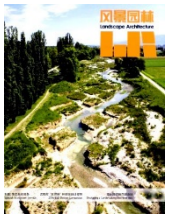


风景园林
Landscape Architecture
ISSN 1673-1530,CN 11-5366/S

《风景园林》网络首发论文

题目：基于多模态大模型的城市空间感知评估框架构建与实证研究
作者：王磊，郭家乐，白钊成，何捷
收稿日期：2025-11-08
网络首发日期：2026-03-20
引用格式：王磊，郭家乐，白钊成，何捷．基于多模态大模型的城市空间感知评估框架构建与实证研究[J/OL]．风景园林．
https://link.cnki.net/urlid/11.5366.S.20260319.1732.003

中图分类号：TU984
文献标识码：A
文章编号：1673-1530(2025)00-0000-00
DOI：10.3724/j.fjyl.LA20250702
收稿日期：2025-11-08

基于多模态大模型的城市空间感知评估框架构建与实证研究

Construction and Empirical Research on an Evaluation Framework for Urban Spatial Perception Based on Large Multimodal Models

王磊 郭家乐 白钊成 何捷*
WANG Lei, GUO Jiale, BAI Zhaocheng, HE Jie*

摘　要：【目的】在“以人为本”和“人工智能”的时代背景下，城市空间感知是精细化评估建成环境具体体现。针对传统空间感知评价方法在客观性和效率上的局限，本文旨在探索一种基于多模态大模型的城市空间感知评价框架的新范式。【方法】首先，融合可持续发展三重底线理论与风景园林学科特点，构建了包含“景观、环境、经济、社会”四个维度的感知评价框架。其次，以北京市五环内区域为案例地采集街景图像，构建系统化的提示词工程，使用通义千问多模态大模型推理生成结构化的文本描述。通过训练的BERT+LSTM模型将描述文本量化为感知分数。【结果】在北京市五环区域范围内开展实证分析，将基于多模态大模型与基于MIT Place Pulse 2.0数据集的传统感知评价方法进行外部一致性验证。通过皮尔逊相关系数r=0.61,P<0.001等分析可以证明，本研究所提出的感知评价方法灵活且有效。【结论】本研究所提出方法为大模型在风景园林学科领域应用提供了标准范式和思路，为城市景观评估、建成环境后评估等工作提供了智能化工具和思路。

关 键 词：多模态大模型；城市空间感知；街景图像；提示词工程；可持续发展三重底线；北京

基金项目：国家自然科学基金“地理空间人工智能与知识图谱支持的尼泊尔上木斯塘南亚丝绸之路文化景观遗产识别与价值评估”（编号 52478049）

文章亮点：

1. 研究融合可持续发展理论，提出基于多模态大模型的感知评估新范式。有效突破了传统方法成本高、样本小的局限。

2. 利用提示词工程引导大模型解析海量街景图像。该方法无需大规模人工标注，即可精细化地量化复杂的城市空间感知。

3. 北京五环内的实证结果与国际传统基准方法高度一致。这为城市规划与景观设计提供了科学、智能化的分析工具。

Abstract: [Objective] As the urbanization rate reached 66.16% by the end of 2023, enhancing urban spatial quality and improving human settlements have become core agendas for high-quality urban development. Urban spatial perception serves as a crucial bridge connecting people with the built environment, which is vital for human-centric planning. However, conventional spatial perception evaluation methods, such as questionnaires and expert scoring, suffer from small sample sizes, high financial costs, and extended research cycles, rendering them inadequate for large-scale, fine-grained urban studies. Furthermore, existing methods relying on traditional deep learning and computer vision require expensive, large-scale manual annotation for supervised training. These traditional models also struggle

to interpret higher-level, complex semantic imagery like "sense of place" or "cultural atmosphere" and lack interpretability due to their inherent "black-box" decision-making nature. The recent emergence of Large Multimodal Models (LMMs), such as Qwen2-VL, offers a breakthrough opportunity, possessing powerful cross-modal understanding and zero-shot reasoning capabilities derived from massive pre-training. Therefore, this study aims to develop a novel, comprehensive evaluation framework based on LMMs and the Triple Bottom Line (TBL) theory. By doing so, it explores an annotation-light, interpretable, reproducible, and highly scalable new paradigm for fine-grained urban perception evaluation.**[Methods]** The empirical study was conducted within the highly heterogeneous urban environment of Beijing's Fifth Ring Road, an area of approximately 667 square kilometers. To construct the database, 157,731 sampling points were generated at 50-meter intervals along the OpenStreetMap road network, and multi-directional Baidu Street View images were collected. To rigorously control for systematic biases caused by seasonal defoliation, imagery was strictly filtered for the months of May to October spanning from 2013 to 2023, ultimately yielding a clean dataset of 122,264 valid sampling points. The evaluation framework was driven by the TBL theory and expanded through a landscape-architecture lens into four primary dimensions: landscape, environment, economy, and society. This framework was further distilled into 6 secondary indicators and 30 tertiary indicators, encompassing both observable physical elements and abstract subjective perceptions. From a technical standpoint, the methodology centered on a systematic prompt engineering pipeline—integrating six structural elements including background, goal, style, tone, audience, and output—to guide the Qwen2-VL-72B model, which was deployed in a distributed manner across four A100 GPUs. The model generated structured JSON semantic descriptions, which were subsequently mapped into quantitative perception scores using a BERT+LSTM sentiment classification model fine-tuned on the ChnSentiCorp and weibo_senti_100k datasets. To robustly validate the external consistency of this novel LMM-based paradigm, the results were benchmarked against a traditional Support Vector Machine (SVM) approach trained on the internationally recognized MIT Place Pulse 2.0 crowdsourced dataset.**[Results]** The large-scale batch inference processed the 122,264 street-view samples in approximately 306 hours, successfully generating 430 MB of structured semantic JSON data across 30 dimensions. The LMM demonstrated exceptional proficiency in identifying explicit visual elements while accurately conducting higher-order reasoning; for instance, it successfully inferred "economic vitality" from commercial signs and identified a "lack of resting spaces" based on the absence of public facilities. Geospatially, when the perception scores were aggregated to the Transportation Analysis Zone (TAZ) level, the distribution exhibited a distinct concentric pattern characterized by higher scores in the urban core and lower scores toward the periphery, perfectly aligning with Beijing's historical radial development. High-value areas (scores > 0.90) were highly concentrated within the core 2nd and 3rd Rings, such as the Forbidden City and the CBD; medium-value areas (0.85–0.90) were widely distributed between the 3rd and 4th Rings; and low-value areas (< 0.85) were predominantly found outside the 4th Ring, particularly in the southern and southwestern peripheral regions. When benchmarked against the MIT Place Pulse 2.0 dataset, both approaches displayed highly consistent macro-level spatial patterns. However, systematic numerical differences emerged: the LMM approach yielded a higher mean score of 0.849 and a lower standard deviation of 0.079, compared to the traditional method's mean of 0.537 and standard deviation of 0.154, likely reflecting the LMM's standardized modern aesthetic bias versus the diverse, culturally varied perspectives of global crowdsourced volunteers. Consistency tests strongly confirmed the framework's reliability, showing a moderate-to-strong agreement with a Pearson correlation coefficient of 0.6095 and a Spearman rank correlation of 0.6510 ($P < 0.001$). Comprehensive residual diagnostics further indicated randomness, homoscedasticity, and approximate normality, providing robust statistical validation for the proposed methodology.**[Conclusion]** This study successfully constructs and rigorously validates an innovative, LMM-based framework for evaluating perceived street-level spatial quality in complex urban environments. By seamlessly integrating a theory-driven evaluation indicator system with meticulously iteratively refined prompt engineering, the proposed methodology achieves effective, reliable, and highly granular quantification of urban spatial perception without the debilitating cost of large-scale manual labeling. The outputs demonstrated remarkable macro-spatial alignment with an internationally established traditional

methodology, highlighting the immense practical value of AI-driven tools in landscape architecture for evidence-based fine-grained environmental diagnosis, predictive design scheme simulation, and collaborative public participation. Despite its strong potential, the study thoughtfully notes inherent limitations, including the risk of cultural biases or aesthetic homogenization introduced by the model's training corpora, as well as the inability of static street-view imagery to capture dynamic temporal rhythms like day-night variations or holiday crowd activities. Looking forward, future research should focus on adopting explainable AI techniques (e.g., Class Activation Mapping) to demystify the model's decision-making process, fusing dynamic multi-source spatial data, and fine-tuning domain-specific foundational models with localized corpora to better resonate with Chinese urban contexts and landscape aesthetics.
Keywords: large multimodal models; urban spatial perception; street view images; prompt engineering; triple bottom line; Beijing

随着我国城镇化率在2023年底达到66.16%[1]，城市空间品质的提升和人居环境的改善已成为城市高质量发展的核心议题。城市空间感知，即人对城市物质空间环境的主观感受和认知评价[2]，是连接“人”与“建成环境”的桥梁，对实现“以人为本”的规划设计至关重要。然而，传统的空间感知研究方法，如问卷访谈、专家打分等，虽能直接获取主观反馈，但存在样本量小、成本高、周期长等局限，难以支撑大尺度、精细化的城市研究[3]。

近年来，以街景图像为代表的城市大数据[4]和以深度学习为核心的人工智能技术，为城市空间感知的定量化研究带来了机遇。研究者们利用计算机视觉技术从街景图像中自动提取绿视率、天空开敞度等客观环境[5]，并通过机器学习模型预测公众的安全感、美学评价等主观感知[6]，极大地提升了研究的效率和空间分辨率。其中，麻省理工学院的MIT Place Pulse 2.0项目通过大规模在线众包平台，开创性地构建了全球多城市的街景感知数据库，成为该领域的标杆[7]。

城市空间感知相关研究尽管取得了显著进展[8–10]，但现有基于传统深度学习模型的方法仍存在不足：一是模型通常需要大量人工标注数据进行监督训练，成本高昂[11]；二是在处理如“场所感”、“文化氛围”等更高级、更综合的“意象”感知时能力有限[12]；三是模型的“黑箱”特性使其决策过程缺乏可解释性[13]。幸运的是，以GPT-4V[14]、通义千问[15]为代表的新一代多模态大模型的出现为此带来了突破性契机。这类模型凭借其在海量图文数据上的预训练，获得了强大的跨模态理解和零样本推理能力，能够像人类一样直接解读图像并输出富有逻辑的文本描述，展现出解决复杂城市感知任务的巨大潜力[16]。

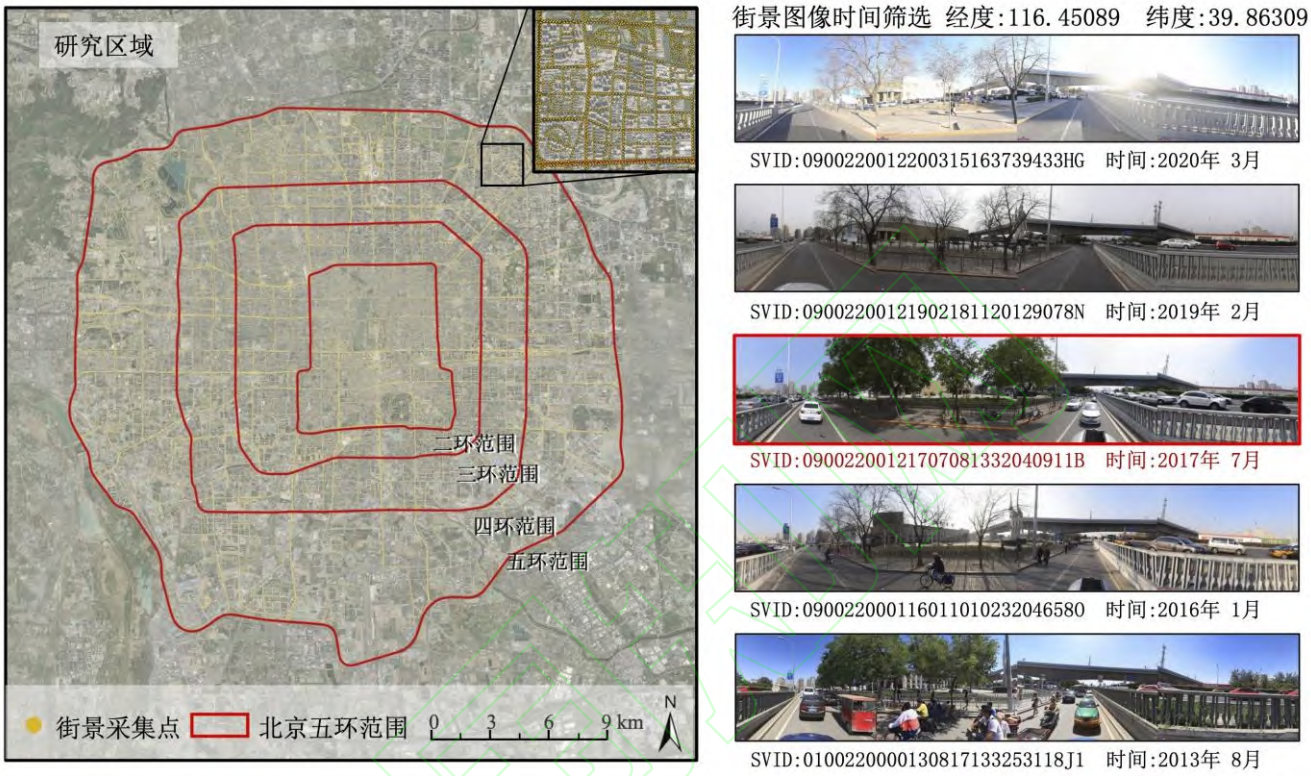
基于此，本研究旨在探索并构建一套基于多模态大模型的城市空间感知评价新框架。该框架以可持续发展三重底线理论为指导，通过系统化的提示词工程，引导大模型对北京市五环内的海量街景图像进行多维度、精细化的感知评价，并与传统方法进行外部一致性验证。与传统的计算机视觉或深度学习感知模型不同，本研究创新性地利用预训练的多模态大模型进行城市感知评价，使模型在无需大规模标注的条件下直接从街景图像中提取语义信息并生成连贯的文本描述。通过系统化的提示词设计，我们的框架引导大模型对街景图像进行结构化描述，实现了从“图像-视觉”到“文字-语义”的自动映射过程。这一方法自动化且高度灵活，不仅将风景园林领域的专业知识嵌入到生成中，还大幅提升了结果的可解释性和可复现性。此外，将城市空间“翻译”为文本描述的过程，有效避免了不同城市建筑风格或主观标注带来的偏见，使得评价结果更加客观透明。本研究通过引入多模态提示词工程和指导理论框架，显著拓展了城市感知的评价维度、提升了模型的可解释性，并为感知评价提供了一种可复制、可扩展的新范式。

1 研究区域与数据

1.1 研究区域

本研究选取北京市五环路以内区域作为实证分析区域（图1）。该区域面积约667平方公里，是中国的政

治、文化和经济中心，涵盖了从历史悠久的胡同街巷到现代化的中央商务区CBD等高度异质性的城市建成环境。空间特征从二环内的核心历史城区到五环的新兴居住区呈现出明显的圈层式梯度变化，为检验和展示空间感知方法的多场景适应性提供了理想的实验场所。



1 研究区域及街景图像时间筛选

Study area and temporal filtering of street view images

1.2 数据来源与处理

（1）街景图像数据

基于 OpenStreetMap 提供的北京市五环内道路网络，沿路网每隔 50 米设置一个采样点，共计 157,731 个。通过百度地图获取了每个采样点四个方向的人眼视角图像（图 1 右）。考虑到冬季植被落叶会显著改变街景中的绿量与视觉要素，从而对城市感知结果产生系统性偏差[17]，本研究将历史街景影像作为候选数据源，通过季节一致性约束提高感知测度的可比性。具体而言，街景时间筛选限定为 5-10 月；在满足月份约束的前提下，采用“最新优先、逐年回溯匹配”的原则选取影像。优先选择年份最新的街景，若该年份缺少 5-10 月影像，则向上一年依次匹配直至获得满足条件的街景。但每个采样点仅保留一组满足规则的街景图像用于后续分析。最终获得 122,264 个有效采样点的街景图像数据。

（2）感知基准对照数据集

采用国际公认的 MIT Place Pulse 2.0 数据集作为基准对照[18]。该数据集包含全球 56 个城市超过 10 万张街景图像以及由在线志愿者通过成对比较产生的六个维度(安全、美丽、活力、富裕、无聊、压抑)的约 110 万次感知评分。采用该数据和机器学习算法，以北京为实例计算出一套可供基准对照的传统感知分数[19]。

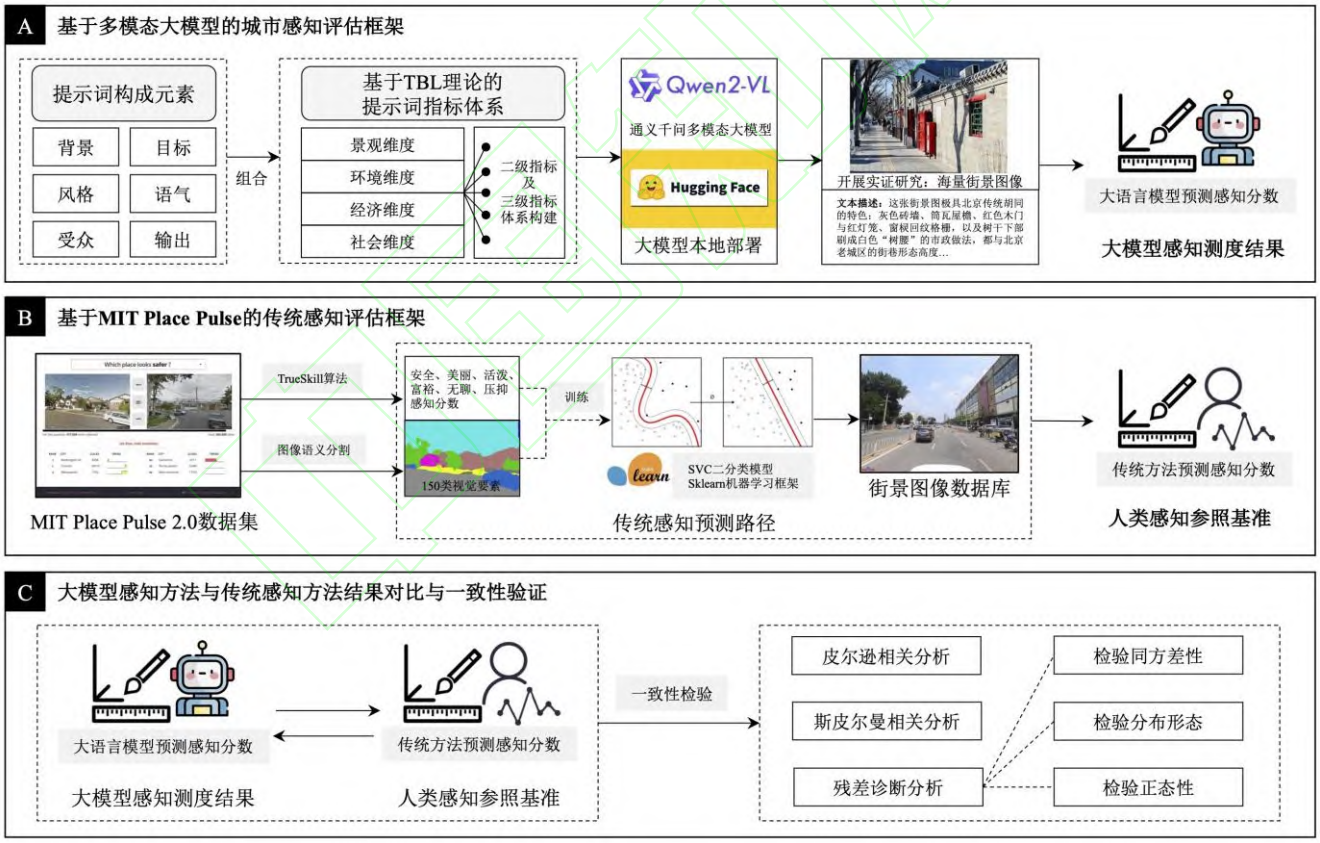
（3）多模态大模型

本研究选用了阿里巴巴开源的通义千问作为核心推理模型[20]。研究通过 Hugging Face 平台的开源工具链[21]，租用高性能 GPU 的云服务器环境计算资源，共计 4 块 A100GPU 分布式部署大模型。实现了 Qwen2-VL-

72B 大语言模型的配置与部署。数据处理与推理方式中，将街景图像通过预处理管道转换为模型可接受的输入格式。利用模型的多模态能力生成描述性文字，并进一步开展感知评价。

2 研究方法

本研究构建了一套以多模态大模型为核心的城市空间感知测度框架（图 2），该框架整合了两种技术路径并进行外部一致性验证。框架的 A 部分为本研究提出的新范式：首先，通过系统化的提示词工程向部署的 Qwen2-VL 多模态大模型发出提问；模型对输入的街景图像进行推理，生成结构化的描述性文本；随后，该文本被送入预先训练好的 BERT+LSTM 自然语言处理模型进行感知评价，最终计算出量化的感知分数。框架的中间部分为数据预处理环节，即通过图像语义分割神经网络，将原始街景图像解析为天空、道路、建筑、植物等基本视觉要素。框架的 B 部分为用于外部一致性验证的传统范式：以国际公认的 MIT Place Pulse 2.0 数据集为基础，训练一个支持向量机分类模型，该模型基于街景的视觉要素来进行感知分数的预测。框架的 C 部分展示了两种范式的最终对比：通过相关性分析与残差诊断分析，系统性地证明了本研究提出的新范式与传统范式结果的一致与差异性。



2 多模态大模型城市空间感知测度分析框架

A measurement and analysis framework for urban spatial perception based on large multimodal models

2.1 理论框架与指标体系构建

为确保感知评价的全面性和科学性，本研究引入可持续发展三重底线理论（Triple Bottom Line, TBL）[22]作为顶层设计框架。该理论强调应从经济(Profit)、社会(People)和环境(Planet)三个维度综合评估发展绩效。结合风景园林学科特点，本研究将其拓展为“景观、环境、经济、社会”四维评价框架。在此框架下，进一

步设计了包含6个二级指标和30个三级指标的精细化评价体系（表1）。该体系不仅涵盖了绿化覆盖率、空间开阔度等传统客观指标[23,24]，更包含了美感评价、社区归属感、文化符号等难以通过传统计算机视觉直接量化的主观感知维度。

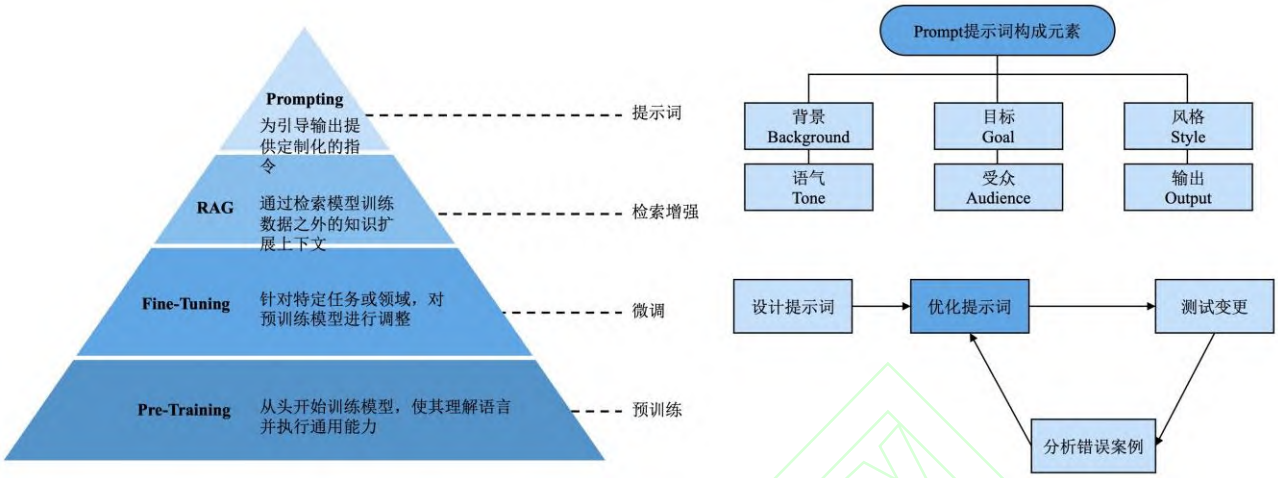
表 1 基于 TBL 理论的城市空间感知评价指标体系

Tab. 1 An urban spatial perception evaluation index system based on the triple bottom line (TBL) theory

一级指标	二级指标	三级指标	一级指标	二级指标	三级指标
景观维度	景观整体质量与视觉吸引力	景观特色度	环境维度	生态环境质量与居民心理健康	污染可见度
		空间开阔度			噪音干扰度
		空间整洁度			热岛效应迹象
		视觉多样性			海绵城市元素
		美感评价			负面情绪度
	绿色基础设施与自然环境要素	绿化覆盖率		交通系统与宜居性	交通可达性
		植被健康度			交通秩序度
		水体可见度			社会与空间公平
		生态廊道联通性			休憩与交往空间
		自然舒适感			无障碍设施
经济维度	经济活力及商业氛围	商业多样性	社会维度	社会文化及场所感	文化符号
		街道繁荣度			历史文脉
		摊贩活动			公共艺术
		商业环境			社区归属感
		经济活力			多元文化融合度

2.2 提示词工程设计与模型选择

提示词工程是将理论框架转化为机器可执行任务的关键环节。如图3左金字塔模型所示，与需要海量计算资源的预训练和微调相比，提示词是一种轻量化、高效率的应用策略，它能充分利用预训练模型的现有能力而无需修改其参数[25]。因此，本研究选择提示词工程作为核心技术路径。研究采用了一套包含“背景、目标、风格、语气、受众、输出”六要素的提示词结构化框架（图3右上），确保指令的全面性与精确性。为达到最优效果，提示词的构建遵循了“设计、优化、测试、分析”的闭环迭代流程（图3右下），通过不断分析错误案例并优化指令，最终得到能够引导模型输出专业、准确且结构化的感知评价文本①。



3 多模态大模型提示词应用框架与构成元素

Application framework and constituent elements of prompts for multimodal large models

表2详细展示了这六个构成元素的具体示例，最终形成的提示词不仅为模型设定了“城市科学研究专家”的角色、明确输出用于感知评价的学术化文本的目标、并规范了输出的专业风格、严谨语气和目标受众，要求其针对30个三级指标逐一进行深入分析，并以中文JSON格式输出，极大地提升了输出结果的可用性和规范性。

表 2 多模态大模型提示词构建示例

Tab. 2 Examples of prompt construction for multimodal large models

提示词构成元素	提示词构建示例
背景（Background）	你是一位城市科学研究专家，正在使用多模态大模型来研究从街景图像中提取对城市空间的感知。
目标（Goal）	依据输入的街景图像，输出可供后续感知评价使用的学术化描述文本，包括对街景特征的深度描述。
风格(Style)	采用学术且专业的表达方式，引用或融合城市规划和社会学中的关键术语，确保对城市特征的描述客观而详实。
语气(Tone)	请使用学术且专业的表达方式，采用正式、严谨的语气，避免夸张或情绪化字眼，注重逻辑与条理性。
受众(Audience)	面向景观规划、地理信息系统和社会学等领域的科研学者，以及从事感知评价与数据建模的工程师。
输出(Output)	每个维度的描述都应该是独立且完整的分析。所有输出必须完全使用中文，包括 JSON 字段名称，不要出现任何英文。 输出格式示例： {"视觉多样性": "详细的描述...", "空间开阔度": "详细的描述...", "空间整洁度": "详细的描述..."}

2.3 感知分数量化与基准对照

（1）模型推理

将122,264张街景图像及最终版提示词输入部署好的大模型中进行批量推理。整个过程耗时约306小时，生成了一个包含所有街景点30个维度文本描述的、大小为430MB的JSON数据集。在多模态大模型完成推理后，对生成的数据进行系统化清理与验证，以确保数据的完整性、准确性和可用性。由于模型推理生成的描述数据需与实际存储的图像数据一一对应，因此首先进行数量匹配检查。通过比对文件系统中的街景图片与JSON记录中的数据条目，识别是否存在缺失或多余的数据，即街景图像未能生成对应的JSON描述。确认某字段是否为空或缺少有效描述，如“未知”或“无数据”。确认部分字段是否存在过短的描述，如少于10个字符。对于上述检查发现的异常数据，统一采取剔除处理，并记录剔除的街景图像序号。随后再次执行多模态大模型的推理，重新处理剔除的错误数据。通过以上数据清理流程，确保推理后的数据既符合研究标准，又具备较高的可靠性和一致性。

（2）感知分数计算

为将多模态大模型输出的 30 个感知维度描述文本转化为可用于空间分析的定量指标，本研究采用 BERT[26]结合 LSTM[27]的感知评价模型实现“文本-分数”映射。首先，该模型先在通用中文情感数据集 ChnSentiCorp[28]与社交媒体语料 weibo_senti_100k[29]上进行训练与微调，以兼顾较正式的空间描述语体与口语化体验表达，从而降低语域差异可能带来的系统性偏差。在推理阶段，对大模型生成的 JSON 文本进行系统化清理与一致性校验，如数量匹配、字段缺失、无效值、过短文本等，对异常记录统一剔除并记录编号后执行二次推理补全，以保证样本完整性与数据质量。随后，将每个维度的描述文本输入感知评价模型，得到属于“积极感知”的概率值 $p \in [0,1]$，并将该概率定义为对应维度的感知分数，分数越高表示感知越积极。最后，针对每一张街景图像，大模型分别生成 30 条针对不同三级指标的独立描述文本，感知评价模型对文本逐一推理，获得 30 个独立维度的感知分数。基于 TBL 理论框架的整体性原则各感知维度对城市空间品质的贡献具有同等重要性，将 30 个维度的感知分数进行等权平均，从而获得该点的综合感知得分。

（3）外部一致性验证

为验证本方法的有效性，将多模态大模型计算出的综合感知分数与基于MIT Place Pulse 2.0数据集的传统方法计算出的分数进行对比。值得注意的是 MIT Place Pulse 2.0不被视为与本研究30指标体系一一对应的“同构量表”，而是作为一个国际通用的外部参照，用于检验本方法在总体感知排序与宏观空间格局层面的外部一致性。传统方法首先通过图像语义分割提取街景的视觉要素占比，然后训练一个支持向量分类器来预测其积极或消极感知。通过计算两种方法得分的皮尔逊相关系数r和斯皮尔曼秩相关系数ρ，并进行空间分布格局对比，来评估新方法的一致性和可靠性。

3 结果与分析

3.1 多模态大模型识别生成感知文本

通过对街景图像进行批量推理，本研究生成了包含30个感知维度文本描述的结构化数据集。为了直观展示该方法的输出结果，图4以一张典型的城市街景图像及其对应的多模态大模型感知文本为例，揭示了本框架进行城市空间感知评价的内在逻辑与细节特征。

如图4所示，该街景图像展现了一个典型的现代城市街道场景：道路宽阔、绿化良好且环境整洁。多模态大模型生成的感知文本精确地捕捉了这些显性的物理环境特征。例如，在“空间开阔度”维度下，模型准确识别出“街景视野开阔，建筑物之间的间距较大，没有明显的视觉遮挡”；在“空间整洁度”与“绿化覆盖率”维度下，模型分别给出了“地面、立面和公共设施保持整洁”以及“植被密度较高，树冠覆盖率良好”的细致描述。这表明大模型具备优秀的视觉要素提取与场景理解能力。

研究区域街景图像示例

基于多模态大模型识别的城市感知文本

```
“视觉多样性”：“街景中展示了丰富的颜色、材质和纹理，建筑物的设计风格为现代的高层建筑。",
"空间开阔度"："街景视野开阔，建筑物之间的间距较大，没有明显的视觉遮挡，街道宽阔，视野良好。",
"空间整洁度"："地面、立面和公共设施保持整洁，没有明显的破损或杂乱，整体环境干净有序。",
"景观特色度"："街景中没有明显的地标性元素或地域文化符号，建筑物和设施较为常规。",
"美感评价"："街景整体视觉舒适感良好，建筑布局合理，绿化与建筑和谐，给人以宁静和有序的美感。",
"绿化覆盖率"："植被密度较高，树冠覆盖率良好，街道两旁的树木提供了良好的遮阴效果。",
"植被健康度"："植被种类多样，色彩饱和度高，显示出良好的健康状况。",
"水体可见度"："街景中没有明显的河流、湖泊或湿地等水景元素。",
"生态廊道连通性"："绿地之间的连接状况良好，形成了连续的绿色空间。",
"自然舒适感"："主观上感受到的放松度较高，亲自然感明显，绿化和开阔的空间提供了良好的隔音降噪效果。",
"污染可见度"："街景中没有明显的垃圾、工业设施或其他潜在污染源。",
"噪音干扰度"："可视线索范围内没有明显的交通流量、工地施工或商业嘈杂，整体环境较为安静。",
"热岛效应迹象"："建筑密度适中，遮阴状况良好，地面铺装材质对热环境的影响较小。",
"海绵城市元素"："街景中可见雨水花园和下沉式绿地等可持续元素，显示了海绵城市的设计理念。",
"负面情绪度"："街景没有带来压抑、沉闷或疲惫的心理感受，整体氛围积极。",
"交通可达性"："人行道和自行车道连续性良好，宽度适宜，公共交通站点分布合理。",
"交通秩序度"："车流和人流有序，没有明显的拥堵或混行现象。",
"社会与空间公平"："街景中没有表现出明显的空间不平等或设施分配不均。",
"休憩与交往空间"："街边长椅、绿荫空间和凉亭分布合理，质量较高。",
"无障碍设施"："街景中配置了缘石坡道、无障碍通道和电梯等无障碍设施。",
"商业多样性"："商铺类型多样，包括餐饮、零售和服务业，分布均匀。",
"街道繁荣度"："人流量和车流量适中，店铺门庭若市，显示出一定的街道活力。",
"摊贩活动"："街景中没有明显的街头文化活动或灵活经营形态。",
"商业环境"："招牌设计和橱窗布置合理，店铺整洁度高。",
"经济活力"："通过地标性商场和品牌密集度可推断出较高的消费水平和商圈影响范围。",
"文化符号"："街景中没有明显的具有地方特色的雕塑、建筑或传统图案。",
"历史文脉"："老建筑保护情况良好，文物古迹和历史街区完整度高。",
"公共艺术"："街景中没有明显的壁画、装置艺术或小品等公共艺术元素。",
"社区归属感"："街景中没有明显的社区布告栏、社区活动场所或邻里交往空间。",
"多元文化融合度"："街景中没有明显的各类族群、文化活动痕迹等多元文化融合的迹象。
```

4 城市空间感知街景图像及对应的多模态大模型感知文本示例

Street view images of urban spatial perception, along with corresponding example perceptual descriptions from a multimodal large model

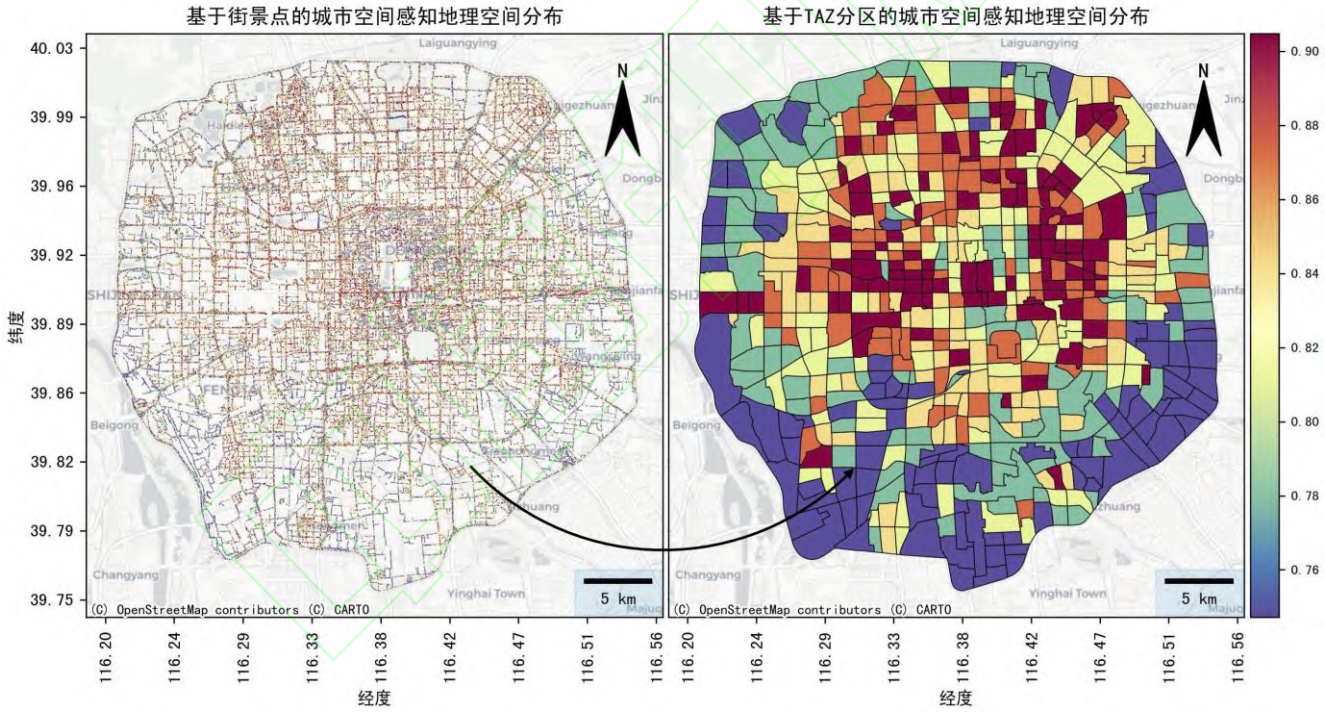
除了客观环境要素，模型同样展现出对主观感受和高级意象的出色评价能力。在“美感评价”维度中，模型综合各项视觉元素，给出了“绿化与建筑和谐，给人以宁静和有序的美感”的评价；在“自然舒适感”和“负面情绪度”维度下，模型进一步将其转化为人类的心理感受，描述为“主观上感受到放松度较高，亲自然感明显”以及“街景没有带来压抑、沉闷或疲惫的心理感受，整体氛围积极”。这些描述不仅客观翔实，还带有强烈的情境感知色彩，与人类的主观体验过程高度一致。

值得注意的是，多模态大模型不仅能识别图像中存在的显性元素，还能结合背景知识进行高阶推理，并准确判定特定元素的缺失。例如，模型能够基于画面中的建筑和商业线索，在“经济活力”维度推断出“较

高的消费水平和商圈影响范围”；同时，在“景观特色度”和“摊贩活动”维度中，也能客观地指出“没有明显的地标性元素或地域文化符号”以及“没有明显的街头文化活动”。这充分证明了本研究提出的基于多模态大模型与系统化提示词工程的方法，能够有效引导模型从多维度、专业化且拟人化的视角，对复杂的城市空间开展精细化的感知评价。

3.2 城市空间感知结果空间分布

利用多模态大模型计算得城市空间感知分数，图5所示本研究绘制了北京市五环内的城市空间感知地图。基于街景采样点的精细化评价结果直观地揭示了感知分数的微观空间异质性（图5左）。为了更好地解读宏观空间格局并与城市管理单元相衔接，本研究进一步将街景点感知分数通过空间连接的方法，聚合到交通分析小区（Transportation Analysis Zones，TAZ）层面，计算每个TAZ单元内的感知均值，从而生成了基于TAZ分区的城市空间感知地理空间分布图（图5右）[30–32]。从整体空间格局来看，感知分数呈现出明显的“中心高，外围低”的圈层式分布特征，这与北京“单中心”放射状的城市发展历史高度吻合。这一格局不仅是物理建成环境的直观反映，更是城市功能、经济活动与历史文脉共同作用下，在居民感知层面的综合体现。



5 基于多模态大模型的感知结果地理空间分布

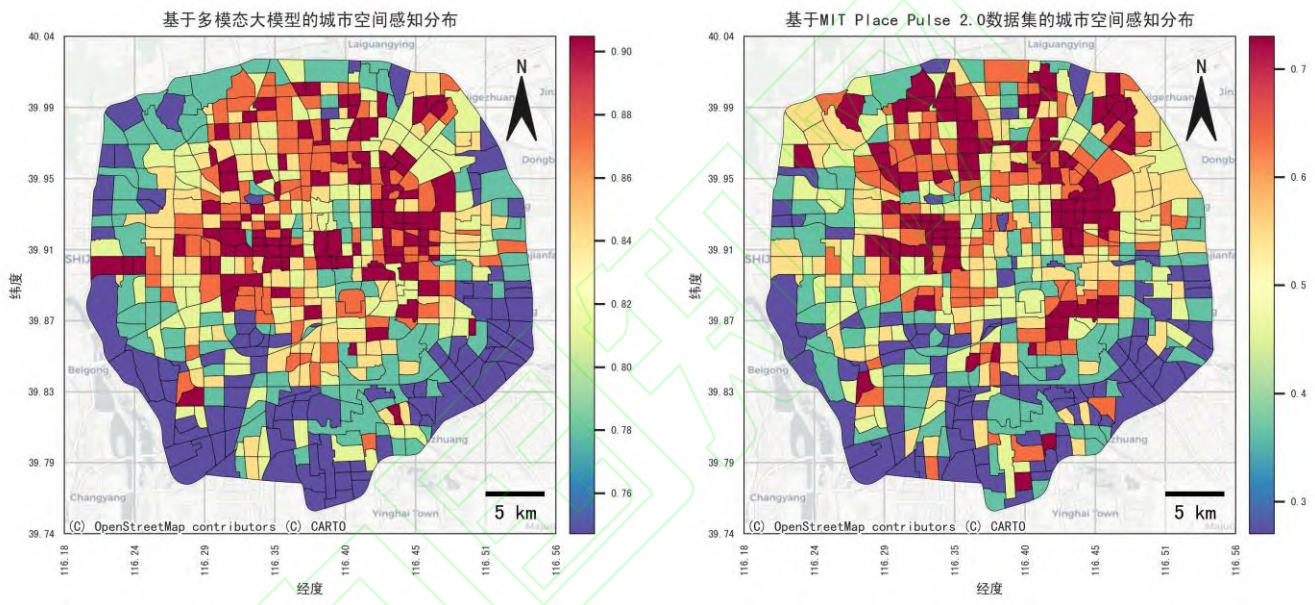
Geospatial distribution of perception results based on large multimodal models

高值区（得分>0.90）主要集中在二环至三环内的核心区域，如东城、西城的历史文化街区故宫周边、什刹海等，以及朝阳区的中央商务区。这些区域不仅是城市功能的核心承载区，更在长期的历史积淀与高强度的维护更新中，形成了具有独特景观特色、丰富文化符号和高经济活力的建成环境，因此获得了模型的高度评价。中值区（得分0.85-0.90）广泛分布于三环至四环之间，涵盖了中关村等科技园区、望京等国际化社区以及大型城市公园周边。这些区域代表了北京现代化建设的成果，其较高的绿化覆盖率、开阔的街道空间和完善的现代设施，共同营造了良好的整体感知。低值区（得分<0.85）则主要分布在四环以外至五环沿线，特别是在南部和西南部区域。这些区域多为新兴的大型居住区、部分待转型的老旧工业区或城乡结合部，可能存在景观风貌单调、公共服务设施配套不足或交通拥挤等问题，导致感知评价相对较低。

这一空间分异格局深刻地揭示了城市发展不均衡在城市感知层面的投射。它超越了传统的土地利用或经济密度分析，从体验品质的维度为我们描绘了一幅生动的城市软实力地图，为识别城市更新的重点区域和制定差异化的空间品质提升策略提供了科学依据。

3.3 传统感知方法结果对比与一致性验证

为验证本研究所提出感知测度方法的可靠性，将本研究的感知结果与基于MIT Place Pulse 2.0的传统方法计算结果进行了一致性验证对比。图6所示在宏观空间格局上，两种方法的结果高度一致，均清晰地揭示了北京城市空间感知的核心、边缘结构。历史文化街区、核心商务区等重点区域在两种方法中均被识别为感知高地，验证了本研究方法在捕捉城市核心空间品质方面的准确性。



6 多模态大模型与MIT Place Pulse 2.0数据集计算感知结果地理空间分布差异对比

Comparison of geospatial distribution differences between perception results computed by a multimodal large model and those from the MIT Place Pulse 2.0 dataset

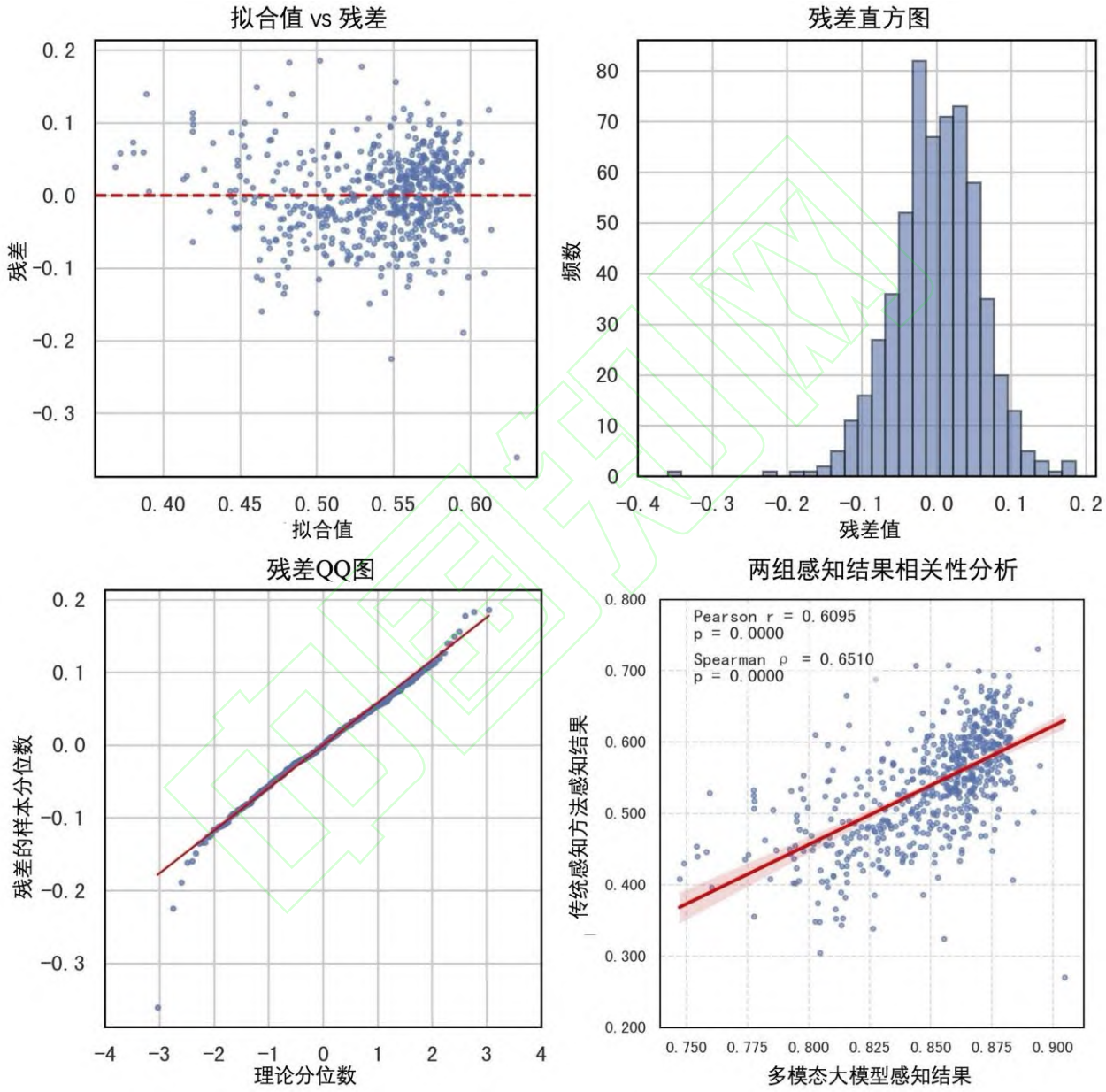
在数值分布上，两种方法存在系统性差异。表3所示多模态大模型的评分整体偏高（均值为0.849），且分布更为集中（标准差为0.079），而传统方法的评分均值较低（0.537），离散度更大（标准差为0.154）。这可能源于评分主体的不同：大模型基于其庞大的预训练知识库形成了一个相对统一的审美标准，倾向于对现代化、秩序化的景观给出积极评价；而传统方法依赖于全球多文化背景的众包志愿者投票，其评价标准更多元，也更容易识别出极端负面的场景。

为进一步量化一致性，本研究对两种方法的122,264个对应站点的感知分数进行了相关性与残差分析。如图7右下的散点图所示，两种方法计算的感知分数呈现出清晰的正相关趋势。经计算，两者的皮尔逊相关系数r为0.6095，表明存在中等偏强的线性关系；斯皮尔曼秩相关系数ρ为0.6510，数值更高，说明两种方法在对街景进行优劣排序时具有更强的一致性。两个相关系数均在P<0.001水平上高度显著，证明了本研究提出的新范式与传统方法在感知评价上具有很强的一致性。

表3 两组感知结果描述性统计与总体差异

Tab. 3 Descriptive statistics and overall differences between the two groups' perception results

指标	多模态大模型感知结果	MIT Place Pulse 2.0感知结果	差值
均值	0.849	0.537	+0.312
标准差	0.079	0.154	-0.075
中位数	0.869	0.538	+0.331



7 多模态大模型与MIT Place Pulse 2.0数据集计算感知结果相关性分析

Correlation analysis between perception results computed by large multimodal models and the MIT Place Pulse 2.0 dataset

为检验该相关关系的稳健性，本研究进一步开展了残差诊断。拟合值vs残差图图7左上显示，残差散点随机、均匀地分布在零线两侧，未呈现出明显的曲线或喇叭状形态，表明误差是随机的，且方差齐性。残差直方图图7右上呈现出以零为中心的近似正态钟形分布，而残差QQ图图7左下中的样本点也基本落在理论分位数直线上。这三张残差诊断图共同表明，两种方法之间的误差是随机、无偏且近似正态分布的。这些统计

结果有力地证明了，本研究提出的基于多模态大模型的方法能够有效、可靠地模拟人类对城市空间的主观感知，为其在更大范围内的应用提供了坚实的统计学保障。

4 讨论

4.1 城市空间感知方法学差异与分析

表 4 显示了两组感知结果在方法学上带来的差异与偏差来源。两套方法的本质差异体现在数据获取、评分主体、情境依赖以及可扩展性几个方面。数据输入与评分机制方面，多模态大模型基于街景图像加预设的 30 个感知子维度提示词，通过百亿级参数的预训练模型直接推理评分。MIT Place Pulse 2.0 数据集依赖在线两两对比问卷，由不同受访者投票得出相对分数。

评分主体与文化视角方面多模态大模型评分来源于模型内部对文本与图像联合语义的统一理解，因训练语料多来自欧美互联网，对现代化、高绿化景观往往打出偏高评价。MIT Place Pulse 2.0 数据集则囊括了全球各地参与者超过百万次投票，评判标准受文化、地理背景影响显著，能捕捉地方性审美差异与生活体验。

情境依赖性与偏差风险方面多模态大模型的输出高度依赖提示词设计与底层语料库，提示词一旦调整即可影响整体偏好。MIT Place Pulse 2.0 数据集评分则易受到问卷样本结构，以及受访者对低收入或欠发达地区的刻板印象所左右。例如年轻化、网络活跃用户偏多。

扩展能力方面多模态大模型推理成本较低，可在短期内批量化、自动化地处理大量图像。MIT Place Pulse 2.0 数据集依赖持续的问卷收集与运营维护，短期内难以显著扩容。

表 4 两组感知结果方法学差异与偏差来源

Tab. 4 Methodological differences and sources of bias between the two sets of perceptual results

维度	多模态大模型感知结果	MIT Place Pulse 2.0 感知结果
信息输入	图像+提示词（30 个感知子维度）	街景图像+在线对比式问卷
主体	预训练模型（参数≈百亿）	全球受访者（>100 万次两两对比）
情境依赖性	严重依赖提示词设计与模型训练语料，可随提示词迭代优化	受文化/地理背景影响，具有地域性
偏差风险	训练数据多来自欧美互联网，可能高估现代化景观	投票平台用户年轻化，或对低收入/欠发达地区有刻板印象
扩展性	推理成本低，可覆盖亿级街景	数据收集成本高，问卷扩容周期长
指标体系	基于可持续发展三重底线理论构建评价框架	由全球在线参与者通过成果对比投票产生评价无固定理论指导

4.2 人工智能在景观规划领域的实践启示

本研究为风景园林规划与设计实践提供了直接的应用价值，主要体现在以下几个方面：

（1）实现建成环境的精细化诊断与循证评估。传统的场地分析多依赖定性判断和抽样调查，效率和精度有限。本研究提出的范式，能够实现对城市建成环境感知质量的大尺度、高分辨率的自动化评估，形成量化的城市感知体检报告。通过对各分项指标的空间分布进行分析，规划师可以精准识别出城市空间品质的短板区域。例如，定位出缺乏绿色基础设施的街道、缺少公共交往空间的社区或风貌特色不足的商业区。这为城市更新、社区营造和景观提升项目中的资源优先配置和靶向性干预提供了直接的实证依据。

（2）赋能设计方案的前瞻性模拟与比选。传统的设计评价多为建成后的后评估，周期长且难以补救。

本研究证实了多模态大模型模拟人类感知的潜力，这为设计过程的前评估提供了新的可能。在方案设计阶段，设计师可将不同方案的效果图、意向图等视觉材料输入模型进行预测性分析，量化比较各方案在美学、安全、活力等感知维度上的潜在表现。这一应用能够辅助设计师在早期阶段进行科学决策，从多个备选方案中择优，从而提升设计方案的综合效益和前瞻性。

（3）辅助公众参与及协同规划。本研究以模拟大众感知为核心，其成果具有内在的“以人为本”属性。生成的感知地图可作为公众参与过程中直观的沟通工具，帮助居民理解其生活环境的现状品质。此外，该方法也为处理和分析公众意见提供了新的技术路径。例如，可以利用该模型对公众提交的带有文字描述的现场照片进行快速、标准化的感知评价和问题归类，从而更高效地识别社区共识与矛盾焦点，为协同规划提供量化数据支持。

综上所述，本研究不仅提供了一种技术工具，更重要的是，它为规划设计实践提供了向数据驱动和循证设计转型的方法论支持，有助于提升人居环境品质改善工作的科学性与精准度。

4.3 研究局限性与未来展望

本研究也存在一定的局限性。首先，大模型的感知结果高度依赖其训练数据和提示词设计，可能存在潜在的文化偏见或审美趋同。其次，本研究采用了实验进行时性能最优的Qwen2-VL-72B作为核心模型，但多模态大模型技术正处在高速迭代期，新模型层出不穷。未来更新的模型可能在感知精度、推理效率和多语言文化理解上表现更佳。因此，本研究的结论具有一定的技术时效性，后续研究需持续关注并整合最新的模型进展。最后，街景数据是静态的，无法完全捕捉城市空间的动态变化，如昼夜、季节、节假日的人流活动。

未来研究可以从以下几个方面展开：（1）引入可解释性AI技术，如利用CAM（Class Activation Map）可视化模型在做判断时重点关注的图像区域，以打开“黑箱”，增强模型决策的透明度。（2）融合多源动态数据，如手机信令、社交媒体签到数据等，以捕捉城市空间的动态活力。（3）构建领域专属大模型，通过在中国本土的、包含丰富风景园林专业知识的图文语料上进行微调，构建更懂中国城市和园林美学的专家模型。

5 结论

本研究成功构建并验证了一套基于多模态大模型的城市街道空间感知评价框架。结果表明通过系统化的理论构建和精密的提示词工程，前沿的AI大模型能够有效、可靠且精细化地量化复杂的城市空间感知。在北京的实证研究中，该方法不仅揭示了城市空间感知的宏观格局，其结果也与国际公认的传统方法高度一致。

这项工作展示了新一代人工智能技术在风景园林和城市科学领域的巨大应用潜力，为实现更高效、更科学、更人性化的城市规划与景观设计提供了强有力的智能化工具。随着技术的不断进步，一个由数据和AI驱动的、能够深度理解并优化人居环境的新时代正在到来。

注释(Notes)：

① 完整提示词模板：

背景（Background）你是一位城市科学研究专家，正在使用多模态大模型来研究从街景图像中提取对城市空间的感知。

目标（Goal）依据输入的街景图像，输出可供后续感知评价使用的学术化描述文本，包括对街景特征的深度描述。

风格（Style）采用学术且专业的表达方式，引用或融合城市规划和社会学中的关键术语，确保对城市特征的描述客观而详实。

语气（Tone）请使用学术且专业的表达方式，采用正式、严谨的语气，避免夸张或情绪化字眼，注重逻辑与条理性。
受众（Audience）面向景观规划、地理信息系统和社会学等领域的科研学者，以及从事感知评价与数据建模的工程师。
输出（Output）每个维度的描述都应该是独立且完整的分析。所有输出必须完全使用中文，包括 JSON 字段名称，不要出现任何英文。
输出格式示例：
{"视觉多样性": "详细的描述...",
"空间开阔度": "详细的描述...",
"空间整洁度": "详细的描述..."}
请从以下 30 个维度进行深入分析并以中文 JSON 格式输出：
1. 视觉多样性：街道空间颜色、材质、纹理的丰富程度
2. 空间开阔度：街景是否通透，视野是否受限
3. 空间整洁度：地面、立面、公共设施等是否整洁、无明显破损或杂乱
4. 景观特色度：是否具有独特的地标性元素、地域文化符号
5. 美感评价：对于街景整体的视觉舒适感、艺术性、和谐度的主观评价
6. 绿化覆盖率：视野内植被密度、树冠覆盖率
7. 植被健康度：植被种类、多样性、色彩饱和度
8. 水体可见度：是否存在河流、湖泊、湿地等水景元素
9. 生态廊道连通性：绿地或水体之间的连接状况
10. 自然舒适感：主观上感受到的放松度、亲自然感、隔音降噪效果等
11. 污染可见度：是否能观察到垃圾、工业设施或其他潜在污染源
12. 噪音干扰度：可视线索范围内是否有交通流量、工地施工、商业嘈杂等
13. 热岛效应迹象：建筑密度、遮阴状况、地面铺装材质对热环境的影响
14. 海绵城市元素：雨水花园、下沉式绿地、立面绿化、屋顶花园等可持续元素可见度
15. 负面情绪度：街景是否带来压抑、沉闷或疲惫的心理感受
16. 交通可达性：人行道、自行车道的连续性与宽度，公共交通站点分布
17. 交通秩序度：是否有秩序的车流、人流，是否有拥堵或混行现象
18. 社会与空间公平：是否表现出某种空间不平等或设施分配不均
19. 休憩与交往空间：街边长椅、绿荫空间、凉亭或遮阳蓬等分布与质量
20. 无障碍设施：缘石坡道、无障碍通道、电梯、盲道等配置情况
21. 商业多样性：商铺类型数量及分布
22. 街道繁荣度：人流量、车流量、店铺门庭若市程度
23. 摊贩活动：是否有街头文化活动，是否显示出灵活经营形态
24. 商业环境：招牌设计、橱窗布置、店铺整洁度等
25. 经济活力：通过地标性商场或品牌密集度可推断的消费水平、商圈影响范围
26. 文化符号：具有地方特色的雕塑、建筑、传统图案等
27. 历史文脉：老建筑保护情况、文物古迹、历史街区完整度
28. 公共艺术：壁画、装置艺术、小品等对城市氛围的提升
29. 社区归属感：可视化的社区布告栏、社区活动场所、邻里交往空间
30. 多元文化融合度：是否能在街景中观察到各类族群、文化活动痕迹等

参考文献(References)：

[1] 【高国力：深入实施新型城镇化战略 稳步提高城镇化水平和质量】-国家发展和改革委员会[EB/OL]. [2025-08-12]. https://www.ndrc.gov.cn/wsdwhfz/202408/t20240819_1392442.html.
Deeply Implement the New-type Urbanization Strategy and Steadily Improve the Level and Quality of Urbanization[EB/OL]. National Development and Reform Commission. [2025-08-12]. https://www.ndrc.gov.cn/wsdwhfz/202408/t20240819_1392442.html.
[2] LYNCH K. The Image of the City[M]. Cambridge, MA, USA: MIT Press, 1964.

[3] GU Y, QUINTANA M, LIANG X, et al. Designing effective image-based surveys for urban visual perception[J/OL]. Landscape and Urban Planning, 2025, 260: 105368. DOI:10.1016/j.landurbplan.2025.105368.
[4] BILJECKI F, ITO K. Street view imagery in urban analytics and GIS: A review[J/OL]. Landscape and Urban Planning, 2021, 215: 104217. DOI:10.1016/j.landurbplan.2021.104217.
[5] 郑屹, 杨俊宴. 基于大规模街景图片人工智能分析的精细化城市修补方法研究[J/OL]. 中国园林, 2020, 36(8): 73-77.
ZHENG Y, YANG J Y. Research on refined urban mending methods based on AI analysis of large-scale street view images[J/OL]. Chinese Landscape Architecture, 2020, 36(8): 73-77.
[6] 曹越皓, 杨培峰, 李敏敏, 等. 基于视觉融合模型的城市景观精细化感知方法研究——以重庆市主城区为例[J]. 中国园林, 2025, 41(3): 76-83.
CAO Y H, YANG P F, LI M M, et al. Refined urban landscape perception method based on a visual fusion model: A case study of Chongqing main urban area[J]. Chinese Landscape Architecture, 2025, 41(3): 76-83.
[7] NAIK N, PHILIPOOM J, RASKAR R, et al. Streetscore - Predicting the Perceived Safety of One Million Streetscapes[C/OL]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2014: 779-785[2025-10-13]. https://www.cv-foundation.org/openaccess/content_cvpr_workshops_2014/W20/html/Naik_Streetscore_-_Predicting_2014_CVPR_paper.html.
[8] WANG L, HOU C, ZHANG Y, et al. Measuring solar radiation and spatio-temporal distribution in different street network direction through solar trajectories and street view images[J/OL]. International Journal of Applied Earth Observation and Geoinformation, 2024, 132: 104058. DOI:10.1016/j.jag.2024.104058.
[9] FAN Z, ZHANG F, LOO B P Y, et al. Urban visual intelligence: Uncovering hidden city profiles with street view images[J/OL]. Proceedings of the National Academy of Sciences, 2023, 120(27): e2220417120. DOI:10.1073/pnas.2220417120.
[10] ZHANG F, FAN Z, KANG Y, et al. “Perception bias”: Deciphering a mismatch between urban crime and perception of safety[J/OL]. Landscape and Urban Planning, 2021, 207: 104003. DOI:10.1016/j.landurbplan.2020.104003.
[11] ZHANG F, LIU Y. 街景影像——基于人工智能的方法与应用[J]. 遥感学报, 2021.
ZHANG F, LIU Y. Street view imagery: AI-based methods and applications[J]. Journal of Remote Sensing, 2021.
[12] LIU Y, CHEN M, WANG M, et al. An interpretable machine learning framework for measuring urban perceptions from panoramic street view images[J/OL]. iScience, 2023, 26(3): 106132. DOI:10.1016/j.isci.2023.106132.
[13] 刘智谦, 吕建军, 姚尧, 等. 基于街景图像的可解释性城市感知模型研究方法[J]. 地球信息科学学报, 2022, 24(10): 2045-2057.
LIU Z Q, LV J J, YAO Y, et al. Research methods for interpretable urban perception models based on street view images[J]. Journal of Geo-information Science, 2022, 24(10): 2045-2057.
[14] OPENAI, ACHIAM J, ADLER S, et al. GPT-4 Technical Report[A/OL]. arXiv, 2024[2025-02-05]. http://arxiv.org/abs/2303.08774. DOI:10.48550/arXiv.2303.08774.
[15] WANG P, BAI S, TAN S, et al. Qwen2-VL: Enhancing Vision-Language Model’s Perception of the World at Any Resolution[A/OL]. arXiv, 2024[2025-10-13]. http://arxiv.org/abs/2409.12191. DOI:10.48550/arXiv.2409.12191.
[16] HOU C, ZHANG F, LI Y, et al. Urban sensing in the era of large language models[J/OL]. The Innovation, 2025, 6(1): 100749. DOI:10.1016/j.xinn.2024.100749.
[17] WANG L, ZHANG T, HE J, et al. Cross-platform complementarity: Assessing the data quality and availability of Google Street View and Baidu Street View[J/OL]. Transactions in Urban Data, Science, and Technology, 2025: 27541231241311474. DOI:10.1177/27541231241311474.
[18] SALESSES P, SCHECHTNER K, HIDALGO C A. The Collaborative Image of The City: Mapping the Inequality of Urban Perception[J/OL]. PLoS ONE, 2013, 8(7): e68400. DOI:10.1371/journal.pone.0068400.

[19] ZHANG F, ZHOU B, LIU L, et al. Measuring human perceptions of a large-scale urban region using machine learning[J/OL]. Landscape and Urban Planning, 2018, 180: 148-160. DOI:10.1016/j.landurbplan.2018.08.020.
[20] TEAM Q. Qwen2-VL: 更清晰地看世界[EB/OL]. (2024-08-29)[2025-02-05]. https://qwenlm.github.io/zh/blog/qwen2-vl/.
TEAM Q. Qwen2-VL: See the World More Clearly[EB/OL]. (2024-08-29)[2025-02-05]. https://qwenlm.github.io/zh/blog/qwen2-vl/.
[21] Hugging Face – The AI community building the future.[EB/OL]. (2025-10-14)[2025-10-16]. https://huggingface.co/.
[22] ELKINGTON J. Cannibals with Forks: The Triple Bottom Line of 21st Century Business[M]. Capstone, 1997.
[23] 郑凌予, 杨淑梅, 伍夏, 等. 基于空间感知的社区绿色暴露指数对活动行为影响研究[J/OL]. 中国园林, 2023, 39(1): 92-97.
ZHENG L Y, YANG S M, WU X, et al. Effects of a community green exposure index based on spatial perception on activity behavior[J/OL]. Chinese Landscape Architecture, 2023, 39(1): 92-97.
[24] 邱瑶, 罗涛, 王艳云, 等. 基于视觉关注度与审美偏好的城市景观元素感知特征研究[J/OL]. 中国园林, 2023, 39(6): 82-87.
QIU Y, LUO T, WANG Y Y, et al. Perceptual characteristics of urban landscape elements based on visual attention and aesthetic preference[J/OL]. Chinese Landscape Architecture, 2023, 39(6): 82-87.
[25] 巴泽智, 张辉, 谢铮涵, 等. 大语言模型自动化提示工程技术研究综述[J]. 计算机科学与探索, 2025, 19(12): 3131-3152.
BA Z Z, ZHANG H, XIE Z H, et al. A Review of Automated Prompt Engineering Techniques for Large Language Models[J]. Journal of Frontiers of Computer Science and Technology, 2025, 19(12): 3131-3152.
[26] google-bert/bert-base-chinese ·Hugging Face[EB/OL]. [2025-02-05]. https://huggingface.co/google-bert/bert-base-chinese.
[27] VENNERØD C B, KJÆRRAN A, BUGGE E S. Long Short-term Memory RNN[A/OL]. arXiv, 2021[2025-02-08]. http://arxiv.org/abs/2105.06756. DOI:10.48550/arXiv.2105.06756.
[28] DATASETS AT HUGGING FACE. ChnSentiCorp 中文情感分析数据集[CP/OL]. [2025-02-16]. https://huggingface.co/datasets/lansinuote/ChnSentiCorp.
DATASETS AT HUGGING FACE. ChnSentiCorp: A Chinese Sentiment Analysis Dataset[CP/OL]. [2025-02-16]. https://huggingface.co/datasets/lansinuote/ChnSentiCorp.
[29] DATASETS AT HUGGING FACE. weibo_senti_100k[CP/OL]. (2024-08-14)[2025-02-16]. https://huggingface.co/datasets/dirtycomputer/weibo_senti_100k.
[30] YAO Y, LI X, LIU X, et al. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model[J/OL]. International Journal of Geographical Information Science, 2017, 31(4): 825-848. DOI:10.1080/13658816.2016.1244608.
[31] HUANG Z, QI H, KANG C, et al. An Ensemble Learning Approach for Urban Land Use Mapping Based on Remote Sensing Imagery and Social Sensing Data[J/OL]. Remote Sensing, 2020, 12(19): 3254. DOI:10.3390/rs12193254.
[32] HOSSEINZADEH A, ALGOMAIAH M, KLUGER R, et al. Spatial analysis of shared e-scooter trips[J/OL]. Journal of Transport Geography, 2021, 92: 103016. DOI:10.1016/j.jtrangeo.2021.103016.

图表来源(Sources of Figures and Tables)：

图表均由作者自行绘制。

作者简介（中英文）：

王磊/男/北京大学地球与空间科学学院博雅博士后/研究方向为城市计算、景观感知

WANG Lei, Boya Postdoctoral Fellow, School of Earth and Space Sciences, Peking University. His research focusing on urban computing and landscape perception.

郭家乐/男/米兰理工大学土木与环境工程系地理信息工程专业硕士在读/研究方向为 GeoAI、时空数据智能分析建模与模拟、计算地理与可持续城市建模

Guo Jiale, Master student in Geoinformation Engineering in the Department of Civil and Environmental Engineering at Politecnico di Milano. His research focuses on GeoAI, spatiotemporal data intelligence analysis, modeling and simulation, computational geography, and sustainable urban modeling.

白钊成/男/同济大学建筑与城市规划学院在读博士研究生/研究方向为数字景观

Bai Zhaocheng, PhD candidate at the College of Architecture and Urban Planning, Tongji University. His research focuses on a digital landscapes.

何捷/男/哈尔滨工业大学(深圳)建筑学院教授/博士生导师/研究方向为空间历史大数据、景观考古学与文化景观遗产、大数据与空间行为、地理设计与户外游憩

通信作者邮箱（Corresponding author Email）：hejie2021@hit.edu.cn

HE Jie, Professor at the School of Architecture, Harbin Institute of Technology (Shenzhen), and a doctoral supervisor. His research focuses on spatial historical big data, landscape archaeology and cultural landscape heritage, big data and spatial behavior, and geographic design and outdoor recreation.