

# 关于基础模型在作物类型制图中的泛化能力

Yi-Chia Chang<sup>1</sup>, Adam J. Stewart<sup>2,3</sup>, Favyen Bastani<sup>4</sup>, Piper Wolters<sup>4</sup>,

Shreya Kannan<sup>1</sup>, George R. Huber<sup>1</sup>, Jingtong Wang<sup>1</sup>, Arindam Banerjee<sup>1</sup>

<sup>1</sup>University of Illinois Urbana-Champaign, <sup>2</sup>Technical University of Munich,

<sup>3</sup>Munich Center for Machine Learning, <sup>4</sup>Allen Institute for AI

{yichia3, shreya28, ghuber3, jw132, arindamb}@illinois.edu,

adam.stewart@tum.de, {favyenb, piperw}@allenai.org

**摘要**—基于自监督学习预训练的基础模型在各种下游任务中展示了强大的迁移学习能力，包括语言理解、文本生成和图像识别。

地球观测（EO）领域已经产生了几个直接基于多光谱卫星图像预训练的基础模型，这些模型应用于精准农业、野火和干旱监测以及自然灾害响应等领域。

然而，很少有研究探讨了这些模型向新地理区域推广的能力，而且潜在的地理空间偏差问题——在数据丰富的发达国家训练的模型无法很好地转移到数据稀缺的发展中国家——仍然存在。

我们调查了流行的 EO 基础模型在农业领域向新的地理区域转移的能力，在这个领域中，耕作实践差异和类别不平衡使得迁移学习特别具有挑战性。

首先，我们在五大洲选择了五个作物分类数据集，标准化数据集大小并协调类别以专注于四种主要谷物：玉米、大豆、水稻和小麦。

然后，我们比较了三个流行的预训练模型，这些模型分别基于 SSL4EO-S12、SatlasPretrain 和 ImageNet 进行预训练，并使用分布内（ID）和分布外（OOD）评估进行对比。

实验表明，专门为 Sentinel-2 设计的预训练权重，如 SSL4EO-S12，优于像 ImageNet 这样的通用预训练权重。

此外，尽管只需 100 张标记图像就足以实现总体高准确率，但由于类别不平衡，需要 900 张图像才能达到平均高准确率。

所有协调的数据集和实验代码都是开源的，并可供下载。

**Index Terms**—遥感，机器学习，作物类型制图，迁移学习，基础模型

地图，准确率超过 80% [10, 11]，但大多数其他地区由于农作物自我报告政策较弱、政府资源较少以及小农户耕作比例较高导致难以绘制，其作物类型地图更新不规律或不完整。即使在这些地区，高精度的稀疏作物类型标签也很罕见。因此，尽管已提出多种通过遥感来绘制作物类型的方案 [12–15]，但由于世界上大多数地方不存在高质量的训练数据，这些方法很少被全球部署。

迁移学习已被应用于遥感技术中的农业监测和土地覆盖制图 [16, 17]，包括在西班牙和法国检测水稻以及在荷兰检测夏季大麦 [18]、农业地块边界划定 [19] 和树种检测的地理泛化能力 [20]。预训练的基础模型已在与遥感数据相关的下游任务中显示出良好的泛化性能 [21–23]。像 SSL4EO-S12 [24] 和 SatlasPretrain [25] 这样基于全球可用数据（如 Sentinel-1 和 Sentinel-2）预训练的遥感基础模型，为提高全球作物类型制图精度提供了关键机遇，通过提升下游任务中训练模型跨区域泛化的能力。然而，与这些基础模型结合进行作物类型映射的迁移学习尚未在大规模上得到研究，部分原因是缺乏统一的全球作物类型标签数据集。

在本文中，我们首先提供了一个融合了五大洲五个区域数据集的全球作物类型映射数据集。我们在每个大陆识别出具有最高标签质量的数据来源，并将其作物类型标签与各自地区生长季节高峰时期拍摄的一系列无云 Sentinel-2 图像统一配对。我们专注于四种主要谷物并围绕这些类别调和各个数据集的分类标签。然后，我们试验了三种流行的预训练模型，并发现 SSL4EO-S12 提供了最佳结果。进一步使用不同数量的分布内和分布外数据进行的实验突显了在非洲和南美洲等数据稀缺地区需要额外的数据集。所有实验均使用 TorchGeo [26]

## I. 介绍

作物类型地图在粮食安全 [1] 和保护方面有许多关键的下游应用，包括作物产量预测 [2–5]、理解野生动物与农田之间的相互作用 [6, 7] 以及区域作物损失评估 [8, 9]。然而，不同地区的作物类型地图的可用性和准确性差异很大：美国和欧盟维护着大型且定期更新的

实现,所有预处理后的数据集均可从 Hugging Face<sup>1</sup> 下载,所有代码可在 GitHub<sup>2</sup> 上获取。

## II. 数据整理

文献中存在数十个现有的作物类型分类数据集,其标签质量和作物类型类别各不相同。虽然有一些小规模的数据集是由专家手动标注的,但大多数大规模数据集都是由机器学习模型生成的,在可以可靠地用于进一步模型训练之前必须首先经过专家的审查。为了促进全球作物类型分类模型的训练,我们手动检查并验证了所有公开可用的作物类型分类数据集的有效性,并从五个大洲中选择了五个因其规模和标签准确性而值得注意的数据集。表 I 列出了这五个数据集以及原始作者发布的关于数据集大小和估计准确水平的统计信息。

由于每个数据集包括不同数量的作物类别且粒度水平各不相同,我们首先将所有掩码标签统一为六类:0: 无数据(未知),1: 玉米,2: 大豆,3: 水稻,4: 小麦,5: 其他(已知)。玉米、大豆、水稻和小麦被选为我们研究中的四大主要谷物作物。这四种农作物代表了价值最高的四种作物,占全球谷物产量的绝大多数以及全球大约一半的农业用地 [32]。它们在我们选择的所有数据集中普遍存在,尽管频率差异很大。我们还为“其他”类别保留了位置,包括其他形式的农业和非农业土地使用,以及“无数据”,即地表覆盖类型未知且可能或不可能包含我们的四种感兴趣的作物之一。本研究计算指标时忽略了所有无数据像素,以避免不公平地惩罚模型。

课程调和后,我们将每个数据集切分为  $256 \times 256$  px 的块。由于这些数据集中大多数以栅格和矢量掩模层的形式发布,并且没有相应的影像,我们从 Google Cloud<sup>3</sup> 下载了每个标注区域的自己的 Sentinel-2 影像。对于每个掩模,我们在掩模获取年份生长季节高峰时期(根据纬度而定)下载同一位置的一张无云 Sentinel-2 L1C 图像块。这也适用于那些随附影像的一些数据集,以确保所有 Sentinel-2 光谱带都存在于图像中。然后将所有图像和掩模扭曲为分辨率为 10 m/px 的 Web Mercator 投影。虽然时间序列信息可能会提高模型性能,但我们在这项研究中并未考虑这一点,因为并非所有模型都支持卫星影像时间序列。为了计算可行性及更好地代表

现实世界有限的可用数据,我们对每个区域进行了 1,000 张图像的子采样,总共得到了 5,000 张图像。

## III. 方法论

### A. ResNet-50 与 U-Net

一个具有 U-Net 架构的 ResNet-50 骨干网被提出用于提取卫星图像的空间特征以进行作物类型映射。ResNet [33], 由卷积层和残差块组成,是计算机视觉任务中常用的一种深度神经网络,用于特征提取。我们选择了 ResNet-50,因为它具有增加的模型复杂性、捕捉表示的能力以及可用的预训练权重。使用 ResNet-50 作为编码器来提取特征可以提高下游领域的泛化能力和训练效率,特别是在标记数据有限的情况下。U-Net [34] 的解码器在由 ResNet-50 骨干网提取的特征图之上操作。在整个实验过程中,我们冻结了 ResNet-50 编码器权重,并微调了 U-Net 解码器权重,这使得训练速度更快,并符合我们的迁移学习目标。

### B. 预训练权重

本研究评估了以下预训练权重:

1) *SSL4EO-S12* [24]: 提供了使用 MoCo-v2 [35] 和 DINO [36] 进行预训练的 ResNet 和 Vision Transformer (ViT) [37] 主干网络。全球预训练数据集包括 Sentinel-1 GRD 和 Sentinel-2 L1C/L2A 影像。

Satlas 预训练 [25]

2) *Satlas* 预训练 [25]: 提供了在全局 Landsat 8/9、Sentinel-1/2 和国家农业影像计划 (NAIP) 图像上预训练的 ResNet 和 Swin Transformer [38] 主干网络。

3) *ImageNet* [39]: 作为基线是因为其在计算机视觉研究中的易用性和广泛使用。这一基线使我们能够将性能与未在卫星图像上训练的模型进行比较。

### C. 迁移学习

我们进行了若干迁移学习实验,包括域内和域外评估。在这里,我们将域内 (ID) 定义为与模型训练所用的相同地理区域的数据,将域外 (OOD) 定义为在训练期间未见过的不同地理区域的数据。在域内实验中,我们试图了解性能如何随着增加的域内数据量而提高,范围从 10 到 900 个域内样本。在域外实验中,我们研究了结合不同数量的域内数据与域外数据的影响。在这项实验中,我们使用所有域外数据,意味着最多可使用来

<sup>1</sup>[https://huggingface.co/datasets/torchgeo/harmonized\\_global\\_crops](https://huggingface.co/datasets/torchgeo/harmonized_global_crops)

<sup>2</sup><https://github.com/yichi/crop-type-transfer-learning>

<sup>3</sup><https://cloud.google.com/storage/docs/public-datasets/sentinel-2>

表 I: 用于本研究的数据集列表。所有数据集都经过了协调和子采样，每类包含 1,000 个斑块，以实现公平比较。类别数量包括“其他”但不包括“无数据”。分辨率以每像素的米数来衡量。通过地面调查或自我声明收集的标签被认为是接近 100% 准确的，而机器学习标注的数据集则列出了它们在主要作物类别上的报告总体准确性。我们基于所有数据集的平均值计算协调数据集的准确性。数据集按字母顺序列出。

| Dataset                                 | Region        | # Classes | Resolution (m) | Accuracy      | License      |
|-----------------------------------------|---------------|-----------|----------------|---------------|--------------|
| Cropland Data Layer (CDL) [27]          | USA           | 134       | 30             | 84.1%         | CC0-1.0      |
| EuroCrops [28]                          | Europe        | 331       | -              | Self-declared | CC-BY-SA-4.0 |
| Northeastern China Crop Map (NCCM) [29] | China         | 4         | 10             | 87.0%         | CC-BY-4.0    |
| South Africa Crop Type (SACT) [30]      | South Africa  | 9         | 10             | Ground survey | CC-BY-4.0    |
| South America Soybean (SAS) [31]        | South America | 2         | 30             | Ground survey | Unknown      |
| Harmonized                              | Global        | 5         | 10             | 94.2%         | CC-BY-SA-4.0 |

自其他地区的 4,000 个样本。我们改变域内训练样本的数量以分析它们对模型性能的贡献情况，特别是在域外数据可用时的情况。所有实验均采用 90-10 的训练-测试划分。

#### IV. 结果与讨论

表 II: 不同预训练权重和 ID 样本量的平均 F1 分数。SSL4EO-S12 在所有地区始终优于其他两种预训练权重。增加 ID 数据可以提升模型性能。

| Weights   | ImageNet | SatlasPretrain | SSL4EO-S12 |        |             |
|-----------|----------|----------------|------------|--------|-------------|
| # Samples | 900 ID   | 900 ID         | 10 ID      | 100 ID | 900 ID      |
| CDL       | 0.58     | 0.57           | 0.28       | 0.38   | <b>0.77</b> |
| EuroCrops | 0.39     | 0.35           | 0.26       | 0.30   | <b>0.48</b> |
| NCCM      | 0.59     | 0.37           | 0.17       | 0.38   | <b>0.74</b> |
| SACT      | 0.32     | 0.43           | 0.18       | 0.16   | <b>0.70</b> |
| SAS       | 0.82     | 0.74           | 0.22       | 0.27   | <b>0.85</b> |

##### A. 预训练权重比较

在第一次实验中，我们探索了各种预训练权重，以找到一种适合 ID 设置下作物类型映射的优秀权重。表 II 显示，在所有研究区域中，SSL4EO-S12 始终比其他所有权重高出 3-27% 的表现。相比之下，SatlasPretrain 的表现与其他模型相比处于次优状态。这可能是因为该模型仅接受 13 个光谱带中的 9 个作为输入，而 SSL4EO-S12 接收全部 13 个，并且可以轻松使用 timm [40] 等库重复 ImageNet 权重。所有模型在 EuroCrops 和 SACT 上的表现较低，在这些区域无数据像素最多。然而，SSL4EO-S12 在这类数据稀缺的地区表现出稳健性。

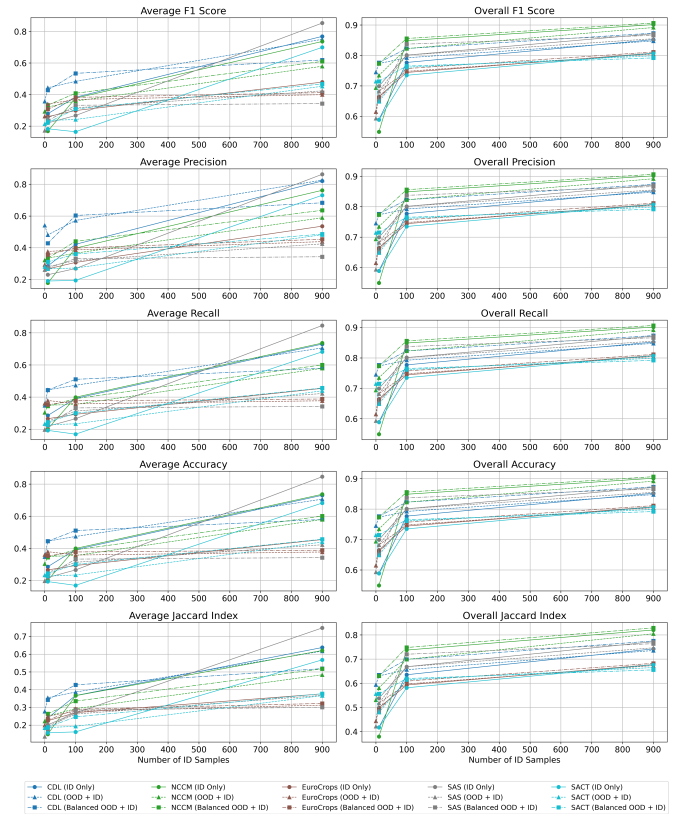


图 1: 报告使用 SSL4EO-S12 预训练权重的 ID、OOD + ID 和平衡 OOD + ID 指标。给出了 F1 分数、精度、召回率、准确率和 Jaccard 指数 (IoU) 的平均值和总体指标。对于所有指标，数值越高越好。

## B. 域内评估

我们使用表现最佳的预训练权重 (SSL4EO-S12) 来探索 ID 训练数据量对模型性能的影响。表 II 和图 1 显示, 所有区域都受益于增加 ID 数据, 正如预期的那样。然而, 虽然 100 个样本足以实现高总体准确率, 但需要 900 个样本才能达到高平均值准确率。额外的样本似乎对于模型克服所有数据集中存在的极端类别不平衡是必要的。

## C. 少样本学习评估

我们在 OOD 数据上进行了三项实验: 零样本学习, 仅使用 OOD 数据进行训练; 少样本学习, 将 OOD 数据与不断增加数量的 ID 数据结合使用; 以及一种平衡变化的少样本学习, 其中所有 ID/OOD 划分使用的总训练数据量相同。这些实验的结果如表 III 所示。

在有限的 ID 样本场景中, 对 OOD 数据进行预训练的优势尤为显著。通过比较表 II 和 表 III, 我们可以看到, 在没有使用任何 ID 样本 (零样本) 的情况下用 OOD 样本训练的模型已经能够超越使用 10 个 ID 样本训练的模型。这表明了利用 OOD 数据预训练权重的好处, 可以在目标区域无法获取 ID 数据时缓解困难。

然而, 结果也显示当更多的 ID 样本可用时, OOD 可能导致性能更差。例如, 在 4,000 个 OOD + 900 个 ID 样本上训练的模型的表现不如仅在 900 个 ID 样本上训练的模型。这是由于 ID 和 OOD 数据之间的分布变化, 因为与目标区域相比, OOD 数据通常具有不同的类别平衡和领域形状。我们建议使用 OOD 数据进行预训练, 但仅针对 ID 数据进行微调以处理这种分布变化。

虽然在遥感中最常见的场景是一个大型固定大小的 OOD 数据集和不同数量的 ID 数据, 但很难区分由于额外的 ID 数据带来的准确性提升与由于额外训练数据 (OOD + ID) 带来的提升。我们在固定的训练规模下重复上述实验, 并改变 ID 和 OOD 数据的比例。尽管平均 F1 分数有些微变化, 总体趋势保持不变。

## D. 定性评估

图 2 显示了每个数据集的示例预测。掩码和预测之间最明显的视觉差异是失去了像道路这样的细尺度特征, 这在使用基础模型进行 CDL 作物分类的先前工作中已经可以看到 [41]。这一点在 NCCM 中尤为明显, 在那里可以看到玉米和大豆之间的套种 (即在同一区域内

种植不同作物的行)。在 EuroCrops 和 SACT 预测中, 我们可以看到模型难以区分小麦和其他作物, 这可能是由于杂草与它们外观相似的缘故。

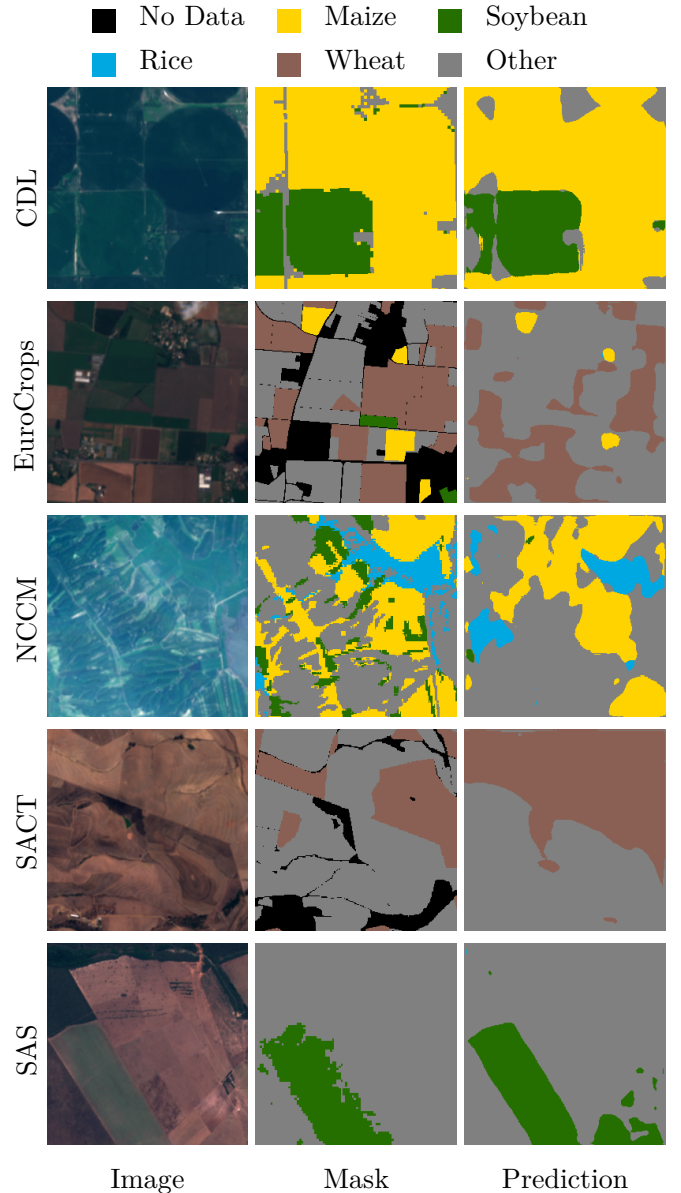


图 2: 示例输入 Sentinel-2 图像、地面真实掩码和模型预测的可视化, 使用 SSL4EO-S12 预训练权重。总体结果令人满意, 模型捕捉到了一般的类别分布, 并正确识别了大多数区域。

## V. 结论

在这项工作中, 我们提出了一个通过协调来自五大洲的五个作物分类数据集而创建的全球语义分割数据集用于作物类型映射。我们研究了各种预训练基础模型



表 III: ID-OOD 样本使用 SSL4EO-S12 预训练权重的平均 F1 分数。结果显示了用有限的 ID 数据进行 OOD 预训练的好处，以及当有更多的 ID 样本可用时优先考虑 ID 样本的重要性。

| Dataset   | 4000 OOD | 4000 OOD |         |             | 3990 OOD | 3900 OOD | 3100 OOD    |
|-----------|----------|----------|---------|-------------|----------|----------|-------------|
|           |          | +10 ID   | +100 ID | +900 ID     | +10 ID   | +100 ID  | +900 ID     |
| CDL       | 0.36     | 0.44     | 0.48    | <b>0.75</b> | 0.43     | 0.53     | <b>0.62</b> |
| EuroCrops | 0.26     | 0.34     | 0.37    | <b>0.40</b> | 0.31     | 0.38     | <b>0.41</b> |
| NCCM      | 0.29     | 0.32     | 0.36    | <b>0.58</b> | 0.33     | 0.41     | <b>0.61</b> |
| SACT      | 0.21     | 0.24     | 0.24    | <b>0.45</b> | 0.22     | 0.31     | <b>0.47</b> |
| SAS       | 0.17     | 0.25     | 0.32    | <b>0.42</b> | 0.23     | 0.33     | <b>0.34</b> |

在 ID 和 ID + OOD 样本设置下执行作物分类的能力。我们发现，如 SSL4EO-S12 这样的基于 Sentinel-2 影像所有光谱带进行预训练的模型优于其他竞争模型。我们还发现在数据有限的地区，OOD 数据可以提升性能。然而，如果不谨慎处理，在 ID 和 OOD 数据上同时训练实际上可能会损害模型性能。

更大的数据集仍然需要，特别是对于非洲和南美洲这样的地区，在这些地方大型多类数据集非常稀缺。然而，数据集的大小并不是唯一重要的因素，所有用于本研究的数据集都受到类别不平衡的影响。为了更精确地预测不常见的类别，特别是在模型从最重要的四种谷物转移到数百个农业类别时，需要解决类别不平衡的问题，可以通过加权数据集采样或加权损失函数来实现。

#### 参考文献

- [1] E. A. Kebede, H. Abou Ali, T. Clavelle, H. E. Froehlich, J. A. Gephart, S. Hartman, M. Herrero, H. Kerner, P. Mehta, C. Nakalembe *et al.*, “Assessing and addressing the global state of food production data scarcity,” *Nature Reviews Earth & Environment*, vol. 5, no. 4, pp. 295–311, 2024.
- [2] S. Khaki and L. Wang, “Crop yield prediction using deep neural networks,” *Frontiers in plant science*, vol. 10, p. 452963, 2019.
- [3] P. C. Doraiswamy, S. Moulin, P. W. Cook, and A. Stern, “Crop yield assessment from remote sensing,” *Photogrammetric Engineering & Remote Sensing*, vol. 69, no. 6, pp. 665–674, 2003.
- [4] T. Van Klompenburg, A. Kassahun, and C. Catal, “Crop yield prediction using machine learning: A systematic literature review,” *Computers and Electronics in Agriculture*, vol. 177, p. 105709, 2020.
- [5] J. Fan, J. Bai, Z. Li, A. Ortiz-Bobea, and C. P. Gomes, “A GNN-RNN approach for harnessing geospatial and temporal information: Application to crop yield prediction,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 11, pp. 11 873–11 881, Jun. 2022. [Online]. Available: <https://ojs.aaai.org/index.php/AAAI/article/view/21444>
- [6] E. M. Gross, B. P. Lahkar, N. Subedi, V. R. Nyirenda, L. L. Lichtenfeld, and O. Jakoby, “Seasonality, crop type and crop phenology influence crop damage by wildlife herbivores in africa and asia,” *Biodiversity and Conservation*, vol. 27, pp. 2029–2050, 2018.
- [7] R. F. Pywell, M. S. Heard, B. A. Woodcock, S. Hinsley, L. Ridding, M. Nowakowski, and J. M. Bullock, “Wildlife-friendly farming increases crop yield: evidence for ecological intensification,” *Proceedings of the Royal Society B: Biological Sciences*, vol. 282, no. 1816, p. 20151740, 2015.
- [8] N. Silleos, K. Perakis, and G. Petsanis, “Assessment of crop damage using space remote sensing and gis,” *International Journal of Remote Sensing*, vol. 23, no. 3, pp. 417–427, 2002.
- [9] M. S. Rahman and L. Di, “A systematic review on case studies of remote-sensing-based flood crop loss assessment,” *Agriculture*, vol. 10, no. 4, p. 131, 2020.
- [10] A. Verhegghen, R. d’Andrimont, F. Waldner, and M. Van der Velde, “Accuracy assessment of the first Eu-wide crop type map with lucas data,” in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 2021, pp. 1990–1993.
- [11] D. Luman and T. Tweddale, “Assessment and potential of the 2007 usda-nass cropland data layer for statewide annual land cover applications,” *Technical Report INHS 2008 (49)*, 2008.
- [12] A. J. Prins and A. Van Niekerk, “Crop type mapping using lidar, sentinel-2 and aerial imagery with machine learning algorithms,” *Geo-Spatial Information Science*, vol. 24, no. 2, pp. 215–227, 2021.
- [13] J. Adrian, V. Sagan, and M. Maimaitijiang, “Sentinel sar-optical fusion for crop type mapping using deep learning

- and google earth engine,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 175, pp. 215–235, 2021.
- [14] P. Tang, J. Chanussot, S. Guo, W. Zhang, L. Qie, P. Zhang, H. Fang, and P. Du, “Deep learning with multi-scale temporal hybrid structure for robust crop mapping,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 209, pp. 117–132, 2024.
- [15] Y. Yuan, L. Lin, Z.-G. Zhou, H. Jiang, and Q. Liu, “Bridging optical and sar satellite image time series via contrastive feature extraction for crop classification,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 195, pp. 222–232, 2023.
- [16] Y. Ma, S. Chen, S. Ermon, and D. B. Lobell, “Transfer learning in environmental remote sensing,” *Remote Sensing of Environment*, vol. 301, p. 113924, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425723004765>
- [17] C. Broni-Bediako, J. Xia, J. Song, H. Chen, M. Siam, and N. Yokoya, “Generalized few-shot semantic segmentation in remote sensing: Challenge and benchmark,” *IEEE Geoscience and Remote Sensing Letters*, vol. 21, pp. 1–5, 2024.
- [18] A. Koukos, H.-W. Jo, V. Sitokoustantinou, I. Tsoumas, C. Kontoes, and W.-K. Lee, “Towards global crop maps with transfer learning,” in *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, 2024, pp. 1540–1545.
- [19] H. Kerner, S. Chaudhari, A. Ghosh, C. Robinson, A. Ahmad, E. Choi, N. Jacobs, C. Holmes, M. Mohr, R. Dodhia, J. M. L. Ferres, and J. Marcus, “Fields of The World: A machine learning benchmark dataset for global agricultural field boundary segmentation,” 2024. [Online]. Available: <https://arxiv.org/abs/2409.16252>
- [20] S. Sachdeva, I. Lopez, C. Biradar, and D. Lobell, “A distribution shift benchmark for smallholder agroforestry: Do foundation models improve geographic generalization?” *The Twelfth International Conference on Learning Representations 2024 Machine Learning for Remote Sensing (ML4RS) Workshop*, 2024.
- [21] G. Tseng, R. Cartuyvels, I. Zvonkov, M. Purohit, D. Rolnick, and H. Kerner, “Lightweight, pre-trained transformers for remote sensing timeseries,” 2024. [Online]. Available: <https://arxiv.org/abs/2304.14065>
- [22] G. Tseng, I. Zvonkov, C. Nakalembe, and H. R. Kerner, “CropHarvest: A global dataset for crop-type classification,” in *NeurIPS Datasets and Benchmarks*, 2021. [Online]. Available: <https://api.semanticscholar.org/CorpusID:248529758>
- [23] M. Allen, F. Dorr, J. A. Gallego-Mejia, L. Martínez-Ferrer, A. Jungbluth, F. Kalaitzis, and R. Ramos-Pollán, “Fewshot learning on global multimodal embeddings for earth observation tasks,” 2023. [Online]. Available: <https://arxiv.org/abs/2310.00119>
- [24] Y. Wang, N. A. A. Braham, Z. Xiong, C. Liu, C. M. Albrecht, and X. Zhu, “SSL4EO-S12: A large-scale multimodal, multitemporal dataset for self-supervised learning in earth observation [software and data sets],” *IEEE Geoscience and Remote Sensing Magazine*, vol. 11, pp. 98–106, 2023. [Online]. Available: <https://api.semanticscholar.org/CorpusID:262975520>
- [25] F. Bastani, P. Wolters, R. Gupta, J. Ferdinando, and A. Kembhavi, “SatlasPretrain: A large-scale dataset for remote sensing image understanding,” *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 16 726–16 736, 2022. [Online]. Available: <https://api.semanticscholar.org/CorpusID:258947021>
- [26] A. J. Stewart, C. Robinson, I. A. Corley, A. Ortiz, J. M. Lavista Ferres, and A. Banerjee, “TorchGeo: Deep learning with geospatial data,” in *Proceedings of the 30th International Conference on Advances in Geographic Information Systems*, ser. SIGSPATIAL ’22. Seattle, Washington: Association for Computing Machinery, Nov. 2022, pp. 1–12. [Online]. Available: <https://dl.acm.org/doi/10.1145/3557915.3560953>
- [27] USDA NASS, “Cropland Data Layer,” USDA NASS Marketing and Information Services Office, Washington, D.C., 2024. [Online]. Available: <https://croplandcros.scinet.usda.gov/>
- [28] M. Schneider, T. Schelte, F. Schmitz, and M. Körner, “EuroCrops: The largest harmonized open crop dataset across the European Union,” *Scientific Data*, vol. 10, no. 1, p. 612, Sep. 2023. [Online]. Available: <https://doi.org/10.1038/s41597-023-02517-0>
- [29] N. You, J. Dong, J. Huang, G. Du, G. Zhang, Y. He, T. Yang, Y. Di, and X. Xiao, “The 10-m crop type maps in northeast china during 2017–2019,” *Scientific data*, vol. 8, no. 1, p. 41, 2021. [Online]. Available: <https://doi.org/10.1038/s41597-021-00827-9>
- [30] Western Cape Department of Agriculture and Radiant Earth Foundation, “Crop type classification dataset for western cape, south africa,” Radiant MLHub, 2021, version 1.0. [Online]. Available: <https://doi.org/10.34911/rdnt.j0co8q>
- [31] X.-P. Song, M. C. Hansen, P. Potapov, B. Adusei, J. Pickering, M. Adami, A. Lima, V. Zalles, S. V. Stehman,

- C. M. Di Bella *et al.*, “Massive soybean expansion in south america since 2000 and implications for conservation,” *Nature sustainability*, vol. 4, no. 9, pp. 784–792, 2021.
- [32] A. R. Martin, M. W. Cadotte, M. E. Isaac, R. Milla, D. Vile, and C. Violle, “Regional and global shifts in crop diversity through the anthropocene,” *PLoS One*, vol. 14, no. 2, p. e0209788, 2019.
- [33] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [34] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.
- [35] X. Chen, H. Fan, R. Girshick, and K. He, “Improved baselines with momentum contrastive learning,” *arXiv preprint arXiv:2003.04297*, 2020.
- [36] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, and A. Joulin, “Emerging properties in self-supervised vision transformers,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021, pp. 9650–9660.
- [37] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, “An image is worth 16x16 words: Transformers for image recognition at scale,” 2021. [Online]. Available: <https://arxiv.org/abs/2010.11929>
- [38] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, “Swin Transformer: Hierarchical vision transformer using shifted windows,” in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 9992–10 002.
- [39] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.
- [40] R. Wightman, “PyTorch Image Models,” <https://github.com/rwightman/pytorch-image-models>, 2019.
- [41] A. Stewart, N. Lehmann, I. Corley, Y. Wang, Y.-C. Chang, N. Ait Ali Braham, S. Sehgal, C. Robinson, and A. Banerjee, “SSL4EO-L: Datasets and foundation models for Landsat imagery,” *Advances in Neural Information Processing Systems*, vol. 36, 2024.