

咖啡厅入口看起来方便进出吗？门在哪里？ 迈向用于视觉查询的地理空间 AI 代理

Jon E. Froehlich^{1,2} Jared Hwang¹ Zeyu Wang¹ John S. O'Meara¹ Xia Su¹ William Huang³
Yang Zhang³ Alex Fiannaca⁴ Philip Nelson² Shaun Kane²

¹University of Washington ²Google Research ³UCLA ⁴Google DeepMind

jonf@cs.uw.edu



图 1. 我们介绍了我们的愿景——地理视觉代理，即能够通过分析大规模的地空图像和传统 GIS 数据源来理解和回应关于世界的复杂视觉空间询问的多模态 AI 代理。例如，街景人工智能 [14]（上图）将街景提供给盲人用户，通过将地理环境、用户信息以及动态街景图像结合到一个 MLLM 中实现，该系统可以通过人工智能聊天界面和可访问屏幕阅读器控制来使用。

Abstract

交互式数字地图彻底改变了人们旅行和了解世界的方式；然而，它们依赖于预先存在的地理信息系统数据库中的结构化数据（例如，道路网络、POI 索引），限制了它们解决与什么世界的样子相关的地理视觉问题的能力。

我们提出了我们对地理视觉代理的愿景——能够理解和响应关于世界细微视觉空间查询的多模态人工智能代理，通过分析大规模的地理空间图像存储库，包括街景（例如，Google Street View）、基于地点的照片（例如，TripAdvisor、Yelp）和航拍图像（例如，卫星

照片），以及传统的 GIS 数据源。

我们定义了我们的愿景，描述了感知和交互方法，提供了三个示例，并列举了未来工作的关键挑战和机遇。

1. 介绍

在过去的二十年里，精确的位置感知、无处不在的互联网连接以及交互式数字地图已经将人类的移动性从旅行规划转变为原位导航。尽管有了这些进步，目前的地图系统仍然局限于现有的结构化地理空间数据，大量的视觉信息——潜藏在街景、航空和用户贡献的

图像中——未被挖掘且无法用于回答我们所说的地视问题。也就是说，关于某个地点或路线的以视觉为导向的地理问题。例如，想象一个轮椅使用者询问“去 35 楼的图书馆有没有楼梯？”或一个盲人旅行者询问“咖啡馆的门在哪里，它看起来是什么样子的？”

在本次研讨会论文中，我们介绍了对地理视觉代理——多模态 AI 代理的愿景，这些代理能够通过分析大规模的地理空间图像库（例如，街景和航拍图像）结合传统的 GIS 数据库（例如，道路网络、POI 数据库、交通时刻表），理解和回应关于世界的细微视觉-空间查询。我们设想 Geo-Visual Agents 在从先验的旅行规划到原位导航等各种情境中充当“视觉空间副驾”的角色。至关重要，尽管我们预期许多高价值的用户场景中，一个地理视觉代理通过 AR 眼镜 [12, 33, 34, 63] 或智能手机摄像头 [40, 50, 57] 实时感知和处理可视空间数据，但同样大量的问题可以通过分析现有的（且大部分未被利用的）地理相关图像库来解答——无论是按需（例如，启动一个 AI 代理进行查询和分析来源）还是通过预计算。

我们的愿景超越了当前地理空间人工智能 (GeoAI) [17, 30, 36] 如计算机辅助放射治疗优化人工智能 [7] 和超图 [52] 的范式，这些主要集中在为领域专家进行大规模数据分析。类似地，我们的工作与 GIS 研究中出现的范式如“自主 GIS”——基于 AI 的科学助手，帮助“理由、推导、创新并推进解决紧迫的全球地理空间挑战的方法”[39] 的工作相关但不同。此外，由于我们设想的代理主要通过多模态对话式 AI 工作，我们从近期如地空视觉问答 (GVQA) 的研究中汲取灵感，例如 MQVQA [62] 和 TAMMI [6]，这些研究试图赋予多模态 LLM 特定领域的地理知识；然而，这些系统主要面向分析师，并且主要处理远程航空影像。尽管相关，我们的重点在于解决个人在规划旅行或实际导航空间时的个性化、互动性以及往往即时的需求。

以下，我们扩展了我们的愿景，包括视觉空间查询的分解、感知和交互的方式，并给出了三个新兴的例子：街景人工智能 [14]，访问性侦察队 [27] 和自行车管家。整个过程中，我们强调了关键机遇与开放挑战。

2. 跨旅行阶段的地理可视化查询

我们设想地理视觉代理在整个出行周期中提供价值，从出行前规划到原位导航。以下是四个旅行阶段以

及其中的地理视觉代理的机会，重点是可访问性，但也包括更广泛的用户场景，如驾驶和骑行。选择和融合数据源将取决于用户的任务和数据可用性。例如，出行前规划可能依赖于街道景观图像、用户贡献的照片和基于地点的评论，而在现场导航可能会结合这些来源与来自用户实时摄像头（例如，来自 AR 眼镜）的视觉内容以及环境感知（例如，旅行模式推断，位置）。

出行前规划。在此阶段，用户并未实际出现在某个地点，而是计划未来的访问。代理充当远程互动向导，使详细调查成为可能，并在旅行前减少不确定性。例如：(1) 一个盲人家长计划去公园旅行时可能会询问“游乐场有哪些设备，看起来安全吗？”(2) 一个行动不便的人虚拟地勘察路线并询问“从我所在的位置到我的医生办公室的路上都有无障碍坡道吗？”(3) 一位潜在的购房者可能会问一些与社区相关的问题，如“街道看起来是什么样子的？”，“有树木遮阴的人行道吗？”和“有多少涂鸦？”

在导航时。在旅行过程中，用户处于认知和身体负荷之下，导航环境、选择路线并动态避开障碍物。在这里，代理提供有关目的地或即将进行的操作的前瞻信息，增强情境意识，并促进原位旅行决策。例如：(1) 接近交叉路口的司机询问，“你说在下一个红绿灯处左转。有什么地标吗？”(2) 快要到达决策点的骑车人查询，“下一个路口有自行车专用道吗？在道路的哪一侧？”(3) 下火车的铁路用户问，“哪个出口离图书馆的无障碍入口最近？”

目的地到达。当到达目的地时，用户会面临一系列与目的地外观、入口路径和位置以及障碍物或安全问题相关的“最后 10 米”问题。例如，(1) 接近目的地时，送货司机可能会询问“这个建筑的装卸区在哪里？”；(2) 在繁忙的广场遇到朋友的人可能会问“我在寻找那家咖啡店；你能描述一下它的店面吗，这样我可以更容易地找到它？”。(3) 盲人乘客的拼车服务在一个繁忙的机场接他们，并问道“你能帮我找到车牌号为 KNI667 的银色丰田凯美瑞吗？”。

室内探索。最终，当进入目的地时，代理的角色可以转变为支持在机场、商店或办公大楼等复杂的室内环境中的微导航。这一阶段面临着显著的数据挑战，因为全面的视觉和地图数据集对于室内空间来说较为罕见 [13]。例如，(1) 一个试图找到五金店中特定商品位置的顾客可能会询问“根据过道的指示牌，我应该朝哪

个方向去寻找水管部门？”(2) 一位视力不佳的旅行者查看机场出发信息牌：“你能告诉我达美航空 850 航班是从哪个登机口出发的吗？”；(3) 在大型会议中心使用轮椅的人：“你能指引我去最近的无障碍卫生间吗？”

这些场景共同展示了 Geo-Visual Agents 如何改变我们导航和理解地点的方式，增强可达性，提供基于地标的导航，改善个人安全，甚至带来意外的发现。以下我们将描述潜在的数据来源，然后概述交互模式。

3. 感知与数据源

地理视觉代理的强大之处在于其能够综合异构数据源，将视觉证据与结构化地理空间数据融合，形成对某个地点或路线的整体且准确的理解。我们在下面重点关注与地理相关的图像来源，而不是结构化的 GIS 数据。

街道景观图像。街景图像 (SVI) [25, 38]—例如谷歌街景视图 (GSV)，循环媒体，卡塔视图和地图力里—提供了世界范围内的丰富、大规模图像档案。仅 GSV 自身就有超过 2200 亿张图像，跨越 100 个国家的 1000 万英里 [19]。此类数据可用于分析道路状况 [3]、街道标记 (人行横道 [2, 35]、自行车道 [46])、人行道基础设施 (人行道材质 [23]、缘石坡道 [21, 42])、公交站 [32]、建筑立面 [31]、涂鸦 [53]、树木和植被 [37]、社区健康指标 [55, 64] 等。主要限制包括图像的近期性 [56]，由于 SVI 摄像头前方物体 (如例如，一辆公交车) 引起的遮挡 [48]，以及地理分布 (图像沿道路每 10-15 米分布，但不覆盖人行道、公园或建筑物内部)。

用户贡献的照片。基于地点的平台如谷歌地方，耶尔普和 TripAdvisor 包含了大量的与特定兴趣点关联的照片，这些照片来源于群众贡献，为 SVI 提供了有用的补充，包括建筑物内部、经过策划 (商家上传) 的店面照片和菜单、食品 [16] 以及社交活动 (例如，[61])—所有这些都是通常伴随着用户贡献的文字 (例如，评论)。然而，我们发现，在文献中对这种多模态数据的分析较少见。这里的关键限制是数据可用性，特别是对于不受欢迎或最近才开放的地方，以及在谁上传和为什么 (例如，见 [4, 59]) 中的社会偏见。

航空影像。来自卫星、飞机或无人机的航空影像可以提供高分辨率的从上方或倾斜 (45 度角) 视角的空间结构视图，包括建筑物轮廓、停车场、植被和行人基础设施 [24]。虽然遥感和摄影测量研究已经存在了

数十年—例如，用于土地利用分类、农业、灾害响应和军事分析 [29, 60]—但这些技术尚未应用于 Geo-Visual Agent 的情境 (例如，回答终端用户关于停车场位置、屋顶餐厅露台或未绘制的行人捷径的问题)。与街道景观类似，航空影像可能会受到遮挡 (来自树冠、云层)、高大建筑物的阴影以及缺乏可用性的困扰。在美国，高分辨率航空影像通常由联邦政府提供，如 USGS [54] 和 NASA [41]。

机器人扫描。自动驾驶车辆、地面配送机器人和无人机等机器人 [49, 51] 装备了传感器套件 (摄像头、激光雷达) 可以生成环境的高保真扫描，不仅产生图像，还产生具有量测功能的 3D 重建 [26]。尽管是一个潜在有前景的未来数据来源，目前缺乏开放的数据和 API。

基于基础设施的相机。基础设施安装的用于交通、天气、安全和监控的摄像头提供城市的实时视图，并且独特地提供了关于行人和汽车运动、人类活动、天气条件和临时障碍物的动态信息 [28, 44, 47]；然而，尽管某些摄像头的视频流是公开的—例如，交通部的交通摄像头—大多数并不公开，隐私是一个关键考虑因素。此外，还存在密度和可用性 (例如，在农村地区) 不足的问题。

第一人称相机流。最后，来自 AR 眼镜 [12, 33, 63]、智能手机摄像头 [5, 40, 50, 57] 和行车记录仪 [43, 58] 的第一人称摄像流对于现场旅行阶段至关重要，提供了实时的自我中心视角用于导航、识别瞬时障碍物和读取标志。虽然这些视频流主要用于即时辅助，但它们也可能有助于在持续反馈循环中更新或纠正现有的地理空间数据集 (例如，[58])。然而，关键考虑因素包括高计算和功耗要求、强大的网络连接以及用户和旁观者的隐私问题。

4. 处理和解读与人工智能相关的事务

我们的愿景不仅依赖于多样化的地理空间图像和现有的 GIS 数据，还依赖于多模态 AI 的进步 (例如，场景理解 [9, 11]、对象可利用性 [22, 33] 和空间推理 [8, 10, 15, 45]) 来提取语义信息和对象关系。虽然可以预先计算某些已知高价值实体的信息 (如例如，人行道斜坡的存在和位置 [21, 42])，但我们预计会有大量的定制查询，这将需要一个 Geo-Visual Agent 在实时中寻找、分析并将基于图像的来源与 GIS 数据库中的预存元数据进行综合。

5. 交付答案

最后，我们愿景的一个关键方面是如何代理传输信息的方式，这取决于用户的技能、他们当前的环境以及数据的复杂性和类型。无论采用何种传递方式，代理都需要报告不确定性及数据来源以建立信任并减少错误。

以音频为主的界面：对于无需手持和/或无需注视的操作——这对驾驶员、骑车人以及视障和低视力用户来说是必不可少的——音频界面至关重要（例如，使用耳塞或智能扬声器）。然而，挑战在于提供结构良好的口头描述以传达复杂的视觉信息而不使用户感到不知所措。

多模态界面：代理应选择并展示相关的图像。例如，在描述入口之后，代理可以显示一扇门的照片（例如，来自 SVI 或 Yelp）。挑战在于 AI 从大量档案中选择最适当的照片——适当地裁剪——的能力。

人工智能生成的抽象可视化：对于高度复杂的空间信息，一张原始照片或冗长的文字描述可能不足以表达。一个令人兴奋的研究前沿是代理程序生成简化、抽象图表的能力——类似于现代的直线驱动器系统 [1]。使这些抽象内容可访问，也许是通过触觉方式，也是一个重要的开放研究领域。

6. 案例研究应用

为了展示和具体化我们的愿景，我们突出三种新兴的 Geo-Visual Agent 原型。

StreetViewAI. 当前的 SVI 工具对盲人用户来说是无法访问的。我们的团队通过设计街景人工智能 [14](Figure 1)，正在解决这个问题，该工具使用上下文感知、实时 AI 来支持虚拟探索路线、检查目的地或甚至远程参观如大峡谷 [18] 等旅游地点。StreetViewAI 为盲人用户提供可访问的交互控制，使他们可以在全景图像之间平移和移动，并与实时多模态 AI 代理动态对话讨论场景和本地地理信息。在实验室研究中，盲人用户有效地使用了 StreetViewAI 进行虚拟导航街景。关键挑战：协调用户对 SVI 的心理模型、过度信任 AI 的倾向以及将丰富的视觉数据综合成简洁音频的困难。

AI 代理。街景视图 AI 采用三个独立的 AI 子系统。最相关的是人工智能聊天代理，它允许用户与当前和过去的街景以及附近的地理环境进行对话互动。该代理

使用谷歌的多模态实时 API [20]，支持实时交互、函数调用，并保留单个会话内所有互动的记忆。当用户通过打字或说话启动聊天时，我们会将每个 GSV 互动连同用户的当前视图和地理环境（例如，附近的地点，当前朝向）一起传输。因此，用户可以询问当地地理、当前和过去的视图以及物体关系（例如，“入口在哪里？”）。

可访问性侦察兵。评估不熟悉环境的可访问性对残疾人来说是一项重要但往往费力的工作。虽然存在标准化检查表，但它们通常无法考虑个人的独特和变化的需求。访问性侦察队 [27] 是一个基于大语言模型的系统，旨在通过从图像生成个性化的可访问性扫描来解决这一问题——例如，从旅行顾问网, Yelp 和爱彼迎——根据自我报告的能力和兴趣识别潜在的问题。在用户研究中，我们发现可访问性侦察的个性化扫描比通用扫描更有用，并且其协作的人工智能方法有效并建立了信任。

人工智能代理。可访问性侦察管道首先通过从用户对其能力和偏好的纯文本描述中初始化创建一个结构化的用户模型，该模型采用 JSON 格式。为了评估环境，代理模拟了用户评估环境可访问性的过程，先分析图像和用户的意图（例如，“去约会”）以识别用户可能执行的潜在任务，例如“就餐”或“如厕”。然后，代理将这些任务分解为完成它们所需的原始动作如“抓取”。对于每个任务，代理分析用户模型、任务信息和分段图像以识别并描述环境问题。至关重要的是，该系统设计为人与 AI 协作；用户可以提供对已识别问题的反馈，代理将使用这些反馈更新用户模型。

自行车管家。现有的映射工具使用距离和海拔等客观数据定义最佳自行车路线，但通常忽略了与骑车者舒适度和感知安全相关的主观品质。然而，理想的自行车路线取决于标准 GIS 数据库中找不到的因素，例如街道是否有树木、路面质量或自行车道的宽度。自行车管家是一个早期原型 Geo-Visual Agent，它通过融合来自开放街道地图的结构化数据与 SVI 的视觉分析来生成个性化的骑行路线。该系统为用户的特定配置文件（例如，初学者、专家）创建优化的路线，并允许他们对路线段进行评分，从而形成一个反馈循环，以改进他们未来行程的偏好。

7. 讨论与结论

在本文中，我们介绍了对 Geo-Visual Agents 的愿景，这是一种动态且具有对话能力的 AI 副驾，能够实时观察和推理世界。我们的设想中的代理可以回答关于视觉世界的细微问题——从盲人用户导航复杂交叉口到骑车者寻找最安全、最舒适的路线。我们的原型提供了一个初步的窗口来实现这一愿景，提供了个性化的交互体验，远远超越了当前的地图服务。

仍然存在显著的挑战，包括：(1) 动态信息综合：创建能够智能选择、融合并推理异构实时和存档数据源的代理；(2) 信任和透明度：传达不确定性和数据来源；(3) 语音用户界面：通过文本或语音简洁地表达复杂视觉信息；(4) 个性化：从用户的独特需求和偏好中学习；(5) 空间推理：准确跟踪并建模物体和场景之间的空间关系；(6) 生成的空间抽象：动态生成空间可视化以帮助理解。(7) 数据源可用性：高保真地理空间图像的可获得性，包括户外图像（例如，公园、行人专用道路的街景图像）和室内图像（例如，公共建筑物内部）以及结构化的 GIS 数据；(8) 数据新鲜度和正确性：所有技术都依赖于最新和准确的数据。

应对这些挑战将需要计算机视觉、人机交互、无障碍技术和地理空间科学等多学科的努力。我们期待在 ICCV 研讨会上与跨学科的参会者讨论我们的 Geo-Visual Agent 的愿景。

参考文献

- [1] Maneesh Agrawala and Chris Stolte. Rendering effective route maps: improving usability through generalization. In Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, page 241 – 249, New York, NY, USA, 2001. Association for Computing Machinery. 4
- [2] Dragan Ahmetovic, Roberto Manduchi, James M. Coughlan, and Sergio Mascetti. Mind your crossings: Mining gis imagery for crosswalk localization. ACM Trans. Access. Comput., 9(4), 2017. 3
- [3] Shazab Ali, Meng Xu, and Daehan Kwak. Smart roadway monitoring: Pothole detection and mapping via google street. In Internet Computing and IoT and Embedded Systems, Cyber-physical Systems, and Applications: 25th International Conference, ICOMP 2024, and 22nd International Conference, ESCS 2024, Held as Part of the World Congress in Computer Science, Computer Engineering and Applied Computing, CSCE 2024, Las Vegas, NV, USA, July 22–25, 2024, Revised Selected Papers, page 151. Springer Nature, 2025. 3
- [4] V. Antoniou and A. Skopeliti. Measures and indicators of vgi quality: An overview. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, II-3/W5:345–351, 2015. 3
- [5] Apple. Detect doors around you using Magnifier on iPhone. <https://support.apple.com/guide/iphone/detect-doors-around-you-iph35c335575/ios>, 2025. Accessed: 2025 年 8 月 23 日. 3
- [6] Hichem Boussaid, Lucrezia Tosato, Flora Weissgerber, Camille Kurtz, Laurent Wendling, and Sylvain Lobry. Visual question answering on multiple remote sensing image modalities. In Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR) Workshops, pages 2319–2328, 2025. 2
- [7] CARTO. Genai | ai-powered spatial insights. <https://carto.com/gen-ai>, 2025. Accessed: 2025 年 8 月 23 日. 2
- [8] Boyuan Chen, Zhuo Xu, Sean Kirmani, Brain Ichter, Dorsa Sadigh, Leonidas Guibas, and Fei Xia. Spatialvlm: Endowing vision-language models with spatial reasoning capabilities. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 14455–14465, 2024. 3
- [9] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4):834–848, 2018. 3
- [10] An-Chieh Cheng, Hongxu Yin, Yang Fu, Qiushan Guo, Ruihan Yang, Jan Kautz, Xiaolong Wang, and Sifei Liu. Spatialrgpt: Grounded spatial reasoning in vision-language models. In Advances in Neural Information Processing Systems, pages 135062–135093. Curran Associates, Inc., 2024. 3
- [11] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The

- cityscapes dataset for semantic urban scene understanding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. 3
- [12] Alexander Fiannaca, Ilias Apostolopoulos, and Eelke Folmer. Headlock: a wearable navigation aid that helps blind cane users traverse large open spaces. In Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility, page 19 – 26, New York, NY, USA, 2014. Association for Computing Machinery. 2, 3
- [13] Jon E. Froehlich, Anke M. Brock, Anat Caspi, João Guerreiro, Kotaro Hara, Reuben Kirkham, Johannes Schöning, and Benjamin Tannert. Grand challenges in accessible maps. *Interactions*, 26(2):78 – 81, 2019. 2
- [14] Jon E. Froehlich, Alex Fiannaca, Nimer Jaber, Victor Tsaran, and Shaun Kane. Streetviewai: Making street view accessible using context-aware multimodal ai. In The 38th Annual ACM Symposium on User Interface Software and Technology, page 22, New York, NY, USA, 2025. ACM. 1, 2, 4
- [15] Rao Fu, Jingyu Liu, Xilun Chen, Yixin Nie, and Wenhao Xiong. Scene-llm: Extending language model for 3d visual understanding and reasoning, 2024. 3
- [16] Alessandro Gambetti and Qiwei Han. Aigen-foodreview: A multimodal dataset of machine-generated restaurant reviews and images on social media, 2024. 3
- [17] Google. Google earth ai: Our state-of-the-art geospatial ai models. <https://blog.google/technology/ai/google-earth-ai/>, 2025. Accessed: 2025 年 8 月 23 日. 2
- [18] Google. Treks: Grand canyon. <https://www.google.com/maps/about/behind-the-scenes/streetview/treks/grand-canyon/>, 2025. Accessed: 2025 年 8 月 23 日. 4
- [19] Google. Celebrate 15 years of exploring your world on Street View. <https://www.google.com/streetview/anniversary/>, 2025. Accessed: 2025 年 8 月 23 日. 3
- [20] Google. Vertex ai multimodal live api. <https://cloud.google.com/vertex-ai/generative-ai/docs/multimodal-live-api>, 2025. Accessed: 2025 年 8 月 23 日. 4
- [21] Kotaro Hara, Jin Sun, Robert Moore, David Jacobs, and Jon Froehlich. Tohme: detecting curb ramps in google street view using crowdsourcing, computer vision, and machine learning. In Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology, page 189 – 204, New York, NY, USA, 2014. Association for Computing Machinery. 3
- [22] Mohammed Hassanin, Salman Khan, and Murat Tahali. Visual affordance and function understanding: A survey. *ACM Comput. Surv.*, 54(3), 2021. 3
- [23] Maryam Hosseini, Fabio Miranda, Jianzhe Lin, and Claudio T. Silva. Citysurfaces: City-scale semantic segmentation of sidewalk materials. *Sustainable Cities and Society*, 79:103630, 2022. 3
- [24] Maryam Hosseini, Andres Sevtsuk, Fabio Miranda, Roberto M. Cesar, and Claudio T. Silva. Mapping the walk: A scalable computer vision approach for generating sidewalk network datasets from aerial imagery. *Computers, Environment and Urban Systems*, 101:101950, 2023. 3
- [25] Yujun Hou and Filip Biljecki. A comprehensive framework for evaluating the quality of street view imagery. *International Journal of Applied Earth Observation and Geoinformation*, 115:103094, 2022. 3
- [26] Dingkun Hu and Jennifer Minner. Uavs and 3d city modeling to aid urban planning and historic preservation: A systematic review. *Remote Sensing*, 15(23), 2023. 3
- [27] William Huang, Xia Su, Jon E. Froehlich, and Yang Zhang. Accessibility scout: Personalized accessibility scans of built environments. In The 38th Annual ACM Symposium on User Interface Software and Technology, page 18, New York, NY, USA, 2025. ACM. 2, 4
- [28] Gaurav Jain, Basel Hindi, Zihao Zhang, Koushik Srinivasula, Mingyu Xie, Mahshid Ghasemi, Daniel Weiner, Sophie Ana Paris, Xin Yi Therese Xu, Michael Malcolm, Mehmet Kerem Turkcan, Javad Ghaderi, Zoran Kostic, Gil Zussman, and Brian A. Smith. Streetnav: Leveraging street cameras to support precise outdoor navigation for blind pedestrians. In Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology, New York, NY, USA, 2024. Association for Computing Machinery. 3
- [29] Bhargavi Janga, Gokul Prathin Asamani, Ziheng Sun, and Nicoleta Cristea. A review of practical ai for re-

- remote sensing in earth sciences. *Remote Sensing*, 15 (16), 2023. 3
- [30] Krzysztof Janowicz, Song Gao, Grant McKenzie, Yingjie Hu, and Budhendra Bhaduri. Geoai: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. *International Journal of Geographical Information Science*, 34 (4):625–636, 2020. 2
- [31] Hyejin Kim, Seula Park, and Jiyoung Kim. A study on barrier-free entrance object detection using deep learning in street view imagery. In *2024 IEEE International Conference on Big Data (BigData)*, pages 8716–8718, 2024. 3
- [32] Minchu Kulkarni, Chu Li, Jaye Jungmin Ahn, Katrina Oi Yau Ma, Zhihan Zhang, Michael Saugstad, Kevin Wu, Yochai Eisenberg, Valerie Novack, Brent Chamberlain, and Jon E. Froehlich. Busstopcv: A real-time ai assistant for labeling bus stop accessibility features in streetscape imagery. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*, New York, NY, USA, 2023. Association for Computing Machinery. 3
- [33] Jaewook Lee, Andrew D. Tjahjadi, Jiho Kim, Junpu Yu, Minji Park, Jiawen Zhang, Jon E. Froehlich, Yapeng Tian, and Yuhang Zhao. Cookar: Affordance augmentations in wearable ar to support kitchen tool interactions for people with low vision. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, New York, NY, USA, 2024. Association for Computing Machinery. 2, 3
- [34] Jaewook Lee, Jun Wang, Elizabeth Brown, Liam Chu, Sebastian S. Rodriguez, and Jon E. Froehlich. Gaze-pointar: A context-aware multimodal voice assistant for pronoun disambiguation in wearable augmented reality. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2024. Association for Computing Machinery. 2
- [35] Meiqing Li, Hao Sheng, Jeremy Irvin, Heejung Chung, Andrew Ying, Tiger Sun, Andrew Y Ng, and Daniel A Rodriguez. Marked crosswalks in us transit-oriented station areas, 2007 – 2020: A computer vision approach using street view imagery. *Environment and Planning B: Urban Analytics and City Science*, 50(2): 350–369, 2023. 3
- [36] Wenwen Li and Chia-Yu Hsu. Geoai for large-scale image analysis and machine vision: Recent progress of artificial intelligence in geography. *ISPRS International Journal of Geo-Information*, 11(7), 2022. 2
- [37] Xiaojiang Li, Carlo Ratti, and Ian Seiferling. Quantifying the shade provision of street trees in urban landscape: A case study in boston, usa, using google street view. *Landscape and Urban Planning*, 169: 81–91, 2018. 3
- [38] Yongchang Li, Li Peng, Chengwei Wu, and Jiazhen Zhang. Street view imagery (svi) in the built environment: A theoretical and systematic review. *Buildings*, 12(8), 2022. 3
- [39] Zhenlong Li, Huan Ning, Song Gao, Krzysztof Janowicz, Wenwen Li, Samantha T. Arundel, Chaowei Yang, Budhendra Bhaduri, Shaowen Wang, A-Xing Zhu, Mark Gahegan, Shashi Shekhar, Xinyue Ye, Grant McKenzie, Guido Cervone, and Michael E. Hodgson. Giscience in the era of artificial intelligence: A research agenda towards autonomous gis, 2025. 2
- [40] Alice Lo Valvo, Daniele Croce, Domenico Garlisi, Fabrizio Giuliano, Laura Giarré, and Ilenia Tinnirello. A navigation and augmented reality system for visually impaired people. *Sensors*, 21(9), 2021. 2, 3
- [41] National Aeronautics and Space Administration and U.S. Geological Survey. Landsat data access. <https://landsat.gsfc.nasa.gov/data/data-access/>, 2025. Free access to Landsat satellite imagery archive dating back to 1972. Joint NASA-USGS program providing continuous Earth observation data. 3
- [42] John S. O’Meara, Jared Hwang, Zeyu Wang, Michael Saugstad, and Jon E. Froehlich. Rampnet: A two-stage pipeline for bootstrapping curb ramp detection in streetscape images from open government metadata. In *Workshop on Vision Foundation Models and Generative AI for Accessibility: Challenges and Opportunities at ICCV 2025*. IEEE, 2025. Workshop Paper. 3
- [43] Sangkeun Park, Joohyun Kim, Rabeb Mizouni, and Uichin Lee. Motives and concerns of dashcam video sharing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, page 4758 – 4769, New York, NY, USA, 2016. Association for Computing Machinery. 3
- [44] Yurii Piadyk, Joao Rulff, Ethan Brewer, Maryam

- Hosseini, Kaan Ozbay, Murugan Sankaradas, Srimat Chakradhar, and Claudio Silva. Streetaware: A high-resolution synchronized multimodal urban scene dataset. *Sensors*, 23(7), 2023. 3
- [45] Kanchana Ranasinghe, Satya Narayan Shukla, Omid Poursaeed, Michael S. Ryoo, and Tsung-Yu Lin. Learning to localize objects improves spatial reasoning in visual-llms. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12977–12987, 2024. 3
- [46] Luís Rita, Ricky Nathvani, Miguel Peliteiro, Tudor-Codrin Bostan, Emily Muller, Esra Suel, A. Barbara Metzler, Tiago Tamagusko, and Adelino Ferreira. Using deep learning and google street view imagery to assess and improve cyclist safety in london. *Sustainability*, 15(13), 2023. 3
- [47] Joao Rulff, Giancarlo Pereira, Maryam Hosseini, Marcos Lage, and Claudio Silva. Towards data-informed interventions: Opportunities and challenges of street-level multimodal sensing, 2024. 3
- [48] Manaswi Saha, Michael Saugstad, Hanuma Teja Madali, Aileen Zeng, Ryan Holland, Steven Bower, Aditya Dash, Sage Chen, Anthony Li, Kotaro Hara, and Jon Froehlich. Project sidewalk: A web-based crowdsourcing tool for collecting sidewalk accessibility data at scale. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, page 1 – 14, New York, NY, USA, 2019. Association for Computing Machinery. 3
- [49] Hunsoo Song, Joshua Carpenter, Jon E. Froehlich, and Jinha Jung. Accessible area mapper for inclusive and sustainable urban mobility: A preliminary investigation of airborne point clouds for pathway analysis. In *1st ACM SIGSPATIAL Workshop on Sustainable Mobility (SuMob 2023)*, 2023. 3
- [50] Xia Su, Han Zhang, Kaiming Cheng, Jaewook Lee, Qiaochu Liu, Wyatt Olson, and Jon E. Froehlich. Rassar: Room accessibility and safety scanning in augmented reality. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2024. Association for Computing Machinery. 2, 3
- [51] Xia Su, Ruiqi Chen, Jingwei Ma, Chu Li, and Jon E. Froehlich. Flymethrough: Human-ai collaborative 3d indoor mapping with commodity drones. In *The 38th Annual ACM Symposium on User Interface Software and Technology*, page 14, New York, NY, USA, 2025. ACM. 3
- [52] SuperMap. Ai gis. <https://www.supermap.com/en-us/key-technologies/ai-gis.html>, 2025. Accessed: 2025 年 8 月 23 日. 2
- [53] Eric K. Tokuda, Roberto M. Cesar, and Claudio T. Silva. Quantifying the presence of graffiti in urban environments. In *2019 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 1–4, 2019. 3
- [54] U.S. Geological Survey. Earthexplorer. <https://earthexplorer.usgs.gov/>, 2025. Query and order satellite images, aerial photographs, and cartographic products. Provides access to over 40 years of Landsat data and various aerial photography collections. 3
- [55] Zeyu Wang, Koichi Ito, and Filip Biljecki. Assessing the equity and evolution of urban visual perceptual quality with time series street view imagery. 145: 104704. 3
- [56] Zeyu Wang, Yingchao Jian, Adam Visokay, Don MacKenzie, and Jon E. Froehlich. Street view for whom? an initial examination of google street view’s urban coverage and socioeconomic indicators in the us. Under review, 2025. Submitted for review. 3
- [57] Chris Yoon, Ryan Louie, Jeremy Ryan, MinhKhang Vu, Hyegi Bang, William Derksen, and Paul Ruvolo. Leveraging augmented reality to create apps for people with visual disabilities: A case study in indoor navigation. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, page 210 – 221, New York, NY, USA, 2019. Association for Computing Machinery. 2, 3
- [58] Aziza Zhanabatyrova, Clayton Frederick Souza Leite, and Yu Xiao. Automatic map update using dashcam videos. *IEEE Internet of Things Journal*, 10(13): 11825–11843, 2023. 3
- [59] Guiming Zhang and A-Xing Zhu. The representativeness and spatial bias of volunteered geographic information: a review. *Annals of GIS*, 24(3):151–162, 2018. 3
- [60] Lefei Zhang and Liangpei Zhang. Artificial intelligence for remote sensing data analysis: A review of challenges and opportunities. *IEEE Geoscience and Remote Sensing Magazine*, 10(2):270–294, 2022. 3

- [61] Mengxia Zhang and Lan Luo. Can consumer-posted photos serve as a leading indicator of restaurant survival? evidence from yelp. *Management Science*, 69(1):25–50, 2023. [3](#)
- [62] Meimei Zhang, Fang Chen, and Bin Li. Multistep question-driven visual question answering for remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–12, 2023. [2](#)
- [63] Yuhang Zhao, Elizabeth Kupferstein, Brenda Veronica Castro, Steven Feiner, and Shiri Azenkot. Designing ar visualizations to facilitate stair navigation for people with low vision. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, page 387 – 402, New York, NY, USA, 2019. Association for Computing Machinery. [2](#), [3](#)
- [64] Shengyuan Zou and Le Wang. Detecting individual abandoned houses from google street view: A hierarchical deep learning approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 175:298–310, 2021. [3](#)