Marginal representation of streamed multi-source remote sensing data using Gaussian process prior variational auto-encoder

Mathieu Fauvel[1], Nicolas Dobigeon[2], and Julien Michel[1]

[1]Centre d'Etudes Spatiales de la Biosphère - CESBIO
[2]Institut de recherche en informatique de Toulouse - IRIT

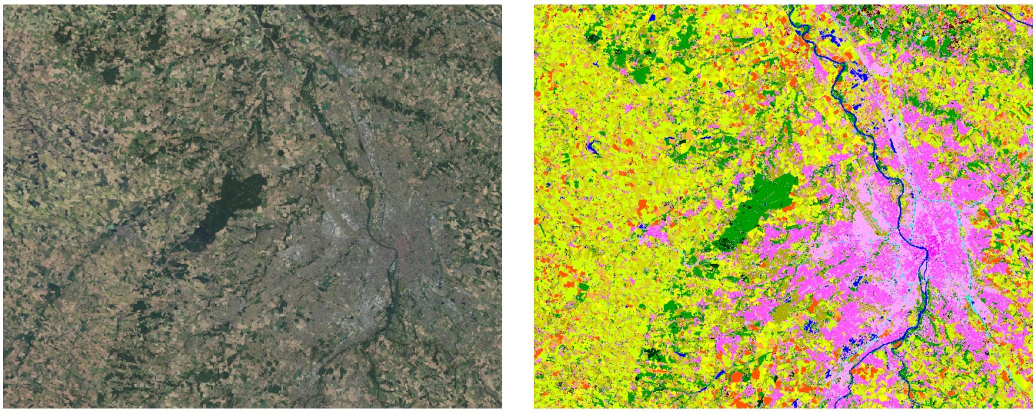

Figure 1: Land cover maps of the French territory produces by the CESBIO-lab using machine learning algorithms and satellite Earth observations, extracted over the city of Toulouse.

1 Subject

This offer is part of the RELEO (REpresentation Learning for Earth Observation) of ANITI-2, the follow-on of the Interdisciplinary Artificial Intelligence Institute in the frame of the French ANR "AI Clusters".

Over the last ten years, Earth Observation (EO) has made enormous progress in terms of spatial and temporal resolution, data availability and open policies for end-users. The increasing availability of complementary imaging sensors allows to observe land ecosystems state variables and processes at different spatio-temporal scales. Big EO data can thus enable the design of new land monitoring systems providing critical information in order to guide climate change monitoring, mitigation and adaptation. One main information is the land cover state and trend of continental surfaces, see Figure 1 for an example of land cover map.

Conventional machine learning methods are not well adapted to the complexity of multi-modal, multi-resolution Satellite Image Time Series (SITS) with irregular sampling, and therefore not suitable for extracting and processing all the relevant information. On the other hand, methods based on deep neural networks have shown to be very effective to learn low-dimensional representation of complex data for several tasks and come with high potential for EO data, but they often come from the Computer Vision (CV) and natural language processing (NLP) communities and need to be extended to handle the specificities of Earth Observation data.

Previous works at the CESBIO-lab have shown that generative encoder-decoder architectures such as the Variational Auto-Encoder (VAE) models or the U-NET models perform very well for a variety of EO tasks: estimation of biophysical parameters or Sentinel-1 to Sentinel-2 translation, to cite a few.

However, such approaches appear to be inadequate to handle data coming from more than 2 sources and acquired at different time and spatial resolutions, as prioritized in the RELEO chair within ANITI. In particular, the generative capability of these models may generalize poorly to unseen region or temporal period. Processing such streams of data requires to jointly encode all source into a structured latent space where each complementary information carried by each source can be embedded while ensuring long-term encoding of newly acquired data (from possibly new sensor).

Besides, VAEs usually assume independence between samples and require that all latent variables are generated even if some data source is missing. These assumptions are generally made for sake of simplicity and computational efficiency of the training and inference steps. However, assuming independence of samples amount to ignoring the correlation between adjacent pixels in the spatial and/or temporal domains. The second requirements is commonly addressed thanks to masking strategies. Because of the very deterministic nature of such neural networks architectures, they do not properly encode uncertainty related to missing data. Furthermore, they are not able to impute the resulting missing information in the latent space.

The objective of this PhD is to learn a low-dimensional probabilistic representation of multi-source EO data addressing the problem of streamed multi-source data. Specifically, in this PhD, the candidate will investigate the relevance of the Gaussian process (GP) prior. Adopting this GP prior is expected to model correlations for multiple (possibly infinite) sources of data in the latent space, i.e., when the number of EO sources may vary during training or inference. Furthermore, building on the conditioning property inherent to GP, it might be possible to reconstruct missing data in the learned latent space. Lastly, as for any generative deep model, sampling from the latent space is straightforward.

The methodological challenges that will be considered during the PhD are three-fold:

1. Defining a probabilistic generative model with GP prior for the latent space. This prior should encode properly time and space for any sensors. A possible starting solution will be based on an adaptation of existing works from multi-view objects reconstruction, such as [1] or [2].

2. Developing a scalable training algorithm. Conventional GP are known to scale poorly but solutions based on sparse and variational approximations have been successfully employed (see [3] for a past work co-funded by the CNES).

3. Maintaining fast inference as with VAE. This requires to *amortize* parameters of the model during the training process.

This model will be evaluated and compared to state of the art generative representation learning algorithms for the different scenarios identified in the RELEO chair hosted by ANITI. In particular, the learned latent representations will be used to generate Essential Climate Variables in order to monitor land uses and land cover changes (see Figure 1), as well as vegetation state and trend, carbon cycle and water cycle.

2 Scientific environment

The PhD. student will benefit from a favorable context and will be able to rely on the most recent results and advances in machine learning and Earth observation signal & image processing. He/she will be mainly co-advised by the following researchers within the CESBIO-lab:

- Mathieu Fauvel, INRAe Researcher
- Nicolas Dobigeon, Professor at Toulouse INP
- Julien Michel, CNES Engineer

He/She will take advantage from the lab scientific actitivies (e.g. scientific seminar: `https://src.koda.cnrs.fr/activites-ia-cesbio/ds-cb`) as well as the ANITI (`https://aniti.univ-toulouse.fr/`) dynamic in Toulouse. Also, the PhD is funded jointly by the CNES (`https://cnes.fr/en`) and CS-Group (`https://www.csgroup.eu/en/`) and, as such, the recruit will benefit of their expertise during the PhD.

3 Candidate background

Master or Engineering school students with major in applied mathematics, computer science or electrical engineering. The candidate must have a solid background in at least one of the following subjects:

- Statistical signal and image processing,
- Machine learning or data science,
- Remote sensing data processing.

A good knowledge of English and scientific programming skills (Python, git) are required. Broader interest in Earth observation will be appreciated.

4 Contact & application procedure

Applicants are also invited to send (as pdf files)

- A detailed curriculum,
- Official transcripts from each institution you have attended (in French or English),
- (optional) link to code repository or recommendation letter.

to the co-advisors

- Mathieu Fauvel, mathieu.fauvel@inrae.fr
- Nicolas Dobigeon, nicolas.dobigeon@toulouse-inp.fr
- Julien Michel, julien.michel4@univ-tlse3.fr

You will be contacted if your profile meets the expectations. Review of applications will be closed on mid-March 2024.

References

[1] F. P. Casale, A. V. Dalca, L. Saglietti, J. Listgarten, and N. Fusi, "Gaussian process prior variational autoencoders," 2018.

[2] M. Jazbec, M. Ashman, V. Fortuin, M. Pearce, S. Mandt, and G. Rätsch, "Scalable gaussian process variational autoencoders," 2021.

[3] V. Bellet, M. Fauvel, and J. Inglada, "Land Cover Classification with Gaussian Processes using spatio-spectro-temporal features," *IEEE Transactions on Geoscience and Remote Sensing*, Jan. 2023. [Online]. Available: https://hal.science/hal-03781332