PRACTICAL EMBEDDING WORKFLOWS WITH TERRA-TORCH, THE GEOSPATIAL FINE-TUNING TOOLKIT

Anonymous authors
Paper under double-blind review

ABSTRACT

Geospatial foundation models pretrained on large-scale Earth observation archives offer strong transfer capabilities across remote sensing tasks, but practical adoption remains challenging due to heterogeneous data formats, complex fine-tuning pipelines, and inconsistent evaluation protocols. We present TerraTorch, a configuration-driven toolkit for reproducible adaptation and benchmarking of geospatial foundation models.

This workshop contribution complements earlier TerraTorch system work by focusing specifically on embedding-centric workflows: (i) generic embedding generation from pretrained encoders and (ii) downstream learning on top of frozen embeddings for semantic segmentation. All demonstrations are linked to executable repository examples, lowering the barrier for ML4RS researchers and practitioners to apply foundation models in real-world Earth observation settings.

1 INTRODUCTION

Machine learning for remote sensing (ML4RS) presents unique challenges due to the high dimensionality, spectral complexity, and spatial–temporal structure of Earth observation data. Large geospatial foundation models (GFMs) pretrained on vast unlabeled satellite and climate datasets offer a promising foundation for downstream tasks in segmentation, classification, regression, and related applications. However, adapting these models remains non-trivial due to heterogeneous preprocessing, model configuration complexity, and evaluation heterogeneity.

In the ML4RS workshop at ICLR 2026, we emphasize translating research into practice through reproducible workflows that make foundation models accessible to the broader community. TerraTorch addresses these challenges by offering a flexible, configuration-driven toolkit for fine-tuning GFMs and benchmarking them on standardized tasks described in GeoBenchV2 (Simumba et al., 2025) and Pangea (Marsocci et al., 2024). TerraTorch builds on established deep learning abstractions such as PyTorch Lightning (Falcon & The PyTorch Lightning team, 2019) and domain-specific tooling such as TorchGeo (Stewart et al., 2022).

Importantly, this workshop contribution is not intended as a repetition of the full TerraTorch system description introduced in earlier work (Gomez, 2025). Instead, we focus on a specific emerging paradigm in ML4RS: embedding-first workflows, where pretrained geospatial encoders are used to generate reusable representations that support efficient downstream adaptation.

2 BACKGROUND AND RELATED WORK

Foundation Models for Earth Observation. Geospatial foundation models are large neural networks pretrained on multi-modal remote sensing data to acquire general representations useful across tasks. Examples include Prithvi-EO (Szwarcman et al., 2026) and TerraMind (Jakubik et al., 2025). Despite their promise, applying these models requires substantial effort for data handling, fine-tuning, and evaluation.

Existing Frameworks. Prior efforts for geospatial ML infrastructure include TorchGeo (Stewart et al., 2022) for domain data loaders and architectures. TerraTorch distinguishes itself by combining modular model instantiation, task engines, and end-to-end workflows under a single config-driven interface, with recent emphasis on embedding-centric pipelines.

Benchmarking and Reproducibility. Standardized evaluation remains challenging in geospatial ML. Benchmark suites such as GEO-Bench support reproducible comparison. TerraTorch adopts these standards directly, enabling systematic evaluation.

Ecosystem Tools. TerraTorch is part of the broader TerrastackAI ecosystem: (i) TerraKit (TerraKit, 2025) generates ML-ready datasets from sources such as SentinelHub (Sinergise, 2024), NASA EarthData (NASA, 2024), and STAC catalogs (STAC, 2021); (ii) Geospatial Studio (Geospatial Studio, 2025) provides scalable training, inference, and visualization interfaces; (iii) TerrastackAI promotes common interfaces and governance across these components.

3 TerraTorch System Design

3.1 Architecture Overview

TerraTorch is structured around three primary components that enable both end-to-end fine-tuning and embedding-based workflows:

Data Modules. Built on PyTorch Lightning DataModules and TorchGeo abstractions, these handle multi-band spatiotemporal ingestion, augmentation, batching, and integration with common geospatial formats such as GeoTIFF (GeoTIFF, 2019), netCDF (netCDF, 2023), and GeoParquet (GeoParquet, 2023).

Modular Model Factory. A factory pattern enables dynamic composition of pretrained backbones and task-specific heads. Supported backbones include Prithvi, TerraMind, Clay (Clay, 2024), DOFA (Xiong et al., 2024), DINOv3 (Chen et al., 2025), and generic models from timm. This modularity is equally important for fine-tuning and for representation extraction, where the backbone is used as a frozen embedding generator.

Task Engines and CLI. TerraTorch provides LightningModule task wrappers encapsulating losses, metrics, and training logic. Experiments are defined via YAML/JSON configuration files specifying data sources, model components, optimization, and evaluation metrics. The same configuration interface is used for embedding export workflows, enabling reproducible embedding generation at scale.

4 Tutorial Workflows and Demonstrations

The tutorial session accompanying this workshop paper is organized around two embedding-centric workflows. Together, they demonstrate how TerraTorch supports (i) generic embedding generation from pretrained geospatial foundation encoders and (ii) efficient downstream learning on top of frozen embeddings. All resources for the tutorials can be found in (Embedding Examples, 2025).

4.1 Tutorial 1: Generic Embedding Generation with TerraTorch

A central tutorial workflow focuses on embedding generation as a generic and reusable capability of TerraTorch. Instead of immediately fine-tuning foundation models end-to-end for a specific task, TerraTorch supports an embedding-first paradigm in which pretrained geospatial foundation models are used as representation engines.

In this workflow, participants learn how to extract embeddings from arbitrary Earth observation imagery (e.g., Sentinel-2, HLS, or custom multispectral data) and store them as structured intermediate features. These embeddings can then serve as a common substrate for many downstream use cases, including semantic segmentation, regression, retrieval, anomaly detection, or lightweight probing tasks.

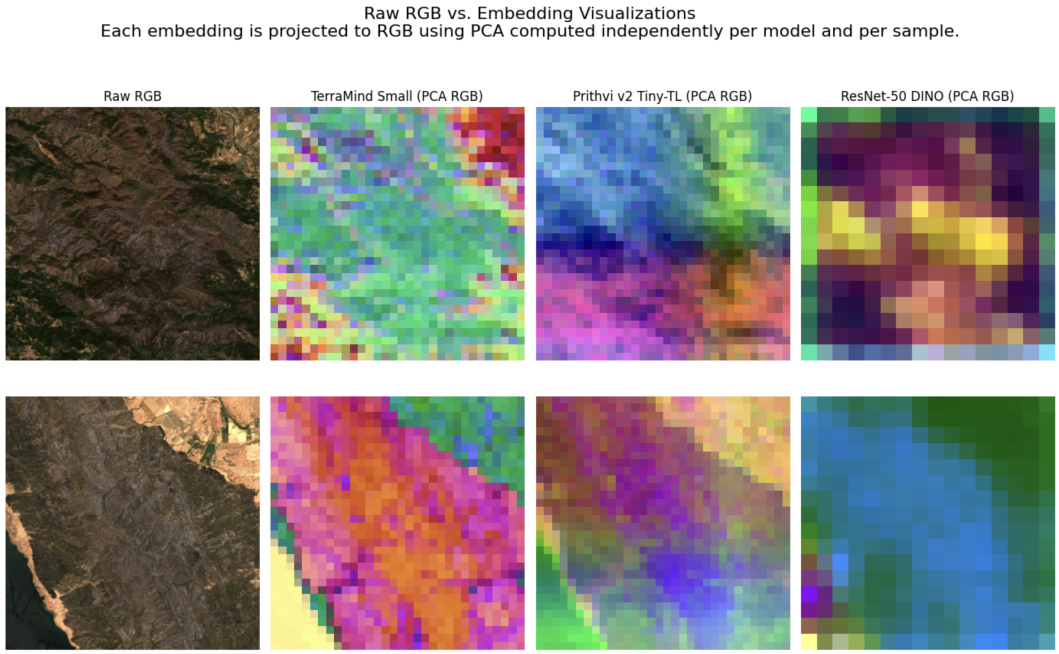

Figure 1: PCA visualizations comparing embeddings of various backbones in TerraTorch.

The tutorial demonstrates how TerraTorch treats embedding extraction as a first-class workflow, fully integrated into its configuration-driven interface. Using a YAML configuration, users specify:

- the input data module and preprocessing pipeline,
- the pretrained foundation backbone used for feature extraction,
- the embedding granularity (patch-level or tile-level pooling),
- the export format and output structure for downstream reuse.

A key practical motivation is that embedding extraction can be performed once at scale, decoupling the expensive encoder forward pass from later task-specific training. This enables rapid experimentation across multiple downstream models without repeatedly recomputing representations. For example, the different embeddings illustrated in Figure 1 have been created by simply changing a single string selecting the backbone in the TerraTorch configuration file.

Participants will follow the provided example notebook `embedding_generation_burnscars.ipynb`, which serves as a concrete template for generic embedding generation. The full pipeline is specified declaratively in the companion configuration file `embedding_generation_burnscars.yaml`.

The resulting embeddings are stored in GeoParquet format, preserving geospatial indexing while remaining efficient for large-scale machine learning workflows. Overall, this tutorial establishes the foundation for embedding-centric ML4RS pipelines, where a single pretrained encoder can support many downstream applications through reusable representations.

4.2 Tutorial 2: Embedding Application for Downstream Burn Scar Segmentation

The second tutorial builds directly on the embeddings generated in the first workflow and demonstrates one concrete downstream application: semantic segmentation of burn scars. Burn scar mapping is a highly relevant remote sensing task, since burned areas are characterized by subtle spectral changes and spatial context that benefit from foundation model representations.

Rather than training a full segmentation network from raw imagery, TerraTorch is configured to consume the precomputed embeddings as the primary input modality. A lightweight segmentation decoder - in this case a MLP - is then trained on top of these frozen representations. This separation provides substantial practical advantages:

- reduced compute requirements, since the backbone is not re-executed,
- faster convergence, as only task-specific head parameters are updated,
- rapid benchmarking of multiple decoder architectures on the same embedding set.

Participants will run the notebook `downstream_segmentation_burnscars.ipynb`, with the experiment fully specified in `downstream_segmentation_burnscars.yaml`. The configuration defines:

- the embedding dataset loading strategy,
- the downstream segmentation head and loss functions,
- evaluation metrics appropriate for burn scar detection,
- training and validation splits aligned with reproducible benchmarking.

This tutorial highlights how TerraTorch enables efficient adaptation workflows: once embeddings are available, downstream segmentation models can be trained and compared quickly, without repeating costly feature extraction.

Together, the two tutorials provide an end-to-end demonstration of TerraTorch's support for generic embedding generation and practical embedding-based downstream learning in ML4RS.

5 Credits

The embedding workflow integration and maintenance in TerraTorch are carried out as part of the Embed2Scale project (Earth Observation & Weather Data Federation with AI Embeddings), funded by the EU's Horizon Europe programme (Grant Agreement No. 101131841), with additional support from SERI and UKRI.

6 Conclusion

TerraTorch provides a reproducible and extensible toolkit for applying geospatial foundation models across remote sensing workflows. This workshop contribution complements earlier TerraTorch system work by focusing on practical embedding-centric demonstrations: generic embedding generation and downstream segmentation on top of frozen representations. By linking these workflows directly to executable repository examples, TerraTorch lowers the barrier for researchers and practitioners to adopt foundation models in real-world Earth observation applications.

References

Wei Chen, Ravi Patel, Angel Gómez, Tian Li, Akshay Gupta, Sarah More, Daniel Jones, Lei Wang, Rodrigo Sánchez, and Lukas Müller. Satcls: A standardized benchmark for satellite image classification. *arXiv preprint arXiv:2508.10104*, 2025. doi: 10.48550/arXiv.2508.10104. URL `https://arxiv.org/abs/2508.10104`. A satellite image classification benchmark for evaluating geospatial foundation models.

Clay. Clay foundation model for remote sensing representations. `https://clay-foundation.github.io/model/`, 2024. Model release.

Embedding Examples. Terratorch embedding examples: Embedding generation and downstream segmentation. `https://github.com/terrastackai/terratorch/tree/main/examples/embeddings`, 2025. Example notebooks and configs for embedding workflows, accessed 2026.

William Falcon and The PyTorch Lightning team. PyTorch Lightning, March 2019. URL https://github.com/Lightning-AI/pytorch-lightning.

GeoParquet. Geoparquet: A columnar storage format for geospatial data. https://geoparquet.org/, 2023. Open specification.

Geospatial Studio. Geospatial studio: Scalable geospatial ai training & visualization platform. https://github.com/terrastackai/geospatial-studio, 2025. GitHub repository, accessed 2026.

Open Geospatial Consortium GeoTIFF. Geotiff format specification. https://www.ogc.org/standards/geotiff, 2019. Geospatial raster standard.

Gomez. Terratorch: A configuration-driven toolkit for fine-tuning geospatial foundation models. *arXiv preprint arXiv:2503.20563*, 2025. doi: 10.48550/arXiv.2503.20563. URL https://arxiv.org/abs/2503.20563. Toolkit for geospatial foundation model adaptation and benchmarking.

Johannes Jakubik, Felix Yang, Benedikt Blumenstiel, Erik Scheurer, Rocco Sedona, Stefano Maurogiovanni, Jente Bosmans, Nikolaos Dionelis, Valerio Marsocci, Niklas Kopp, Rahul Ramachandran, Paolo Fraccaro, Thomas Brunschwiler, Gabriele Cavallaro, Juan Bernabé-Moreno, and Nicolas Longépé. Terramind: Large-scale generative multimodality for earth observation. *arXiv preprint arXiv:2504.11171*, 2025. doi: 10.48550/arXiv.2504.11171. URL https://arxiv.org/abs/2504.11171. Accepted at ICCV 2025; Multimodal foundation model for Earth observation.

Valerio Marsocci, Yuru Jia, Georges Le Bellier, David Kerekes, Liang Zeng, Sebastian Hafner, Sebastian Gerard, Eric Brune, Ritu Yadav, Ali Shibli, Heng Fang, Yifang Ban, Maarten Vergauwen, Nicolas Audebert, and Andrea Nascetti. Pangaea: A global and inclusive benchmark for geospatial foundation models. *arXiv preprint arXiv:2412.04204*, 2024. doi: 10.48550/arXiv.2412.04204. URL https://arxiv.org/abs/2412.04204. A standardized evaluation benchmark for geospatial foundation models (PANGAEA).

NASA. Nasa earthdata: Distributed active archive centers (daacs). https://earthdata.nasa.gov/, 2024. Earth observation data portal.

netCDF. Network common data form (netcdf). https://www.unidata.ucar.edu/software/netcdf/, 2023. Scientific data format.

Naomi Simumba, Nils Lehmann, Paolo Fraccaro, Hamed Alemohammad, Geeth De Mel, Salman Khan, Manil Maskey, Nicolas Longepe, Xiao Xiang Zhu, Hannah Kerner, Juan Bernabé-Moreno, and Alexander Lacoste. Geo-bench-2: From performance to capability, rethinking evaluation in geospatial ai. *arXiv preprint arXiv:2511.15658*, 2025. doi: 10.48550/arXiv.2511.15658. URL https://arxiv.org/abs/2511.15658. Accessed 2026.

Sinergise. Sentinel hub: Access to sentinel and other earth observation data. https://www.sentinel-hub.com/, 2024. Data platform.

STAC. Spatiotemporal asset catalogs (stac) specification. https://stacspec.org/, 2021. Open geospatial metadata standard.

Caleb Stewart, Caleb Robinson, Isaac Corley, Anthony Ortiz, and Juan M. Lavista Ferres. Torchgeo: Deep learning with geospatial data. *Proceedings of the 30th ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL)*, 2022.

Daniela Szwarcman, Sujit Roy, Paolo Fraccaro, THorsteinn Eli Gislason, Benedikt Blumenstiel, Rinki Ghosal, Pedro Henrique de Oliveira, Joao Lucas De Sousa Almeida, Rocco Sedona, Y. Kang, S. Chakraborty, Sizhe Wang, Carlos Gomes, Ankur Kumar, Vishal Gaur, Myscon Truong, Denys Godwin, Sam Khallaghi, Hyunho Lee, Chia Yu Hsu, Ata Akbari Asanjan, Besart Mujeci, Disha Shidham, Rufai Omowunmi Balogun, Venkatesh Kolluru, Trevor Keenan, Paulo Arévalo, Wenwen Li, Seyed Hamidreza Alemohammad, Pontus Olofsson, Timothy J. Mayer, Christopher Hain, Robert Kennedy, Bianca Zadrozny, David Bell, Gabriele Cavallaro, Campbell

Watson, Manil Maskey, Rahul Ramachandran, and Juan Bernabe Moreno. Prithvi-eo-2.0: A versatile multi-temporal foundation model for earth observation applications. *IEEE Transactions on Geoscience and Remote Sensing*, 64(1):1–20, jan 2026. doi: 10.1109/tgrs.2025.3642610. URL https://ieeexplore.ieee.org/document/11296896/.

TerraKit. Terrakit: Ml-ready geospatial dataset generation toolkit. https://github.com/terrastackai/terrakit, 2025. GitHub repository, accessed 2026.

Zhitong Xiong, Yi Wang, Fahong Zhang, Adam J. Stewart, Joëlle Hanna, Damian Borth, Ioannis Papoutsis, Bertrand Le Saux, Gustau Camps-Valls, and Xiao Xiang Zhu. Neural plasticity-inspired multimodal foundation model for earth observation. *arXiv preprint arXiv:2403.15356*, 2024. URL https://arxiv.org/abs/2403.15356. Dynamic One-For-All (DOFA) model for multimodal Earth observation.