

面向通过基础模型组合实现可扩展和通用的地观测数据挖掘

Man Duc Chuc

University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam

ARTICLE HISTORY

Compiled 2025 年 6 月 28 日

摘要

基础模型正通过提供可泛化和可扩展的解决方案，迅速改变地球观测数据挖掘的方式，这些方案适用于场景分类和语义分割等关键任务。虽然在地理空间领域，大多数努力都集中在使用大规模地球观测数据集从头开始训练大型模型上，但另一种尚未充分探索的策略是复用和组合现有的预训练模型。在这项研究中，我们探讨了是否可以有效地结合基于遥感和通用视觉数据集预训练的基础模型，以提高在一系列关键地球观测任务中的性能。使用 GEO-Bench 基准测试，我们在涵盖多种空间分辨率、传感器模式和任务类型的 11 个数据集上评估了几种著名模型，包括 Prithvi、Hiera 和 DOFA。结果显示，较小的预训练模型的特征级集成可以匹配或超过大型模型的性能，同时需要更少的训练时间和计算资源。此外，研究还强调了应用知识蒸馏将集合的优势转移到更为紧凑的模型中的潜力，为在实际地球观测应用中部署基础模型提供了实用路径。

KEYWORDS

基础模型；普里特维；希拉；SAM，DOFA；集成；知识蒸馏；地球观测

1. 介绍

基础模型正在改变地球观测（EO）数据的分析。在过去几年中，许多基础模型被引入以从大量的 EO 数据中学习可泛化的表示，在各种任务上取得了显著性能，包括场景分类和语义分割 Cong et al. (2022); Jakubik et al. (2023); Hong et al. (2024); Xiong et al. (2024); Szwarcman et al. (2024); Duc and Fukui (2025)。这些模型通过自监督学习在庞大的无标签遥感数据集上进行了预训练，然后针对下游任务进行微调。一个例子是 Prithvi-EO-2.0，这是一个由 IBM 和 NASA 开发的基于变压器的基础模型 Szwarcman et al. (2024)。Prithvi-EO-2.0 在一个超过 420 万个全球时间序列样本的数据集上进行了预训练，这些样本来自多光谱图像的协调 Landsat – Sentinel-2 (HLS) 数据档案。该数据集涵盖了美国大陆多年、30 米分辨率的图像。模型使用掩码自动编码框架进行预训练以学习丰富的时空特征。经过微调后的模型可以应用于多种任务，如洪水制图、野火烧伤痕迹分割和多时态作物类型映射。发现对 Prithvi-EO-2.0 进行微调比从随机初始化开始的训练更快收敛，并且预训练模型在许多任务上被报告超过了最先进的（SOTA）方法。这代

表了从特定任务模型向加速下游学习的预训练模型的转变。

这些改进突显了大规模预训练和模型扩展在关键遥感任务性能中的重要影响。从头开始使用遥感数据训练大型模型的趋势推动了专门针对地球观测领域的新架构的发展。例如，研究人员正超越单一模态的主干网络，转向统一多传感器模型。一个例子是动态一网打尽 (Dynamic One-For-All, DOFA) 架构，它采用一种基于输入传感器模态 Xiong et al. (2024) 动态调整其权重的动态超网络。这使得单个主干能够处理来自五种不同类型的传感器的数据，每种传感器具有不同的光谱特性。DOFA 在五个卫星数据源的组合上进行联合训练，学习了一种高度适应性的表示方法，可以泛化到 14 项多样化的地球观测任务中，甚至包括以前未见过的传感器的数据。这些结果突显了遥感领域向通用基础模型发展的趋势，这些模型是在 PB 级档案上训练的，并可以根据从城市制图到灾害响应等高影响应用进行微调。

尽管近期有所进展，地球观测领域的主要焦点仍然是开发专门的遥感基础模型，如 Prithvi 和 DOFA，这些模型仅在特定领域的数据集上进行训练。然而，一种具有显著潜力的替代范式仍处于探索不足的状态。这涉及到利用来自地球观测领域甚至是更广泛的计算机视觉领域的现有预训练模型，而不是从头开始训练新模型。在通用 AI 中广泛采用的技术，如模型集成 Jiang et al. (2023); Huang et al. (2024) 和知识蒸馏 Hinton et al. (2015); Gou et al. (2021)，在这些技术中信息是从一个或多个大型教师模型转移到一个小的学生成型器，仍然相对少见于地球观测研究中。

在主流计算机视觉领域，像 Meta 的 Segment Anything Model (SAM) 这样的大规模基础模型展示了显著的零样本能力，几乎可以对任何对象生成准确的分割掩模，只需最少的提示 Kirillov et al. (2023)。在大约 1100 万自然图像和超过 10 亿个分割掩模的前所未有的数据集上进行训练，SAM 已经成为一个能够捕捉细粒度视觉特征的高度通用模型。尽管 SAM 最初并不是为遥感应用开发的，但在土地覆盖分类等 EO 任务中使用它的兴趣正在增加。初步研究表明，无需特定任务培训，SAM 就可以确定卫星图像中的建筑物、道路和水体等功能。通过微调该模型 Osco et al. (2023); Ren et al. (2024)，准确性可以进一步提高。这些努力突显了在遥感领域复用大型预训练视觉模型的潜力，提供了一种替代方案，即无需从头开始训练具有数亿参数的模型。

探索这种跨域适应的研究数量有限，例如将 SAM 应用于 EO 数据已经报告了有希望的结果，包括改进的零样本分类和更快的标注。这些发现表明，EO 社区可以从更广泛的 AI 社区对基础模型的投资中显著受益。弥合特定领域与通用模型之间的差距很可能是未来研究的一个有前景的方向。在这项研究中，我们探讨现有预训练的基础模型，特别是来自一般计算机视觉和 EO 领域的那些模型，是否可以有效地结合以改进各种遥感任务的性能，而无需从头开始训练更大的模型。这种方法有可能为未来的模型集成和知识蒸馏策略提供信息，使构建新的 SOTA 基础模型变得可能，并且计算成本显著降低。这样做有助于推进遥感及更广泛领域中的基础模型的发展。

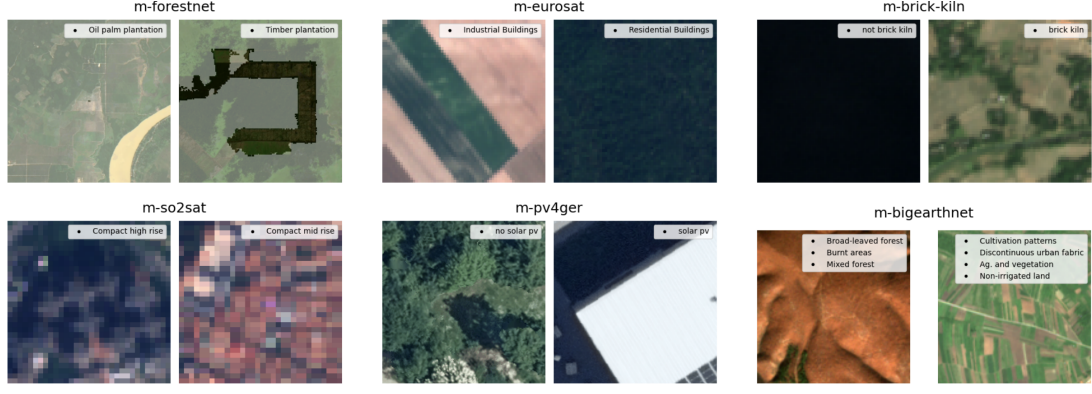
2. 方法论

2.1. 数据集

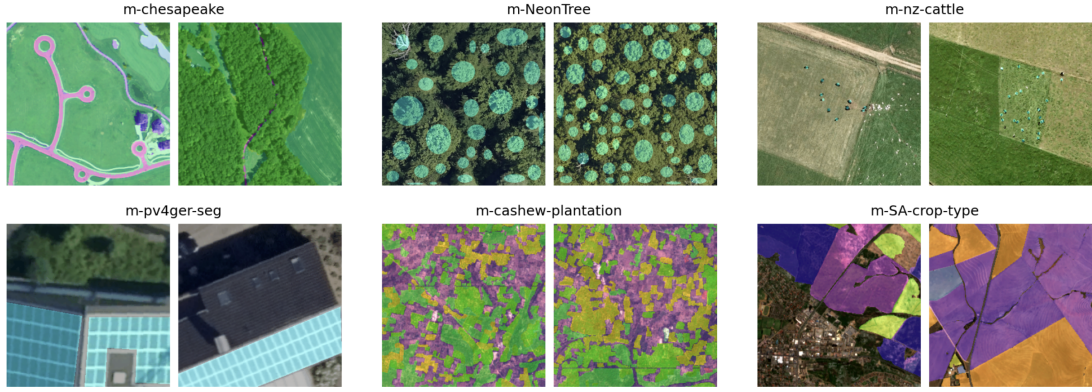
我们使用了 GEO-Bench，这是最广泛采用且严谨的基准测试框架之一，用于评估地球观测（EO）基础模型 Lacoste et al. (2023)。该框架包括六个场景分类数据集和六个语义分割数据集，涵盖了一系列的空间分辨率、数据集大小和应用领域（参见表 1和图 1）。GEO-Bench 支持多种下游任务，包括单标签分类、多标签分类和语义分割（即土地覆盖分类）。然而，在我们的实验中，我们无法获取 m-eurosat 数据集的标签文件，因此未能将其纳入评估。在实现方面，我们采用了 TerraTorch 库 Gomes et al. (2025)，这是一个开源的 Python 框架，允许用户轻松修改或扩展现有代码库。TerraTorch 内置了多个预集成的骨干网络，包括 Prithvi 的两个版本、DOFA、ViT 以及各种基于 CNN 的预训练模型。它还支持多种数据集，包括 GEO-Bench 中的那些。在 TerraTorch 中训练模型非常简单。用户只需定义一个配置文件，指定骨干网络、数据集、训练参数和其他相关设置。在这项研究中，除了现有的骨干网络外，我们还将两个新骨干（希拉_200M 和希拉_普里特维_500M）整合到了 TerraTorch 库中。这些将在以下各节详细描述。

表 1.: GEO-Bench 数据集的特征 Lacoste et al. (2023)。

分类								
名称	图像大小	# 类别	训练	值	测试	# 能带	RGB 颜色分辨率	传感器
m-bigearthnet	120 × 120	43	20,000	1,000	1,000	12	10.0	Sentinel-2
m-so2sat	32 × 32	17	19,992	986	986	18	10.0	Sentinel-2 + Sentinel-1
m-brick-kiln	64 × 64	2	15,063	999	999	13	10.0	Sentinel-2
m-forestnet	332 × 332	12	6,464	989	993	6	15.0	Landsat-8
m-eurosat	64 × 64	10	2,000	1,000	1,000	13	10.0	Sentinel-2
m-pv4ger	320 × 320	2	11,814	999	999	3	0.1	RGB
分割段落								
名称	图像大小	# 类别	训练	值	测试	# 能带	RGB 颜色	传感器
m-pv4ger-seg	320 × 320	2	3,000	403	403	3	0.1	RGB
m-chesapeake-landcover	256 × 256	7	3,000	1,000	1,000	4	1.0	RGBN
m-cashew-plantation	256 × 256	7	1,350	400	50	13	10.0	Sentinel-2
m-SA-crop-type	256 × 256	10	3,000	1,000	1,000	13	10.0	Sentinel-2
m-nz-cattle	500 × 500	2	524	66	65	3	0.1	RGB
m-NeonTree	400 × 400	2	270	94	93	5	0.1	RGB + Hyperspectral + Elevation



(a)



(b)

图 1.: (a) 分类任务和 (b) 划分任务在 GEO-Bench 数据集中的典型示例。

2.2. 基础模型

2.2.1. 地球

Prithvi 模型代表了通用地理空间人工智能的重要进步。该第二代模型 Prithvi-EO-2.0 Szwarcman et al. (2024) 于 2024 年 12 月发布，改进了其前身 Prithvi-EO-1.0 Jakubik et al. (2023)，通过引入架构增强并利用规模更大、分布全球的训练数据集。Prithvi-EO-2.0 基于视觉变压器架构构建，并融入时间与位置嵌入，使其能够更有效地捕捉地球观测数据中复杂的时空模式。这些改进在一系列地球观测任务中的性能方面带来了显著提升。该模型在一个包含 420 万时间序列样本的大型数据集上进行了预训练，该数据集来自 HLS 档案，提供 30 米分辨率的影像。发布了两种主要模型大小：Prithvi 300M (~ 300 百万参数) 和 Prithvi 600M (~ 630 百万参数)，每种都有基础版本和增强版，后者包括时间与位置嵌入。基准测试结果显示，在 GEO-Bench、野火疤痕映射以及燃烧强度映射等任务上，600M 变体通常优于较小的模型。训练这些模型需要大量的计算资源：300M 模型在 80 个 GPU 上进行了大约 21,000 GPU 小时的训练，而 600M 模型则需要 240 个 GPU 和大约 58,000 GPU 小时。在此研究中，我们主要使用了融入时间与位置嵌入的变体模型，特别是在包含时间和/或空间元数据的下游任务中。这些模型分别称为地球_300M 和地球_600M。

_600M。

2.2.2. 任何片段模型

由 Meta AI 于 2023 年 4 月推出的 Segment Anything Model 引入了可提示的图像分割功能 Kirillov et al. (2023)。其架构包括一个图像编码器；一个用于处理点、框、掩码提示以及一些初始文本提示的提示编码器；以及一个轻量级的掩码解码器，实现了快速和交互式的分割。在大规模 SA-1B 数据集上进行训练，该数据集包含超过 10 亿个掩码跨越了 1100 万张图像，SAM 展示了令人印象深刻的零样本泛化能力，能够无需额外微调就能分割之前未见过的对象。在此基础上，Meta 于 2024 年 8 月发布了 SAM 2，将分割功能扩展到了视频 Ravi et al. (2024)。SAM 2 采用 Hiera 架构 Ryali et al. (2023) 作为其图像编码器，并引入了一个记忆银行模块，实现了对图像和视频的实时分割。这一增加使模型能够保留跨帧的信息，有效地管理遮挡和对象再现。为了支持训练，Meta 开发了 SA-V 数据集，这是迄今为止最大的视频分割数据集之一，包含跨越 50900 个视频的 3550 万个掩码。评估结果显示，SAM 2 显著提高了视频分割精度，同时将所需用户交互次数减少了三倍与早期方法相比。在图像分割方面，它优于原始 SAM，实现了六倍更快的推理速度和更高的准确性，这主要归功于 Hiera 编码器的效率。训练 SAM 2 需要大量的计算资源，据报道涉及 256 个 A100 GPU 运行了 108 小时。在我们的工作中，我们使用了 SAM 2 的大版本，特别提取了 Hiera 图像编码器 (~2.12 亿参数)，称为希拉_200M，这构成了完整 SAM 2 模型的约 95% (~2.24 亿参数)，并以与其他骨干架构相同的方式使用它作为骨干。

2.2.3. 权限分配框架

DOFA (动态一应俱全) 模型于 2024 年 6 月推出，是一种受神经可塑性生物学概念启发的多模态 EO 基础模型 Xiong et al. (2024)。它旨在使深度学习模型能够在统一框架内自适应地整合各种数据模式。与传统的专门针对特定模式（光学、雷达或高光谱）的 EO 模型不同，DOFA 利用了一种动态超网络架构，根据输入的光谱特性调整其内部参数，特别是中心波长。核心在于，DOFA 具备一个共享视觉变换器主干和一个受波长条件制约的动态补丁嵌入模块，这使它能够处理来自各种传感器的不同通道数目的输入。这种灵活的设计使模型能够为每种模式动态生成定制化的权重和偏置，促进有效的跨模态表示学习。该模型是在超过 800 万张图像的大规模多模态 EO 数据集上进行训练的，这些图像来源于 Sentinel-1、Sentinel-2、NAIP、高分卫星和 EnMAP 等来源，涵盖了 SAR、RGB、多光谱和高光谱领域。DOFA 在包括 GEO-Bench 基准套件在内的分类和分割等 14 个下游任务上进行了评估，在这 14 项任务中的 13 项中表现优于或匹配了其他最新预训练模型如 GFM、SatMAE、Scale-MAE 和 SpectralGPT 的表现。在我们的研究中，我们使用了 DOFA-large 变体 *Dofa_300M*，它包含大约 3.3 亿个参数。

2.3. 模型集成

模型集成是机器学习中广泛使用的一种技术，它通过结合多个模型来增强泛化能力、鲁棒性和整体性能，适用于各种任务。然而，在地理空间基础模型的背景下，这种方法仍

然很少被探索。将如 Prithvi 这样的模型（该模型在多光谱、中分辨率时间序列数据上进行训练）与 SAM 模型相结合（该模型主要在高分辨率自然图像和视频上进行训练），是一种利用它们在不同模态和空间分辨率上的互补优势的有前途的方法。如果成功，集成基础模型可以开启一个新的研究方向，允许开发强大的基础模型而不必承担从头开始训练的巨大计算成本，这目前是主导方法。此外，了解哪些模型组合效果最好可能指导知识蒸馏策略，使创建有效保留集合优势的同时更轻量级和更具成本效益的模型成为可能。在我们的研究中，我们采用了特征向量连接作为集成策略。具体来说，该集成结合了地球_300M 和希拉_200M，并被称为希拉_普里特维_500M。这种方法保留了每个单独模型丰富的高维表示，尽管与简单的平均技术相比其计算成本更高。具体而言，我们从每个模型中提取特征嵌入，在传递到轻量级任务特定头部（如分割解码器或分类模块）之前将它们连接起来。不同于输出级别的集成可能遭受模态不匹配的问题，特征级别连接允许更深入和有效的多模态信息整合。由于计算资源有限，我们无法探索其他模型和组合。这仍是一个未来研究的领域。

3. 实验与结果

3.1. 实验设置

本研究考虑了三个下游任务：单标签分类、多标签分类和语义分割。对于每个任务，我们设计了一种统一的任务特定架构，所有主干网络都可以无缝集成到该架构中。具体来说，对于分类任务（包括单标签和多标签），我们在主干网络的最后一层输出上应用了简单的线性投影层。这些分类头针对每个数据集进行了调整以适应类别数量的差异，但在所有主干网络中保持一致。对于语义分割，我们跨所有数据集使用了 UPerNet 架构 Xiao et al. (2018)。虽然网络结构稍作调整以适应每个数据集的具体情况，但总体上在所有骨干模型中保持统一。所有实现均使用 TerraTorch 框架高效完成。

在使用 GEO-Bench 数据集时，我们遵循了推荐的评估协议以确保公平和可重复的模型比较 Lacoste et al. (2023); Szwarcman et al. (2024)。该过程从超参数调整开始，在我们的案例中，每个数据集为每个模型分配了一个固定的试验预算，即 16 次。然后，在验证集上确定的最佳超参数被用于运行多次实验，我们在研究中的每次数据集都进行了 10 次实验，并使用了不同的随机种子。这种方法考虑到了训练 AI 模型时固有的随机性，并使性能比较更加可靠。超参数调整是通过 Optuna 进行贝叶斯优化完成的¹。在所有实验中，均采用了 AdamW 优化器，并在一个包含学习率和权重衰减的共享搜索空间上进行了调优。为了可比性，每个数据集的批量大小都被固定为一个合理的值，并且在整个模型中保持一致。此外，我们将实验限制在光学传感器数据上，只向每个主干网络输入其预训练过的光谱带。所有输入图像均统一调整为 224×224 像素，适用于所有数据集。

¹<https://optuna.readthedocs.io/en/stable/>

3.2. 结果与讨论

图 2展示了使用各种基础模型在 11 个遥感数据集上的基准评估结果。顶部行 (a) 显示了分类准确率，除了 m-bigeearthnet 这个多标签分类任务，它使用 F1 分数。底部行 (b) 显示了语义分割任务的平均交并比 (mIoU)。每个箱形图总结了在不同随机种子下进行的 10 次独立运行的表现，以捕捉变化性。总体而言，结果显示希拉_普里特维_500M 和地球_600M 是最优模型，各自在四个数据集上取得了最佳结果。随后的是 Dofa_300M 和希拉_200M，前者在两个数据集 (m-forestnet, m-so2sat) 上表现最优，后者在一个数据集 (m-pv4ger-seg) 上表现最优。值得注意的是，尽管只使用了三个可见波段作为输入（见表 2），希拉_200M 在七项任务上的表现优于如地球_300M 等较大的模型。

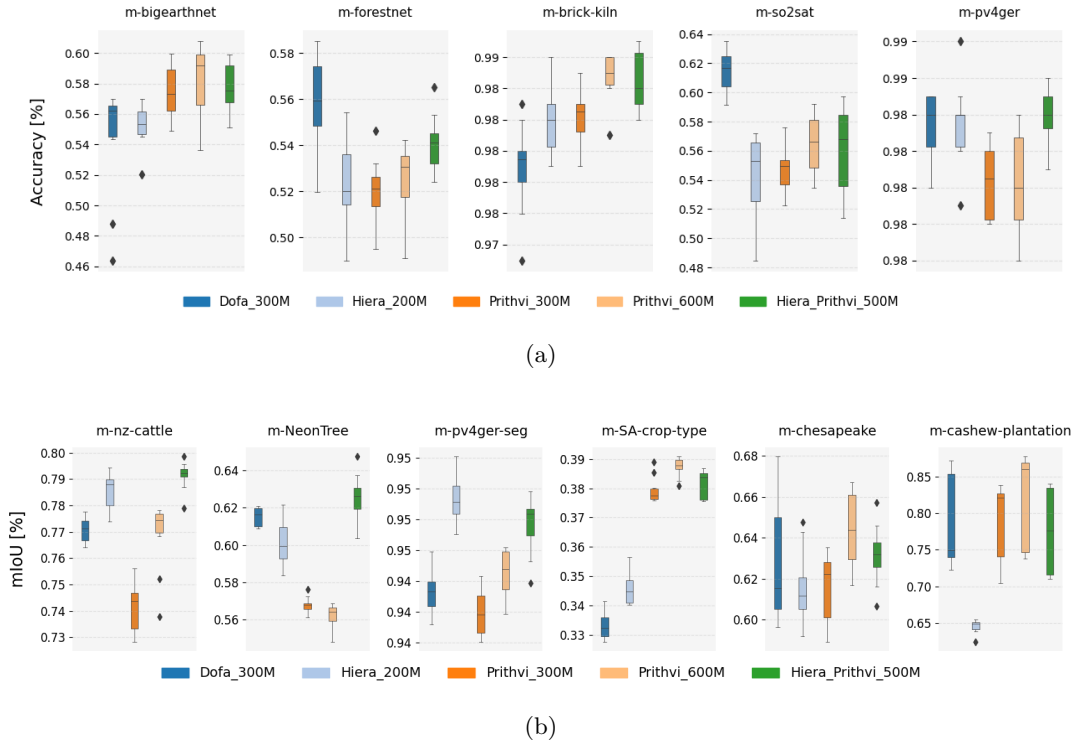


图 2.: 所有模型在 GEOBench 上的 10 次重复运行中性能分布，分别针对 (a) 分类任务和 (b) 分割任务的准确率和平均 IoU。我们列出了每个数据集的输入传感器和分辨率。

特征级别的集成，如希拉_普里特维_500M 所示，提高了许多案例中的准确性和一致性。希拉_200M 和普里特维_300M 的集成在 11 个数据集集中的 10 个上始终优于单独的模型，例外的是 m-pv4ger-seg 数据集。这突显了通过集成利用预训练模型以实现遥感任务中卓越性能的潜力。操作自由度也作为强有力的竞争者出现，在六个任务上超越了类似规模的模型如普里特维_300M 和希拉_200M。这表明在未来的工作中，进一步探索涉及 DOFA 的模型集成或知识蒸馏可能是有价值的。虽然更大的模型（例如，地球_600M）通常表现更强，但它们对资源的需求显著更高。总体而言，基础模型的综合优势指向了地空人工智能的有希望方向。结果支持这样一种观点：模型集成可以匹敌甚至超越大型单个模型，提供了训练大规模模型的一种计算效率更高的替代方案。

表 2.: 输入每个模型的光谱波段。为了更好的表格格式，数据集名称已缩短。RGB 代表红、绿和蓝。RGBN 代表红、绿、蓝和近红外 (N)。RGBNS1S2 代表红、绿、蓝、近红外、短波红外 1 (S1) 和短波红外 2 (S2)。

姓名	大地球网	二氧化硫满足	砖窑	森林网络	pv4ger	pv4ger-分段	切萨皮克	腰果	SA-作物类型	nz-奶牛	霓虹树
希拉_200M	RGB	RGB	RGB	RGB	RGB	RGB	RGB	RGB	RGB	RGB	RGB
Dofa_300M	RGBNS1S2	RGBNS1S2	RGBNS1S2	RGBNS1S2	RGB	RGB	RGBN	RGBNS1S2	RGBNS1S2	RGB	RGB
地球_300M	RGBNS1S2	RGBNS1S2	RGBNS1S2	RGBNS1S2	RGB	RGB	RGBN	RGBNS1S2	RGBNS1S2	RGB	RGB
地球_600M	RGBNS1S2	RGBNS1S2	RGBNS1S2	RGBNS1S2	RGB	RGB	RGBN	RGBNS1S2	RGBNS1S2	RGB	RGB
希拉_普瑞特维_500M	RGBNS1S2	RGBNS1S2	RGBNS1S2	RGBNS1S2	RGB	RGB	RGBN	RGBNS1S2	RGBNS1S2	RGB	RGB

额外的分析表明，基于 Hiera 的主干网络可能受益于更大的原始输入尺寸。在具有更大图像维度的数据集上，如 m-pv4ger (320×320)、m-NeonTree (400×400) 和 m-nz-cattle (500×500)，希拉_200M 和希拉_普里特维_500M 通常优于其他主干网络。此外，使用不同随机种子的实验显示，在这些数据集上的性能波动比具有较小输入尺寸的数据集要低。另外，基于 Hiera 的模型在高分辨率数据集上表现更好，包括具有 0.1 米分辨率 (m-pv4ger、m-pv4ger-seg、m-nz-cattle 和 m-NeonTree) 和 1 米分辨率 (m-chesapeake-landcover) 的数据集。这种行为可能归因于 Hiera 模型的原始预训练分辨率，这些模型是在主要由自然相机图像组成的 1024×1024 图像上进行训练的。这一特征在未来的应用中非常重要，特别是在为微调或其他下游任务准备数据集时。

4. 结论

在这项研究中，我们探索了地理空间基础模型开发的一个新方向，通过利用现有的预训练模型来实现最先进的性能。我们对来自地理空间和通用计算机视觉领域的近期代表性视觉基础模型进行了全面的基准评估。此外，我们还评估了两个代表性模型（即 Hiera 和 Prithvi）在特征级上的集成效果。这些模型使用 GEO-Bench 框架，在多种地球观测任务上进行了评估。结果显示，如地球_600M 等近期模型在许多分类和分割任务中表现出顶级性能。值得注意的是，较小模型如希拉_200M 和地球_300M 的特征级集成证明非常有效，在多个数据集上取得了具有竞争力的结果，并与较大的独立模型相匹敌。这突显了结合预训练模型以提高准确性和鲁棒性的潜力。

我们的分析进一步揭示，基于 Hiera 的模型在具有更高空间分辨率和更大输入尺寸的数据集上表现出更强的性能，这可能是由于它们对 1024×1024 自然图像进行了预训练。这一见解对于下游应用具有重要意义，表明将目标数据集的特点与模型的预训练配置相匹配可以增强迁移学习的效果。

虽然较大的模型如地球_600M 表现出强劲的性能，但它们需要大量的计算资源。我们的研究表明，模型集成提供了一个有前途且更节约资源的选择，能够匹配甚至超越这些较大独立模型的表现。这突显了灵活和模块化模型设计在推进地理空间人工智能中的重要性。未来的研究可能将重点放在探索知识蒸馏技术上，以压缩集成模型为适用于实际部署的轻量级模型。

致谢

本工作得到了日本中部大学国际数字地球应用科学研究中心，一个国际联合使用/研究中心的支持。我们衷心感谢日本信息通信技术研究所的 Tran Van Hien 博士对本研究中使用的模型训练提供的宝贵支持。

参考文献

- Yezhen Cong, Samar Khanna, Chenlin Meng, Patrick Liu, Erik Rozi, Yutong He, Marshall Burke et al. 2022. “SatMAE: Pre-training Transformers for Temporal and Multi-Spectral Satellite Imagery”. *Advances in Neural Information Processing Systems 35*(NeurIPS), 1–15.
- Chuc Man Duc, Hiromichi Fukui. 2025. “SatMamba: Development of Foundation Models for Remote Sensing Imagery Using State Space Models”. *arXiv*, 1–5.
- Carlos Gomes, Benedikt Blumenstiel, Joao Lucas de Sousa Almeida, Pedro Henrique de Oliveira, Paolo Fraccaro, Francesc Marti Escofet, Daniela Szwarcman et al. 2025. “TerraTorch: The Geospatial Foundation Models Toolkit”. *arXiv*, 1–5.
- Jianping Gou, Baosheng Yu, Stephen John Maybank, Dacheng Tao. 2021. “Knowledge Distillation: A Survey”. *International Journal of Computer Vision* 129(6), 1789–1819.
- Geoffrey Hinton, Oriol Vinyals, Jeff Dean. 2015. “Distilling the Knowledge in a Neural Network”. *arXiv*, 1–9.
- Danfeng Hong, Bing Zhang, Xuyang Li, Yuxuan Li, Chenyu Li, Jing Yao, Naoto Yokoya et al. 2024. “SpectralGPT: Spectral Remote Sensing Foundation Model”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46(8), 5227–5244.
- Yichong Huang, Xiaocheng Feng, Baohang Li, Yang Xiang, Hui Wang, Ting Liu, Bing Qin. 2024. “Ensemble Learning for Heterogeneous Large Language Models with Deep Parallel Collaboration”. *Advances in Neural Information Processing Systems* 37, 119838–119860.
- Jakubik et al. 2023. “Foundation Models for Generalist Geospatial Artificial Intelligence”. *arXiv*, 1–26.
- Dongfu Jiang, Xiang Ren, Bill Yuchen Lin. 2023. “LLM-BLENDER: Ensembling Large Language Models with Pairwise Ranking and Generative Fusion”. *Proceedings of the Annual Meeting of the Association for Computational Linguistics* 1, 14165–14178.
- Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao et al. 2023. “Segment Anything”. *arXiv*, 1–30.
- Alexandre Lacoste, Nils Lehmann, Pau Rodriguez, Evan David Sherwin, Hannah Kerner, Björn Lütjens, Jeremy Andrew Irvin et al. 2023. “GEO-Bench: Toward Foundation Models for Earth Monitoring”. *Advances in Neural Information Processing Systems* 36, 1–21.
- Lucas Prado Osco, Qiusheng Wu, Eduardo Lopes de Lemos, Wesley Nunes Gonçalves, Ana Paula Marques Ramos, Jonathan Li, José Marcato Junior. 2023. “The Segment Anything Model (SAM) for remote sensing applications: From zero to one shot”. *International Journal of Applied Earth Observation and Geoinformation* 124(November).
- Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma,

- Haitham Khedr et al. 2024. “SAM 2: Segment Anything in Images and Videos” . *arXiv*, 1–42.
- Simiao Ren, Francesco Luzi, Saad Lahrichi, Kaleb Kassaw, Leslie M. Collins, Kyle Bradbury, Jordan M. Malof. 2024. “Segment anything, from space?” *2024 IEEE Winter Conference on Applications of Computer Vision, WACV 2024 1*, 8340–8350.
- Chaitanya Ryali, Yuan-Ting Hu, Daniel Bolya, Chen Wei, Haoqi Fan, Po-Yao Huang, Vaibhav Aggarwal et al. 2023. “Hiera: A Hierarchical Vision Transformer without the Bells-and-Whistles” . *Proceedings of the 40th International Conference on Machine Learning*.
- Daniela Szwarzman, Sujit Roy, Paolo Fraccaro, Porsteinn Elí Gíslason, Benedikt Blumenstiel, Rinki Ghosal, Pedro Henrique de Oliveira et al. 2024. “Prithvi-EO-2.0: A Versatile Multi-Temporal Foundation Model for Earth Observation Applications” . *arXiv*, 1–31.
- Tete Xiao, Yingcheng Liu, Bolei Zhou, Yuning Jiang, Jian Sun. 2018. Unified Perceptual Parsing for Scene Understanding. *Proceedings of the European Conference on Computer Vision (ECCV)*, 418–434.
- Zhitong Xiong, Yi Wang, Fahong Zhang, Adam J. Stewart, Joëlle Hanna, Damian Borth, Ioannis Papoutsis et al. 2024. “Neural Plasticity-Inspired Multimodal Foundation Model for Earth Observation” . *arXiv*, 1–36.