Modèles de diffusion à Haute Résolution appliqué à l'Observation de la Terre.

V. Sanchez[1,3], J. Inglada[1], T. Oberlin[3], M.Fauvel[1,2]

[1] Univ Toulouse, CESBIO, [2]INRAe, [3]ISAE-SUPAERO, Toulouse, France

Context

Monitoring Essential Variables

- Monitoring Essential Climate and Biodiversity Variables (EVs) provides key information to **understand climate, biodiversity and environmental changes** using Earth Observation (EO) systems.
- **Retrieving EVs from multi-source data** is **challenging due to the singularities of EO data**, such as indirect observation of interest variables, **varying spatial resolution and irregularly sampled time series**.

Deep Learning

- **Learn complex patterns** from huge amounts of data.
- Most of the **recent models lack physical consistency and interpretability**. They are **not able to process data with irregular and unaligned sampling, which is common in multi-modal EO**.
- **Training** also **requires** large amounts of **labeled data**, which are scarce and noisy in EO. Consequently, **current models have a restricted usage in large scale EO systems**.

RELEO: REpresentation Learning for Earth Observation

Framework

- Development of new **self-supervised representation learning**[3] methods to produce **semantically meaningful probabilistic representations**[2] from **high-dimensional multi-modal EO data**.
- Novelty lies on the **use of prior knowledge from physical models into DL** and thus proposing **advances in uncertainty estimation and interpretability**.
- The hybrid AI system will **blend physical priors and DL to pre-train models**. The system will process **multi-modal data to capture complementary spatio-temporal patterns**. **Physics-guided DL methods** will be designed to condition the decoding of generic embeddings to retrieve and forecast EVs and their uncertainties.

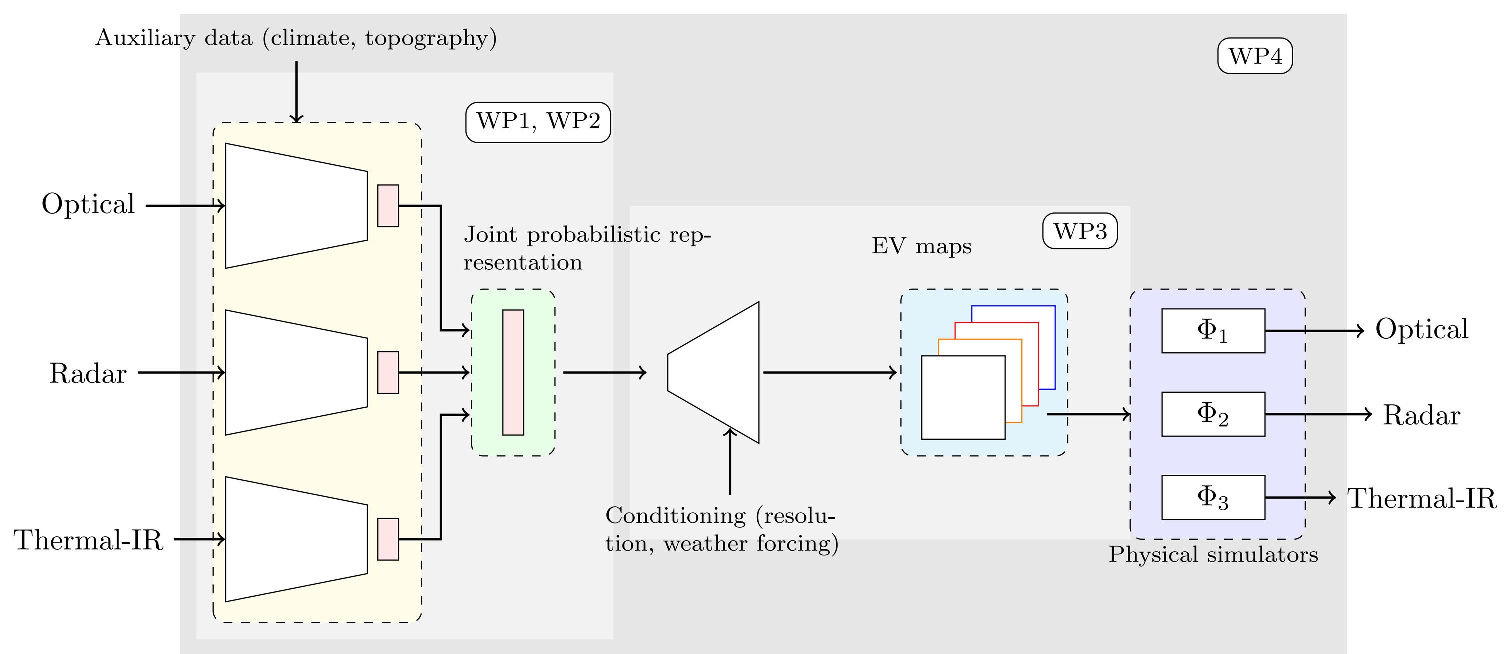

Fig. 1: The RELEO framework combines probabilistic neural encoders, on-demand decoding into EVs and self-supervised learning through physical simulators and data assimilation.

One for all, and all for one : Perceiver IO [5]

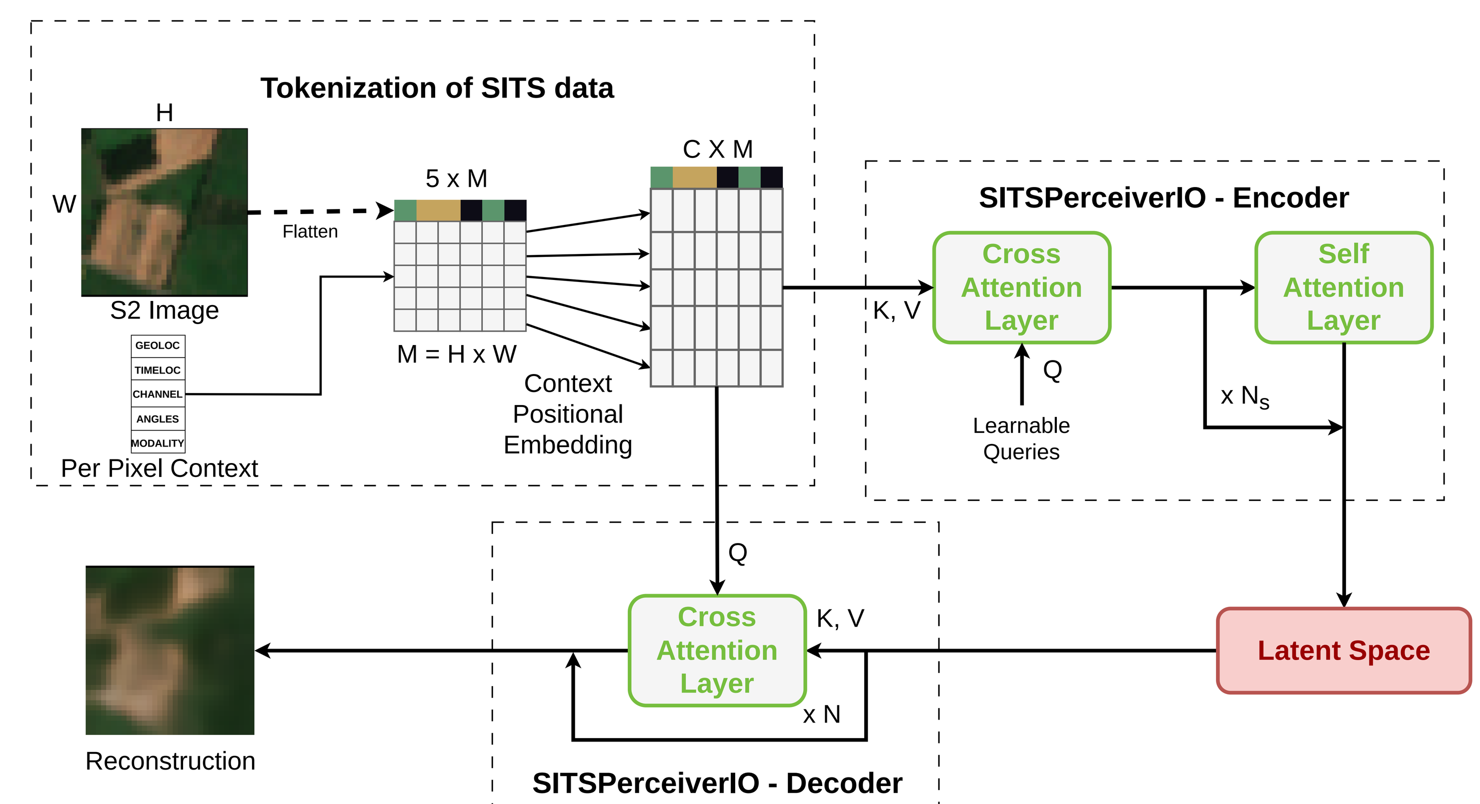

Fig. 2: SITS Perceiver IO Architecture for a reconstruction task

Industrial Partners

- CS Sopra Steria / Magellium / Thales Alenia Space / Thales Services Numériques

Perceiver IO

Why Perceiver IO ?

- **No need for different encoding heads for different kinds of data**
- **Leverages cross-attention to reduce the cost of the deeper layers of the network**
- **Handle general purpose inputs and outputs while scaling linearly in both input and output size**
- **Multi-modal Masked Autoencoder (MAE) training easily implemented through cross-attention with a single decoder**

Reconstruction task with Perceiver IO

Will you find which is the TARGET and which is the PREDICTION ?

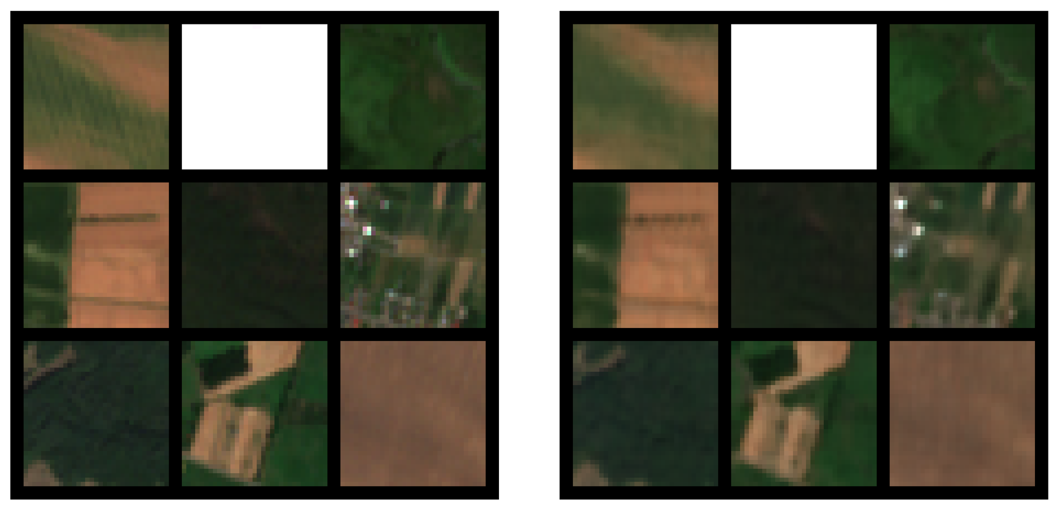

Fig. 3: Results of SENTINEL-2 reconstruction with Perceiver IO on validation set.

Diffusion Models (DMs) for SITS

What is a Diffusion Model ?

- Ability to **generate diverse, high quality, samples**, and the relative **simplicity of the training scheme**, which allows very large models to be trained at scale. [4]
- "***The basic idea behind these models is based on the observation that it is hard to convert noise into structured data, but it is easy to convert structured data into noise.***" Kevin P. Murphy in [8]

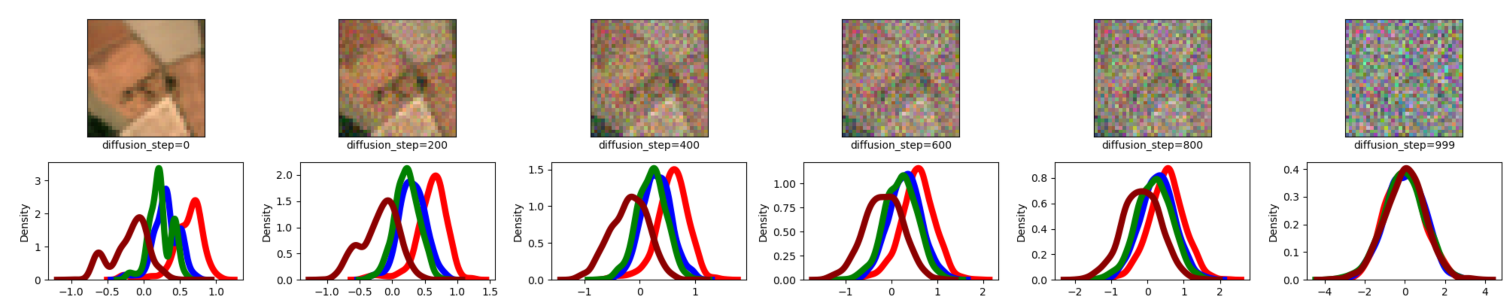

Fig. 4: Forward Sampling of SENTINEL-2 data

PhD Perspectives : Diffusion Model with Perceiver-IO [10, 9]

- **Meteorological conditioning of Diffusion Model [7] :** Learning the impact of meteorology on vegetation.
- **Temporal Diffusion Models [6]** : Learning Sequences of satellite images are irregular.
- **Multimodal/Joint Diffusion Models [1]** : Majority of generative models tends to learn and generate a conditional model $p(x|y)$. Perceiver-IO with DMs allow to obtain a *joint model* : $p(x, y)$.

References

References

[1] Fan Bao, Shen Nie, Kaiwen Xue, Chongxuan Li, Shi Pu, Yaole Wang, Gang Yue, Yue Cao, Hang Su, and Jun Zhu. One transformer fits all distributions in multi-modal diffusion at scale. In *International Conference on Machine Learning*, pages 1692–1717. PMLR, 2023.

[2] Iris Dumeur, Silvia Valero, and Jordi Inglada. Paving the Way Toward Foundation Models for Irregular and Unaligned Satellite Image Time Series. *CoRR*, 2024.

[3] Iris Dumeur, Silvia Valero, and Jordi Inglada. Self-Supervised Spatio-Temporal Representation Learning of Satellite Image Time Series. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, pages 1–18, 2024.

[4] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[5] Andrew Jaegle, Sebastian Borgeaud, Jean-Baptiste Alayrac, Carl Doersch, Catalin Ionescu, David Ding, Skanda Koppula, Daniel Zoran, Andrew Brock, Evan Shelhamer, et al. Perceiver io: A general architecture for structured inputs & outputs. *arXiv preprint arXiv:2107.14795*, 2021.

[6] Samar Khanna, Patrick Liu, Linqi Zhou, Chenlin Meng, Robin Rombach, Marshall Burke, David Lobell, and Stefano Ermon. Diffusionsat: A generative foundation model for satellite imagery. *arXiv preprint arXiv:2312.03606*, 2023.

[7] Johannes Leonhardt, Jürgen Gall, and Ribana Roscher. Climsat–a diffusion autoencoder model for climate-conditional satellite image editing. *Science of Remote Sensing*, page 100235, 2025.

[8] Kevin P Murphy. *Probabilistic machine learning: Advanced topics.* MIT press, 2023.

[9] Yunyi Shen and Alexander Gagliano. Diffusion autoencoders with perceivers for long, irregular and multimodal astronomical sequences. *arXiv preprint arXiv:2510.20595*, 2025.

[10] Y Wang, AA Elhag, N Jaitly, JM Susskind, and MA Bautista. Swallowing the bitter pill: Simplified scalable conformer generation. arxiv 2024. *arXiv preprint arXiv:2311.17932.*