Joint encoding of multi-spectral images

Master internship

General information

- Duration: 6 months (standard stipend). To start between February and April 2024.

- Institutes: Université Paris Cité, Laboratoire d'Informatique Paris Descartes (LIPADE), team Systèmes Intelligents de Perception and team EVERGREEN (Inria, INRAE, Cirad)

- Location: 45 rue des Saints-Pères, 75006, **Paris** (LIPADE)
  or 500, rue Jean François Breton, 34090, **Montpellier** (EVERGREEN)

- Supervision: Sylvain Lobry, Camille Kurtz, Laurent Wendling (LIPADE), Diego Marcos (Inria), Dino Ienco (INRAE)

- **Application: Please apply on TOCOME. The position is open until filled, and full consideration will be given to application received before 15/12/2023.**

Proposed topic

Context

By using location on the Earth's surface as the common link between different modalities, a geo-spatial foundation model would be able to incorporate a variety of data sources, including remote sensing imagery, textual descriptions of places, and features in maps. Leveraging the large amounts of available unlabeled geo-spatial data from these different sources, the GEO-ReSeT [1] (Generalized Earth Observation with Remote Sensing and Text) ANR project has the objective to learn a better representation of any geo-spatial location and convey a semantic representation of the information. Such a foundation model has the potential to revolutionize Earth observation by allowing for few or zero-shot solutions to classical problems such as land-cover and land-use mapping, target detection, and visual question answering. It will also be useful for a wide range of applications with a geo-spatial component, including environmental monitoring, urban planning and agriculture. By leveraging several data modalities, this foundation model could provide a more comprehensive and accurate understanding of the Earth's surface, enabling more informed decisions and actions. This will be particularly valuable for new potential users in sectors such as journalism, social sciences or environmental monitoring, who may not have the resources or expertise to collect their own training datasets and develop their own methods, thus moving beyond open Earth observation data and democratizing the access to Earth observation information.

Work to be done

The work to be conducted during the proposed M2 internship will contribute to the ambition of the GEO-ReSeT ANR project by studying a model that is robust to different multi-spectral modalities. Different sensors measure different spectral bands, at different spatial resolutions, which can capture different information about the target. For instance, Sentinel-2 (multi-spectral satellite from the Copernicus program of the European Union) measures 13 spectral bands at resolutions ranging from 10 to 60m. On the other hand, Landsat 9 measures 11 bands at resolutions ranging from 15 to 100m. In addition, hyperspectral sensors which measures hundreds of different spectral bands can be used.

Currently, several approaches exist to jointly work on data obtained from different multi-spectral instruments. One of the most classical one is to train different feature extractors for each modality and to fuse the obtained latent representation. Another approach is to fuse the data at the input level. Finally, it is also possible to make a prediction from each modalities and do a fusion at the prediction level [1]. These approaches tend to perform well. However, they require to train one model for each modality, which generally requires an important amount of supervision and is computationally heavy. A different approach is to translate different modalities to the input space of one of them [2]. This approach has the advantage of reducing the number of different models to learn. However, it will also remove the particularities (in our case in both spatial and spectral resolution) of the other modalities. Recent remote sensing based foundation models can be interpreted as from this last category [3], even though no explicit conversion is performed.
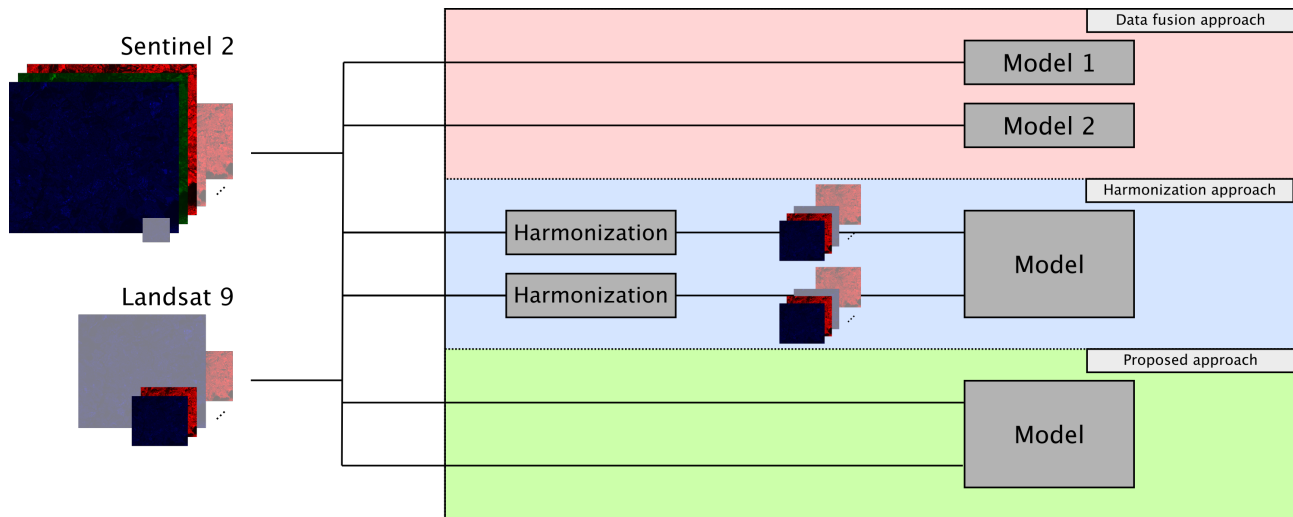
---

[1] https://geo-reset.sylvainlobry.com/

Figure 1: Comparison of the data fusion approach, the harmonization approach and our proposed approach. Our objective is to have a single model able to account for the specific properties of each sensors.

In this work, our objective is to design and train a model that is able to take as input any multi-spectral acquisition while keeping the physical measurements (i.e. spectral bands and spatial resolution), as shown in Figure 1. The work to be performed in this internship will lead to the following three contributions:

- Contribution A: the candidate will review the state of the art for the fusion of multi-spectral data and will implement a baseline. Furthermore, the candidate will implement a base line taking a unified representation of different multi-spectral sensors as an input.

- Contribution B: the candidate will propose and design an architecture that takes as an input a multi-spectral image and a description of the meta-data of the image (in particular the spectral information and the spatial resolution). For this purpose, a possible research path is to use a transformer-based methodology.

- Contribution C: the proposed architecture will be compared to the baselines on a downstream task to demonstrate the relevance of the proposed approach.

In this project, we will evaluate the approach on a setting restricted to Landsat 8/9 and Sentinel-2. We will exploit the Harmonized Landsat and Sentinel-2 product[2] for comparison with a method taking as input a unified representation.

Desired background

We are looking for a Master 2 student or final year of MSc, or engineering school in computer science. The ideal candidate should have knowledge in image processing, computer vision, natural language processing, geo-information sciences, Python programming and an interest in handling large amount of data, in particular remote sensing.

Bibliography

[1] Mauro Dalla Mura et al. "Challenges and opportunities of multimodality and data fusion in remote sensing". In: *Proceedings of the IEEE* 103.9 (2015), pp. 1585–1601.

[2] Sani M Isa et al. "Supervised conversion from Landsat-8 images to Sentinel-2 images with deep learning". In: *European Journal of Remote Sensing* 54.1 (2021), pp. 182–208.

[3] Favyen Bastani et al. "SatlasPretrain: A Large-Scale Dataset for Remote Sensing Image Understanding". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 16772–16782.

---

[2]https://hls.gsfc.nasa.gov