Foundation Models for Earth Observation

EarthPulse (Barcelona)

Description

Foundation models are large neural networks that are trained in a self-supervised way on a diverse range of tasks and datasets, and are then fine-tuned on a specific task of interest. These models have been shown to achieve state-of-the-art performance on a wide range of tasks, and have the **potential to unlock the full potential of Earth Observation data**.

Self-supervised learning is a type of unsupervised learning where the training data is automatically labeled by the algorithm. This is achieved by defining and solving a **pretext task**, which is a task that is easier to solve than the ultimate task of interest. The model is trained to solve the pretext task, and the **learned representations** are then used to solve the ultimate task with improved efficiency.

Earth Observation data is particularly well-suited for self-supervised learning, as it is abundant and diverse, and it is often expensive to obtain labeled data. Self-supervised learning has the potential to unlock the full potential of EO data by learning representations that are useful for a wide range of tasks.

Research Topics

The following are a list of topics related to Foundation Models that the student can choose from to work during the project:

- Develop and evaluate **Pretext Tasks** specifically designed for EO data.
- Pre-training models with self-supervised learning for **different EO data types** (e.g. optical, SAR, hyperspectral, etc.)
- Pre-training models including the **temporal dimension** (image time series, videos).

In all cases, **transfer learning** from self-supervised pre-trained models to downstream tasks, and **evaluation** of self-supervised learning methods on EO data is required.

Objectives

Due to the available time and resources, the research is focused on **small-scale demonstrations** that can be scaled in future works using existing datasets.

An important goal of the work is to **generate a scientific paper** describing the methodology and reporting results.

The student will implement their code in the open source PytorchEO library[1].

Required skills

The student should be comfortable working with:

- Python and Pytorch.
- Ubuntu (terminal) and Jupyter Notebooks.
- Satellite data and processing pipelines.

[1] https://github.com/earthpulse/pytorchEO

Willingness to work at our office at least 2 days a week and a startup mindset is a plus.

About us

EarthPulse is a startup based in Barcelona working in the intersection of Artificial Intelligence and Earth Observations. EarthPulse was funded in 2020 and, since then, it has been a leader in the European AI4EO community, with projects such as the AI4EO challenges platform[2] and the EOTDL[3] for the European Space Agency (ESA). The company is driven by the mission of making Earth Observation data more accessible and useful, leveraging AI as a key enabler, generating analytics that can be useful for monitoring, vulnerability, and impact assessment in different industrial sectors (such as utility managers, insurance companies, etc).

Our office is located at San Juan de la Salle 42, Barcelona. We are a team of 10 including software engineers and data scientists in the technical team from who you can learn from. We own a workstation with 2x3090 RTX NVIDIA GPUs that the student will use during the project. Additional cloud resources can also be leveraged if necessary. Our main interest with this collaboration is to generate quality scientific publications.

Learn more about EarthPulse at https://earthpulse.ai/ or contact at juan@earthpulse.es

[2] https://platform.ai4eo.eu/
[3] https://www.eotdl.com/