Self-supervised learning for satellite image time series interpretation

Master internship

General information

- **Keywords :** remote sensing, satellite image time series, computer vision, deep learning, self-supervised learning
- **Duration :** 6 months (standard stipend). To start between February and April 2025.
- **Institute :** Université Paris Cité, Laboratoire d'Informatique Paris Descartes (LIPADE), team Systèmes Intelligents de Perception
- **Location :** 45 rue des Saints-Pères, 75006, Paris
- **Supervision :** Ayoub Karine, Camille Kurtz and Laurent Wendling
- **Applications :** send an email with CV and cover letter to ayoub.karine@u-paris.fr, camille.kurtz@u-paris.fr and laurent.wendling@u-paris.fr with subject "[Internship ANR M2-BDA]"

Note : Depending on the progress and obtained results, **this internship may be followed by a PhD thesis in Computer Science in our team**. The funding is already secured through the ANR M2-BDA project. The PhD thesis topic will be an extension of this internship subject.

Proposed topic

Context

This internship is part of the M2-BDA project, funded by the ANR, aiming to investigate recent advancement in deep learning methods to extract meaningful knowledge from large heterogeneous earth observation (EO) data. It has the ambition to propose new generic methods to deal such data for monitoring environment changes. This project gathers researchers whit complementary expertise in computer science and environmental geography. As a leader of a work-package in the framework of the M2-BDA project, the SIP team of LIPADE has the ambition to propose innovative Self-Supervised Learning (SSL) methods trained on large volumes of unlabeled EO data. We will focus, in this internship, to the Satellite Image Time Series (SITS).

Work to be done

SITS are composed by successive satellite images captured in the same area. Consequently, it provides crucial information about the evolution of a sensed scene over time (2D+t). Deep learning methods hold a great potential for automatically extracting features from this spatiotemporal remote sensing data [1]. Despite the availability of several SITS data, the performance of supervised deep learning is far from satisfactory. That is mainly due to the lack of sufficient labels which conditions the efficacy of model training. Indeed, the labeling task is time-consuming and requires expertise. To overcome this limit, SSL is a promising alternative to tackle the challenge of insufficient labels. In this paradigm, the data labels are generated automatically from the data itself and the model is trained to solve a pretext task. The resulting pre-trained model is then fine-tuned on the downstream tasks (e.g. environmental ones). Among the most adopted pretext task adopted in the literature is the distinction between different augmentations of the same image (positive samples) and other images (negative samples) [2] using a contrastive loss. The data augmentation adopted for natural images can not be applied for SITS. We expect that data augmentation through spatio-temporal planar representations [3] will enhance the performance of SSL on SITS especially when the transformer is adopted. In fact, the local views extracted from theses representations would be considerate as a good candidate for the patchs of transformer architecture. The proposed SSL method will be tested in different SITS available dataset [1]. The different steps of the internship are (c.f figure 1) :

▷ **Step 1 – Literature review :** the candidate will review the supervised and the self-supervised methods for SITS interpretation. An implementation of some baseline methods, especially the self-supervised ones is to be realized.

▷ **Step 2 – Spatio-temporal representations of SITS :** the SITS data can be viewed as 2D+t datacubes. From it, multiple 2D planar representations can be extracted to model spatio-temporal information. The basis for this step will be a DeepSTAR [3] method proposed in part by our team.

1. https://github.com/corentin-dfg/Satellite-Image-Time-Series-Datasets

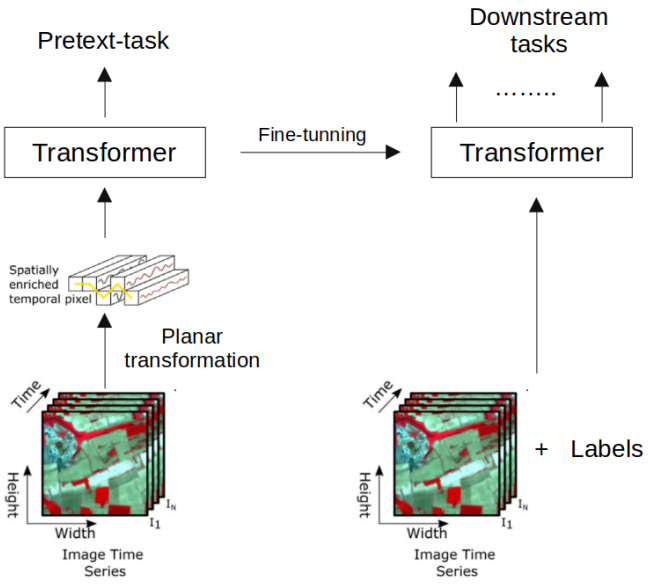

FIGURE 1 – Process of self-supervised learning from satellite image time series.

▷ **Step 3 – Self-supervised representation learning from SITS :** to deal with unlabeled data in SSL approach, a data augmentation is often used to replace the label information. To achieve this goal, we hypothesis that the spatio-temporal planar representations constructed in previous step is a fruitful candidate since it is dedicated to SITS. After that, different strategies for SSL will be tested using positive and negative samples [2] or only the positive ones [4]. We will focus on transformer model [5].

▷ **Step 4 – Test on down-stream remote sensing tasks :** the learned representations of SITS data (step 3) can then be fine-tuned on down-stream remote sensing tasks (e.g. classification, segmentation, change detection) that suffer from a large amount of annotated data.

Desired background for the candidate

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 deep learning, computer vision, Python programming and an interest in handling remote sensing images.

Bibliography

[1] Lynn MILLER, Charlotte PELLETIER et Geoffrey I WEBB. "Deep Learning for Satellite Image Time-Series Analysis : A review". In : *IEEE Geoscience and Remote Sensing Magazine* (2024).

[2] Ting CHEN et al. "A simple framework for contrastive learning of visual representations". In : *International conference on machine learning*. PMLR. 2020, p. 1597-1607.

[3] Mohamed CHELALI et al. "Deep-STaR : Classification of image time series based on spatio-temporal representations". In : *Computer Vision and Image Understanding* 208-209 (2021), p. 103221.

[4] Mathilde CARON et al. "Emerging properties in self-supervised vision transformers". In : *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, p. 9650-9660.

[5] Xin GUO et al. "Skysense : A multi-modal remote sensing foundation model towards universal interpretation for earth observation imagery". In : *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024, p. 27672-27683.