Towards an Efficient Foundation Model for Very-high-Resolution Satellite Imagery

PhD proposal within UBS-IRISA/CNES

Expected starting: September/October 2024
Application deadline: **March 15, 2024**

Context

Remote sensing imagery for Earth observation (EO) has emerged as a dynamic research area, enabling precise identification, characterization, and interpretation of objects and materials on the Earth's surface. The ongoing progress in satellite technology has led to the availability of numerous very-high-resolution (VHR) optical satellites, facilitating daily acquisitions. This enables the creation of highly detailed maps with sub-meter spatial resolution, benefiting various essential EO applications such as urban planning, swift disaster mapping, natural resource management, and wildlife monitoring.

In recent years, deep learning (DL) has found success in various machine learning and computer vision domains, including remote sensing (RS). Despite this, applying DL to real-world scenarios using VHR satellite images for operational purposes faces numerous challenges. The foremost challenge involves the difficulty of annotating domain-specific data, particularly in EO applications that demand expert knowledge. Generating precise and comprehensive labeled datasets for training deep models is a time-consuming and expensive endeavor. For instance, in rapid disaster mapping, acquiring accurate labels is nearly unfeasible due to the infrequent and unique nature of catastrophic events. Another challenge comes from the significant domain shifts inherent in RS data, arising from diverse sensor characteristics (i.e., spatial resolution and spectral bands) and varied acquisition conditions. Lastly, the exploding size of deep models, with millions (or even billions) of parameters, raises concerns. Not only do they demand substantial computational and storage resources, but they also cause negative environmental issues.

Therefore, designing efficient models while maintaining high accuracy becomes essential in every EO application to reduce energy cost and more importantly, to minimize the environmental impact. These models are expected to be reusable or transferred with low resources. Recent studies have showed that self-supervised pretraining with unlabeled RS images outperforms popular ImageNet-pretrained models in EO downstream tasks, especially when labels are scarce [1]. By leveraging self-supervised learning (SSL) on the abundance of multi-source unlabeled data, foundational models (FMs) have started their era by providing high performance on a wide range of downstream tasks. However, current trends of FMs in EO only focus on creating large vision FMs using substantial multi-source images (RingMo [3], Billion-scale ViT [4]), or large vision-language FMs (RemoteCLIP [5]). These models demand significant computational resources for training and deployment. Therefore, developing resource-efficient foundation models in the context of EO is imperative to mitigate environmental concerns in the future.

Objectives

This PhD topic aims to develop efficient foundation models with a focus on EO applications using VHR satellite imagery. The main objectives are the three-folds:

- Developing effective and sustainable vision foundation models by leveraging SSL from vast amount of unlabeled VHR satellite images. We consider publicly available images with a spatial resolution less than 2m/pixels such as SPOT-6/7, Pleiades, Pleiades-Neo, Rapid-Eyes, etc. from THEIA platform [1], as well as those provided by the CNES. For methodological aspect, we will adopt the current trends in SSL including contrastive learning and masked autoencoders adapted to the specifications of VHR images (multispectral bands, fine-grain textures, etc.).

- Distilling compact and lightweight models in downstream tasks to tackle two opted applications including: (1) rapid mapping of wildfires and (2) detection of marine megafauna. Our goal is to develop deep models that guarantee good prediction on these two tasks (precision above 85%) with low resources: limited labels for training, low storage memory, rapid prediction at inference time. To do that, we plan to exploit the self-supervised and task-specific knowledge distillation during the training of downstream tasks.

- Closing the gaps in case of domain shifts in multimodal learning of multi-sensor satellite images. In most real-world EO applications, deep models should be able to deal with any available VHR sensor at prediction time. For example, in the context of rapid disaster mapping, with the urgency for timely decision-making, leveraging all available sources of VHR satellite images during the time of an on-going event is indispensable.

The proposed topic is expected to provide cutting-edge solutions for large-scale EO applications using multi-source VHR satellite images in the context of sustainable development. All datasets, codes and pretrained models will be published for reproducibility and reuse purposes.

Required background and skills

- MSc or Engineering degree with excellent academic track and proven research experience in one of the following fields: computer science, applied maths, signal and image processing;

- Experience with machine learning, in particular deep learning;

- Interests for environment and earth observation applications;

- Skills and interest in programming (Python and frameworks such as Pytorch/Tensorflow will be appreciated);

- Excellent communication skills (spoken/written English) is required ;

Supervision and Funding

The expected PhD candidate will join the OBELIX research group[2] from IRISA (UMR 6074) is located in the UBS (Université Bretagne Sud) campus in Vannes 56000, France. He/She will be jointly supervised by **Sébastien Lefèvre**[3] (Professor at UBS), **Minh-Tan Pham**[4] (Associate Professor at UBS) and **Stéphane May** (Senior Researcher at CNES). This PhD track will co-funded by the UBS's doctoral program and the CNES (The National Centre for Space Studies).

[1] `https://www.theia-land.fr/en/satellite-data/`
[2] `http://www-obelix.irisa.fr/`
[3] `https://people.irisa.fr/Sebastien.Lefevre/`
[4] `https://sites.google.com/site/mtanpham89`

Application

Position to be filled as soon as possible. To apply, please prepare your **Detailed CV + Master transcripts + Reference letter(s)** (in English or French) and :

- submit their application to the CNES website **HERE**[5]
- send their application to **minh-tan.pham@irisa.fr** and **sebastien.lefevre@irisa.fr**

Deadline for application: **March 15, 2024**. Potential candidates will be contacted for interview.

References

[1] Berg, P., Pham, M. T., & Courty, N. (2022). Self-supervised learning for scene classification in remote sensing: Current state of the art and perspectives. Remote Sensing, 14(16), 3995.

[2] Wang, Y., Albrecht, C. M., Ait Ali Braham, N., Mou, L., & Zhu, X. X. (2022). Self-supervised Learning in Remote Sensing: A Review. IEEE Geoscience and Remote Sensing Magazine (GRSM).

[3] Sun, X., Wang, P., Lu, W., Zhu, Z., Lu, X., He, Q., ... & Fu, K. (2022). RingMo: A remote sensing foundation model with masked image modeling. IEEE Transactions on Geoscience and Remote Sensing.

[4] Cha, K., Seo, J., & Lee, T. (2023). A billion-scale foundation model for remote sensing images. arXiv preprint arXiv:2304.05215.

[5] Liu, F., Chen, D., Guan, Z., Zhou, X., Zhu, J., & Zhou, J. (2023). RemoteCLIP: A Vision Language Foundation Model for Remote Sensing. arXiv preprint arXiv:2306.11029.

[6] Mai, G., Huang, W., Sun, J., Song, S., Mishra, D., Liu, N., ... & Lao, N. (2023). On the opportunities and challenges of foundation models for geospatial artificial intelligence. arXiv preprint arXiv:2304.06798.

[7] Xiong, Z., Wang, Y., Zhang, F., & Zhu, X. X. (2024). One for All: Toward Unified Foundation Models for Earth Vision. arXiv preprint arXiv:2401.07527.

[5] https://recrutement.cnes.fr/fr/annonce/2701619-24-272-towards-an-efficient-foundation-model-for-vhr-imagery-