遥感基础模型：从单模态走向多模态

Danfeng Hong，*IEEE* 高级会员，Chenyu Li、Xuyang Li、Gustau Camps-Valls，*IEEE* 会士，以及 Jocelyn Chanussot，*IEEE* 会士

arXiv:2603.00988v1 [cs.CV] 1 Mar 2026

摘要—**遥感（RS）技术对于加深我们对地球的认识日益至关重要。随着遥感数据的数量和多样性持续呈指数级增长，我们迫切需要先进的数据建模与理解能力，以有效管理和解读这些海量数据集。基础模型为遥感领域带来了显著的新发展机遇，并蕴藏着颠覆性变革的巨大潜力。本文对遥感领域的基础模型进行了全面的技术综述，并通过探讨其从单模态向多模态的演进，提供了全新的视角。我们希望这项工作能为同时关注基础模型与遥感的研究人员提供有价值的入门参考，帮助他们在这一快速发展的领域开展新项目或探索新的研究课题。本综述围绕以下三个关键问题展开：**

- **什么是遥感中的基础模型?**
- **为什么遥感领域需要基础模型?**
- **我们如何有效地指导青年研究人员，帮助他们全面、实用地了解基础模型在遥感应用中的作用?**

具体而言，我们首先概述背景与动机，强调基础模型在遥感领域的重要性。随后，我们回顾现有的遥感基础模型，并将其系统地分为单模态方法和多模态方法。此外，我们还设置了教程式章节，指导研究人员，尤其是初学者，如何在遥感领域训练基础模型，并将其应用于现实世界任务。本综述旨在帮助遥感领域的研究人员更深入、更高效地理解基础模型，使他们能够轻松入门，并有效地将这些模型应用于各种遥感应用。

索引词——**人工智能、基础模型、多模态、单模态、遥感、地球观测。**

一、引言

本研究得到国家自然科学基金委员会项目（批准号：42271350）以及中国科学院国际伙伴计划项目（项目编号：313GJHZ2023066FN）的支持。GCV 衷心感谢欧洲研究理事会（ERC）通过 USMILE 项目（资助协议编号：855187）以及欧盟委员会通过 HORIZON 项目 ELIAS（资助协议编号：101120237）和 ELLIOT（资助协议编号：101214398）提供的支持。

D. Hong 任职于中国南京 211189 东南大学自动化学院。（电子邮箱：danfeng.hong@seu.edu.cn）

C. Li 任职于中国南京 211189 东南大学数学学院。（电子邮箱：chenyuli.erh@gmail.com）

X. Li 任职于中国北京 100094 中国科学院空天信息创新研究院，以及中国北京 100049 中国科学院大学电子、电气与通信工程学院。

G. Camps-Valls 任职于图像处理实验室（IPL）、瓦伦西亚大学，地址为西班牙帕特纳瓦伦西亚 46980。（电子邮箱：gustau.camps@uv.es）

J. Chanussot 任职于格勒诺布尔阿尔卑斯大学、INRIA、法国国家科学研究中心、格勒诺布尔国立理工学院、LJK（法国格勒诺布尔 38000）。（电子邮箱：jocelyn.chanussot@grenoble-inp.fr）

RECENT 地球观测（EO）技术，尤其是遥感（RS）技术的进步，彻底改变了我们监测地球表面及环境各个方面的能力，例如地表特征、地下状况、空气和水质以及生物的健康状况 [1]。EO 技术是获取有关地球物理属性、环境状况和动态变化等关键信息的重要工具。从太空监测这些方面，有助于更准确、及时地作出决策，从而改善自然资源管理、灾害防备和环境保护。

这些能力推动了对地球系统的科学认识，并在应对气候变化、生物多样性丧失和可持续发展等全球性挑战方面发挥着至关重要的作用。遥感数据能够以独特的视角在不同尺度上研究地球，提供通过传统地面观测方法难以或无法获得的洞见。整合这些洞见可以推动改善人类福祉、促进社会稳定并确保环境可持续性的行动。

A. 遥感基础模型日益重要的作用

然而，遥感数据在规模和多样性方面的快速增长已经超出了传统模型的能力，因此有必要开发更复杂的方法来处理和分析高维数据。随着遥感数据集变得更加复杂，其中包含卫星影像、光谱数据、时间序列以及激光雷达等传感器模态，传统方法难以捕捉这些数据源中丰富的信息。这凸显了对能够处理多模态、时空和高维数据的新一代人工智能模型的迫切需求。

基础模型的出现为这一挑战提供了有效的解决方案 [2]。这类大规模模型可以在海量、多样的数据集上进行预训练，并针对特定任务进行微调，因此在遥感应用中具有巨大潜力。它们能够利用遥感数据的规模与复杂性，学习可迁移的表示，从而适用于多种地理空间任务，例如土地覆盖分类、气候监测、灾害管理和农业。基础模型的多功能性在标注数据稀缺或获取成本高昂的场景中尤为宝贵；这在遥感领域十分常见，因为对大范围区域进行像素级标注既耗费人力，也需要大量资源。基础模型采用自监督预训练与微调的方法，使海量无标签数据能够用于学习通用特征，并将其应用于各种遥感任务。

这一框架提升了遥感分析的可扩展性和效率，使我们能够更好地捕捉地球观测数据复杂、多模态且动态的特性 [3]–[6]。

B. 挑战与困难

尽管取得了这些令人振奋的进展，基础模型在遥感领域的应用仍相对缺乏探索。虽然这些模型潜力巨大，但在该领域的推广和集成仍面临若干障碍，我们将在下文概述这些障碍以及目前正在兴起的解决方案。

- 首要挑战是许多从事遥感、地球观测和地球科学研究的研究人员对基础模型缺乏认识和了解。这些领域的大多数研究人员并不熟悉基础模型的底层原理，例如自监督学习和迁移学习；或者缺乏在工作中应用这些先进技术的专业知识。这种知识差距严重阻碍了基础模型在实际遥感任务中的成功应用，因为研究人员可能难以理解这些模型的工作原理，也不知道如何根据自身的具体需求对其进行调整。为缓解这一问题，近期的一些教程类论文和基准研究（如 BigEarthNet [18],SpectralGPT [19]）开始提供实践资源，帮助遥感研究人员了解预训练和微调策略在实际中的应用。

- 其次，尽管一些综述尝试从特定角度整理并分类遥感基础模型（如表 I 所示），但目前仍缺乏一种全面、类别清晰且以教程为导向的模型分类体系。现有基础模型数量众多，其中一些模型之间仅存在渐进式差异，这让研究人员感到困惑和不确定。由于缺乏用于评估和分类这些模型的标准化框架，研究人员难以针对特定应用选择最合适的模型，也难以比较不同模型在各种任务中的表现。这种不明确的状况加剧了遥感社区在充分发挥基础模型潜力时所面临的挑战。为此，社区推动的举措应运而生，例如 HuggingFace 上专门面向遥感的模型库，以及标准化评估数据集，如 SEN12MS [20] 或 C2Seg [21],，它们为模型比较提供了更透明的基准。

- 最后，遥感研究人员对于基础模型的技术使用指南仍存在显著的不确定性。即使这些模型已经可以使用，许多研究人员仍不清楚如何对其进行微调，或如何使其适用于特定应用。关于模型在真实遥感场景中的性能、可扩展性和鲁棒性等问题，仍大多没有答案。由于缺乏适当的指南和评估基准，研究人员可能会犹豫是否采用这些模型，担心其性能无法满足实际应用的需求。此外，缺少将基础模型与现有遥感系统集成的标准化协议，也使其难以得到广泛采用和部署。

尽管如此，近期已有若干研究展示了实用的适配流程，例如微调用于作物分类、灾害监测或洪水制图的预训练模型。这些研究为如何将基础模型整合到真实世界的遥感任务中提供了具体指导。

这些挑战严重阻碍了基础模型在遥感领域的发展与应用。随着遥感数据的规模和复杂性不断增长，解决这些障碍对于释放基础模型在地球观测领域的全部潜力至关重要。为此，本综述旨在提供一份全面且易于理解的参考资料，阐明基础模型的核心概念，并探讨阻碍其融入遥感领域的关键问题。

C. 与现有综述的比较

近期，已有多篇综述探讨了遥感、地球科学和地球观测领域的基础模型，尽管这一领域仍未得到充分研究。表 I 按发表年份总结了这些研究。

Jiao 等人（2023）综述了遥感领域受脑启发的基础模型，探讨了它们在处理复杂任务方面的独特优势，并分析了其实施所面临的挑战 [7]。大约在同一时期，Tan 等人（2023）通过案例研究，探讨了多模态基础模型在地理学、环境、农业和城市规划领域的潜力与挑战，并强调了其适应性 [8]。

Li 等人（2024）聚焦于遥感领域的视觉语言模型（VLMs），概述了其最新进展、应用和未来方向 [9]。与此同时，Zhu 等人（2024）将关注点转向地球与气候基础模型，探讨其理论基础，并提出了一个用于推动其在地球科学领域发展的框架 [10]。Zhou 等人（2024）提出了地理视觉语言基础模型这一概念，强调将地理空间先验与多模态学习相结合，以用于地理信息处理 [11]。

Lu 等人（2024）对遥感领域的人工智能基础模型进行了广泛综述，系统分析了它们在不同遥感任务中的应用与挑战 [12]。Zhang 等人（2024）则采取了不同的思路，展望一种旨在实现通才智能的通用遥感基础模型，探讨“单模型、多任务”范式及其解决任务碎片化问题的潜力 [13]。Xiao 等人（2024）对遥感基础模型进行了结构化综述，涵盖其研究动机、关键方法、现有基准以及尚未解决的挑战 [14]。

Zhang 等人（2024）将讨论扩展到地球科学基础模型，探讨其在地球系统建模中的作用，并指出关键构建技术、进展以及未来研究方向 [15]。进入2025年，Tao 等人（2025）深入探讨视觉语言模型在遥感领域的最新进展，重点介绍针对遥感任务优化的新型多模态数据集，并将这些模型的关注点从描述性分析转向决策支持 [16]。最后，另一项（2025年）综述考察了潜力与局限性

表 I 遥感、地球科学和地球观测领域基础模型综述论文列表，列出了标题、发表年份、期刊、描述和参考链接。

标题	Year	期刊	描述	Ref.
脑启发遥感感知基础模型与未解决问题：全面综述	2023	IEEE《地球科学与遥感学报》	本综述考察了遥感领域的脑启发式基础模型，并重点介绍其复杂任务中的应用、挑战与优势。	[7]
多模态基础模型的前景与挑战用于地理、环境、农业与城市规划应用的应用	2023	arXiv	本综述通过案例研究，重点介绍多模态基础模型的潜力、挑战与灵活性，涉及地理学、环境、农业和城市规划领域。	[8]
遥感中的视觉语言模型：当前进展与未来趋势	2024	IEEE GRSM	本综述探讨遥感领域的视觉语言模型，重点介绍其进展、应用——化及未来趋势。	[9]
论地球与气候基础模型	2024	arXiv	本文综述探讨了地球与气候模型背后的理论，并提出了一个地球与气候科学的基础框架。	[10]
迈向视觉-语言地理基础模型：综述	2024	arXiv	本综述介绍了地理视觉-语言模型，重点强调地理空间先验知识及其在地理信息处理中的多模态融合能力。	[11]
遥感中的人工智能基础模型：综述	2024	arXiv	本文综述了遥感领域的人工智能基础模型，分析了其在不同任务中的应用及面临的挑战。	[12]
通用遥感智能基础模型：潜力与前景	2024	《科学通报》	本综述展望了通用遥感模型，具备“一个模型，多项任务”架构，以应对任务碎片化，并探讨高分辨率图像处理中的瓶颈。	[13]
《遥感与地球观测基础模型：综述》《遥感与地球观测基础模型：综述》	2024	arXiv	本综述回顾了遥感基础模型，涵盖其研究动机、关键概念、现有研究、基准、挑战及未来方向。	[14]
当地球科学遇上基础模型：迈向通用地球科学人工智能系统	2024	IEEE GRSM	本文综述探讨了地球科学基础模型在推动地球系统建模方面的潜力，并重点介绍其优势、构建技术、最新进展与未来挑战。	[15]
遥感领域视觉语言模型的进展：数据集、能力与增强技术	2025	RS	本综述涵盖了视觉语言模型的最新进展，并介绍了遥感优化的多模态数据集，推动模型从描述性分析转向决策支持。	[16]
When 遥感感知遇见基础模型：综述与展望	2025	RS	本综述探讨了遥感基础模型的潜力与挑战，重点介绍了近期进展、开放性问题以及弥合差距的未来方向自然图像和遥感图像之间。	[17]
遥感基础模型：演进：从单模态到多模态（本文）	2025	–	本综述全面梳理了遥感领域的基础模型，独特之处在于以教程式的实用指南为重点，阐述其从单模态向多模态的演进，为研究人员提供指导。	–

遥感基础模型的相关内容，重点在于弥合自然图像和遥感图像之间的差距，同时识别关键的开放性挑战和未来研究方向 [17]。

尽管这些综述为了解遥感领域的基础模型提供了宝贵见解，但它们主要聚焦于特定子领域、方法或应用。相比之下，我们的综述从更广阔的视角探讨从单模态向多模态的转变。我们着重指出遥感领域的关键挑战与未解问题，阐明基本概念和学习范式，探讨新兴趋势，并提供实用指导，尤其关注遥感基础模型的未来发展。

D. 目标与贡献

本文的主要目标是为遥感研究人员和从业者提供清晰的路径，帮助他们理解并有效应用基础模型。我们介绍这些模型，阐释其独特的学习范式，并展示它们在遥感领域中的实际应用。本综述采用直观、易用的方式，旨在推动基础模型在遥感研究与应用中的广泛普及。

本文的贡献可概括为三个主要目标：

- 我们的目标是提供一份详细且易于理解的综述，专门介绍为遥感领域量身打造的基础模型。尽管基础模型在自然语言处理和计算机视觉等多个领域已受到广泛关注，但其在遥感领域的应用及潜力仍相对缺乏探索。本文旨在弥补这一空白，深入概述基础模型背后的核心原理、其在遥感领域中的运作方式，以及它们与遥感数据处理与分析传统方法的比较。通过清晰阐释这些模型的学习范式，我们希望帮助遥感领域的研究人员、学生和工程师更好地理解基础模型与传统方法之间的根本差异。
- 基础模型在遥感领域应用中最显著的趋势之一，是从单模态模型（处理一种类型的数据）向多模态模型（集成并分析多种类型的数据，如光学数据、激光雷达数据和合成孔径雷达数据）演进。这一演进代表着遥感数据处理与分析方式的根本转变，正朝着更加全面、集成的方法发展。在这篇综述中，我们系统回顾了该领域的关键研究，为基础模型如何演进以应对遥感领域多模态数据集成的挑战提供了全新而全面的视角。通过突出这一转变，我们旨在让读者能够直观、易懂地认识到基础模型在遥感领域日益提升的复杂性与能力。
- 尽管遥感社区对基础模型的认识日益增强，许多研究人员和从业者仍难以理解如何在真实遥感任务中有效应用这些模型。为此，我们提供了一个专门章节，作为实用教程，帮助希望将预训练基础模型纳入自身工作的用户。本教程

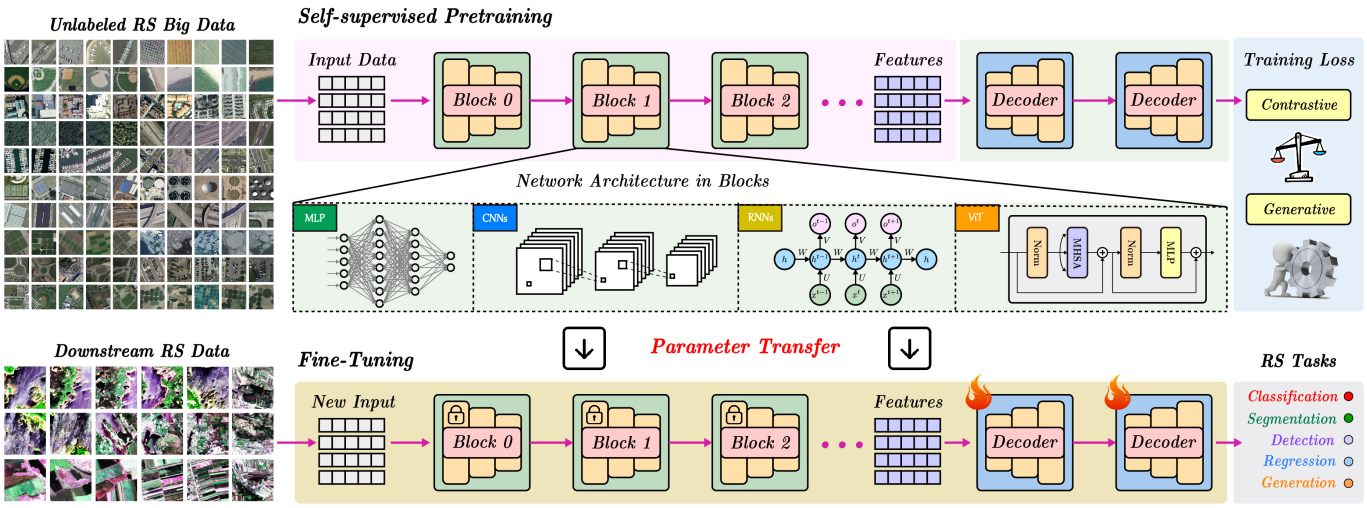


图1. 遥感基础模型的一般学习范式：自监督预训练 + 微调。以视觉相关示例说明：模型首先使用大量无标注遥感数据进行预训练，所用骨干架构包括 MLP、多层感知机、卷积神经网络、RNNs 或 ViTs。预训练期间通常采用对比损失和生成（重建）损失。随后，预训练权重被迁移至下游遥感任务中，既可以保持冻结，也可以进行微调（通常侧重于解码器权重），从而支持分类、分割、检测、回归和生成等应用。

将引导读者了解应用这些模型所需的步骤，从针对特定遥感任务微调预训练模型，到将其集成到现有遥感工作流中。通过提供教程式指导，我们旨在帮助研究人员和从业者在各自领域有效利用遥感基础模型，从而应对下一代对地观测日益复杂的挑战。

E. 论文结构

本文其余部分安排如下：第二节介绍遥感基础模型的概念，并简要概述这些模型在遥感领域中的运作方式。第三节回顾遥感基础模型的发展历程，追溯其从单模态模型向多模态模型的转变，并重点介绍关键进展。第四节以教程形式指导读者如何在遥感领域应用预训练基础模型，帮助用户了解将这些模型用于特定任务的实际要点。最后，第五节讨论遥感基础模型面临的未决挑战，探讨潜在解决方案，并概述未来的研究方向，以进一步提升这些模型在该领域的效能与应用。本文从理论和实践两个方面探讨基础模型，旨在为遥感社区提供有价值的参考，推动这些强大工具在地球观测领域的采用与应用。

II. 基础模型

本节从概念角度简要概述基础模型，并介绍其全新的学习范式。本节将探讨以下问题：“什么是基础模型？”以及“它们的学习范式与以往有何不同？”

A. 基础模型简要认识

1) 机器学习范式的转变：基础模型的特点是规模庞大，并通过在海量、多样化的数据集上进行预训练来学习泛化表示。传统模型通常针对特定任务设计，需要大量特定领域的训练；而基础模型则旨在泛化应用于广泛的任务和领域。其架构通常基于 Transformer 模型或类似的深度学习框架，使其能够捕捉数据中的复杂模式与关系。这种泛化能力源于大量的参数化，以及从异构数据源中学习的能力，因此只需少量额外数据，就能针对特定下游应用对这些模型进行微调。

基础模型最显著的优势之一是其可扩展性。它们能够处理大规模数据集和复杂的数据结构，因而适用于从自然语言处理（如 GPT、BERT）到计算机视觉（如 CLIP、DINO）等多种领域。此外，其可迁移性使其能够在有限的领域特定适配下适应新任务，从而大幅减少对大规模标注数据集的需求。这在标注数据稀缺或获取成本高昂的领域尤为有益。此外，它们处理多模态数据（如文本、图像、音频等）的能力，有助于更全面地理解并表征复杂现象。

2) 遥感基础模型推动地球观测变革：遥感基础模型是变革性工具，旨在应对地球观测数据所面临的独特挑战。遥感数据本质上十分复杂，包含光学 RGB、合成孔径雷达（SAR）、光谱数据、激光雷达和时空序列等多种模态。传统遥感模型往往难以整合这些多模态数据集，并且针对每项特定任务都需要大量标注数据

表 II 遥感基础模型的代表性预训练数据集。

Name	领域	模态	分辨率	图像数量
SeCo	通用场景	光谱	10 m	约 100 万
SSL4EO-S12	土地覆盖	光谱、合成孔径雷达	10–60 米	1,004,316
fMoW	分类	RGB、光谱	0.3–10 m	1,047,691
MillionAID	分类	RGB	0.5–11.4 米	400,695
SSL4EO-L	通用场景	光谱	30 m	500 万
SatLas	目标检测、变化检测、分类、语义分割	RGB、光谱	1–10 米	856K
GeoPile（GFM）	通用场景	RGB	0.1–30 米	587,954
DOFA-data	通用场景	RGB、合成孔径雷达、光谱	0.5–30 米	8,081,411
Prithvi	通用场景	光谱	30 m	420万
Skysense	通用场景	RGB、合成孔径雷达、光谱	0.31–20 米	2,150 万
MMEarth	通用场景	RGB、光谱、合成孔径雷达、数字高程模型	10 m	120万

应用。基础模型通过利用多种遥感数据集进行大规模预训练，克服了这些局限，使模型能够学习可应用于广泛地理空间任务的通用表示。表II总结了广泛用于遥感基础模型预训练的数据集，并重点介绍其领域侧重、传感模态、空间分辨率、数据规模和下游适用性等关键特征。具体而言，*fMoW* 和*MillionAID* 提供丰富的类别标注，以及中高分辨率 RGB 影像，但光谱数据覆盖有限（0.3–11.4 米），因此尤其适用于图像级或场景级分类以及目标存在识别。*SSL4EO-S12*（Sentinel-1/2，10–60 米）[22]和*SSL4EO-L*（Landsat，30 米）[23] 兼具广泛的光谱覆盖和规律的重访周期，支持大规模语义分割和土地覆盖制图，适用于中等分辨率即可满足需求的场景。 *Satlas* 专为目标检测、变化检测、分类和语义分割而设计，利用 NAIP/RGB/Sentinel 影像（1–10 米）处理需要精细空间细节和跨源一致性的任务。诸如*SeCo*、*GeoPile*（*GFM*）、*Prithvi*、*DOFA-data*、*Skysense*和*MMEarth*等广泛的“通用场景”语料库涵盖多种模态（RGB、合成孔径雷达、光谱数据、数字高程模型）和分辨率（0.1–30 米），有助于提升模型在各类下游任务中的稳健可迁移性。尤其是，*DOFA-data*、*Skysense*和*MMEarth* 纳入了合成孔径雷达和/或数字高程模型图层，为需要全天候传感（合成孔径雷达）或地形背景（数字高程模型）的应用提供了优势。

这些模型在预训练阶段采用自监督或无监督学习技术，从海量无标注遥感数据中提取有意义的模式和特征。这种方法显著降低了对标注数据集的依赖；由于在广阔且复杂的地理区域进行像素级标注面临诸多挑战，遥感领域的标注数据集往往十分有限。基础模型能够实现高效的迁移学习，只需少量标注数据进行微调，即可执行土地覆盖分类、气候监测、灾害应急响应和农业评估等特定任务。

遥感基础模型的适应性使其能够处理不同的数据分辨率、传感器类型和时间频率，从而适用于动态且异构的地球观测应用。其可扩展性确保它们能够处理大规模数据集，例如连续的卫星影像流或全球气候数据，为科学研究和实际应用提供稳健且及时的见解。

B. 遥感领域的学习范式

1) 传统遥感方法的瓶颈： 遥感图像处理与分析中的传统学习范式通常包含两种主要方法。第一种方法侧重于从遥感图像中手动或自动提取特征表示，然后应用分类或回归算法解读这些特征，以完成土地覆盖分类、目标检测或环境监测等特定任务。第二种方法采用端到端模型，直接将输入数据映射为输出标签，省去显式特征提取步骤，从而简化学习过程。尽管这些传统方法在某些应用中效果显著，但其根本局限在于标注数据的可用性。在遥感领域，标注尤为困难，因为需要对海量数据进行像素级标注，而这些数据通常以鸟瞰视角覆盖广阔的地理区域。这一过程劳动密集且耗时，导致大规模标注数据集稀缺，因而大量遥感数据仍未标注、未得到充分利用。

2) 基础模型正在革新遥感： 相比之下，基础模型的出现引入了一种变革性的学习范式，其特点是**“自监督预训练 + 微调”**。这种方法已日益成为遥感领域的主导框架，为传统方法的局限提供了解决方案。自监督预训练使模型能够从海量无标注遥感数据中学习丰富且通用的表示，方法是识别数据本身固有的模式、结构和关系。这一过程利用遥感数据集的天然冗余性和复杂性，例如时间序列、多模态输入（如光学、合成孔径雷达、激光雷达）以及光谱变化，无需人工标注即可构建稳健的特征表示。

在预训练阶段之后，通过微调使用相对少量的标注数据，使预训练模型适应特定的下游任务。这显著

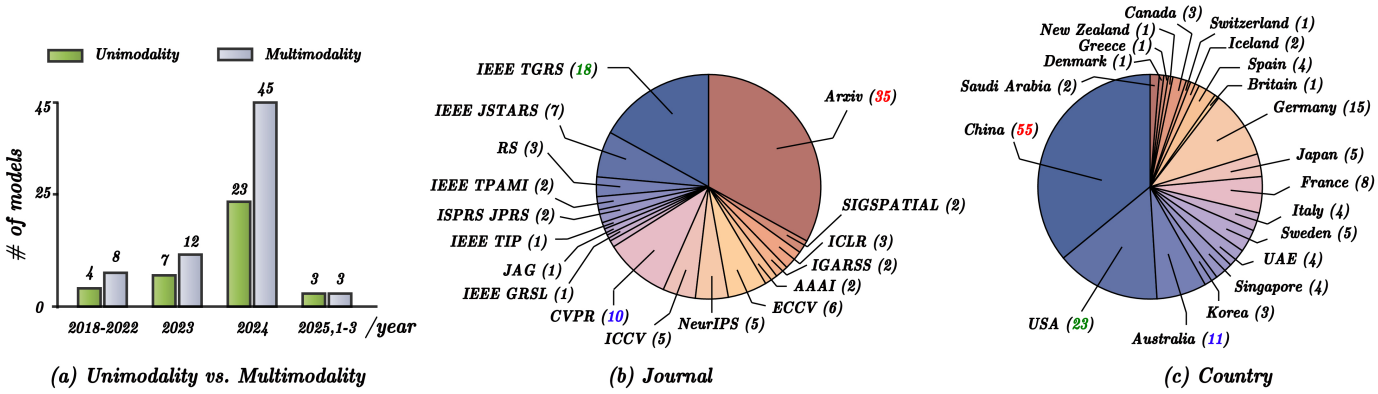


图2. 遥感相关领域基础模型的统计概览。(a) 对遥感中的单模态和多模态基础模型进行了定量比较。(b) 使用饼图展示了遥感基础模型在不同期刊中的分布。(c) 展示了不同国家开发的遥感基础模型数量。两张饼图中数量排名前三的类别分别以红色、绿色和蓝色突出显示。

减少了对大规模标注数据集的依赖，同时在各种应用中保持出色的性能。微调过程会优化模型的表示，使其符合特定任务的目标，例如土地覆盖分类、变化检测或目标识别，从而提升准确率和泛化能力。

基于基础模型的范式在遥感场景中尤其具有优势，这是因为来自多种传感器和平台的无标签数据极为丰富。通过有效利用这些数据，基础模型能够克服传统方法固有的数据稀缺问题，从而实现更高效、更具可扩展性的解决方案。此外，基础模型能够泛化至不同的遥感任务和数据集，因此具有很强的通用性，可支持从环境监测到城市规划等广泛应用。

图1 展示了遥感领域基础模型的一般学习范式，凸显了从传统监督式方法向更灵活、更强大的自监督预训练与微调框架的转变。这一转变不仅提高了遥感数据的利用效率，也为更加先进、泛化能力更强且更稳健的遥感应用铺平了道路。

III. 从单模态到多模态的演进

本节直观概述了遥感领域基础模型的发展，追溯其从单模态到多模态的演变。通过这一视角，我们探讨这些模型如何日益整合多种数据类型，即单模态数据和多模态数据。图3 展示了2018年至2025年遥感领域基础模型从单模态向多模态发展的路线图。如图所示，与单模态模型相比，多模态基础模型受到的关注和青睐日益增加，尤其是在2024年之后，其发展显著加速。值得注意的是，融合了大型语言模型（LMMs）的视觉基础模型越来越受到研究人员的关注。

A. 概述：统计分析

我们收集了105篇遥感相关领域基础模型的论文，并进行了统计分析，以更好地了解其趋势和影响。图2 从三个角度概述了这些模型：模态分布、发表期刊和各国贡献。

图 2(a) 对比了单模态与多模态遥感基础模型，展现出向多模态方法转变的明显趋势。随着遥感基础模型的快速发展，各类数据源（如光学影像、合成孔径雷达和光谱数据）的整合不断增加，这一趋势在 2024 年尤为显著。人们对多模态模型日益浓厚的兴趣，反映出遥感应用对更全面、更稳健解决方案的需求。

关于出版物分布（见图 2(b)），arXiv 收录的论文数量最多，凸显了研究成果在正式同行评审前处于早期传播阶段。在同行评审出版物中，IEEE TGRS 位居第二，体现了地球科学与遥感社区的强大影响力；作为顶级人工智能与计算机视觉会议的 CVPR 位列第三。这表明遥感与前沿人工智能及视觉研究的交集日益扩大。

图2(c) 展示了遥感基础模型研究的地理分布。中国贡献最多，其次是美国和澳大利亚，这表明传统上占主导地位的人工智能研究中心以及新兴的遥感人工智能领军者都积极参与其中。这一分布凸显了遥感基础模型发展的全球性，其驱动力是地球观测和地理空间分析对先进人工智能解决方案的需求。

总体而言，这些统计数据凸显了遥感基础模型的快速进展、向多模态的转变，以及人工智能驱动的方法日益深入的参与。这一趋势反映出研究界正不断加大投入，以拓展遥感智能的边界，并开发更具泛化能力、可扩展且高性能的模型，服务于多样化应用。

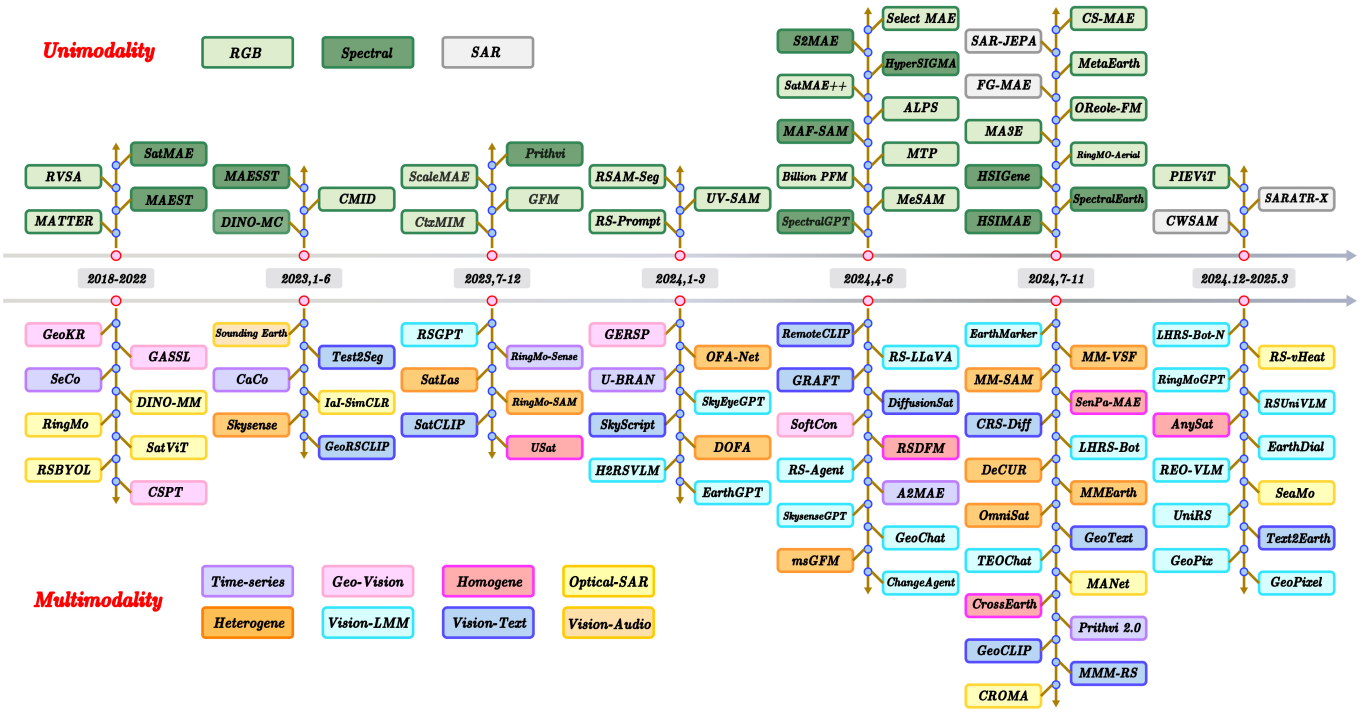


图3. 2018年至2025年遥感基础模型路线图：遥感从单模态方法到多模态方法的演进。单模态数据包括RGB、光谱和SAR数据。多模态集成分为同质数据（例如多光学）、光学-SAR和异构数据（例如光学-SAR-光谱-数字表面模型、视觉-文本、视觉-音频、地理空间-视觉、时间序列和视觉大模型的组合）。不同模型类别采用颜色编码，以区分各种数据模态和集成类型，模型在每个时间段内按时间顺序排列（垂直方向）。

B. 遥感中的单模态基础模型

遥感中的单模态基础模型旨在对来自单一来源或传感器类型的数据进行预训练，并利用特定数据模态的独特特征。示例包括高分辨率 RGB 影像、雷达数据、点云、光谱数据[1]，以及气象数据。这些模型旨在从各自模态中提取最具信息量的特征与模式，从而优化其应对不同特性与挑战的能力。这种有针对性的方法为在专门的数据领域中推进遥感任务奠定了基石。

1) 光学 ***RGB*** 图像：光学 RGB 图像中的高分辨率空间和语义信息可以通过自监督和对比学习方法得到良好表征[24]。例如，MATTER [25] 采用对比学习来强调遥感图像中的纹理信息，而RVSA [26] 采用受掩码自编码器启发的掩码策略，在MillionAID [27] 数据集上进行预训练，随后使用其RVSA模块进行微调。CMID [28]通过将对比学习（CL）与掩码图像建模（MIM）相结合，提出了一种对比掩码图像蒸馏方法，以实现全局语义可分性和局部空间感知性。CtxMIM [29] 引入一个上下文增强生成分支，通过一致性约束为逐块重建提供上下文信息。GFM模型 [30] 利用大规模公共数据集和师生双流网络，并以掩码图像重建作为其目标，在场景分类、变化检测和语义分割任务中取得了显著成功。

Scale-MAE [31]和 SatMAE++ [32] 等模型都针对RS中的图像尺寸和分辨率问题展开研究：Scale-MAE专注于尺度不变表征学习，而 SatMAE++则增强了分辨率感知特征提取。因此，Scale-MAE有利于跨尺度泛化任务，而 SatMAE++则适用于目标检测和细粒度分割等高分辨率应用。SelectiveMAE [33]根据语义丰富度动态编码并重建所选图像块，采用渐进式语义词元选择模块，从重建语义相似词元逐步发展到编码互补的语义依赖关系。MA3E [34] 提出了一种基于旋转角变化的掩码自监督学习框架，用于解决遥感图像与自然图像之间的差异。RingMo-Aerial [35] 将光谱信息作为辅助信息，并通过对比学习进行训练。Billion PFM [36], 和 OReole-FM [37]采用MAE方法，将高分辨率遥感图像的预训练扩展到十亿参数规模的模型。文献[38]提出了一种分辨率引导的自级联生成框架，称为MetaEarth，能够生成具有广泛地理分辨率的任意区域图像。Cross-Scale MAE [39] 在不同尺寸的图像上执行对比掩码建模，形成用于预训练的统一优化框架。PIEViT [40] 是一种新颖的遥感图像自监督学习框架，采用师生架构来处理图像级和

[1]我们将遥感中的多光谱和高光谱数据统称为光谱数据。

块级任务。

此外，分割一切模型（SAM）的分割能力在遥感领域引起了广泛关注。MeSAM [41]为遥感光学图像设计了一种新的适配器，并将其嵌入SAM编码器中，利用多尺度卷积核更有效地处理图像，同时保留高频特征。RSAM-Seg[42] 通过融入领域知识和缩放模块增强SAM，执行云、建筑物、田地和道路分割等任务。Rsprompter [43] 引入了可学习锚点提示和掩码解码器，使SAM能够处理用于实例分割的类别输入。UV-SAM [44]将类激活图生成的粗掩码细化为伪标签，用于在遥感图像上训练语义分割模型。ALPS [45] 提出了一种使用SAM进行全自动密集标注的框架，无需先验知识或提示。

2）光谱数据：多种模型利用多时相和多波段数据来改进空间与光谱信息重建 [46]。SatMAE [47]引入了针对多时相、多通道数据定制的分组掩码策略，从而更好地学习通道与空间关系。DINO-MC [48] 采用对比学习框架，Prithvi [49]通过掩码单个通道来增强光谱空间表征学习。MAF-SAM [50]提出了一种多阶段自适应微调策略，使用前缀适配器将光谱数据压缩为三个通道，应用LoRA微调将作物特异性知识注入图像编码器，并生成带有精确类别信息的多类别分割掩码。在以往方法之外，SpectralGPT [19] 是首个专门针对光谱遥感数据定制的基础模型。通过显式建模多/高光谱观测固有的空谱耦合与光谱序列性，SpectralGPT引入了三维掩码策略以及光谱一致性约束。该设计能够有效学习光谱序列和结构化光谱表征，从而大幅推进光谱遥感数据的建模范式与实际理解。MAEST[51] 和MAESST [52] 引入了针对光谱特征定制的掩码自监督学习算法，以提高土地覆盖分类准确率。Wang等人 [53]在大规模光谱立方体上进行了自监督掩码建模预训练。SpectralEarth [54] 引入了一种自监督学习算法，为EnMap数据的预训练以及一系列微调下游应用提供了坚实的数据基础和算法支持。HyperSIGMA [55] 引入了一种双流框架，分别建模空间与光谱信息。在该方法中，输入光谱数据通过随机选择光谱波段并打乱光谱序列，被统一为固定数量的波段。Pang等人 [56] 提出了一种生成基础模型，该模型基于潜扩散模型并支持多条件控制。它还设计了两阶段超分辨率框架，以提升所生成图像的真实感。

3) 合成孔径雷达（***SAR***）： FG-MAE [57] 利用方向梯度直方图（HOG）作为重建特征，利用 SAR 数据中固有的空间信息。SAR-JEPA [58] 采用联合嵌入预测架构，预测梯度信息，以对 SAR 图像进行预训练。CWSAM [59] 通过引入特定任务的输入模块和分类掩码解码器，并为分割任务注入 SAR 低频信息，将 SAM 框架适配用于 SAR 图像中的土地覆盖分类。SARATR-X [60]采用多尺度掩码策略对 SAR 图像进行预训练，实现目标检测和分类任务。

C. 遥感多模态基础模型

遥感单模态模型通常侧重于单一数据类型，例如光学、雷达或激光雷达，而集成多模态数据则因遥感数据具有异构性而增加了复杂性。这些数据通常来自不同的传感器、平台和尺度，因此在空间、时间和光谱特征方面存在差异。例如，光学、雷达和激光雷达数据分别提供了地球表面的不同视角，但要将它们集成起来，就需要处理分辨率、数据格式和噪声特征方面的差异。此外，同时结合结构化数据（如时间序列或地理空间坐标）与非结构化数据（如文本或图像），会进一步增加处理难度。多模态集成的一大挑战在于，如何将这些不同类型的数据协调统一到一个模型中，使其能够准确捕捉潜在现象，同时保留每种模态的优势。此外，这种集成还必须考虑每种数据类型固有的不确定性和不完整性程度各异的问题，这使得开发稳健可靠的模型更具挑战性。

遥感领域的多模态基础模型通过整合不同数据类型中的互补信息，突破了单一模态的局限。这些模型能够处理同质和异构数据，并将视觉数据与其他模态相结合，例如时间序列、地理空间数据、文本、LMM 和音频。与强调不同遥感数据源的多源数据不同，多模态更侧重于整合不同的数据类型或形式，从而更全面地理解地球观测数据。这种方法通过捕捉每种模态的独特特征及其关系来提升模型性能，最终提高遥感应用的准确率和鲁棒性。在处理多模态遥感数据时，区分同质来源和异构来源至关重要。同质数据（来自相同来源但使用不同传感器的数据，例如Sentinel-2、Landsat、EnMap、高分一号/二号/五号/六号）通常可以在数据层面直接融合，方法包括波段堆叠、重采样或分辨率增强 [61]。相比之下，异构数据（例如光学数据与合成孔径雷达数据，或光学数据与数字高程模型）通常因成像原理、分辨率和噪声特征不同而无法在原始数据层面进行组合。在这种情况下，宜采用特征级融合，包括早期特征拼接、中期特征交互，以及后期决策级组合。近年来，人们还提出了更先进的交互式融合策略，例如跨模态注意力、基于图的推理和协同学习机制，以

动态捕捉跨模态互补性，从而提升模型在复杂下游任务中的表现。关于最佳实践的详细讨论，读者可参阅 [62]。

1) 异构数据 $\mapsto$ 时间序列： 在视觉数据与时间数据的融合中，多个模型提出了新颖策略，以增强遥感中的时空表征。SeCo [63]通过从不同季节的图像中构建正负样本对，明确考虑遥感数据的季节性特征，从而提高模型对季节变化的敏感度。然而，SeCo 难以识别长期时间不变性。为解决这一问题，CACo [64] 采用对比损失函数增强时间一致性，从时间序列数据中提取更稳健的不变特征。

相关进展也延伸至遥感视频数据。RingMo-Sense [65] 针对航空视频，采用多策略掩码进行预训练。与此同时，U-BARN [66] 采用类 BERT 生成式预训练任务，重建光学卫星图像序列中被掩码的部分。

其他模型通过先进的掩码策略改进时空编码。A^2-掩码自编码器 [67]结合了锚点感知掩码策略与地理编码模块，有效捕获空间、时间和光谱变化。最后，Prithvi-EO-2.0 [68] 利用来自 Landsat 和 Sentinel-2 的时间序列数据，通过通道掩码预训练从同质数据集中提取时间特征，进一步强化遥感中的时间序列建模。

*2)*异构数据 $\mapsto$ 地理空间-视觉： GeoKR[69] 利用制图信息作为标签，结合蒸馏与对比学习进行半监督预训练。GASSL [70] 将对比学习应用于经过统一增强的图像数据，并将地理位置作为辅助监督。CSPT [71]提出了一种对比式空间预训练方法，采用双编码器分别对图像和地理位置进行编码。通过对比目标，它从图像中提取有效的位置信息，用于图像分类等下游监督任务。Mai 等 [72] 通过引入用于地理空间视觉表征的自监督对比学习框架，提出了 CSP 模型。GERSP [73] 采用教师-学生蒸馏网络，在地理层面对齐自然图像和遥感图像，以增强预训练模型对一般土地覆盖知识的感知能力。SoftCon [74]引入软对比学习，利用从土地覆盖数据中提取的多标签监督信号优化跨场景软相似度指标，从而增强模型在复杂遥感图像上的泛化能力。

3) 同质数据：对于同质数据，USat[75] 模型采用独特的块投影层和位置编码，使其能够处理来自多种传感器的光谱数据，并适应不同的空间尺度。RSDFM[76] 采用 BEVFormer 架构，基于来自不同平台、角度和高度的数据构建统一视角。SenPa-MAE [77] 将光谱波长信息作为网络的先验嵌入，并采用掩码建模自监督学习，使模型能够出色地处理来自不同光谱传感器的数据。CrossEarth [78] 提出一种领域泛化方法，使基础模型能够在风格各异的下游数据集上执行语义分割。AnySat [79] 是一种多模态模型，探索了分辨率自适应空间编码器。通过采用尺度自适应分块编码模块，AnySat 能有效提升模型对分辨率的敏感度，尤其适用于分辨率各异的光学遥感数据。CSMAE [80] 训练了一个用于多传感器遥感图像检索的跨传感器掩码自编码模型，以便有效检索不同图像模态中语义相似的图像。

*4)*异构数据 $\mapsto$ 光学-***SAR***： RingMo [81]采用了针对异构遥感图像特性设计的稀疏掩码策略和 Transformer架构。这些模型，例如 IaI-SimCLR [82], DINO-MM [83], RS-BYOL [84],，在多模态、多源卫星影像场景中采用了对比学习范式。SatViT [85] 直接将掩码自编码框架迁移至光学和SAR数据源，以学习异构数据。RS-BYOL[84] 通过结合单通道和三通道特征提取，从光谱和SAR数据中学习不变特征嵌入。MANet [86]在SAM编码器中引入两个适配器，一个用于提取单模态信息，另一个用于融合多模态信息。CROMA [87]采用CoCa框架，通过对比式掩码预训练融合光谱和雷达数据，从而提升模型处理数据异构性的能力。RS-vHeat [88]探索使用基于热传导的并行计算模型，来模拟高分辨率遥感图像中的局部区域相关性。近期，Li 等人 [89] 提出了一种新颖的遥感多模态基础模型SeaMo，通过嵌入季节性知识，提升其在广泛下游应用中的性能。

5) 异构数据 $\mapsto$ 光学/***SAR***/光谱/数字表面模型：在多模态遥感领域，已有多种模型被开发出来，以有效整合多种数据源。Sky- Sense [90] 采用分解式多模态时空编码器来处理光学和 SAR 时间序列数据，并利用多粒度对比学习增强其对复杂遥感数据集的理解能力。类似地，SatLas[91]利用 Swin Transformer 在 NAIP、激光雷达和 Sentinel-2 数据上进行大规模监督式预训练，从而提升了处理多样化地理空间输入的能力。

对于分割任务，RingMo-SAM [92] 通过冻结 SAM 编码器并采用 LoRA 进行参数高效微调（PEFT），同时为 RGB、SAR、全极化合成孔径雷达、数字表面模型和多光谱影像数据引入模态专用解码器，从而增强多模态分割能力。在多模态词元化方面，OFA-Net [93] 为不同输入类型（RGB、光谱和雷达）引入专用分词器，确保模型能够感知多样的光谱和空间信息。与此同时，DOFA [94] 采用动态权重生成器，对不同光谱波段的特征进行自适应嵌入，以构建支持 RGB、光谱和雷达数据灵活输入的高效基础模型。

为进一步增强跨传感器学习，msGFM [95]采用基于掩码图像建模的跨传感器预训练方法，从而学习光学数据的联合表示，

合成孔径雷达、光谱和数字表面模型数据，进而提升多模态集成能力。同样，MM-VSF [96] 在知识引导框架下融合光谱影像与气象数据，并以可变步长预测作为其预训练目标，从而增强环境预测能力。

对于多模态分割，MM-SAM [97]扩展了 SAM，以支持跨模态和多模态处理，显著提升了其在多传感器遥感应用中的分割性能。同时，DeCUR [98] 引入解耦策略，将共有表征与特有表征分离，从而有效融合不同模态间的互补信息，例如雷达-光学、RGB-高程和RGB-深度。

有几种模型专注于整体多模态融合和数据同化。MMEarth [99]在预训练期间重建 ℓ_2 种不同的像素级和区域级模态及标签，为下游任务提供全面的视角。最后，OmniSat [100]利用异构模块和对比掩码策略进行预训练，整合高分辨率RGB、光谱时间序列和雷达时间序列数据，在广泛的多模态应用中取得了出色的表现。

6) 异构数据 $\mapsto$ 视觉大模型：大型多模态模型的快速发展拓展了遥感任务的范围，使其不再局限于传统视觉模型，并能够实现更复杂的多模态交互。一种常见方法是利用视觉-语言基础模型，将视觉编码器与文本语义对齐，以便与大语言模型集成。例如，RSGPT [101] 使用EVA-CLIP编码器，并通过先进技术在遥感任务上进行训练。RS-LLaVA[102]通过LoRA微调显著提升了LLaVA的性能。EarthGPT [103] 构建了一种视觉增强感知机制，优化了粗尺度语义理解与细尺度细节提取的融合。

除了静态图像分析外，交互式模型和对话模型在遥感应用中也发挥着至关重要的作用。RS-Agent [104]将大语言模型定位为“中央控制器”，整合多种高性能遥感图像处理工具，以支持多工具、多轮交互。GeoChat [105] 率先推出了首个支持视觉地理定位任务的对话式视觉语言地理基础模型，为交互式遥感图像分析开创了新范式。Change-Agent [106] 提出了一种交互式智能体，可根据用户指令执行全面的变化检测、目标计数和因果分析。EarthMarker [107] 设计了一种跨领域阶段式学习策略，以实现针对图像级、区域级和点级遥感解译的多粒度视觉感知。

有几种模型专注于遥感视频分析和时序推理。SkyEyeGPT [108] 是专为遥感视频描述设计的视觉语言地理基础模型，为视频内容的自动描述提供了新颖的解决方案。TEOChat [109]是一种能够与时间序列遥感数据交互的视觉语言模型，在各种时序推理任务中表现出色。UniRS[110] 是面向多时相遥感任务的统一视觉语言模型，支持单幅图像、图像对和视频，并在统一框架下进行全面的时序分析。

为了提升遥感图像理解，LHRS-Bot [111]采用分层视觉语言对齐策略和课程学习方法，为遥感分析提供了新的视角。LHRS-Bot-Nova [112] 通过采用由专家混合模型强化的视觉编码器和全新的桥接模块，增强了 LHRS-Bot，使多模态大语言模型能够更好地理解遥感图像，从而支持交互式自然语言对话。RingMoGPT [113] 针对遥感领域的关键问题，如目标级识别、定位和多时相变化检测，采用具备位置感知和指令感知能力的 Q-Former 及变化检测模块，以实现目标检测和变化描述。RSUniVLM [114] 引入了面向粒度的专家混合模型，增强了模型捕捉不同粒度视觉信息的能力。

新兴模型强调高分辨率和细粒度感知。GeoPix [115] 将多模态大语言模型对遥感图像的理解扩展至像素级。GeoPixel [116] 是一种端到端高分辨率多模态模型，支持像素级定位，并通过对话中的交错掩码实现细粒度视觉感知。该模型支持最高达 4K 的高分辨率遥感图像分析，并兼容任意宽高比。EarthDial [117]支持光谱、多时相和多分辨率图像，可执行分类、检测、描述生成、问答、视觉推理和视觉定位等多种遥感任务。

一些模型将领域专用人工智能与地球观测相结合。REO-VLM [118]将回归能力与传统生成式功能相结合，利用语言驱动推理融入科学领域知识，从而实现对地球观测数据的复杂解读，并支持生物量回归和图像内容解释等任务。SkySenseGPT [119]引入计算机图形学原理，增强模型感知和推理遥感目标之间关系的能力。这些发展表明，人工智能驱动的遥感分析正转向更加交互化、具备领域意识且支持高分辨率的方向，标志着地理空间智能迈入新时代。

7) 异构数据 $\mapsto$ 视觉-文本：在视觉-文本领域，已有多种模型被开发出来，以弥合遥感图像与文本描述之间的差距。Text2Seg[121] 将 SAM 的零样本分割能力与 CLIP 和 Grounding DINO 相结合，并采用两阶段框架（如 PreSAM 和 PostSAM），实现遥感图像的文本引导语义分割。类似地，GeoRSCLIP [122], 与 RS5M数据集一同发布，是一种专为遥感影像设计的视觉语言预训练模型。受 CLIP 启发，Li等人 [123] 基于对比式视觉语言监督，提出了用于遥感场景分类的视觉语言模型（RS-CLIP）；而 SatCLIP [124] 则增强了图像与其地理位置之间的对齐，从而提升模型的地理空间感知能力。

除了直接进行图文对齐外，其他模型还利用外部地理空间信息来丰富对遥感图像的理解。SkyScript [125] 利用 OpenStreetMap 数据，根据地面采样距离（GSD）为遥感图像添加标注，从而实现更具结构化的地理空间标注。与此同时，RemoteCLIP [126] 是一个用于遥感图像语言任务的预训练模型，通过现有数据集构建图文对

表 III 遥感领域现有单模态基础模型（如 RGB、光谱数据、SAR）的简要比较，详细列出了相应的发表论文、预训练架构、下游应用和代码可用性。CL：对比学习，MAE：掩码自编码器，MIM：掩码图像建模，DIFF：扩散模型，JEPA：联合嵌入预测架构，SC：场景分类，SS：语义分割，CD：变化检测，OD：目标检测，SR：超分辨率，AD：异常检测，SU：光谱解混，DN：去噪，IG：图像生成。

模态	模型	发表论文	架构	下游任务	代码
RGB	MATTER [25]	CVPR'23	CL	场景分类、语义分割、变化检测	https://github.com/periakiva/MATTER
	Scale-MAE [31]	ICCV 2023	MAE	场景分类，语义分割	https://github.com/bair-climate-initiative/scale-mae
	RVSA [26]	TGRS 2022	MIM	场景分类，语义分割，目标检测	https://github.com/ViTAE-Transformer/Remote-Sensing-RVSA
	CMID [28]	TGRS 2023	对比学习&掩码图像建模	场景分类、变化检测、目标检测	https://github.com/NJU-LHRS/official-CMID
	GFM [30]	ICCV 2023	MIM	场景分类、语义分割、变化检测、超分辨率	https://github.com/mmendiet/GFM
	Cross-Scale MAE [39]	NeurIPS'23	MAE	场景分类，语义分割	https://github.com/aicip/Cross-Scale-MAE
	MTP [120]	JSTARS'24	监督式	语义分割，变化检测，目标检测	https://github.com/ViTAE-Transformer/MTP
	选择性掩码自编码器 [33]	Arxiv'24	MAE	场景分类、语义分割、目标检测	https://github.com/Fengxiang23/SelectiveMAE
	MA3E [34]	ECCV'24	MAE	场景分类、语义分割、目标检测	https://github.com/benesakitam/MA3E
	SatMAE++ [32]	CVPR'24	MAE	SC	https://github.com/techmn/satmae pp
	RS-Prompter [43]	TGRS'24	监督式	SS	https://github.com/KyanChen/RSPrompter
	UV-SAM [44]	AAAI'24	监督式	SS	https://github.com/tsinghua-fib-lab/UV-SAM
	RSAM-Seg [42]	RS'24	监督式	SS	https://github.com/Chief-byte/RSAM-Seg
	ALPS [45]	TIP'24	监督式	SS	https://github.com/StriveZs/ALPS
	MeSAM [41]	TGRS'24	监督式	SS	https://github.com/Magic-lem/MeSAM
	MetaEarth [38]	TPAMI 2024	Diff	IG	https://jiupinjia.github.io/metaearth/
光谱数据	SatMAE [47]	NeurIPS'22	MAE	场景分类,语义分割	https://sustainlab-group.github.io/SatMAE/
	DINO-MC [48]	CVPRW 2024	CL	场景分类、变化检测	https://github.com/WennyXY/DINO-MC
	SpectralGPT [19]	TPAMI 2024	MAE	场景分类、语义分割、变化检测	https://github.com/danfenghong/IEEE_TPAMI_SpectralGPT
	MAEST [51]	TGRS 2022	MAE	SS	https://github.com/ibanezfd/MAEST
	MAESST [52]	CVPRW'23	MAE	SS	https://github.com/HSG-AIML/MaskedSST/
	HyperSIGMA [55]	TPAMI'25	MAE	语义分割, 异常检测, 目标检测, 变化检测, 光谱解混, 超分辨率, 去噪	https://github.com/WHU-Sigma/HyperSIGMA
	HSIMAE [53]	JSTARS'24	MAE	SS	https://github.com/Ryan21wy/HSIMAE
	HSIGene [56]	Arxiv'24	Diff	IG	https://github.com/LiPang/HSIGene
SAR	FG-MAE [57]	JSTARS'24	MAE	场景分类、语义分割	https://github.com/zhu-xlab/FGMAE
	SAR-JEPA [58]	ISPRS'24	JEPA	OD	https://github.com/waterdisappear/SAR-JEPA
	CWSAM [59]	JSTARS'25	监督式	SS	https://github.com/xypu98/CWSAM
	SARATR-X [60]	TIP'25	MAE	场景分类, 目标检测	https://github.com/waterdisappear/SARATR-X

，以增强检索、检测和分割能力。为应对缺乏文本标注遥感数据的问题，GRAFT [127]将 Flickr 中带有地理参考信息的地面视角图像与对应的遥感图像和图像描述配对，有效扩充了可用的训练数据。

在生成式建模领域，基于扩散的方法已成为遥感应用的有力工具。DiffusionSat [128]将元数据（如地理位置）融入扩散模型，处理时间序列生成、基于光谱输入的超分辨率重建以及大规模高分辨率数据集上的图像复原等任务。在此基础上，CRS-Diff [129]引入了多条件控制机制，通过支持文本、元数据和图像输入实现精确生成。GeoText[130] 进一步通过将土地覆盖目标与文本描述对齐，增强空间感知与细粒度感知，从而提高地理空间数据集的可解释性。

若干模型还探索了图像-位置对齐和多模态数据融合的新策略。GeoCLIP[131],受图像到 GPS 检索策略的启发，利用傅里叶域中的位置编码，将遥感图像与其对应的 GPS 位置作为连续函数进行对齐。MMM-RS [132] 通过在大规模多模态、多分辨率遥感影像上训练基于 CycleGAN 的模型，推动了文本到图像生成的发展。最后，Text2Earth [133],是一种基于扩散框架构建的 13 亿参数生成模型，专为全球尺度的遥感场景建模而设计，进一步拓展了文本条件遥感图像合成的边界。

8) 异构数据 ↦ 视觉-音频： 在视觉-音频融合中，SoundingEarth [134]利用与遥感图像匹配的音频数据构建训练对。在对比学习框架下，该模型对齐音频和图像表示，从而为需要多模态理解的任务实现有效预训练。

D. 遥感基础模型简明综述

我们总结了单模态与多模态遥感基础模型，详述其相应的发表论文、预训练架构、下游应用和代码可用性。表 III和V 分别按单模态和多模态对这些模型进行分类，深入介绍其架构、实际应用及可用实现，以帮助读者有效区分并利用现有基础模型。

我们进一步展示了多模态基础模型相较于单模态基础模型在代表性遥感任务中的性能提升，如表 IV 所示。具体而言，多模态预训练模型在分类、分割和变化检测任务中始终表现更优。为评估这一提升，我们采用了 PANGAEA-Bench，这是一个综合基准，包含 11 个数据集，涵盖语义分割、变化检测和回归。这些数据集涉及多种时相设置（单时相、双时相和多时相）、全球地理分布、每像素 1.5 至 30 米的空间分辨率，以及异构传感器模态（RGB、光谱、合成孔径雷达）。因此，该基准为评估遥感

表 IV 单模态与多模态模型在 PANGAEA 基准上的性能比较。每个数据集上表现最佳的模型以红色标出，次佳的以蓝色标出，第三佳的加下划线。

模态	模型	BurnSr	MADOS	PASTIS	Sen1Fl11	FBP	DEN	CTM-SS	SN7	AI4Farms	Xview2	BioMassters
		野火	海洋	农业	洪水	土地	土地	农业	变化检测	农业	*CD*	回归
单模态	SatlasNet	79.96	55.86	17.51	90.30	50.97	36.31	46.97	61.88	25.13	52.23	41.67
	Prithvi 1.0	83.62	49.98	33.93	90.37	46.81	27.86	43.07	56.54	26.86	49.35	39.99
	SpectralGPT	80.47	57.99	35.44	89.07	33.42	37.85	46.95	58.86	26.75	48.40	36.11
	S.-S12-MoCo	81.58	51.76	34.49	89.26	53.02	35.44	48.58	57.64	25.38	51.59	40.21
	S.-S12-DINO	81.72	49.37	36.18	88.61	51.15	34.81	48.66	56.47	25.62	50.56	41.23
	S.-S12-MAE	81.91	49.90	32.03	87.79	51.92	34.08	45.80	57.13	24.69	50.44	41.07
多模态	RemoteCLIP	76.59	60.00	18.23	74.26	69.19	31.78	52.05	57.76	25.12	57.41	49.79
	CROMA	82.42	67.55	32.32	90.89	51.83	38.29	49.38	59.28	25.65	53.27	36.81
	DOFA	80.63	59.58	30.02	89.37	43.18	39.29	51.33	61.84	27.07	59.64	42.81
	TerraMind	82.42	69.52	40.51	90.62	59.72	37.87	55.80	60.61	28.12	-	-

基础模型的性能提供了一个严谨且统一的框架。如表 IV 所示，多模态模型在几乎所有任务中都占据了性能领先的位置。尤其是在 CTM-SS 等数据集上，它们的表现显著优于单模态模型，凸显了其在遥感应用中的卓越泛化能力和有效性。

IV. 如何在遥感下游任务中使用预训练基础模型：教程

遥感领域的预训练基础模型通过在海量数据集上进行大规模预训练，为下游任务提供强大的能力。如何将这些模型高效应用于各种任务，是遥感研究与实践中的一个重要关注点。本章将逐步介绍模型选择、环境配置、模型加载、微调和部署，并提供使用遥感基础模型的实用方法。图中展示了这些步骤的详细工作流。4。

A. 选择合适的遥感基础模型

不同的遥感基础模型在预训练数据、网络架构和模型特征方面各具特点。因此，选择合适的模型需要根据下游任务的需求进行仔细评估。

首先，应考虑预训练数据的来源和特征，例如所用传感器的类型、卫星平台、图像尺寸、通道数和数据格式（如整数或浮点数）。这些因素有助于评估模型的预训练数据与下游任务的数据需求之间的匹配程度。

其次，模型架构至关重要。评估模型采用的是 Transformer、卷积网络还是混合骨干网络，并考察其参数量和计算复杂度，以确保其符合任务的资源限制。此外，还需评估该架构是否满足轻量化部署或实时性能方面的特定要求。

最后，还应考虑模型的独特特性，例如其整合多模态数据或纳入时间信息的能力。这些特性可以显著提升模型在专业任务中的性能。

B. 模型配置

选择基础模型后，下一步是根据其文档配置相应的运行环境。核心目标是确保模型所需的库和框架得到正确设置。

大多数遥感基础模型依赖 Python 和 PyTorch 等深度学习框架。此外，通常还需要使用处理遥感数据的工具，例如 rasterio[2] 和 GDAL。设置环境时，应特别注意安装与所选模型兼容的框架版本。确保预训练权重已正确下载并存储在指定位置。

正确配置环境对于支持后续任务（如模型加载和微调）至关重要，也能确保整个工作流可靠运行。

C. 加载预训练基础模型

加载预训练模型的过程不仅涉及导入权重，还需要解决遥感数据独特特性所带来的挑战。如果不加以处理，图像尺寸和通道数的差异可能会导致预训练模型无法正常运行。

1) 处理图像尺寸不匹配问题：

- **问题：** 预训练模型通常针对固定图像尺寸设计（例如，$224 \times 224 \times 3$），而下游任务中的图像尺寸可能有所不同（例如，$128 \times 128 \times 3$）。
- **解决方案：** (i) 图像缩放或插值：调整输入图像的尺寸，使其符合预训练模型的要求。(ii) 位置编码插值：对于使用位置编码的模型（例如，Transformer），对编码值进行插值，以适应新的输入尺寸，同时保持空间一致性。*2)* 处理通道维度不匹配问题：

- **问题：** 遥感数据的通道数往往差异显著，例如单通道SAR图像，或具有数十甚至数百个通道的光谱图像。

[2]https://github. com/mapbox/rasterio

表 V

对遥感领域现有多模态基础模型的简明比较，按模态类型分类（例如，时间序列、地理视觉、同质、光学-SAR、异构、视觉大语言模型、视觉-文本和视觉-音频）。该表详述了它们相应的发表论文、预训练架构、下游应用和代码可用性。CL：对比学习，MAE：掩码自编码器，MIM：掩码图像建模，DIFF：扩散模型，JEPA：联合嵌入预测架构，SC：场景分类，SS：语义分割，RE：回归，CD：变化检测，OD：目标检测，SR：超分辨率，IG：图像生成，VP：视觉预测，VG：视觉定位，OT：目标跟踪，IR：图像检索，GM：几何测量，PS：全色锐化，REC：指代表达理解，TRE：时序指代表达，RES：指代表达分割，VQA：视觉问答，CR：去云

模态	模型	发表论文	架构	下游任务	代码
时间序列	SeCo [63]	2021年国际计算机视觉大会	CL	场景分类、变化检测	https://github.com/ServiceNow/seasonal-contrast
	CACo [64]	CVPR'23	CL	场景分类,语义分割, 变化检测	https://github.com/utkarshmall13/CACo
	RingMo-Sense [65]	TGRS 2023	MAE	目标检测,视觉定位,目标跟踪	https://github.com/
	U-BARN [66]	JSTARS'24	MAE	SS	https://src.koda.cnrs.fr/iris.dumeur/ssl_ubarn
	Prithvi2.0 [68]	Arxiv'24	MAE	场景分类、语义分割、视觉定位、地理匹配	https://github.com/NASA-IMPACT/Prithvi-EO-2.0
地理视觉	GASSL [70]	2021年国际计算机视觉大会	CL	场景分类、语义分割、目标检测	https://github.com/sustainlab-group/geography-aware-ssl
	GeoKR [69]	TGRS'21	CL	场景分类,语义分割, 目标检测	https://ieeexplore.ieee.org/abstract/document/9559903
	CSPT [71]	RS'22	MAE	场景分类，目标检测	https://github.com/ZhAnGToNG1/迁移学习 cspt
	GERSP [73]	TGRS'24	CL	语义分割，图像分类，目标检测	https://github.com/floatingstarZ/GeRSP
	SoftCon [74]	TGRS'24	CL	场景分类, 语义分割, 变化检测	https://github.com/zhu-xlab/softcon?tab=readme-ov-file
同质	USat [75]	Arxiv'23	MAE	SC	https://github.com/stanfordmlgroup/USat
	SenPa-MAE [77]	GCPR'25	MAE	SS	https://github.com/JonathanPrexl/SenPa-MAE
	CrossEarth [78]	Arxiv'24	监督式	SS	https://github.com/Cuzyoung/CrossEarth
	AnySat [79]	CVPR'25	JEPA	场景分类、语义分割	https://github.com/gastruc/AnySat
光学-SAR	RingMo [81]	TGRS 2022	MIM	场景分类、语义分割、变化检测、目标检测	https://github.com/comeony/RingMo
	SatViT [85]	GRSL 2022	MAE	SC	https://github.com/antofuller/SatViT
	DINO-MM [83]	IGARSS'22	CL	SC	https://github.com/zhu-xlab/DINO-MM
	RS-BYOL [84]	JSTARS'22	CL	场景分类，语义分割	https://github.com/
	CROMA [87]	NeurIPS'23	对比学习与掩码自编码器	场景分类、语义分割	https://github.com/antofuller/CROMA
	MANet [86]	Arxiv'24	监督式	SS	https://github.com/sstary/SSRS
	CSMAE [80]	TGRS'25	MAE	IR	https://github.com/jakhac/CSMAE
异构	SatLas [91]	ICCV 2023	监督式	场景分类、语义分割、目标检测、关系提取、地理匹配	https://github.com/allenai/satlaspretrain_模型
	DeCUR [98]	ECCV'24	CL	场景分类，语义分割	https://github.com/zhu-xlab/DeCUR
	DOFA [94]	Arxiv'24	MAE	场景分类，语义分割	https://github.com/zhu-xlab/DOFA
	msGFM [95]	CVPR'24	MAE	场景分类、语义分割、像素级分割、变化检测	https://github.com/boranhan/地理空间基础模型
	OmniSat [100]	ECCV'24	对比学习与掩码自编码器	场景分类，语义分割	https://github.com/gastruc/OmniSat?tab=readme-ov-file
	MMEarth [99]	ECCV'24	MAE	场景分类、语义分割	https://github.com/vishalned/MMEarth-train
	MM-SAM [97]	Arxiv'24	监督式	SS	https://xiaoaoran.github.io/projects/MM-SAM
视觉大语言模型	RSGPT [101]	ISPRS'25	监督式	视觉问答、图像分类	https://github.com/Lavender105/RSGPT
	GeoChat [105]	CVPR'24	监督式	视觉问答、图像分类、场景分类、指代表达理解、视觉定位、目标计数	https://github.com/mbzuai-oryx/GeoChat
	SkyEyeGPT [108]	ISPRS'25	监督式	视觉问答、图像分类、指代表达理解、指代表达生成	https://github.com/ZhanYang-nwpu/SkyEyeGPT
	VHM [135]	AAAI'25	监督式	场景分类、视觉问答、图像分类、指代表达理解、指代表达生成、目标计数、地理匹配	https://github.com/opendatalab/VHM
	RS-LLaVA [102]	RS'24	监督式	视觉问答，图像分类	https://github.com/BigData-KSU/RS-LLaVA?tab=readme-ov-file
	LHRS-Bot [111]	ECCV'24	监督式	场景分类，视觉问答，图像分类，指代表达理解	https://github.com/NJU-LHRS/LHRS-Bot
	EarthGPT [103]	TGRS'24	监督式	场景分类、视觉问答、图像分类、指代表达理解、指代表达生成	https://github.com/wivizhang/EarthGPT
	SkysenseGPT [119]	Arxiv'24	监督式	视觉问答、图像分类、指代表达理解、指代表达生成	https://github.com/Luo-Z13/SkySenseGPT
	RS-Agent [104]	Arxiv'24	智能体	视觉问答、目标计数、指代表达生成	https://github.com/IntelliSensing/IntelliSensing.github.io
	EarthMarker [107]	TGRS'24	监督式	视觉问答、场景分类、图像分类、视觉定位、指代表达理解	https://github.com/wivizhang/earthmarker
	LHRS-Bot-Nova [112]	ISPRS'25	监督式	场景分类、视觉问答、图像分类、目标检测、视觉定位、指代表达生成	https://github.com/NJU-LHRS/LHRS-Bot/tree/nova
	UniRS [110]	Arxiv'24	监督式	视觉问答、图像分类、场景分类、文本遥感图像检索	https://github.com/IntelliSensing/UniRS
	ChangeAgent [106]	TGRS'24	监督式	变化检测、文本遥感图像检索	https://github.com/Chen-Yang-Liu/Change-Agent
	RSUniVLM [114]	Arxiv'24	监督式	视觉问答、场景分类、视觉定位、变化检测、遥感图像分割、文本遥感图像检索	https://github.com/xuliu-cyber/RSUniVLM
	TEOChat [136]	ICLR 2025	监督式	视觉问答、指代表达理解、文本遥感图像检索	https://github.com/ermongroup/TEOChat
	GeoPixel [116]	ICML 2025	监督式	遥感图像分割、视觉问答、视觉定位、目标计数	https://github.com/mbzuai-oryx/GeoPixel
视觉-文本	RS-CLIP [123]	JAG'23	CL	SC	https://github.com/lx709/RS-CLIP
	RemoteCLIP [126]	TGRS'24	CL	场景分类, 图像检索, 目标计数	https://github.com/ChenDelong1999/RemoteCLIP
	GeoRSCLIP [122]	TGRS'24	CL	场景分类、图像检索	https://github.com/om-ai-lab/RS5M
	GeoCLIP [131]	NeurIPS'23	CL	场景分类、地理匹配、图像检索	https://vicentevivan.github.io/GeoCLIP/
	SatCLIP [124]	AAAI'25	CL	SC、GM、IR	https://github.com/microsoft/satclip
	SkyScript [125]	AAAI'24	CL	场景分类，图像检索	https://github.com/wangzhecheng/SkyScript
	GeoText [130]	ECCV'24	监督式	图像检索、视觉定位、目标计数	https://multimodalgeo.github.io/GeoText/
	Text2Seg [121]	SIGSPATIAL-W'24	CL	SS	https://github.com/Douglas2Code/Text2Seg
	DiffusionSat [128]	ICLR'24	Diff	图像生成、超分辨率	https://github.com/samar-khanna/DiffusionSat
	CRS-Diff [129]	TGRS'24	Diff	IG	https://github.com/Sonettoo/CRS-Diff
	Text2Earth [133]	GRSM'25	Diff	图像生成，图像检索	https://chen-yang-liu.github.io/Text2Earth
视觉-音频	Sounding Earth [134]	JAG'23	CL	场景分类、语义分割、图像检索	https://github.com/khdlr/SoundingEarth

- **解决方案：** (i) 通道复制或主成分分析（PCA）降维：复制通道，或使用 PCA 降低输入维度，使其符合模型的要求。(ii) 添加卷积层：在输入阶段引入卷积层，将通道数转换为与预训练模型相匹配的数量。(iii) 修改输入层：直接调整模型的输入层以适应新的通道数，确保与特定数据特征兼容。

D. 微调和部署基础模型

加载模型后，需要进行微调，使其适应特定任务。微调策略可以是冻结某些参数、仅训练特定模块（如输入层或输出层），也可以是对整个模型进行全面微调。

1）选择微调策略：

- **资源感知微调：** 对于计算资源有限的场景，冻结预训练参数，仅调整任务特定模块，以降低计算开销。

- **全面微调：** 对于需要高精度的任务，或资源充足时，可对整个模型进行微调，使其针对下游任务实现全面优化。

2) 部署微调后的模型：微调完成后，即可部署模型，对新数据进行预测。微调后的模型兼具预训练阶段带来的泛化能力和微调过程中的任务特定优化，因此在实际应用中通常能实现更高的准确率和鲁棒性。

E. 教程扩展：用于遥感的视觉语言基础模型

为何需要这一扩展？ 上述教程聚焦于纯视觉基础模型。然而，许多实际遥感应用需要结合影像与文本/元数据信息进行联合推理（例如场景描述、传感器元数据、时间或地理空间属性）。为回应审稿人的关切，我们提供了一份实用的分步指南，介绍如何在遥感工作流中采用视觉语言基础模型（VLFMs）。本节沿用前述流程，同时着重介绍多模态方面的具体内容。

1) 步骤 *1*：明确任务类型和数据配对：

- **任务分类体系：** 确定下游任务属于图像到文本（图像描述、报告生成）、文本到图像检索、视觉问答（*VQA*）、指令跟随/对话，还是带文本提示的多模态分类/检测。
- **配对数据构建：** 遥感语料库很少附有丰富的人工撰写图像描述。可以考虑：(i) 利用现有元数据（传感器名称、获取时间、地理位置、土地覆盖标签）自动生成弱描述；(ii) 使用大语言模型将嘈杂的元数据润色为自然语言；(iii) 采用教师模型（如通用视觉语言基础模型）为图像生成伪描述，再对关键数据集进行人工筛选。
- **文本粒度：** 确定是否需要图像级、图块级或目标级文本。细粒度监督通常能提升视觉问答/视觉定位效果，但会增加标注成本。*2*）步骤 *2*：选择/配置遥感视觉语言基础模型：

- **模型系列：**
  1) 编码器-解码器式（例如类似 BLIP-2 的模型）：由冻结的视觉编码器、轻量级连接器（Q-Former / MLP / cross-attn）以及冻结或部分可调的大语言模型组成。2) 统一 *Transformer*（两种模态共用单一骨干网络），联合编码视觉和文本词元。
  3) 遥感专用视觉语言基础模型（例如类似 GeoChat 的模型），已适配地理语义和遥感词表。
- **连接器设计：** 确保连接器能够处理遥感专用视觉词元（例如大量波段、大尺寸图块）。对于光谱丰富的数据，可在前面添加波段文本词元（“Band-11 (SWIR 1.61 μm)”）或波长描述符，以帮助语言端理解光谱语义。
- **分词器和词表：** 确认分词器中包含传感器名称、土地覆盖类别和地理空间术语；如果没有，则扩展词表，或通过别名映射到现有词元。

3) 步骤 *3*：加载并对齐多模态权重：

- **视觉侧不匹配：** 沿用第 IV 节中的图像尺寸/通道处理策略（调整大小、位置编码插值、输入适配器）。对于光谱输入，可考虑 (i) 光谱分组或 1×1卷积，同时 (ii) 将已丢弃/合并波段的文本描述符传递到语言侧，以弥补信息损失。

- **文本侧初始化：** 先加载大语言模型的分词器和权重；然后初始化或映射连接器参数。如果连接器要求固定数量的视觉词元，则对图像块词元进行插值或稀疏采样，使其数量匹配。
- **跨模态对齐头：** 部分检查点带有图像–文本对比学习或匹配头。替换编码器或词元化器时，请验证维度一致性。

4) 步骤 *4*：多模态微调策略：

- **轻量级适配（注重资源效率）：** 冻结视觉编码器和大语言模型，仅微调 (i) 连接器（例如 Q-Former）、(ii) 大语言模型中的 LoRA/适配器层，和/或 (iii) 用于光谱通道的浅层投影。这通常足以完成检索和图像说明生成。

- **指令微调：** 对于视觉问答或对话类任务，整理指令–响应配对数据（关于遥感场景的问题，以及以影像为依据的回答）。进行监督微调 (SFT)；如果有可用的人类偏好数据，还可以进一步进行偏好优化 (DPO/RLHF)。
- **多任务训练计划：** 结合图像—文本对比损失、图像条件语言建模和多语言目标（例如英语/中文），以增强泛化能力。

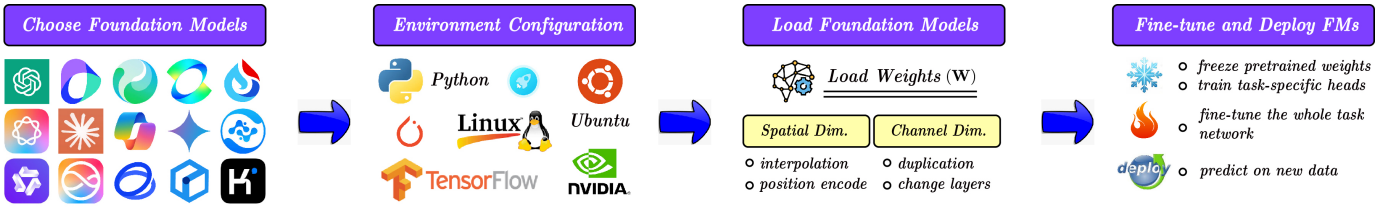


图 4. 将预训练基础模型应用于遥感下游任务的分步工作流。这个教程式流程从模型选择开始，接着进行环境配置和模型加载，最后通过微调和部署所选基础模型，将其用于特定的遥感应用。

5) 步骤 *5*：部署与提示：

- **提示设计：** 提供能够呈现遥感语境的结构化提示："你是一名遥感分析师。给定一幅 Sentinel-2 图像（地面采样距离为 10 米，波段：B2、B3、B4、B8），请回答问题。"此类明确提示可减少幻觉。

- **推理效率：** 对于大型大语言模型，可将其蒸馏为更小的语言头，或缓存键值状态以处理重复查询。在视觉方面，可预先计算图像嵌入，并为检索任务建立索引。

- **评估：** 除了标准自然语言处理指标（BLEU、CIDEr）外，还应采用适用于遥感的衡量指标（土地覆盖精度、地理空间定位准确率），并针对安全关键型任务进行人工专家评估。

要点。将视觉语言基础模型整合到遥感处理流程中，需要（1）构建或整理可靠的图像-文本对，（2）选择合适的多模态架构/连接器，（3）仔细对齐光谱/空间词元与文本描述符，以及（4）采用注重资源效率的微调和提示策略。遵循这些步骤，可在实际遥感应用中实现稳健的跨模态理解与交互。

F. 基础模型的计算资源

为帮助读者或潜在研究人员根据其计算资源选择合适的基础模型，我们在表VI中总结了遥感基础模型的代表性骨干架构。除非另有说明，所有数值均在以下一致假设下计算（或重新计算）：（*i*）单次前向传播，（*ii*）FP32精度，（*iii*）批量大小为1，以及（*iv*）输入分辨率为 $3 \times 224 \times 224$，适用于基于卷积神经网络、视觉变换器和Swin的模型。对于SAM变体，浮点运算次数按其原始输入设置（1024×1024）报告，如SAM论文所述。浮点运算次数表示每次前向传播的浮点运算；当报告乘累加运算（乘累加运算次数）时，我们将其乘以二转换为浮点运算次数。参数量（"参数量"）指可学习权重的总数，而"推理内存"表示在NVIDIA图形处理器上测得的经验范围，会因驱动程序、PyTorch版本和激活检查点的不同而有所变化。对于广泛研究的模型（例如，ResNet-50、ViT-B/L/H、Swin-T/B/L），我们采用官方或普遍接受的参考实现。

我们还报告了遥感领域主流开源基础模型的计算概况，以补充骨干网络层面的总结，详见表 VI。此外，表 VII 列出了各模型的复杂度（以十亿乘累加运算次数计）、可训练参数量和总参数量，数据均来自 PANGAEA-Bench 发布版本。除非另有说明，所有十亿乘累加运算次数均按照基准测试的默认输入配置，测量单次前向传播的乘累加运算次数（1 MAC = 2 浮点运算次数）。「可训练参数」指微调下游任务时更新的权重子集。该表旨在帮助读者在选择当前主流遥感基础模型时，权衡潜在的准确率提升与计算成本。

V. 结论

基础模型通过支持从大规模异构数据源中进行可扩展、可迁移的学习，彻底改变了遥感（RS）领域。早期的遥感基础模型主要聚焦于单模态表征，在各类传感模态内学习通用特征。近年来，这一范式逐渐转向多模态基础模型，联合利用光学 RGB、合成孔径雷达、光谱、激光雷达及辅助数据源中的互补信息。借助大规模预训练和任务自适应微调，这些模型构建了统一且稳健的框架，可用于图像分类、目标检测、语义分割、变化检测等众多下游任务。

此外，本文还从教程的角度出发，系统阐述了在遥感应用中有效使用预训练基础模型所需的关键步骤，包括模型选择、环境配置、模型加载、微调和部署。每个步骤都需要仔细考虑遥感数据的独特特性、计算约束以及领域特定挑战，以确保性能稳健并具备实际适用性。

VI. 未来方向与挑战

尽管取得了显著进展，遥感基础模型的发展仍面临若干亟待进一步研究的开放性挑战：

- **缩放定律与模型增长：** 了解缩放定律是否适用于遥感基础模型，对于指导模型开发、优化数据需求，以及平衡计算效率与性能至关重要。

表 VI 遥感基础模型代表性骨干架构的计算资源。

模型	架构	参数量（M）	浮点运算次数（G）	推理内存（GB）
ResNet-50	ResNet-50	25.6	4.1	1–2
ViT-Base	ViT-B/16	86.0	17.6	2–4
ViT-Large	ViT-L/16	304.0	61.6	4–8
ViT-Huge	ViT-H/14	632.0	190.7	8–12
SAM-B	ViT-B/16 (SAM)	91.0	22.0	2–4
SAM-L	ViT-L/16 (SAM)	308.0	62.0	4–8
SAM-H	ViT-H/14 (SAM)	641.0	200.0	8–12
Swin-T	Swin-T	28.3	4.5	1–2
Swin-B	Swin-B	87.8	15.4	2–4
Swin-L	Swin-L	197.3	34.5	4–6

表 VII 遥感代表性基础模型的计算统计。

模型	浮点运算次数 (G)	可训练参数 (M)	总参数（M）
CROMA	278.46	46.95	350.0
DOFA	130.04	39.35	150.9
GFM-Swin	38.10	33.71	120.4
Prithvi	129.70	39.34	125.7
RemoteCLIP	35.14	39.35	16.8
SatlasNet	33.40	33.25	121.2
Scale-MAE	207.54	46.95	350.1
SpectralGPT	761.38	164.40	249.8
S12-MoCo	93.80	30.89	53.5
S12-DINO	95.86	30.89	53.5
S12-MAE	93.80	30.89	53.5
S12-Data2Vec	93.78	30.89	53.5

- **泛化能力与模型标准的定义：**未来的研究应确立明确的标准，以界定何为通用或规模足够大的遥感基础模型，并阐明此类模型是否必须采用自监督方式。
- **预训练模型的比较分析：**需要深入探究为何在许多情况下，随机权重或在其他领域预训练的模型能够取得与现有遥感专用基础模型相当的结果。
- **灾难性遗忘与知识保留：**解决训练后的遥感基础模型中的灾难性遗忘问题，对于确保长期适应性和高效的持续学习至关重要。
- **数据管理与可扩展性：**大规模遥感数据的管理，包括预处理、标注和高效存储，仍然是一大瓶颈。未来研究应探索标准化框架和策略，以促进遥感基础模型的可扩展开发。
- **多模态适应性与开放协作：**提升模型对不同遥感数据模态的适应能力，并推动共享预训练模型、数据集和基准的协作，将加速整个领域的创新。
- **鲁棒性与不确定性：**未来的遥感基础模型应超越事后鲁棒性评估（例如，REOBench [137]），转向鲁棒性感知与不确定性感知预训练，显式建模传感器噪声、领域偏移和极端条件。

整合不确定性量化（例如概率表示或贝叶斯启发式学习）对于在真实世界的地球观测应用中实现可靠决策和风险感知部署至关重要。

通过应对这些挑战，基础模型将继续推动遥感领域的进步，使环境监测更加准确、促进可持续资源管理，并加深我们对地球动态过程的认识。

VII. 主要缩略语

为便于阅读，现将全文中使用的所有缩略语汇总如下。希望此列表能帮助读者更轻松地查找和理解这些缩略语。

CNNs：卷积神经网络 CROMA：对比式雷达-光学掩码自编码器 CSPT：连续预训练 CWSAM：分类别 SAM DOFA：动态万全 EO：地球观测 FG-MAE：特征引导掩码自编码器 GASSL：地理感知自监督学习 GDAL：地理空间数据抽象库 GeoKR：地理知识驱动表征学习 GFM：地理空间基础模型

GRAFT：地面遥感对齐训练 GSD：地面采样距离 HOGs：方向梯度直方图 LoRA：低秩自适应 LMM：大语言模型 MAE：掩码自编码器 MAF-SAM：多阶段自适应微调分割一切模型 MANet：多模态适配器网络 MATTER：材质与纹理表征学习 MLP：多层感知机 msGFM：多传感器地理空间基础模型 OmniSat：全向卫星 OFA-Net：万全网络 RingMo：遥感基础模型 RNNs：循环神经网络 RS：遥感 RS-BYOL：遥感自举潜在表征 RS-DFM：遥感分布式基础模型 RVSA：旋转可变尺寸注意力 S2MAE：空谱掩码自编码器 SatMAE：卫星掩码自编码器 SatViT：卫星视觉Transformer SpectralGPT：光谱生成式预训练Transformer SAM：分割一切模型 SAR：合成孔径雷达 SARATR-X：SAR 自动目标识别 SAR-JEPA：SAR 联合嵌入预测架构 SeaMo：季节性多模态基础模型 SenPa-MAE：传感器参数掩码自编码器 SoftCon：软对比学习 ViTs：视觉 Transformer

VIII. 致谢

本研究得到中国国家自然科学基金项目（批准号：42271350）以及中国科学院国际合作伙伴计划项目（批准号：313GJHZ2023066FN）的支持。GCV 衷心感谢欧洲研究理事会（ERC）通过 USMILE 项目（资助协议号：855187）以及欧盟委员会通过 HORIZON 项目 ELIAS（资助协议号：101120237）和 ELLIOT（资助协议号：101214398）提供的支持。

参考文献

[1] D. Hong、C. Li、B. Zhang、N. Yokoya、J. A. Benediktsson 和 J. Chanussot，“多模态人工智能基础模型：释放遥感大数据在地球观测中的力量”，*The Innovation Geoscience*，第 2 卷第 1 期，页码 100055，2024。[2] C. Li、D. Hong、B. Zhang、T. Liao、N. Yokoya、P. Ghamisi、M. Chen、L. Wang、J. A. Benediktsson 和 J. Chanussot，“可解释基础模型作为解读地球系统的解码器”，*The Innovation*，第 5 卷第 5 期，2024。[3] C. Bodnar、W. P. Bruinsma、A. Lucic、M. Stanley、A. Vaughan、J. Brandstetter、P. Garvan、M. Riechert、J. A. Weyn、H. Dong、J. K. Gupta、K. Thambiratnam、A. T. Archibald、C.-C. Wu、E. Heider、M. Welling、R. E. Turner 和 P. Perdikaris，“地球系统基础模型”，*arXiv* 预印本 *arXiv:2405.13063*，2024。[4] S. Chen、G. Long、J. Jiang、D. Liu 和 C. Zhang，“用于理解天气与气候数据的基础模型：综合综述”，*arXiv* 预印本 *arXiv:2312.03014*，2023。

[5] E. Rolf、K. Klemmer、C. Robinson 和 H. Kerner，《任务关键型——卫星数据是机器学习中的一种独特模态》，载于机器学习国际会议，PMLR，2024年。[6] C. Li、D. Hong、B. Zhang、Y. Li、G. Camps-Valls、X. X. Zhu 和 J. Chanussot，《Urbansam：为城市建设中的分割一切模型学习受不变性启发的适配器》，arXiv预印本 arXiv:2502.15199，2025年。[7] L. Jiao、Z. Huang、X. Lu、X. Liu、Y. Yang、J. Zhao、J. Zhang、B. Hou、S. Yang、F. Liu 等，《脑启发式遥感基础模型与开放问题：综合综述》，IEEE《应用地球观测与遥感专题期刊》，第16卷，第10084–10120页，2023。[8] C. Tan、Q. Cao、Y. Li、J. Zhang、X. Yang、H. Zhao、Z. Wu、Z. Liu、H. Yang、N. Wu 等，《多模态基础模型在地理、环境、农业和城市规划应用中的前景与挑战》，arXiv预印本 arXiv:2312.17016，2023。[9] X. Li、C. Wen、Y. Hu、Z. Yuan 和 X. X. Zhu，《遥感中的视觉语言模型：当前进展与未来趋势》，IEEE地球科学与遥感杂志，2024。[10] X. X. Zhu、Z. Xiong、Y. Wang、A. J. Stewart、K. Heidler、Y. Wang、Z. Yuan、T. Dujardin、Q. Xu 和 Y. Shi，《论地球与气候基础模型的基础》，arXiv预印本 arXiv:2405 04285，2024年。。[11] Y. Zhou、L. Feng、Y. Ke、X. Jiang、J. Yan、X. Yang 和 W. Zhang，《迈向视觉-语言地理基础模型：综述》，arXiv预印本 arXiv:2406.09385，2024。[12] S. Lu、J. Guo、J. R. Zimmer-Dauphinee、J. M. Nieusma、X. Wang、P. VanValkenburgh、S. A. Wernke 和 Y. Huo，《遥感中的人工智能基础模型：综述》，arXiv预印本 arXiv:2408.03464，2024。[13] M. Zhang、B. Yang、X. Hu、J. Gong 和 Z. Zhang，《面向通用遥感智能的基础模型：潜力与前景》，《科学通报》，第69卷，第23期，第3652–3656页，2024。[14] A. Xiao、W. Xuan、J. Wang、J. Huang、D. Tao、S. Lu 和 N. Yokoya，《遥感与地球观测基础模型：综述》，arXiv预印本 arXiv:2410.16602，2024。[15] H. Zhang、J.-J. Xu、H.-W. Cui、L. Li、Y. Yang、C.-S. Tang 和 N. Boers，《地球科学遇上基础模型：迈向通用地球科学人工智能系统》，IEEE地球科学与遥感杂志，2024。[16] L. Tao、H. Zhang、H. Jing、Y. Liu、D. Yan、G. Wei 和 X. Xue，《遥感视觉语言模型的进展：数据集、能力与增强技术》，Remote Sensing，第17卷，第1期，第162页，2025年。[17] C. Huo、K. Chen、S. Zhang、Z. Wang、H. Yan、J. Shen、Y. Hong、G. Qi、H. Fang 和 Z. Wang，《遥感遇上基础模型：综述及后续进展》，remote sensing，第17卷，第2期，2025年。[18] G. Sumbul、M. Charfuelan、B. Demir 和 V. Markl，《BigEarthNet：用于遥感图像理解的大规模基准档案》，载于 IGARSS 2019——2019 IEEE国际地球科学与遥感研讨会，第5901–5904页，IEEE，2019年。[19] D. Hong、B. Zhang、X. Li、Y. Li、C. Li、J. Yao、N. Yokoya、H. Li、P. Ghamisi、X. Jia 等，《SpectralGPT：光谱遥感基础模型》，IEEE模式分析与机器智能汇刊，第46卷，第8期，第5227–5244页，2024。[20] M. Schmitt、L. H. Hughes、C. Qiu 和 X. X. Zhu，《Sen12ms——为深度学习与数据融合整理的地理参考多光谱 Sentinel-1/2 影像数据集》，arXiv预印本 arXiv:1906.07789，2019年。[21] D. Hong、B. Zhang、H. Li、Y. Li、J. Yao、C. Li、M. Werner、J. Chanussot、A. Zipf 和 X. X. Zhu，《跨城市至关重要：用于利用高分辨率领域自适应网络进行跨城市语义分割的多模态遥感基准数据集》，Remote Sensing of Environment，第299卷，第113856页，2023。[22] Y. Wang、N. A. A. Braham、Z. Xiong、C. Liu、C. M. Albrecht 和 X. X. Zhu，《Ssl4eo-s12：用于地球观测自监督学习的大规模多模态、多时相数据集 [软件与数据集]》，IEEE地球科学与遥感杂志，第11卷，第3期，第98–106页，2023。[23] A. Stewart、N. Lehmann、I. Corley、Y. Wang、Y.-C. Chang、N. A. Ait Ali Braham、S. Sehgal、C. Robinson 和 A. Banerjee，《Ssl4eo-l：Landsat影像数据集与基础模型》，神经信息处理系统进展，第36卷，59807，2023 59787页–。[24] S. Dou、Y. Wu、H. Jia、Y. Zhou、Y. Liu 和 Y. Liu，《Cc2vec：结合类型化词元与对比学习，实现有效的代码克隆检测》，《ACM软件工程会刊》，第1卷，FSE期，第1564–1584页，2024。

[25] P. Akiva、M. Purri 和 M. Leotta，《面向遥感任务的自监督材料与纹理表征学习》，载于 CVPR 会议录，第8203–8215页，2022年。[26] D. Wang、Q. Zhang、Y. Xu、J. Zhang、B. Du、D. Tao 和 L. Zhang，《推动纯视觉变换器迈向遥感基础模型》，IEEE地球科学与遥感汇刊，第61卷，第1–15页，2022年。[27] Y. Long、G.-S. Xia、S. Li、W. Yang、M. Y. Yang、X. X. Zhu、L. Zhang 和 D. Li，《关于创建航空影像解译基准数据集：综述、指南与 Million-AID》，IEEE《应用地球观测与遥感专题期刊》，第14卷，第4205–4230页，2021年。[28] D. Muhtar、X. Zhang、P. Xiao、Z. Li 和 F. Gu，《Cmid：用于遥感图像理解的统一自监督学习框架》，IEEE地球科学与遥感汇刊，第61卷，第1–17页，2023。[29] M. Zhang、Q. Liu 和 Y. Wang，《Ctxmim：用于遥感图像理解的上下文增强掩码图像建模》，arXiv预印本 arXiv:2310.00022，2023。[30] M. Mendieta、B. Han、X. Shi、Y. Zhu 和 C. Chen，《通过持续预训练迈向地理空间基础模型》，载于 ICCV 会议录，第16806–16816页，2023。[31] C. J. Reed、R. Gupta、S. Li、S. Brockman、C. Funk、B. Clipp、K. Keutzer、S. Candido、M. Uyttendaele 和 T. Darrell，《Scale-MAE：面向多尺度地理空间表征学习的尺度感知掩码自编码器》，载于 IEEE/CVF国际计算机视觉会议论文集，第4088–4099页，2023。[32] M. Noman、M. Naseer、H. Cholakkal、R. M. Anwer、S. Khan 和 F. S. Khan，《重新思考多光谱卫星影像的Transformer预训练》，载于 IEEE/CVF计算机视觉与模式识别会议论文集，第27811–27819页，2024。[33] F. Wang、H. Wang、D. Wang、Z. Guo、Z. Zhong、L. Lan、J. Zhang、Z. Liu 和 M. Sun，《在大型遥感数据集上扩展高效掩码自编码器学习》，arXiv电子预印本，第arXiv–2406页，2024。[34] Z. Li、B. Hou、S. Ma、Z. Wu、X. Guo、B. Ren 和 L. Jiao，《面向遥感图像的掩码角度感知自编码器》，载于欧洲计算机视觉会议，第260–278页，Springer，2024。[35] W. Diao、H. Yu、K. Kang、T. Ling、D. Liu、Y. Feng、H. Bi、L. Ren、X. Li、Y. Mao 等，《Ringmo-aerial：采用仿射变换对比学习的航空遥感基础模型》，arXiv预印本 arXiv:2409.13366，2024。[36] K. Cha、J. Seo 和 T. Lee，《面向遥感图像的十亿级基础模型》，IEEE《应用地球观测与遥感专题期刊》，2024。[37] P. Dias、A. Tsaris、J. Bowman、A. Potnis、J. Arndt、H. L. Yang 和 D. Lunga，《Oreole-fm：迈向面向高分辨率卫星影像的十亿参数基础模型——成果与挑战》，载于第32届 ACM地理信息系统进展国际会议论文集，第597–600页，2024。[38] Z. Yu、C. Liu、L. Liu、Z. Shi 和 Z. Zou，《Metaearth：用于全球尺度遥感图像生成的生成式基础模型》，IEEE模式分析与机器智能汇刊，2024。[39] M. Tang、A. Cozma、K. Georgiou 和 H. Qi，《Cross-Scale MAE：遥感领域多尺度利用的故事》，神经信息处理系统进展，第36卷，第20054–20066页，2023。[40] K. Lu、R. Zhang、X. Huang、Y. Xie、X. Ning、H. Zhang、M. Yuan、P. Zhang、T. Wang 和 T. Liao，《用于遥感自监督学习的模式集成与增强视觉变换器》，IEEE地球科学与遥感汇刊，2025年。[41] X. Zhou、F. Liang、L. Chen、H. Liu、Q. Song、G. Vivone 和 J. Chanussot，《Mesam：面向光学遥感图像的多尺度增强分割一切模型》，IEEE地球科学与遥感汇刊，2024。[42] J. Zhang、X. Yang、R. Jiang、W. Shao 和 L. Zhang，《Rsam-seg：一种将先验知识融合用于遥感图像语义分割的SAM方法》，arXiv预印本 arXiv:2402 19004，2024年。。[43] K. Chen、C. Liu、H. Chen、H. Zhang、W. Li、Z. Zou 和 Z. Shi，《Rsprompter：基于视觉基础模型学习提示以实现遥感实例分割》，IEEE地球科学与遥感汇刊，第62卷，第1–17页，2024。[44] X. Zhang、Y. Liu、Y. Lin、Q. Liao 和 Y. Li，《Uv-sam：调整分割一切模型以识别城中村》，载于 AAAI人工智能会议论文集，第38卷，第22520–22528页，2024。[45] S. Zhang、Q. Wang、J. Liu和H. Xiong，“Alps：一种基于分割一切模型的遥感分割自动标注与预训练方案”，arXiv预印本 arXiv:2406.10855，2024。[46] D. Hong、C. Li、N. Yokoya、B. Zhang、X. Jia、A. Plaza、P. Gamba、J. A. Benediktsson和J. Chanussot，“高光谱成像”，Nature Reviews Methods Primers，2026。[47] Y. Cong、S. Khanna、C. Meng、P. Liu、E. Rozi、Y. He、M. Burke、D. Lobell和S. Ermon，“Satmae：面向时序多光谱卫星影像的Transformer预训练”，Proc. NeurIPS，第35卷，第197–211页，2022年。[48] X. Wanyan、S. Seneviratne、S. Shen和M. Kirley，“扩展全局-局部视图对齐以用于遥感影像自监督学习”，载于IEEE/CVF计算机视觉与模式识别会议论文集，第2443–2453页，2024。[49] C.-Y. Hsu、W. Li和S. Wang，“面向图像分析的地理空间基础模型：评估并增强NASA-IBM Prithvi的领域适应性”，《国际地理信息科学杂志》，第1–30页，2024。[50] B. Song、H. Yang、Y. Wu、P. Zhang、B. Wang和G. Han，“一种基于分割一切模型并采用多阶段适应微调的多光谱遥感作物分割方法”，IEEE地球科学与遥感汇刊，2024。[51] D. Ibanez、R. Fernandez-Beltran、F. Pla和N. Yokoya，“用于高光谱图像分类的掩码自编码光谱–空间Transformer”，IEEE地球科学与遥感汇刊，第60卷，第1–14页，2022年。[52] L. Scheibenreif、M. Mommert和D. Borth，“用于高光谱图像分类的掩码视觉Transformer”，载于IEEE/CVF计算机视觉与模式识别会议研讨会（CVPRW）论文集，第2166–2176页，2023。[53] Y. Wang、M. Wen、H. Zhang、J. Sun、Q. Yang、Z. Zhang和H. Lu，“Hsimae：一种面向高光谱图像分类的大规模预训练统一掩码自编码器”，IEEE《应用地球观测与遥感专题期刊》，2024。[54] N. A. A. Braham、C. M. Albrecht、J. Mairal、J. Chanussot、Y. Wang和X. X. Zhu，“Spectralearth：大规模训练高光谱基础模型”，arXiv预印本 arXiv:2408.08447，2024。[55] D. Wang、M. Hu、Y. Jin、Y. Miao、J. Yang、Y. Xu、X. Qin、J. Ma、L. Sun、C. Li等，“Hypersigma：高光谱智能理解基础模型”，arXiv预印本arXiv:2406.11519，2024。[56] L. Pang、X. Cao、D. Tang、S. Xu、X. Bai、F. Zhou和D. Meng，“Hsigene：一种用于高光谱图像生成的基础模型”，arXiv预印本 arXiv:2409.12470，2024。[57] Y. Wang、H. H. Hernandez、C. M. Albrecht和X. X. Zhu，“面向遥感自监督学习的特征引导掩码自编码器”，IEEE《应用地球观测与遥感专题期刊》，2024。[58] W. Li、W. Yang、T. Liu、Y. Hou、Y. Li、Z. Liu、Y. Liu和L. Liu，“预测梯度更优：利用联合嵌入预测架构探索面向SAR自动目标识别的自监督学习”，ISPRS摄影测量与遥感杂志，第218卷，第326–338页，2024。[59] X. Pu、H. Jia、L. Zheng、F. Wang和F. Xu，“Classwise-sam-adapter：参数高效微调使分割一切模型适配SAR领域以进行语义分割”，arXiv预印本arXiv:2401.02326，2024。[60] W. Li、W. Yang、Y. Hou、L. Liu、Y. Liu和X. Li，“Saratr-x：迈向构建用于SAR目标识别的基础模型”，IEEE图像处理汇刊，2025年。[61] G. Vivone、L.-J. Deng、S. Deng、D. Hong、M. Jiang、C. Li、W. Li、H. Shen、X. Wu、J.-L. Xiao等，“遥感图像融合中的深度学习：方法、协议、数据与未来展望”，IEEE地球科学与遥感杂志，2024。[62] D. Hong、L. Gao、N. Yokoya、J. Yao、J. Chanussot、Q. Du和B. Zhang，“更多样即更好：多模态深度学习与遥感影像分类的结合”，IEEE地球科学与遥感汇刊，第59卷，第5期，第4340–4354页，2020。[63] O. Manas、A. Lacoste、X. Gir´o-i Nieto、D. Vazquez和P. Rodriguez，“季节对比：基于未经筛选的遥感数据的无监督预训练”，载于IEEE/CVF国际计算机视觉会议论文集，第9414–9423页，2021年。[64] U. Mall、B. Hariharan和K. Bala，“面向卫星图像的变化感知采样与对比学习”，载于IEEE/CVF计算机视觉与模式识别会议论文集，第5261–5270页，2023。[65] F. Yao、W. Lu、H. Yang、L. Xu、C. Liu、L. Hu、H. Yu、N. Liu、C. Deng、D. Tang等，“Ringmo-sense：通过时空演化解耦实现时空预测的遥感基础模型”，IEEE地球科学与遥感汇刊"，2023年。

[66] I. Dumeur、S. Valero 和 J. Inglada，《卫星影像时间序列的自监督时空表征学习》，IEEE《应用地球观测与遥感专题期刊》，2024年。[67] L. Zhang、Y. Zhao、R. Dong、J. Zhang、S. Yuan、S. Cao、M. Chen、J. Zheng、W. Li、W. Liu 等，《A^2-mae：一种基于锚点感知掩码自编码器的时空光谱统一遥感预训练方法》，arXiv 预印本 arXiv:2406.08079，2024年。[68] D. Szwarcman、S. Roy、P. Fraccaro、O. E. Gislason、B. Blumenstiel、R. Ghosal、P. H. de Oliveira、J. L. d. S. Almeida、R. Sedona、Y. Kang 等，《Prithvi-eo-2.0：面向地球观测应用的通用多时相基础模型》，arXiv 预印本 arXiv:2412 02732，2024年。.[69] W. Li、K. Chen、H. Chen 和 Z. Shi，《遥感图像的地理知识驱动表征学习》，IEEE 地球科学与遥感汇刊，第60卷，第1–16页，2021年。[70] K. Ayush、B. Uzkent、C. Meng、K. Tanmay、M. Burke、D. Lobell 和 S. Ermon，《地理感知自监督学习》，载于 ICCV 论文集，第10181–10190页，2021年。[71] T. Zhang、P. Gao、H. Dong、Y. Zhuang、G. Wang、W. Zhang 和 H. Chen，《连续预训练：一种利用相关无标签数据的遥感领域知识迁移学习策略》，《遥感》，第14卷，第22期，文章编号5675，2022年。[72] G. Mai 、 N. Lao 、 Y. He 、 J. Song 和 S. Ermon，《 Csp：面向地理空间视觉表征的自监督对比式空间预训练》，载于国际机器学习会议论文集，第23498–23515页，PMLR，2023。[73] Z. Huang、M. Zhang、Y. Gong、Q. Liu 和 Y. Wang，《面向遥感图像的通用知识增强预训练》，IEEE地球科学与遥感汇刊，第62卷，13，2024年1pp. – 。[74] Y. Wang、C. M. Albrecht 和 X. X. Zhu，《用于高效地球观测预训练的多标签引导软对比学习》，arXiv 预印本 arXiv:2405.20462，2024年。[75] J. Irvin、L. Tao、J. Zhou、Y. Ma、L. Nashold、B. Liu 和 A. Y. Ng，《Usat：用于多传感器卫星影像的统一自监督编码器》，arXiv 预印本 arXiv:2312.02199，2023年。[76] Z. Wang、P. Cheng、P. Tian、Y. Wang、M. Chen、S. Duan、Z. Wang、X. Li 和 X. Sun，《Rs-dfm：面向多样化下游任务的遥感分布式基础模型》，arXiv 预印本 arXiv:2406 07032，2024年。.[77] J. Prexl 和 M. Schmitt，《Senpa-mae：用于多卫星自监督预训练的传感器参数感知掩码自编码器》，arXiv 预印本 arXiv:2408.11000，2024年。[78] Z. Gong、Z. Wei、D. Wang、X. Ma、H. Chen、Y. Jia、Y. Deng、Z. Ji、X. Zhu、N. Yokoya 等，《Crossearth：用于领域泛化遥感语义分割的地理空间视觉基础模型》，arXiv 预印本 arXiv:2410.22629，2024年。[79] G. Astruc、N. Gonthier、C. Mallet 和 L. Landrieu，《Anysat：适用于任意分辨率、尺度和模态的地球观测模型》，arXiv 预印本 arXiv:2412.14123，2024年。[80] J. Hackstein、G. Sumbul、K. N. Clasen 和 B. Demir，《探索掩码自编码器在遥感传感器无关图像检索中的应用》，IEEE地球科学与遥感汇刊，2024年。[81] X. Sun、P. Wang、W. Lu、Z. Zhu、X. Lu、Q. He、J. Li、X. Rong、Z. Yang、H. Chang 等，《Ringmo：采用掩码图像建模的遥感基础模型》，IEEE地球科学与遥感汇刊，第61卷，第1–22页，2022年。[82] J. Prexl 和 M. Schmitt，《面向 Sentinel-1/2 影像的多模态多目标对比学习》，载于《IEEE/CVF计算机视觉与模式识别会议》论文集，第2136–2144页，2023年。[83] Y. Wang、C. M. Albrecht 和 X. X. Zhu，《用于联合SAR-光学表征学习的自监督视觉变换器》，载于 IGARSS 2022——2022年 IEEE 国际地球科学与遥感研讨会论文集，第139–142页，IEEE，2022年。[84] P. Jain、B. Schoen-Phelan 和 R. Ross，《从多光谱与SAR图像中学习不变表征的自监督学习》，IEEE《应用地球观测与遥感专题期刊》，第15卷，第7797–7808页，2022年。[85] A. Fuller、K. Millard 和 J. R. Green，《Satvit：面向地球观测的Transformer预训练》，IEEE地球科学与遥感快报，第19卷，第1–5页，2022年。[86] X. Ma、X. Zhang、M.-O. Pun 和 B. Huang，《Manet：微调分割一切模型以实现多模态遥感语义分割》，arXiv 预印本 arXiv:2410.11160，2024年。

[87] A. Fuller, K. Millard, and J. Green, “Croma：采用对比式雷达-光学掩码自编码器的遥感表征”，NeurIPS论文集，第36卷，2024年。[88] H. Hu, P. Wang, H. Bi, B. Tong, Z. Wang, W. Diao, H. Chang, Y. Feng, Z. Zhang, Q. Ye等，“Rs-vheat：热传导引导的高效遥感基础模型”，arXiv预印本 arXiv:2411 17984，2024年。[89] X. Li, C. Li, G. Vivone, and D. Hong, “Seamo：面向遥感的季节感知多模态基础模型”，Information Fusion，第125卷，第103334页，2026年。[90] X. Guo, J. Lao, B. Dang, Y. Zhang, L. Yu, L. Ru, L. Zhong, Z. Huang, K. Wu, D. Hu等，“Skysense：迈向地球观测影像通用解译的多模态遥感基础模型”，载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第27672–27683页，2024年。[91] F. Bastani, P. Wolters, R. Gupta, J. Ferdinando, and A. Kembhavi, “Satlaspretrain：用于遥感图像理解的大规模数据集”，载于《IEEE/CVF国际计算机视觉会议论文集》，第16772–16782页，2023年。[92] Z. Yan, J. Li, X. Li, R. Zhou, W. Zhang, Y. Feng, W. Diao, K. Fu, and X. Sun, “Ringmo-sam：用于在多模态遥感图像中分割任何对象的基础模型”，IEEE地球科学与遥感汇刊，第61卷，第1–16页，2023年。[93] Z. Xiong, Y. Wang, F. Zhang, and X. X. Zhu, “One for all：迈向地球视觉统一基础模型”，arXiv预印本 arXiv:2401.07527，2024年。[94] Z. Xiong, Y. Wang, F. Zhang, A. J. Stewart, J. Hanna, D. Borth, I. Papoutsis, B. Le Saux, G. Camps-Valls, and X. X. Zhu, “受神经可塑性启发的跨模态地球观测基础模型”，arXiv电子预印本，第arXiv–2403页，2024年。[95] B. Han, S. Zhang, X. Shi, and M. Reichstein, “利用多传感器地理空间基础模型连接遥感传感器”，载于CVPR论文集，第27852–27862页，2024年。[96] P. Ravirathinam, A. Khandelwal, R. Ghosh, and V. Kumar, “迈向知识引导的时空遥感应用多模态基础模型”，arXiv预印本 arXiv:2407.19660，2024年。[97] A. Xiao, W. Xuan, H. Qi, Y. Xing, N. Yokoya, and S. Lu, “利用多种模态分割任何对象”，arXiv预印本 arXiv:2408 09085，2024年。[98] Y. Wang, C. M. Albrecht, N. A. A. Braham, C. Liu, Z. Xiong, and X. X. Zhu, “解耦多模态自监督学习中的共有与特有表征”，载于《欧洲计算机视觉会议》，第286–303页，Springer，2024年。[99] V. Nedungadi, A. Kariryaa, S. Oehmcke, S. Belongie, C. Igel, and N. Lang, “MMEarth：探索用于地理空间表征学习的多模态前置任务”，载于《欧洲计算机视觉会议》，第164–182页，Springer，2024年。[100] G. Astruc, N. Gonthier, C. Mallet, and L. Landrieu, “OmniSat：用于地球观测的自监督模态融合”，载于《欧洲计算机视觉会议》，第409–427页，Springer，2025年。[101] Y. Hu, J. Yuan, C. Wen, X. Lu, Y. Liu, and X. Li, “RSGPT：遥感视觉语言模型与基准”，ISPRS摄影测量与遥感杂志，第224卷，第272–286页，2025年。[102] Y. Bazi, L. Bashmal, M. M. Al Rahhal, R. Ricci, and F. Melgani, “Rs-llava：用于遥感影像联合图像描述与问答的大型视觉语言模型”，《遥感》，第16卷，第9期，第1477页，2024年。[103] W. Zhang, M. Cai, T. Zhang, Y. Zhuang, and X. Mao, “EarthGPT：面向遥感领域多传感器图像理解的通用多模态大语言模型”，IEEE地球科学与遥感汇刊，2024年。[104] W. Xu, Z. Yu, Y. Wang, J. Wang, and M. Peng, “RS-Agent：通过智能体自动执行遥感任务”，arXiv预印本 arXiv:2406.07089，2024年。[105] K. Kuckreja, M. S. Danish, M. Naseer, A. Das, S. Khan, and F. S. Khan, “GeoChat：用于遥感的具备依据的大型视觉语言模型”，载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第27831–27840页，2024年。[106] C. Liu, K. Chen, H. Zhang, Z. Qi, Z. Zou, and Z. Shi, “Change-Agent：迈向交互式综合遥感变化解译与分析”，IEEE地球科学与遥感汇刊，2024年。[107] W. Zhang, M. Cai, T. Zhang, J. Li, Y. Zhuang, and X. Mao, “Earth-marker：通过视觉提示学习实现区域级与点级遥感影像理解”，arXiv预印本 arXiv:2407 13596，2024年。

[108] Y. Zhan, Z. Xiong, and Y. Yuan, “Skyeyegpt：通过使用大语言模型进行指令微调，统一遥感视觉语言任务”，ISPRS摄影测量与遥感杂志，第221卷，第64–77页，2025年。[109] J. A. Irvin, E. R. Liu, J. C. Chen, I. Dormoy, J. Kim, S. Khanna, Z. Zheng, and S. Ermon, “TEOChat：面向时序地球观测数据的大型视觉语言助手”，载于《第十三届国际学习表征会议》，2025年。[110] Y. Li, W. Xu, G. Li, Z. Yu, Z. Wei, J. Wang, and M. Peng, “UniRS：通过视觉语言模型统一多时相遥感任务”，arXiv预印本 arXiv:2412.20742，2024年。[111] D. Muhtar, Z. Li, F. Gu, X. Zhang, and P. Xiao, “LHRS-Bot：利用VGI增强的大型多模态语言模型赋能遥感”，载于《欧洲计算机视觉会议》，第440–457页，Springer，2024年。[112] Z. Li, D. Muhtar, F. Gu, X. Zhang, P. Xiao, G. He, and X. Zhu, “LHRS-Bot-Nova：改进的遥感视觉语言解释多模态大语言模型”，arXiv预印本 arXiv:2411 09301，2024年。[113] P. Wang, H. Hu, B. Tong, Z. Zhang, F. Yao, Y. Feng, Z. Zhu, H. Chang, W. Diao, Q. Ye等，“RingMoGPT：面向视觉、语言和基于视觉grounding 的任务的统一遥感基础模型”，IEEE地球科学与遥感汇刊，2024年。[114] X. Liu and Z. Lian, “RSUniVLM：通过面向粒度的专家混合构建的遥感统一视觉语言模型”，arXiv预印本 arXiv:2412.05679，2024年。[115] R. Ou, Y. Hu, F. Zhang, J. Chen, and Y. Liu, “GeoPix：用于遥感像素级图像理解的多模态大语言模型”，arXiv预印本 arXiv:2501.06828，2025年。[116] A. Shabbir, M. Zumri, M. Bennamoun, F. S. Khan, and S. Khan, “GeoPixel：用于遥感的像素定位大型多模态模型”，arXiv预印本 arXiv:2501.13925，2025年。[117] S. Soni, A. Dudhane, H. Debary, M. Fiaz, M. A. Munir, M. S. Danish, P. Fraccaro, C. D. Watson, L. J. Klein, F. S. Khan等，“EarthDial：将多传感器地球观测转化为交互式对话”，arXiv预印本 arXiv:2412.15190，2024年。[118] X. Xue, G. Wei, H. Chen, H. Zhang, F. Lin, C. Shen, and X. X. Zhu, “REO-VLM：改造视觉语言模型以应对地球观测中的回归挑战”，arXiv预印本 arXiv:2412.16583，2024年。[119] J. Luo, Z. Pang, Y. Zhang, T. Wang, L. Wang, B. Dang, J. Lao, J. Wang, J. Chen, Y. Tan等，“SkySenseGPT：用于遥感视觉语言理解的细粒度指令调优数据集与模型”，arXiv预印本 arXiv:2406.10100，2024年。[120] D. Wang, J. Zhang, M. Xu, L. Liu, D. Wang, E. Gao, C. Han, H. Guo, B. Du, D. Tao等，“MTP：通过多任务预训练推进遥感基础模型发展”，IEEE《应用地球观测与遥感专题期刊》，2024年。[121] J. Zhang, Z. Zhou, G. Mai, M. Hu, Z. Guan, S. Li, and L. Mu, “Text2Seg：通过文本引导的视觉基础模型实现零样本遥感图像语义分割”，载于ACM SIGSPATIAL地理知识发现人工智能国际研讨会论文集，第63–66页，2024年。[122] Z. Zhang, T. Zhao, Y. Guo, and J. Yin, “Rs5m和GeoRSCLIP：面向遥感的大型视觉语言数据集与大型视觉语言模型”，IEEE地球科学与遥感汇刊，2024年。[123] X. Li, C. Wen, Y. Hu, and N. Zhou, “RS-CLIP：通过对比式视觉语言监督实现零样本遥感场景分类”，《国际应用地球观测与地理信息期刊》，第124卷，第103497页，2023年。[124] K. Klemmer, E. Rolf, C. Robinson, L. Mackey, and M. Rußwurm, “SatCLIP：利用卫星影像构建全球通用的位置嵌入”，arXiv预印本 arXiv:2311.17179，2023年。[125] Z. Wang, R. Prabha, T. Huang, J. Wu, and R. Rajagopal, “SkyScript：面向遥感的大规模语义多样视觉语言数据集”，载于AAAI论文集，第38卷，第5805–5813页，2024年。[126] F. Liu, D. Chen, Z. Guan, X. Zhou, J. Zhu, Q. Ye, L. Fu, and J. Zhou, “RemoteCLIP：面向遥感的视觉语言基础模型”，IEEE地球科学与遥感汇刊，2024年。[127] U. Mall, C. P. Phoo, M. K. Liu, C. Vondrick, B. Hariharan, and K. Bala, “无需标注、通过地面遥感对齐构建的遥感视觉语言基础模型”，arXiv预印本 arXiv:2312 06960，2023年。[128] S. Khanna, P. Liu, L. Zhou, C. Meng, R. Rombach, M. Burke, D. B. Lobell, and S. Ermon, “DiffusionSat：用于卫星影像的生成式基础模型”，载于《第十二届国际学习表征会议》，2024年。

[129] D. Tang、X. Cao、X. Hou、Z. Jiang 和 D. Meng，“Crs-diff：可控生成式遥感基础模型。arxiv 2024”，*arXiv* 预印本 *arXiv:2403.11614*。[130] M. Chu、Z. Zheng、W. Ji、T. Wang 和 T.-S. Chua，“迈向自然语言引导的无人机：具备空间关系匹配能力的 Geotext-1652 基准”，载于欧洲计算机视觉会议，第 213–231 页，Springer，2024。[131] V. Vivanco Cepeda、G. K. Nayak 和 M. Shah，“GeoCLIP：受 CLIP 启发的位置与图像对齐方法，实现有效的全球地理定位”，神经信息处理系统进展，第 36 卷，第 8690–8701 页，2023。[132] J. Luo、Y. Wang、Z. Gu、Y. Qiu、S. Yao、F. Wang、C. Xu、W. Zhang、D. Wang 和 Z. Cui，“MMM-RS：用于文本到图像生成的多模态、多地面采样距离、多场景遥感数据集与基准”，*arXiv* 预印本 *arXiv:2410.22362*，2024。[133] C. Liu、K. Chen、R. Zhao、Z. Zou 和 Z. Shi，“Text2Earth：借助全球尺度数据集和基础模型，释放文本驱动的遥感图像生成潜力”，*arXiv* 预印本 *arXiv:2501.00895*，2025年。[134] K. Heidler、L. Mou、D. Hu、P. Jin、G. Li、C. Gan、J.-R. Wen 和 X. X. Zhu，“遥感数据的自监督视听表征学习”，《国际应用地球观测与地理信息学杂志》，第 116 卷，文章 103130，2023。[135] C. Pang、X. Weng、J. Wu、J. Li、Y. Liu、J. Sun、W. Li、S. Wang、L. Feng、G.-S. Xia、等，“VHM：面向遥感图像分析的通用且可靠的视觉语言模型”，*arXiv* 预印本 *arXiv:2403 20213*，2024。。[136] J. A. Irvin、E. R. Liu、J. C. Chen、I. Dormoy、J. Kim、S. Khanna、Z. Zheng 和 S. Ermon，“TEOChat：面向时序地球观测数据的大型视觉语言助手”，载于第十三届国际学习表征会议，2025年。[137] X. Li、Y. Tao、S. Zhang、S. Liu、Z. Xiong、C. Luo、L. Liu、M. Pechenizkiy、X. X. Zhu 和 T. Huang，“REOBench：地球观测基础模型鲁棒性基准测试”，载于*NeruIPS 2025 D&B* 专题，2025年。



Danfeng Hong （IEEE高级会员）于2019年在德国慕尼黑工业大学（TUM）地球观测信号处理（SiPEO）研究组以最优等荣誉获得工学博士学位。他现任中国南京东南大学正教授，曾任中国北京中国科学院空天信息创新研究院正教授。他的研究兴趣包括人工智能、多模态感知、基础模型、高光谱遥感和地球科学。

Hong博士担任《IEEE模式分析与机器智能汇刊》（TPAMI）、《IEEE图像处理汇刊》（TIP）和《IEEE地球科学与遥感汇刊》（TGRS）的副编辑。自2022年起，他一直被科睿唯安评为高被引研究者。



Chenyu Li 于2018年、2021年和2025年分别在中国南京东南大学获得学士、硕士和博士学位。她的研究兴趣包括可解释人工智能、大地球数据预测、基础模型和高光谱成像。

李博士荣获WHISPERS2025杰出学生论文奖（Paul Gader），并获得IEEE地球科学与遥感汇刊（TGRS）颁发的最佳审稿人奖。



Xuyang Li 于2023年获得中国成都电子科技大学（UESTC）光电科学与工程学院理学学士学位。目前，他正在中国北京中国科学院空天信息创新研究院攻读博士学位。他的研究兴趣包括机器学习、遥感基础模型和多模态解译。



G ustau Camps-Valls（IEEE会士）是西班牙瓦伦西亚大学电气工程正教授。他专攻面向环境与气候挑战的人工智能和机器学习。他领导图像与信号处理（ISP）研究组，专注于服务地球与气候科学的人工智能。他的研究涵盖利用人工智能预测极端事件、改进地球模型，以及了解气候影响对社会造成的影响，例如气候诱发的迁移。他是高被引研究者（2011年、2021—2024年），并曾获得重要认可，相关荣誉包括谷歌经典论文奖（2019年）。他曾担任 IEEE IGARSS 2018 和 AISTATS 2022 的程序委员会主席、五本 IEEE 期刊的副编辑，以及 IEEE 杰出演讲人（2017—2019年）。他是 IEEE、ELLIS、欧洲科学院、欧洲科学院和 AAIA 的会士。他还担任多个顾问职务，包括在 ESA PhiLab 和欧洲气象卫星应用组织任职，并协调 ELLIS 的地球与气候科学人工智能项目。他获得了两项欧洲研究理事会资助（整合者项目和协同项目），以推动地球与气候科学领域的人工智能发展。



Jocelyn Chanussot （IEEE会士）于1995年在法国格勒诺布尔的格勒诺布尔理工学院（格勒诺布尔国立理工学院）获得电气工程理学硕士学位，并于1998年在法国阿讷西的萨瓦大学获得博士学位。1999年至2023年，他任职于格勒诺布尔国立理工学院，担任信号与图像处理教授。目前，他是格勒诺布尔INRIA的研究主任。他的研究兴趣包括图像分析、高光谱遥感、数据融合、机器学习和人工智能。

Chanussot博士是IEEE地球科学与遥感法国分会的创始主席。他曾担任IEEE地球科学与遥感学会负责会议与专题研讨会的副主席。他是IEEE地球科学与遥感汇刊、IEEE图像处理汇刊和《IEEE会刊》的副编辑。他曾担任IEEE《应用地球观测与遥感专题期刊》主编（2011—2015年）。他是IEEE会士、ELLIS会士、AAIA会士、法国大学研究院成员（2012—2017年），也是高被引研究者（科睿唯安／汤森路透，自2018年起）。