Democratizing Foundation Models for Earth Observation Applications

A. Zappacosta[1], R. S. Kuzu[1], C. O. Dumitru[1], D. Espinoza Molina[1], T. Brunschwiler [2], R. Kienzler [2], J. Jakubik [2], B. Blumenstiel[2], G. Cavallaro [3], S. Kesselheim [3], R. Sedona [3], A. Wijata [4], L. Tulczyjew [4], D. Marek [4], J. Nalepa [4]

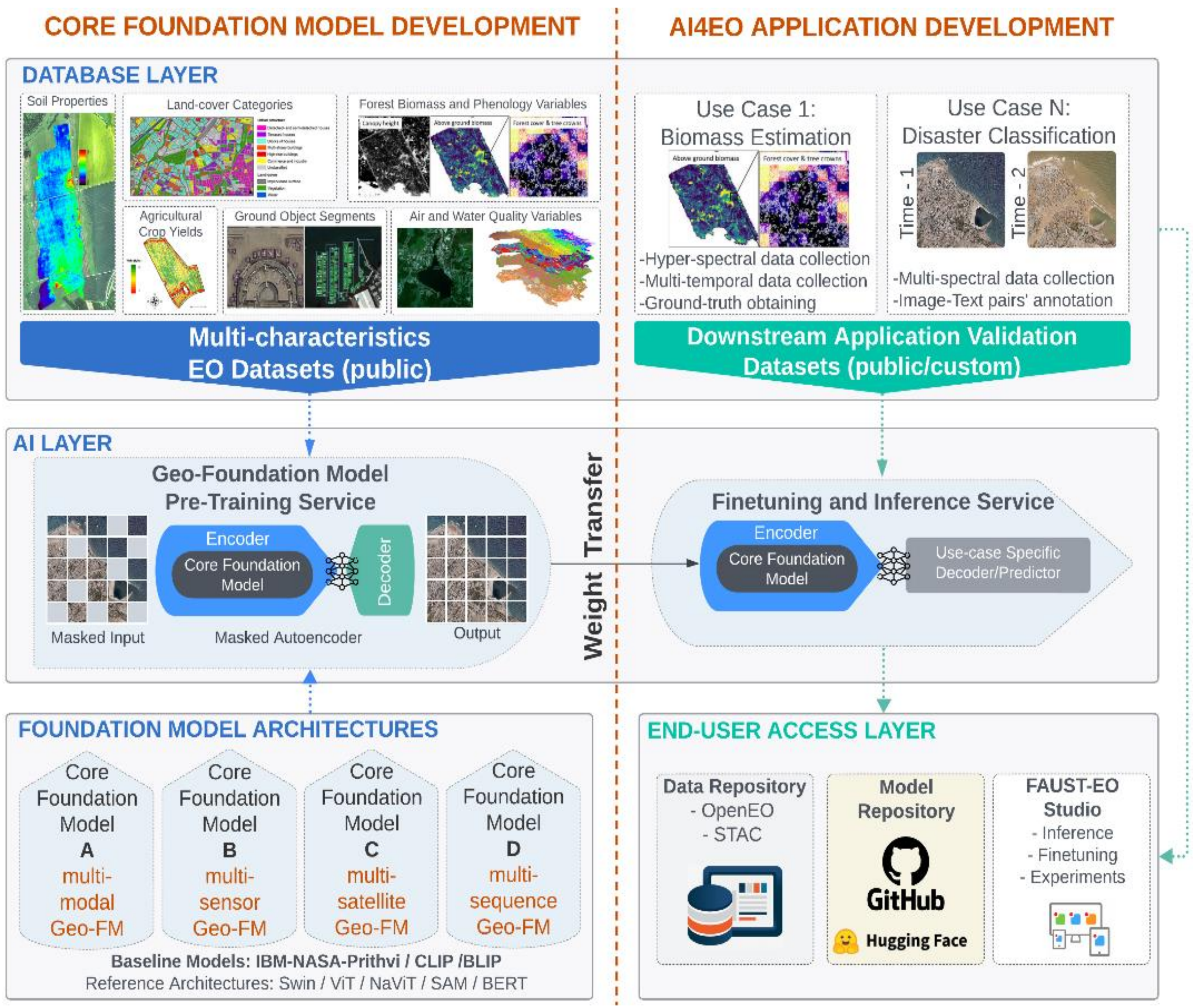
[1] Deutsches Zentrum für Luft- und Raumfahrt (DLR), Remote Sensing Institute (MF), Data Science Department (DAS) – Weßling; [2] IBM Research GmbH – Zurich; [3] Forschungszentrum Jülich GmbH (FZJ) – Jülich; [4] KP Labs - Gliwice

Abstract

In Earth Observation (EO), integrating Foundation Models (FMs) poses a challenge due to disparities with conventional computer vision tasks, particularly in processing multispectral/hyperspectral data. Additionally, the high computational resources needed for FM training limit their widespread use. Leveraging state-of-the-art FMs, we extend their utility to suit diverse sensor types, temporal characteristics, and domains, enhancing their applicability in various EO tasks. Our approach offers a flexible solution, accommodating different multi-modal FMs, and extends beyond specific models for broader EO applications. To demonstrate the effectiveness of this framework, we explore diverse EO use cases (UCs), spanning different thematic areas and geographic locations, showcase the framework's potential for generalization and adaptability across a range of EO challenges, as part of the FAST-EO project.

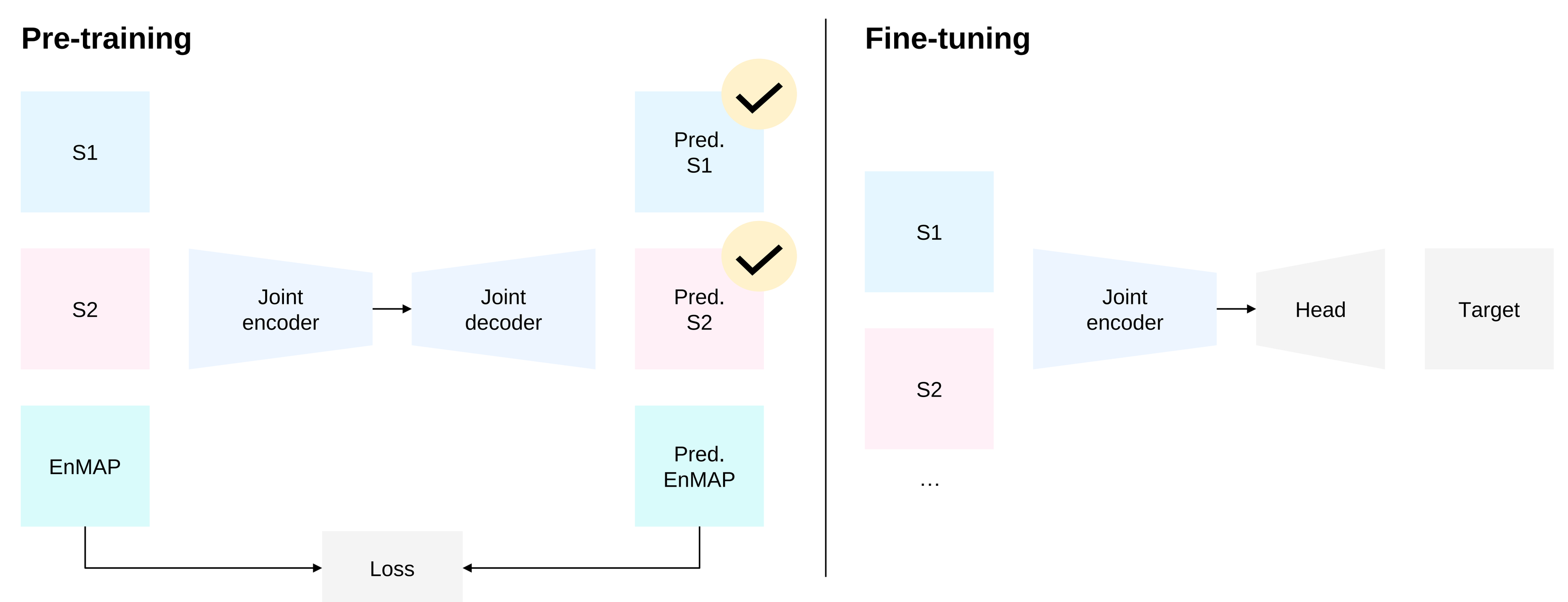
FAST-EO: A Multi-modal Geo-foundation Model

We will extend the Geo-Foundation Model [1-2], pre-trained by IBM and NASA, by including multi-sensor, multi-temporal, and multi-modal capabilities with more efficient backbones, as shown below:



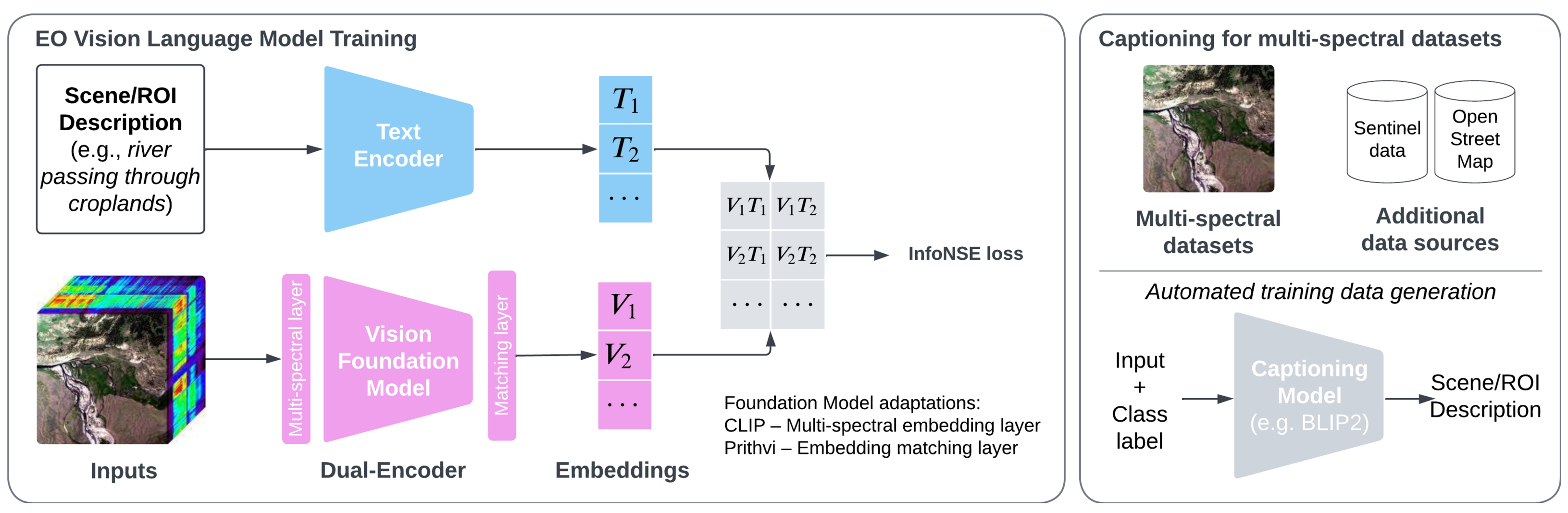
Joint Encoder-decoder Architecture for Vision Modalities

The models incorporating Sentinel-1, Sentinel-2, and EnMAP are being trained using a joint encoder-decoder framework, facilitatating the models to work interchangeably across these sensors, enabling subsequent fine-tuning for specific tasks, as illustrated below.



Contrastive Language and Image Pretraining for Multi-modality

Integration based on CLIP [3] and LLaMA [4] is being conducted for EO data to facilitate natural language interaction with scenes. This integration is supported by a training pipeline utilizing a dual-encoder architecture with multi-spectral image-caption samples. A large-scale dataset is being created, leveraging existing data sources and automated pipelines to ensure its extensive coverage while accommodating multi-spectral imagery and associated captions.



EO Use Cases for Solving Societal Challenges

- UC1: Weather & Climate Disaster Analysis
- UC2: Detection of Methane Leaks
- UC3: Observation of Changes in Forest Above-Ground Biomass
- UC4: Estimation of Soil Properties
- UC5: Detection of Semantic Land Cover Changes
- UC6: Monitoring Expansion of Mining Fields into Farmlands

UC1: Weather & Climate Disaster Analysis

The goal is to fine-tune Geo-Foundation Models (Geo-FMs) for segmenting natural disasters such as floods and wildfire scars at a pixel level using data from multiple satellites on a global scale.

UC2: Detection of Methane Leaks

The goal is to develop an AI-powered system for detecting methane leaks in satellite imagery to ensure global scalability and robustness of the solution against the varying acquisition conditions.

UC3: Observation of Changes in Forest Above Ground Biomass

The goal is to retrieve forest basal area, tree cover density, and tree heights, as well as to detect changes compared to a baseline scenario, including the causes of change, in mountainous regions where employing AI models poses challenges due to poor accuracy and high uncertainty.

UC4: Estimation of Soil Properties

The goal is to develop an AI-powered system for extracting soil parameters from hyperspectral imagery which may be effectively trained from weakly-labelled datasets containing image-level ground-truth information.

UC5: Detection of Semantic Land Cover Changes

The goal is to create reliable and consistent land cover maps to systematically monitor multi-year patterns across Europe, which serve as a tool for detecting dynamic changes resulting from both human activities and natural disasters.

UC6: Monitoring Expansion of Mining Fields into Farmlands

The goal is to map the progression of small-scale mining footprints over time and assessing its impact on environmental pollution and food security in Ghana.

Acknowledgements

FAST-EO (Fostering Advancements in Foundation Models via Unsupervised and Self-supervised Learning for Downstream Tasks in Earth Observation) Project is funded by the European Space Agency (ESA), under Contract No. 4000143501/23/I-DT.

References

[1] Jakubik, Johannes, et al. "Foundation models for generalist geospatial artificial intelligence." *arXiv preprint arXiv:2310.18660* (2023).
[2] Blumenstiel, Benedikt, et al. "Multi-Spectral Remote Sensing Image Retrieval Using Geospatial Foundation Models." *arXiv preprint arXiv:2403.02059* (2024).
[3] Radford, Alec, et al. "Learning transferable visual models from natural language supervision." *International conference on machine learning*. PMLR, 2021.
[4] Touvron, Hugo, et al. "Llama: Open and efficient foundation language models. CoRR, abs/2302.13971, 2023. doi: 10.48550."*arXiv preprint arXiv.2302.13971*(2023).