmental Science & Technology*, 2021, 55(4): 2695-2704.
- [37] SWERDLOW A, XU R, ZHOU B. Street-view image generation from a bird's-eye view layout[J]. *IEEE Robotics and Automation Letters*, 2024, 9(4): 3578-3585.
- [38] PANG H E, BILJECKI F. 3D building reconstruction from single street view images using deep learning[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2022, 112: 102859.
- [39] LIU Z, LI T, REN T, et al. Day-to-night street view image generation for 24-hour urban scene auditing using generative AI[J]. *Journal of Imaging*, 2024, 10(5): 112.
- [40] HOU C, ZHANG F, LI Y, et al. Urban sensing in the era of large language models[J]. *The Innovation*, 2025, 6(1): 100749.
- [41] 周静, 邝远霄, 刘勇, 等. 基于大语言模型的城市街区“人—地”情感关联测度及启示——以上海多伦路街区为例[J]. *上海城市规划*, 2024 (3): 102-108.
- ZHOU Jing, KUANG Yuanxiao, LIU Yong, et al. Measuring human-place emotional connections in urban blocks based on large language models: a case study of Duolun Road Block in Shanghai[J]. *Shanghai Urban Planning Review*, 2024(3): 102-108.
- [42] JANG K M, KIM J. Multimodal large language models as built environment auditing tools[J]. *The Professional Geographer*, 2025, 77(1): 84-90.
- [43] 黄铎, 萧蕾. 大语言模型赋能城市规划的应用分析[J]. *房地产世界*, 2024 (23): 1-4.
- HUANG Duo, XIAO Lei. Application analysis of large language models empowering urban planning[J]. *Real Estate World*, 2024(23): 1-4.
- [44] HUANG W, WANG J, CONG G. Zero-shot urban function inference with street view images through prompting a pretrained vision-language model[J]. *International Journal of Geographical Information Science*, 2024, 38(7): 1414-1442.
- [45] WU M, HUANG Q, GAO S, et al. Mixed land use measurement and mapping with street view images and spatial context-aware prompts via zero-shot multimodal learning[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2023, 125: 103591.
- [46] CHEN M, LI Z, HUANG W, et al. Profiling urban streets: a semi-supervised prediction model based on street view imagery and spatial topology[C]//*Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery, 2024: 319-328.
- [47] XU S, ZHANG C, FAN L, et al. AddressCLIP: empowering vision-language models for city-wide image address localization[C]//LEONARDIS A, RICCI E, ROTH S, et al. *Computer Vision – ECCV 2024*. Cham: Springer Nature Switzerland, 2025: 76-92.
- [48] LIANG H, ZHANG J, LI Y, et al. Automatic estimation for visual quality changes of street space via street-view images and multimodal large language models[J]. *IEEE Access*, 2024, 12: 87713-87727.
- [49] ZHOU Z, WANG Q, LIN B, et al. UNIAA: a unified multi-modal image aesthetic assessment baseline and benchmark[J]. *arXiv*, 2024: 2404.09619.
- [50] BLEČIĆ I, SAIU V, TRUNFIO GIUSEPPE A. Enhancing urban walkability assessment with multimodal large language models[C]//GERVASI O, MURGANTE B, GARAU C, et al. *Computational Science and Its Applications – ICCSA 2024 Workshops*. Cham: Springer Nature Switzerland, 2024: 394-411.
- [51] YU D, BAO R, MAI G, et al. Spatial-RAG: spatial retrieval augmented generation for real-world spatial reasoning questions[J]. *arXiv*, 2025: 2502.18470.