引文格式:刘瑜,陈琳,张帆,等.刍议空间智能与地理空间智能[J].武汉大学学报(信息科学版),2026,51(10):2206-2214.DOI:10.13203/j.whugis20260391

Citation:LIU Yu,CHEN Lin,ZHANG Fan,et al.On Spatial Intelligence and Geo-spatial Intelligence[J].Geomatics and Information Science of Wuhan University,2026,51(10):2206-2214.DOI:10.13203/j.whugis20260391

刍议空间智能与地理空间智能

刘 瑜[1] 陈 琳[2] 张 帆[1] 董 磊[1] 许立言[3] 乐 阳[4] 董卫华[5]
李海峰[6] 吴志峰[7] 柴彦威[8] 孟小峰[9] 林 珲[10] 陈 军[11]

1 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,北京,100871
2 美国东北大学网络科学研究所,波士顿,MA 02115
3 北京大学建筑与景观设计学院,北京,100000
4 香港科技大学(广州)城市治理与设计学域,广东 广州,511453
5 北京师范大学地理科学学部卫星应用前沿交叉研究院,北京,100875
6 中南大学人工智能学院,湖南 长沙,410083
7 广州大学地理科学与遥感学院,广东 广州,510006
8 北京大学城市与环境科学学院,北京,100871
9 中国人民大学信息学院,北京,100872
10 江西师范大学地理与环境学院,江西 南昌,330027
11 国家基础地理信息中心,北京,100830

摘 要:人工智能技术正在改变人类认知与改造空间的方式,当前形成两类内涵迥异的空间智能技术。一类来自计算机与人机交互领域的空间智能,面向机器的具身交互;另一类是地理学发展而来的地理空间智能,侧重宏观尺度地表系统分析模拟。两类技术路径的割裂差异,根源在于其依托的空间本体不同,即微观的具身空间与宏观的离身空间,二者在本体论、认识论层面存在第一性的差异。具身空间立足人体尺度,强调主体体验与社会建构,对智能体的情境感知、实时交互能力提出高要求;离身空间强调客观系统观测,面向区域乃至全球尺度,重点解决海量时空数据处理与复杂过程模拟问题。现有研究往往割裂两类范式,较少讨论二者融合的理论基础。基于此,辨析了具身/离身空间的核心差异,对比两套范式对应的技术体系,提出地理空间数字孪生作为统一集成平台,以多尺度空间智能体实现微观与宏观能力贯通,以期为智慧城市、气候变化等复杂人地系统问题提供新的理论视角与实现路径。

关键词:空间智能;地理空间智能;具身空间;离身空间;本体论;数字孪生;多尺度空间智能体

中图分类号:P208 **文献标识码**:A **收稿日期**:2026-10-06

DOI:10.13203/j.whugis20260391 **文章编号**:1671-8860(2026)10-2206-09

On Spatial Intelligence and Geo-spatial Intelligence

LIU Yu[1] *CHEN Lin*[2] *ZHANG Fan*[1] *DONG Lei*[1] *XU Liyan*[3] *YUE Yang*[4]
DONG Weihua[5] *LI Haifeng*[6] *WU Zhifeng*[7] *CHAI Yanwei*[8]
MENG Xiaofeng[9] *LIN Hui*[10] *CHEN Jun*[11]

1 Institute of Remote Sensing and Geographical Information Systems, School of Earth and Space Sciences, Peking University, Beijing 100871, China
2 Network Science Institute, Northeastern University, Boston MA 02115, USA
3 College of Architecture and Landscape, Peking University, Beijing 100000, China
4 Thrust of Urban Governance and Design, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou 511453, China
5 Advanced Interdisciplinary Institute of Satellite Applications, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
6 School Artificial Intelligence, Central South University, Changsha 410083, China
7 School of Geography and Remote Sensing, Guangzhou University, Guangzhou, 510006, China
8 College of Urban and Environmental Sciences, Peking University, Beijing 100871, China
9 School of Information, Renmin University of China, Beijing 100872, China

基金项目:国家自然科学基金(42430106)。

第一作者:刘瑜,博士,教授,主要从事地理信息科学方向的研究。liuyu@urban.pku.edu.cn

10　School of Geography and Environment, Jiangxi Normal University, Nanchang 330027, China
11　National Geomatics Center of China, Beijing 100830, China

Abstract: Artificial intelligence is transforming how humans perceive and interact with space, giving rise to two technically distinct paradigms of spatial intelligence. Spatial intelligence originates from computer science and human-computer interaction and targets embodied machine-environment interaction. By contrast, geospatial intelligence is rooted in geography and focuses on the analysis and simulation of Earth's surface systems at macro scales. The fragmentation between these two technical approaches stems from divergent spatial ontologies: micro-scale embodied space versus macro-scale disembodied space, which exhibit systematic disparities at both ontological and epistemological levels. Centered on human-body scale, embodied space highlights subjective experience and social construction, imposing stringent requirements on machines for situational awareness and real-time interaction. Disembodied space, however, prioritizes objective systematic observation across regional to global extents, with core challenges lying in massive spatiotemporal data processing and complex-process simulation. Existing studies often treat these two paradigms in isolation, while the theoretical foundation for their integration remains underexplored. Against this backdrop, this paper identifies the fundamental distinctions between embodied and disembodied spaces and compares their corresponding technical stacks. We propose digital twins as a unified integrative infrastructure and adopt multi-scale spatial agents to bridge micro- and macro-level capabilities. This work aims to deliver novel theoretical insights and implementable pathways for tackling complex human-environment challenges such as smart-city governance and climate-change adaptation.

Key words: spatial intelligence; geospatial intelligence; embodied space; disembodied space; ontology; digital twin; multi-scale spatial agent

人工智能(artificial intelligence,AI)技术的快速发展正在把“空间”从一个背景性的存在变成一种可被计算、交互甚至生成的媒介[1],学术界与工业界发展了两种对空间智能的差异化理解与实践。在计算机科学、机器人学与认知科学领域,学者们倡导的空间智能的核心是赋予AI代理在三维物理世界中感知、认知、预测和行动的能力,其应用聚焦于自动驾驶、增强现实与具身机器人[2-3];而在地理信息科学和区域科学领域,经过数十年发展的地理空间分析方法正借助AI实现能力升级,实现对地球表层系统更深度的监测、建模与模拟,服务于从城市规划到全球变化的广泛应用,这方面的技术,通常被称为地理空间智能或时空智能[4-6]。

对比空间智能和地理空间智能的技术目标和实现路径,可发现二者存在较大差异。一个面向小尺度空间内自由导航的服务机器人,与一个模拟城市演化的生成式模型,遵循着截然不同的技术逻辑。这种差异常常被简单归结为应用场景的不同,但相关的理论基础尚未得到充分探讨。我们认为,空间智能与地理空间智能的技术路径分异,是二者所基于的在哲学与科学传统上不同的空间观所决定的。本文将这两种智能对应的空间分别称为具身空间(embodied space)与离身空间(disembodied space),对于前者,嵌入其中的个体通过体验和互动认知该空间的要素配置和空间结构;对于后者,由于其尺度的宏观特性,相应认知方式往往是置身其外的观察与归纳建模。其中,置身其外的观察可以是自顶向下的俯瞰,遥感无疑是这种观察视角的技术实现;而归纳建模过程源于个体具身知识的累积以及群体具身知识的整合[7],例如,个体到达一个陌生环境(如一个大学校园)后,需要在其中生活一段时间,才能形成俯瞰视角的、较为详细准确的认知地图。

这两个空间的本体论差异,奠定了各自技术范式的基础。其中具身空间概念源于人文地理学、哲学现象学与建筑现象学,它是一种以身体尺度、主体体验和社会实践为核心的生活空间。与之相对,离身空间则秉承了自然地理学与空间分析的传统,将空间视为一个可被测量、分析与建模的客观系统[8]。这两种空间观念不仅规定了不同的研究尺度,也蕴含了不同的主客体关系(嵌入参与和外在观察)和价值评判标准(意义通达和规律符合)。本文试图提供一个系统性的理论框架辨析上述差异,并探索有意义的融合路径。首先从本体论与空间认知科学的角度,分析具身空间与离身空间的构成性差异,并阐述这些

差异对于AI技术带来的不同挑战;然后比较空间智能与地理空间智能各自的技术实现路径以及核心技术栈;在此基础上论证在数字孪生技术的框架下,两种空间智能范式从分立走向融合的必然性与可行性,提出多尺度空间智能体作为核心融合愿景;最后对未来研究进行展望,并讨论其伦理意涵。

1 具身空间与离身空间的本体论比较

对空间概念的辨析,是理解相关空间智能技术的前提。本节将构建一个理论框架,阐明具身空间与离身空间在本体论、对应的认知模式及衍生的技术挑战的不同。

1.1 具身空间

具身空间是一个以人体尺度为基准、可通过直接感知并整体把握的空间,如一个房间或者街角。在空间内,认知是通过身体与环境的多通道交互,包括视觉、听觉、触觉等完成的,而其要素配置直接影响个体的体验和行为,例如,空间的边界、物体的布局、光线的明暗都直接作用于感官:过于狭窄封闭的空间让人想尽快离开,开阔明亮的空间则让人愿意停留。感知由此塑造了人在空间中的行动。

从本体论视角审视,具身空间的空间尺度较小,其范围由人类身体或智能体的感知与行动能力所界定。例如,一个智能服务机器人在家庭环境中作业,它所理解的空间是通过激光雷达与视觉传感器"感知"到的、与它的移动能力和机械臂操作范围直接相关的走廊宽度、房门开口以及沙发与茶几之间的可通行间隙。空间的意义正是在这种身体尺度下的移动与交互中展开。进而,具身空间的意义建构性与社会性使得其价值不仅是物理属性的固有产物,也通过人的实践与叙事被持续赋予。一个具体的例子是城市广场,它在市民日常穿行时是通道,在周末市集时可以是市场,在政治集会时则成为表达诉求的论坛,其功能、规则和相应的认知随社会活动而动态重构。最后,具身空间具有情境依赖特征,即其特质依赖于具体的情境。例如,对于自动驾驶汽车而言,道路在普通状态或积雪覆盖条件下,其几何形态未变,但构成了物理上截然不同的驾驶空间;若遇到临时交通管制或大型活动散场的人流,则又叠加了复杂的社会规则,这要求系统必须理解情境的独特约束。

在认识论层面,对此类空间的理解紧密依赖于"具身认知"与"情境认知"的理论框架。如拉考夫与约翰逊在《肉身哲学》中所主张的,心智本质上是具身的,思维大部分是无意识的,抽象概念在很大程度上是隐喻性的[9]。知识的获得是通过身体与环境的实时互动、在"感知-行动"循环中涌现。认知科学中关于"路线知识(route knowledge)"(依赖于连续的地标转向)与"调查知识(survey knowledge)"(依赖于鸟瞰式的认知地图)的区分,正是这种从具身经验到抽象表征的连续的空间认知谱系的体现[7,10]。因此,具身空间并非古典几何学中均质、各向同性的容器,其思想渊源根植于现象学哲学与人文地理学对于生活经验的回归。它呼应了胡塞尔对"生活世界"作为一切科学与意义基础领域的阐述、海德格尔对"在世存在"中人与场所先于反思的融合状态的分析[11-12]。在人文地理学中,段义孚在其经典著作《空间与地方》中提出的地方(place)概念,将空间阐释为依附了人类体验、情感、记忆与意义的地理单元,为理解具身空间的社会文化维度奠定了基石[13]。

对AI系统而言,具身空间下的智能行为需要直面框架问题与符号接地问题的固有困境,也就是如何让智能体习得难以完全形式化的物理常识与社会情境认知[14]。智能体不能仅完成物体类别识别,还需要读懂物体在特定场景下的功能内涵与社会意义:一个杯子既可以作为饮水器具,也可以充当纪念品,还可能临时被当作笔筒,场景改变,物体角色随之变化。与此同时,系统需要内化物理世界运行规则,例如物体刚性、流体运动规律,还要理解排队、人际距离这类不成文社会规范。真实环境是非结构化且动态变化的,在感知常识基础之上,系统还需要具备高鲁棒性的实时决策能力,输出安全且符合社会预期的行动方案,如服务机器人在人流密集餐厅的避障导航,就是该类问题的典型实例。

1.2 离身空间

与具身空间相对,离身地理空间的范围更大,如一个城市或区域,这使得个体难以一次性获得其整体认知。承袭自牛顿的绝对时空观,离身地理空间的认知范式主导了近代科学革命以来的地理学思想,其核心在于将空间视为独立于人类意识的客观实(存)在,并认为其中蕴含的内在规律可通过科学方法被发现与建模[15]。

离身空间具有如下3个方面的特征:首先是尺度的宏观特点。离身空间的分析尺度远超个体尺

度的直接感知范围,从而要求一种外在的、俯瞰式的全局俯瞰视角[16-17]。在离身空间中探讨的人地关系也侧重于抽象的、统计意义上的群体,而非具体的、体验中的个体。其次,离身空间的表达通常是客观的,即该空间及其中的现象可通过标准化工具进行精确测量与表征。空间要素的配置导致了不同地理空间效应的存在[18],如空间异质性与空间依赖性,后者是地理学第一定律所概括的基本事实[19]。一些特定的离身地理空间,如城市,不仅具有客观的物理结构,也受到人类行为和社会活动的持续塑造,具有社会建构的属性。最后,该空间具有系统性与层级性,即该空间是一个由自然与人文圈层紧密耦合的复杂巨系统,强调要素间的物质、能量与信息流动,以及多尺度嵌套的结构与演化规律。在认识论上,对宏观地理空间的理解依赖于离身认知与抽象表征。它通过几何化的要素、地图以及统计模型等符号系统,将无法直接体验的大尺度空间的地理要素和现象压缩并抽象为可计算的模型,这对应于认知科学中所述的高阶调查知识[7,10,20]。

对AI而言,离身空间尺度下的核心挑战在于跨越复杂性鸿沟与驾驭尺度效应,并从多源、多尺度、高维度且非结构化的海量时空数据中提取有价值的模式,并模拟非线性地理过程,进而支持宏观的科学决策[4]。这具体包括:(1)构建可扩展的计算框架,以分布式算法与强大算力处理海量时空数据;(2)实现领域知识嵌入,将地理规律和空间效应融入AI模型,约束其输出符合地理规律[17];(3)发展复杂系统仿真与涌现预测能力,对城市演化、流行病传播、气候变化等宏观过程进行高置信度的动态模拟与情景推演,以支持科学决策[21]。

1.3　两个空间的比较和联系

表1比较了具身空间和离身空间在尺度、观察视角、人与之交互的途径等方面的差异。

表1　具身空间与离身空间的比较

Table 1　Comparison of Embodied Space and Disembodied Space

比较项	具身空间	离身空间
哲学渊源	现象学、人文主义地理学、存在主义	实证主义、空间分析学派、系统论
空间本质	通用的、交互的三维环境,可以是物理世界,也可以是虚拟世界	特指地球表层及其相关空间,具有明确的地理坐标和区域特性
本体论核心	被体验、被建构的“场所”	被测量、被分析的“客观系统”
核心尺度	微观(身体、建筑、街道)	宏观(区域、国家、全球)
视角	平视——人本视角	俯视——全局俯瞰视角
驱动规律	物理规律、社会规律	地理规律
认识论模式	具身认知、在场认知(路线视角)	离身认知、抽象认知(测绘视角)
主导方法论	定性研究、深度描述	定量分析、模型模拟、GIS与遥感
对AI的挑战	情境理解、常识推理、具身交互	海量数据处理、系统建模、尺度转换

这两个尺度的空间特征也体现在不同地理学科的研究范式构建中。自然地理学更多采取了离身空间的视角;传统人文地理学,尤其是在计量革命影响下的空间分析学派,主要采纳了离身空间的视角;而自人本主义地理学兴起以来,对地方感以及身体体验的重视,使得具身空间的视角得到了复兴与发展。这两个空间不是截然对立的,可以认为从微观尺幅到空间尺度,形成了一个渐变的“谱系”。这种连续变化既体现在空间尺度上,也体现在观察视角(从沉浸的第一人称视角到抽离的第三人称或全局俯瞰视角),以及表征方式(从体验性描述到可计算模型)的渐变上,例如中尺度空间,如一个社区乃至城市,兼具了两个空间的特点。这种连续变化的事实使得融合两个空间成为可能。如MONTELLO[7]指出的个体空间认知的发展过程,为两种空间观的融合提供了心理学方面的支持。相关脑影像研究进一步揭示,认知地图在抽象层级与空间尺度上呈现梯度化表征差异,不同尺度空间信息在大脑中形成层级整合[22]。通过大脑中认知地图的构建,个体将局部的具身空间体验逐步整合、抽象为一种离身的、结构化的心理表征。在该过程中,离身空间不是与体验无关的绝对客观领域,而往往源于多主体具身经验的交互与沉淀。这为人们在技术层面设计多尺度空间智能提供了重要启示,即智能体既需要具身交互以获取一手的路线知识,也需要高阶认知模块来整合信息,形成可用于全局推理的鸟瞰模型。近年来,随着图像理解技术的发展,基于具身视角的街景影像,汇总得到离身视角的宏观城市指标的空间分布,也可以视为联系这两个空间的技术路径之一[23]。

2 空间智能与地理空间智能的技术对比

上述两个尺度的空间对应的智能技术侧重解决不同的应用问题,其中具身空间的智能技术核心是支持智能体(包括具身机器人和真实的人)实现高效、自然、安全且符合意图的感知、理解和响应;而针对离身空间的地理空间智能,则以理解、模拟、预测并优化宏观人地系统的结构、过程与规律为目标,从而支持战略决策和系统治理,如表2所示。由于这两个空间的不同理论预设与挑战,空间智能与地理空间智能演化出了两套相对独立的技术体系。

表2 空间智能与地理空间智能的技术栈深度对比

Table 2 An In-Depth Comparison of the Technology Stacks Between Spatial Intelligence and Geospatial Intelligence

比较项	空间智能	地理空间智能
核心任务	环境感知、三维理解、行动规划	数据集成、模式发现、过程模拟
输入数据	实时流数据、RGB-D图像、点云	遥感影像、矢量地图、时空轨迹
核心技术	深度学习、强化学习、神经辐射场	遥感影像分析、时空数据挖掘、图神经网络
主要输出	运动指令、交互动作、生成内容	评估报告、模拟预测结果、决策方案

2.1 空间智能技术栈:具身环境中的感知-行动回路

空间智能的核心目标在于为不同类型智能体构建一个能够适应复杂物理世界的、高效且鲁棒的“感知-行动”回路,旨在模拟乃至超越生物体基于感知进行实时决策与行动的环境适应能力。其技术体系的演进涵盖了从环境感知、三维重建、空间推理到生成交互的流程。

环境感知与三维重建是构建该回路的起点,其已从二维图像识别迈向对三维场景的几何与语义一体化理解。通过融合激光雷达、深度相机等多传感器数据,智能体得以同步获取环境的精确几何信息与丰富的语义信息,其中实例分割与全景分割技术是实现像素级场景解析的基础。更为深刻的进步体现在场景表示范式的转变:传统显式表示正被神经辐射场(neural radiance fields,NeRF)等隐式神经表示技术所补充与超越,NeRF能够从稀疏的二维观测中学习连续的三维场景表示,实现逼真的新视角合成,提升对场景连续性与隐式结构知识的捕获能力[24]。空间推理与决策规划是连接感知与行动的中枢,其核心挑战在于将原始感知数据转化为可支持行动的结构化知识,并将物理常识(如物体的持久性、重力)与社会规则以可解释、可操作的方式注入模型[25]。强化学习是解决序列决策问题的核心,但在真实世界中训练通常成本高昂,因此开发高保真的物理仿真环境,支持智能体能在海量、多样的虚拟场景中通过试错安全、高效地学习复杂技能,是可行的解决途径[26]。最后,智能体需完成从理解世界到生成与具身交互的跨越,具身AI强调智能产生于与环境的持续互动之中,其任务要求智能体通过主动导航与操作来获取信息并解决问题,从而闭合感知-行动回路[27]。同时,生成式AI在三维内容生成与编辑中的应用,要求模型具备对空间结构、功能和物理属性的深刻理解,这进一步推动了空间智能从分析到创造的延伸[28]。

2.2 地理空间智能技术栈:宏观空间的模拟-推演能力

地理空间智能面向大尺度的离身地理空间,其核心技术建立在地理信息系统、遥感科学与空间统计等基础之上,其智能化则贯穿于从数据获取、分析建模、模拟优化到决策支持的全链条。在数据智能获取与预处理层面,基于深度学习的变化检测、地物分类与目标提取技术,已实现对大范围地表覆盖动态的自动化监测,而视觉Transformer等架构的引入,则显著提升了处理高分辨率遥感影像的语义分割精度与效率[29]。平行于遥感手段,社会感知则为汇总个体粒度的微观行为,得到离身的宏观观测提供了支持。在算力方面,以Google Earth Engine为代表的时空大数据云平台提供了全球尺度的地理数据与并行计算能力,使得以往受限于算力与数据访问的大规模分析成为可能[30]。在时空数据挖掘与知识发现层面,不同类型的深度学习模型被广泛用于交通流和犯罪热点等时空分布的建模与预测,其中图神经网络尤其擅长处理道路网络、社交网络等具有显式空间关系结构的数据[31-32],而以AlphaEarth为代表的表示学习技术,使得地理空间智能技术更好地服务于可推理、可泛化的地球智能系统[5]。

在地理过程模拟与系统集成层面,纯数据驱动模型的局限性催生了与地理机理融合的新范式。

通过发展物理信息神经网络或将过程模型作为约束引导AI训练,可显著提升模型在预测外推时的物理一致性与泛化能力[33]。而构建地理基础模型,将地理实体、属性及其复杂关系进行结构化表征,为实现具有逻辑推理能力的认知级智能提供支撑[4]。

然而,地理空间智能仍面临如下的核心挑战。第一,由于地理空间的特殊性所导致的各种效应,需要在AI模型中加以显式表达,尤其需要解决地理空间异质性带来的泛化和可解释性。第二是模型鸿沟,即如何将复杂、模糊且难以量化的人文地理过程(如政策传导、文化偏好)有效嵌入计算模型。第三是对宏观格局的过度关注易导致生态学谬误,即忽视微观个体的异质性行为,从而错误地将群体层面的统计关系和预测结果推论至个体[34]。

3　数字孪生框架下的多尺度空间智能体

面对复杂现实世界中的各类应用,如城市治理和应急响应,由于问题本身的多时空尺度特性,需要集成面向具身空间和离身空间的智能技术加以解决。如前所述,这两个空间在理论上存在融合的可行性,其核心在于实现人本视角和全域观测(全局)视角的变换,以及在个体和群体之间建立统一的表达。数字孪生技术为实现这两个范式的融合提供了理想框架,其目标是构建能够贯通微观和宏观空间的多尺度智能体。

3.1　融合的必然性:复杂人地系统问题的内在要求

宏观地理过程与微观个体行为之间存在着内在的联系。一个宏观地理格局与过程,其驱动力往往源于微观尺度上无数个体的行为及其与局部环境的互动。反之,个体的决策与行动又时刻受到宏观政策、市场力量与社会环境的约束与塑造。这种宏观与微观之间的关联,需要构建双向的分析框架与干预工具。具身空间智能与离身地理空间智能的融合,正是回应这一理论需求与实践挑战的必然方向。

例如,为了实现城市碳中和的宏观目标,需要地理空间智能在离身尺度上对能源结构、绿地系统与交通网络进行系统性模拟与优化[35]。然而,宏观规划的节能降碳目标,最终必须通过微观个体尺度的具体行为来实现:每一栋建筑的暖通空调与照明能耗调控,以及每一位居民的日常出行决策,这亟需空间智能技术增强楼宇自动化系统,并通过对个体低碳行为的感知、引导与激励来落实宏观策略。如果仅依靠宏观地理空间智能开展模拟,缺少对建筑、居民等微观单元行为的精细刻画,规划模拟结果很容易脱离实际;反过来,只关注局部空间的智能调控,缺少全局层面的系统约束,个体层面的低碳行为也难以汇聚成城市尺度的减排成效。因此,两类技术的融合,是搭建感知-模拟-调控闭环、实现城市碳排放精细化管控的必要条件[36-37]。

除了常态化管理决策支持外,应对突发危机的应急管理,如突发公共卫生事件的防控,也需要宏观规划与微观行动之间的集成,从而凸显了衔接这两种空间智能范式的紧迫性。在理论层面,这是对地理学中“地方”与“空间”以及“个体”与“结构”经典二元论在数字时代的方法论回应。通过构建能够贯通宏观与微观的数字孪生环境与多尺度智能体,可以为理解并塑造复杂的人地系统提供高度协同并具有可操作性的方案。

3.2　数字孪生和多尺度智能体:一个融合空间智能框架

数字孪生系统是一个动态、可计算、多时空尺度集成的虚拟空间系统,它为桥接具身空间与离身空间之间的理论分野与实践鸿沟提供了统一的虚拟实验场与协同操作平台[38-39],在地理空间的语境下,LIN等[40]称之为虚拟地理环境。其作为融合框架的核心价值是实现多源异构数据的统一融合与同步,即能够在同一时空基准下,无缝集成宏观的遥感影像、社会经济统计数据与GIS图层,中观的建筑信息模型、倾斜摄影实景三维模型,以及微观的物联网传感器实时流数据、激光点云和监控视频流。例如,在智慧交通管理中,数字孪生可同时接入宏观尺度的区域路网流量、中观尺度的路口信控状态及微观的单个车辆轨迹,从而为从城市战略规划到实时信号优化的多尺度分析构建了权威、一致的单一事实来源[41]。

在数字孪生系统中,需要实现多尺度模型的耦合仿真与交互推演,其中的关键在于解决时间尺度的异构性挑战。面向离身空间的宏观地理过程模拟,如城市形态演化和气候变化影响,常以年、十年甚至世纪为步长,关注长期趋势与稳态规律,侧重战略层的决策;而微观尺度的行为仿真,如交通流和建筑能耗,则需以秒、分钟或日为分辨率,捕捉瞬时动态与短期反馈,主要提供战术层的支持。如何有效耦合这些差异巨大的时间尺度,是融合空间智能面临的挑战[42]。数字孪生系统通过多时间尺度建模方法与异步耦合

技术协调这种差异。具体而言,宏观模型在进行一定步长的推演后,将其输出结果,如更新的土地利用政策和碳排放总量目标,参数化为微观仿真所需的初始条件与边界约束;微观模型则在给定的约束场景下,以高时间分辨率运行,并将其大量个体行为的模拟结果通过聚合与统计分析,如计算区域平均车速和累计能耗,反馈至宏观模型,用于校准宏观参数或触发宏观状态转移。例如,在模拟城市洪涝应急时,宏观水文模型预测的淹没范围与深度可作为输入,驱动微观人群疏散模型;疏散模型中人群的恐慌传播与路径选择导致的拥堵模式又可反馈回宏观模型,修正救援物资投放策略。最终,通过融合多尺度空间智能,从而实现宏观-微观在时空双重维度上的一体化模拟与方案评估[39]。值得指出的是,这种多时空尺度耦合有助于理解地理空间复杂性,即微观具身行为除了受物理约束外,还受心理学、社会博弈的影响,这些行为聚集成宏观的地理现象(如交通拥堵、城市贫民窟扩张),呈现出非线性的复杂涌现,而在数字孪生平台支持下,可以更支持理解地理复杂性产生的机理。

基于上述数据与模型的集成,融合空间智能的最终形态与核心载体是具备尺度弹性的多尺度空间智能体。它是一个分层决策架构,包含高层的"战略规划器"和低层的"战术执行器"。"战略规划器"基于地理空间知识与宏观模型推演结果,在离身空间中进行长时间尺度的任务分解与全局路径规划(如从A区到B区运送应急物资);"战术执行器"则嵌入具身仿真环境,基于实时传感器数据,在平滑的小尺度空间中完成局部指令执行(如绕过前方积水路段)。二者通过共享的时空表征与价值函数对齐进行协同:规划器将全局目标分解为一系列可执行的子目标,并传递给执行器;执行器在实现子目标过程中,将遇到的实时障碍与机遇反馈给规划器,触发规划的重新评估与调整。实现此愿景需突破跨尺度表示学习、层次化强化学习以及跨尺度因果推断等关键技术。这些技术的融合与发展,特别是如何实现智能体在不同层次上的有效协作与学习[43],可推动空间智能从分立走向协同,最终为理解、模拟与干预复杂人地系统提供具有全局视角与局部行动能力的统一工具(图1)。

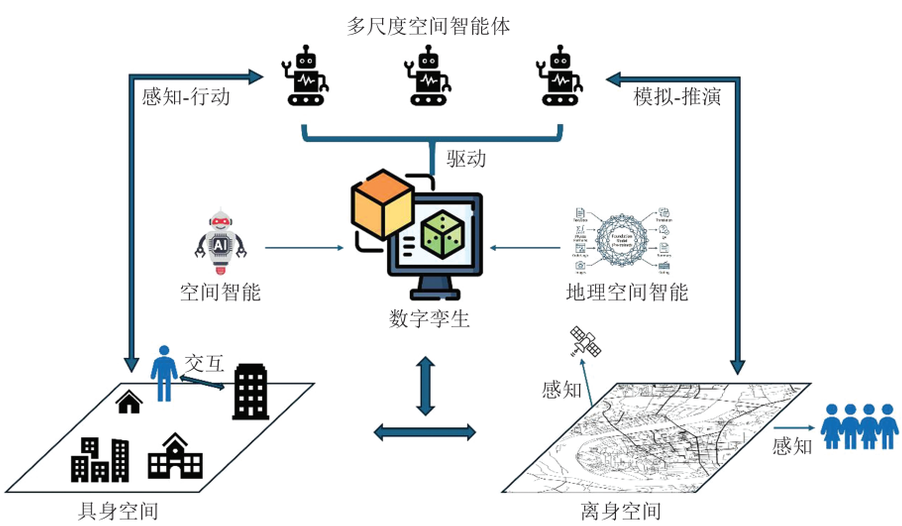


图1 融合空间智能和地理空间智能的数字孪生和多尺度空间智能体技术框架

Fig. 1 Technical Framework of Digital Twin and Multi-scale Spatial Agent that Integrates Spatial Intelligence and Geospatial Intelligence

4 结论和展望

本文比较了空间智能与地理空间智能的技术路径,其根源在于二者所基于的具身空间与离身空间在本体论与认识论上的差异。前者根植于现象学传统,致力于在微观、具身的环境中实现基于个体体验与交互的情境化理解与行动;后者承袭了绝对空间观与分析传统,专注于在宏观的人地耦合系统中对复杂地理过程进行规律性的建模与模拟。而可持续城市、公共卫生这类当代问题大多是宏观-微观交织的,往往同时需要两种能力,单靠一种范式都不够。数字孪生技术为实现二者的有机融合提供了理想的集成框架,它通过构建一个动态的多尺度虚拟空间底座,实现

从宏观数据到微观传感器数据的统一融合，并支持了异质时间尺度模型的耦合仿真。在此基础上，多尺度空间智能体是实现这一融合的载体，其具备在宏观战略规划与微观战术执行间自适应切换的多尺度认知能力。

展望未来，实现多尺度融合空间智能，在理论上需依托地理学、认知科学、计算机科学与复杂系统理论的交叉研究，构建能够统一刻画微观具身交互到宏观地理过程的空间计算理论，为跨尺度模型耦合、时空数据统一表征与智能体的尺度认知迁移筑牢理论根基。进而，需要攻关兼容遥感影像、地图符号、三维点云、视频流与自然语言的多模态空间领域基础模型，兼顾跨尺度语义生成理解能力与精度、计算开销之间的权衡[44]。在应用层面，要面向城市治理、双碳实施、应急管理等真实复杂场景开展落地验证，检验从宏观策略生成、微观具身执行反馈再到宏观模型迭代修正的技术流程。最后，还必须正视融合智能带来的社会风险，围绕算法公平、时空行为数据隐私、人机权责划分、公众可接受度等议题完善包容性治理机制，保障技术真正服务于人地系统可持续发展。

参　考　文　献

[1] FENG J, ZENG J W, LONG Q Y, et al. A Survey of Large Language Model-Powered Spatial Intelligence Across Scales: Advances in Embodied Agents, Smart Cities, and Earth Science[PP/OL]. arXiv (2025-04-14)[2026-09-16]. https://arxiv.org/abs/2504.09848.

[2] YOU H X, ZHOU T Y, ZHU Q, et al. Embodied AI for Dexterity-Capable Construction Robots: DEXBOT Framework[J]. *Advanced Engineering Informatics*, 2024, 62: 102572.

[3] LI Feifei. From Words to Worlds: Spatial Intelligence Is AI's Next Frontier[EB/OL]. 2025. https://drfeifei.substack.com/p/from-words-to-worlds-spatial-intelligence.

[4] JANOWICZ K, GAO S, MCKENZIE G, et al. GeoAI: Spatially Explicit Artificial Intelligence Techniques for Geographic Knowledge Discovery and Beyond[J]. *International Journal of Geographical Information Science*, 2020, 34(4): 625-636.

[5] BROWN C F, KAZMIERSKI M R, PASQUARELLA V J, et al. AlphaEarth Foundations: An Embedding Field Model for Accurate and Efficient Global Mapping from Sparse Label Data[EB/OL]. V2. arXiv:(2025-09-08)[2026-09-16]. https://arxiv.org/abs/2507.22291.

[6] 李德仁，郭昊南，金文杰. 时空智能学：理论与应用[J]. 武汉大学学报（信息科学版），2026，51(9)：1755-1764.
LI Deren, GUO Haonan, JIN Wenjie. Spatiotemporal Intelligence: Theory and Applications[J]. Geomatics and Information Science of Wuhan University, 2026, 51(9): 1755-1764.

[7] MONTELLO D R. Scale and Multiple Psychologies of Space[M]//Spatial Information Theory: A Theoretical Basis for GIS. Berlin, Heidelberg: Springer, 1993: 312-321.

[8] MAZÚR E, URBÁNEK J. Space in Geography[J]. *GeoJournal*, 1983, 7(2): 139-143.

[9] LAKOFF G, JOHNSON M. Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought[M]. New York: Basic Books, 1999.

[10] SIEGEL A W, WHITE S H. The Development of Spatial Representations of Large-Scale Environments[M]. REESE H W, ed. Advances in Child Development and Behavior. New York: Academic Press, 1975, 10: 9-55.

[11] HEIDEGGER M. Being and Time[M]. Macquarrie J, Robinson E, Trans. San Francisco, CA: HarperOne, 1962.

[12] HUSSERL E. The Crisis of European Sciences and Transcendental Phenomenology: An Introduction to Phenomenological Philosophy[M]. CARR D, Trans. Evanston: Northwestern University Press, 1970.

[13] TUAN Y F. Space and Place: The Perspective of Experience[M]. Minneapolis, Minnesota: University of Minnesota Press, 1977.

[14] HARNAD S. The Symbol Grounding Problem[J]. *Physica D: Nonlinear Phenomena*, 1990, 42(1/2/3): 335-346.

[15] HARVEY D. Explanation in Geography[M]. London: Hodder & Stoughton Educ, 1969.

[16] O'SULLIVAN D, UNWIN D J. Geographic Information Analysis[M]. New York: John Wiley & Sons, 2010.

[17] 刘瑜，郭浩，李海峰，董卫华，裴韬，从地理规律到地理空间人工智能[J]. 测绘学报，2022，51(6)，1062-1069.
LIU Yu, GUO Hao, LI Haifeng, Dong, et al. A Note on GeoAI from the Perspective of Geographical Laws[J]. *Acta Geodaetica et Cartographica Sinica*, 2022, 51(6): 1062-1069.

[18] 刘瑜，汪珂丽，邢潇月，等，地理分析中的空间效应[J]. 地理学报，2023，78(3)，517-531.
LIU Yu, WANG Keli, XING Xiaoyue, et al. On Spatial Effects in Geographical Analysis[J]. *Acta*

Geographica Sinica, 2023, 78(3), 517-531.

[19] TOBLER W R. A Computer Movie Simulating Urban Growth in the Detroit Region[J]. *Economic Geography*, 1970, 46(sup1): 234-240.

[20] PEUQUET D J. Representations of Geographic Space: Toward a Conceptual Synthesis[J]. *Annals of the Association of American Geographers*, 1988, 78(3): 375-394.

[21] BATTY M. Digital Twins[J]. *Environment and Planning B: Urban Analytics and City Science*, 2018, 45(5): 817-820.

[22] YANG T Y, ZHAO B, GAO S, et al. Neural Representation of Geoinformation in the Human Brain: Affected by Abstraction Levels and Spatial Scales [J]. *Annals of GIS*, 2025, 31(3): 433-448.

[23] ZHANG F, SALAZAR-MIRANDA A, DUARTE F, et al. Urban Visual Intelligence: Studying Cities with Artificial Intelligence and Street-Level Imagery[J]. *Annals of the American Association of Geographers*, 2024, 114(5): 876-897.

[24] MILDENHALL B, SRINIVASAN P P, TANCIK M, et al. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis [M]//Computer Vision—ECCV 2020. Cham: Springer International Publishing, 2020: 405-421.

[25] GARNELO M, SHANAHAN M. Reconciling Deep Learning with Symbolic Artificial Intelligence: Representing Objects and Relations [J]. *Current Opinion in Behavioral Sciences*, 2019, 29: 17-23.

[26] MAKOVIYCHUK V, et al. Isaac Gym: High Performance GPU-Based Physics Simulation for Robot Learning[PP/OL]. V2. arXiv. (2021-08-25)[2026-09-15]. https://arxiv.org/abs/2108.10470.

[27] DAS A, DATTA S, GKIOXARI G, et al. Embodied Question Answering[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018.

[28] POOLE B, JAIN A, BARRON J T, et al. DreamFusion: Text-to-3D Using 2D Diffusion[C]// The 11th International Conference on Learning Representations, ICLR 2023, New York, USA, 2023.

[29] ALEISSAEE A A, KUMAR A, ANWER R M, et al. Transformers in Remote Sensing: A Survey[J]. *Remote Sensing*, 2023, 15(7): 1860.

[30] GORELICK N, HANCHER M, DIXON M, et al. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone[J]. *Remote Sensing of Environment*, 2017, 202: 18-27.

[31] LI Y G, YU R, SHAHABI C, et al. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting[C]// The 6th International Conference on Learning Representations (ICLR 2018), Vancouver, BC, Canada, 2018.

[32] ZHAO L, SONG Y J, ZHANG C, et al. T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 21(9): 3848-3858.

[33] RAISSI M, PERDIKARIS P, KARNIADAKIS G E. Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations[J]. *Journal of Computational Physics*, 2019, 378: 686-707.

[34] ROBINSON W S. Ecological Correlations and the Behavior of Individuals[J]. *American Sociological Review*, 1950, 15(3): 351.

[35] BATTY M. The New Science of Cities[M]. Cambridge, Massachusetts: MIT Press, 2013.

[36] KENNEDY C, PINCETL S, BUNJE P. The Study of Urban Metabolism and Its Applications to Urban Planning and Design[J]. *Environmental Pollution*, 2011, 159(8/9): 1965-1973.

[37] CHURKINA G. Modeling the Carbon Cycle of Urban Systems[J]. *Ecological Modelling*, 2008, 216(2): 107-113.

[38] 杨元喜. 地理空间数字孪生与时空智能[J]. 测绘学报, 2025, 54(2): 213-220.
YANG Yuanxi. Digital Twin and Spatiotemporal Intelligence of Geospatial Information System[J]. *Acta Geodaetica et Cartographica Sinica*, 2025, 54(2): 213-220.

[39] BATTY M. Inventing Future Cities[M]. Cambridge: The MIT Press, 2018.

[40] LIN H, CHEN M, LU G N, et al. Virtual Geographic Environments (VGEs): A New Generation of Geographic Analysis Tool[J]. *Earth-Science Reviews*, 2013, 126: 74-84.

[41] TAO F, ZHANG H, LIU A, et al. Digital Twin in Industry: State-of-the-Art[J]. *IEEE Transactions on Industrial Informatics*, 2019, 15(4): 2405-2415.

[42] HEPPENSTALL A, CROOKS A, MALLESON N, et al. Future Developments in Geographical Agent-Based Models: Challenges and Opportunities [J]. *Geographical Analysis*, 2021, 53(1): 76-91.

[43] STONE P, VELOSO M. Multiagent Systems: A Survey from a Machine Learning Perspective[J]. *Autonomous Robots*, 2000, 8(3): 345-383.

[44] LI D, WANG M, XIAO J, et al. Perspectives on spatiotemporal Intelligence [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2026, 232, 305-318.