ELECTRONICS & DEFENSE

Safran.AI

Benchmarking Self-supervised Learning Methods in Remote Sensing

Kévin Sanchis

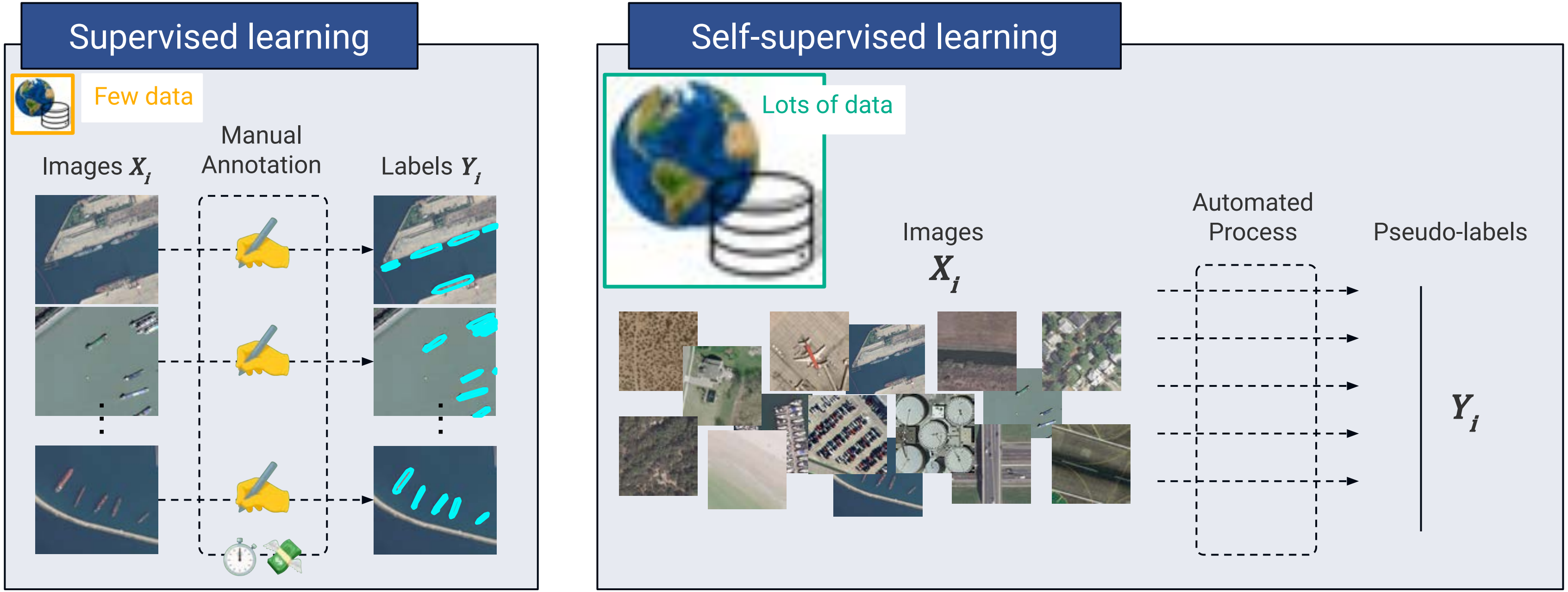
02/10/2024

SAFRAN

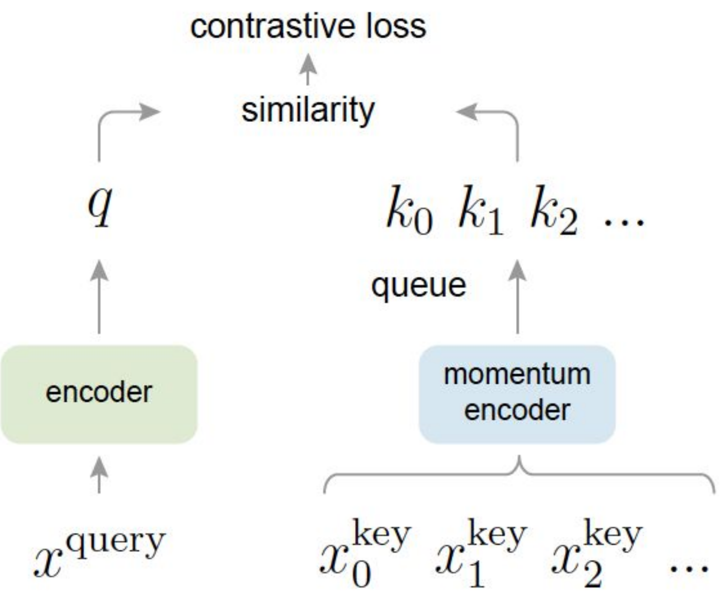
Agenda

Self-Supervised Learning (SSL)

- Reduce the need for large quantities of annotations
- Exploit the massive amount of uncurated, unlabeled remote sensing data for learning good representations

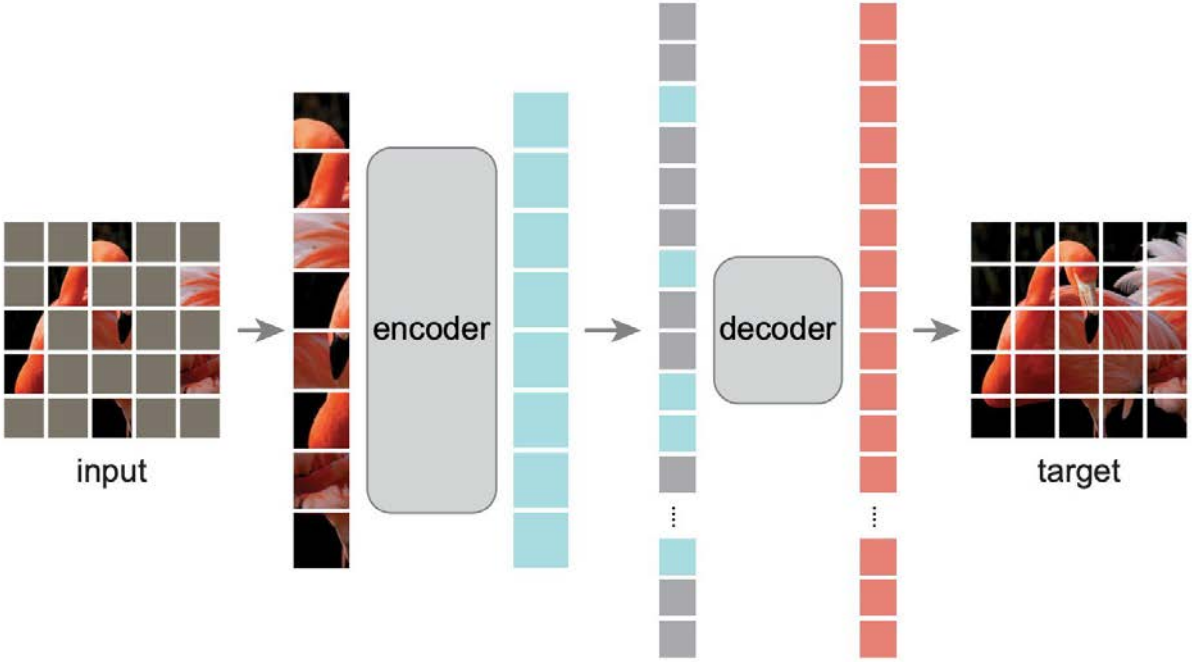


Contrastive Learning Methods



[Momentum Contrast for Unsupervised Visual Representation Learning]

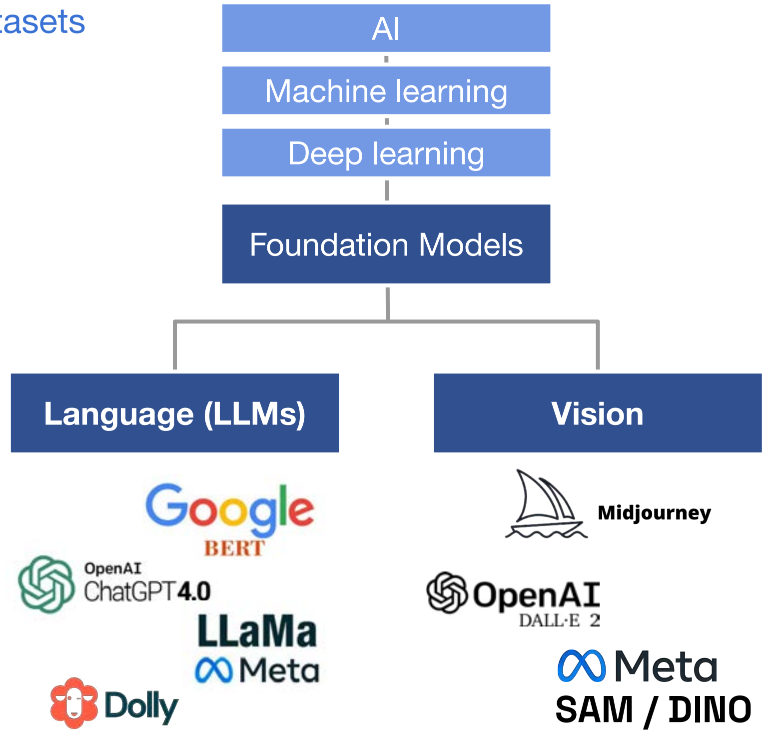
Reconstruction-Based Methods



[Masked Autoencoders Are Scalable Vision Learners]

Foundation Models

- Deep Learning models trained on massive unlabeled datasets
- Usually trained via self-supervised learning
- Handle a wide variety of tasks



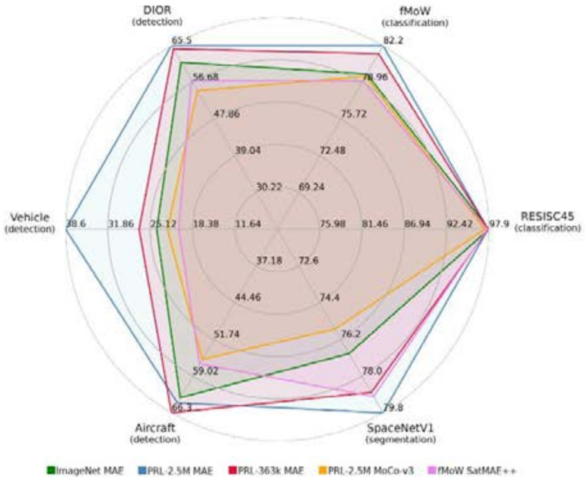
SAFRAN

Can we build a Foundation Model for Remote Sensing applications ?

Benchmarking Self-supervised Learning Methods in Remote Sensing

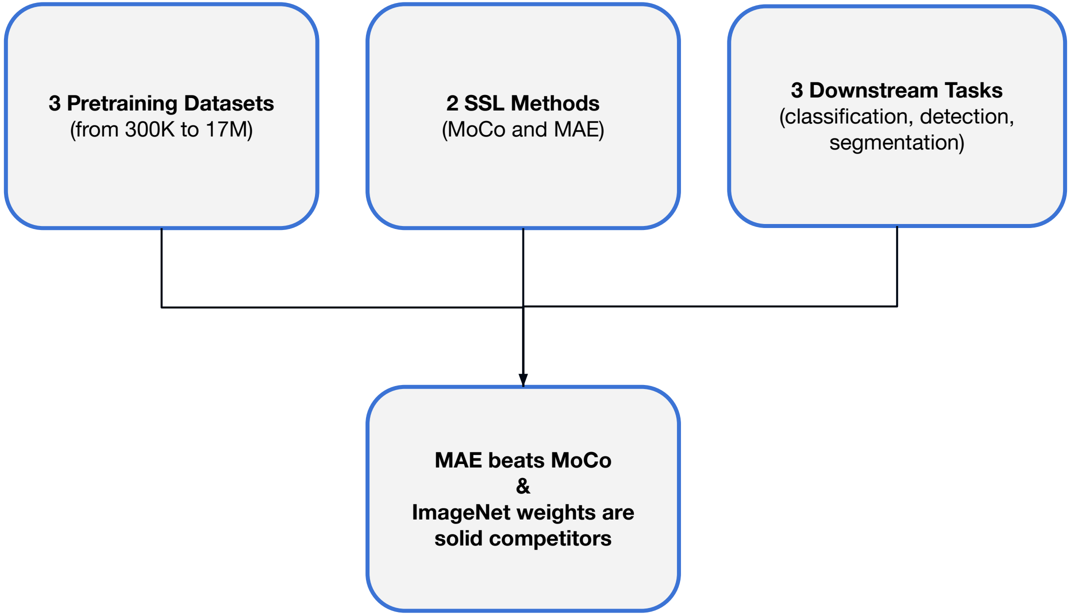
Fabien Merceron, Vincent Partimbene, Gohar Dashyan, Sébastien Saubert,
Guillaume Peltier, Kévin Sanchis and Pierre-Antoine Ganaye
Preligens
Paris, France
Email: <name>.<surname>@preligens.com

***Abstract*—Self-supervised pretraining has proved to be a competitive tool to improve downstream task performance in the field of remote sensing. Attempts to create geospatial foundation models based on such pretraining techniques are increasing in numbers, and are a promising solution to exploit the vast amount of unannotated remote sensing imagery. Due to the widespread availability of various self-supervised techniques, either generic or specific to remote sensing, it becomes of importance for practitioners to find a way to identify the best performing pretraining method based on the downstream task being tackled. In this paper, we present a systematic benchmark of commonly used self-supervised pretraining methods and provide insights into the most appropriate approach depending on the chosen downstream tasks. Our results indicate that Masked Auto Encoders (MAE), a reconstruction-based method, seems to be the overall winner on most use-cases. We also show that ImageNet remains a powerful pretraining dataset and can produce competitive baselines, while building a tailored pretraining dataset using high-resolution satellite images can effectively improve the downstream performance compared to such baselines. Finally, we study the computational efficiency of pretraining methods and provide recommendations**



Submitted to CAID 2024 & Powered by the Jean-Zay HPC

Main Conclusions of the Study



Agenda

01

Introducing the Context

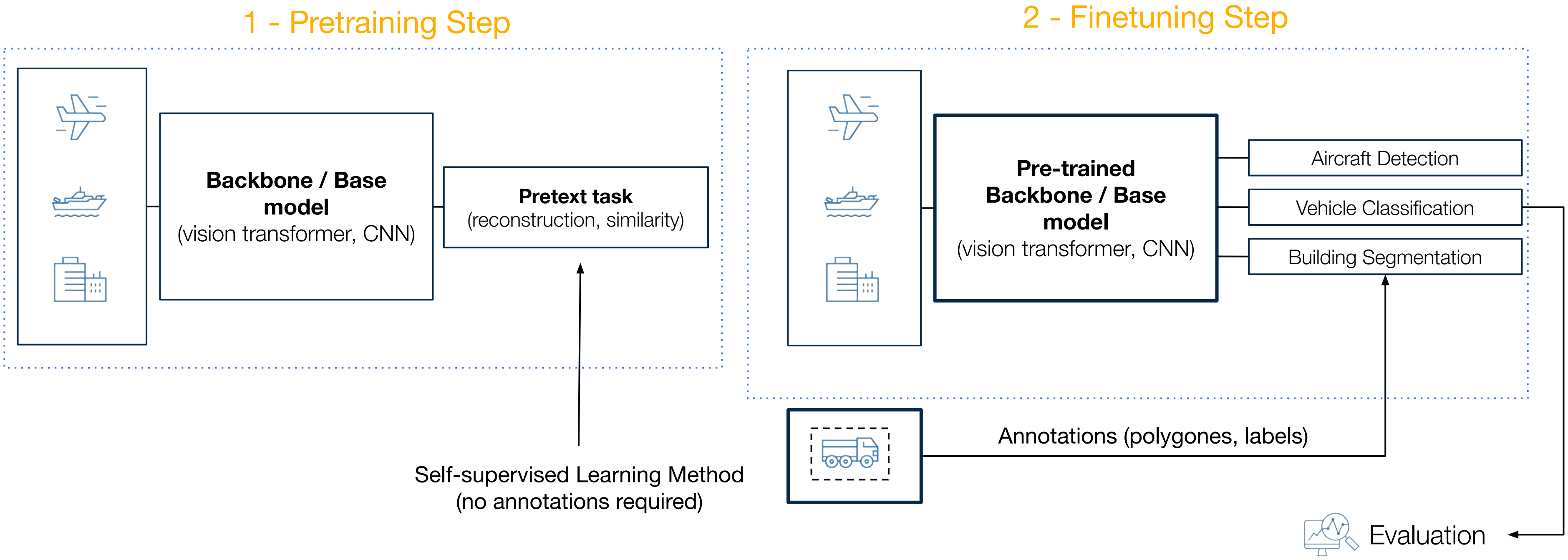
02

Scaling up AI in Remote Sensing

03

Conclusion & Future Directions

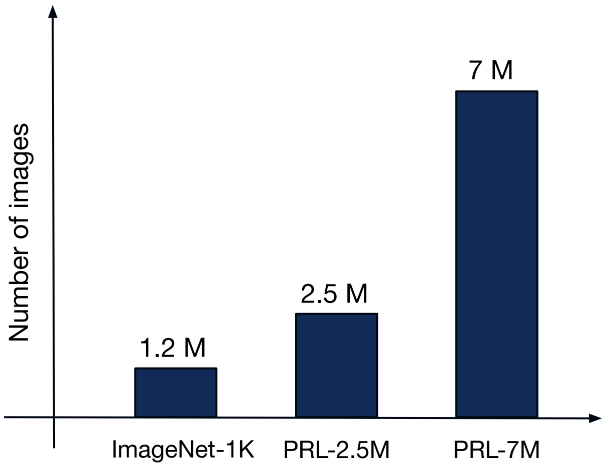
Experimental Protocol



SAFRAN

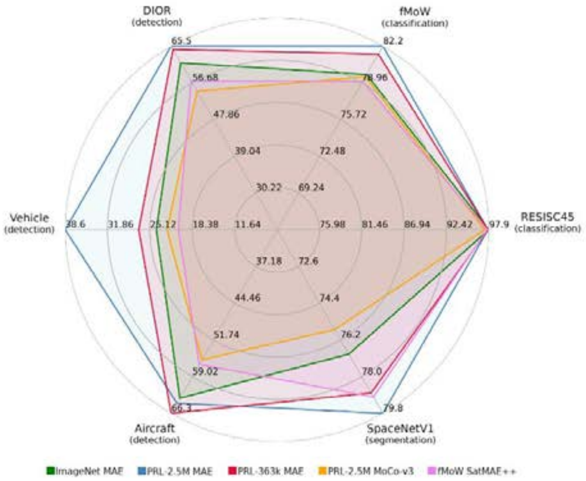
Why do we need a HPC ?

Massive unlabeled datasets



Our internal datasets (we also experimented with 17M)

Wide variety of tasks



SAFRAN

Technical Requirements

Datasets

► Process remote sensing imagery at scale (tiling, data selection)

Experiments

► Hundreds of trainings on large-scale datasets

► Various architectures and sizes

Results

► Compare all trainings by visualizing performance

SAFRAN

Building Pretraining Datasets

Datasets

► Process remote sensing imagery at scale (tiling, data selection)

Thousands of rasters

Source: Maxar

An image in our datasets

Parallel processing on multiple machines

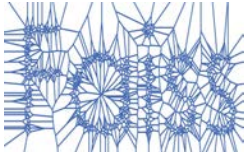
Building Pretraining Datasets

Datasets

► Process remote sensing imagery at scale (tiling, data selection)

Thousands of rasters

Source: Maxar



SAFRAN

Training Models at Scale

Experiments

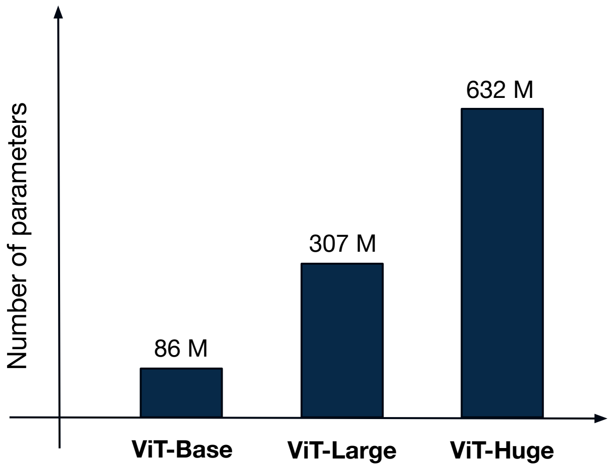
- Hundreds of trainings on large-scale datasets
- Various architectures and sizes

Taking advantage of open-source software to reduce boilerplate and easy access to neural network architectures

Training Models at Scale

Experiments

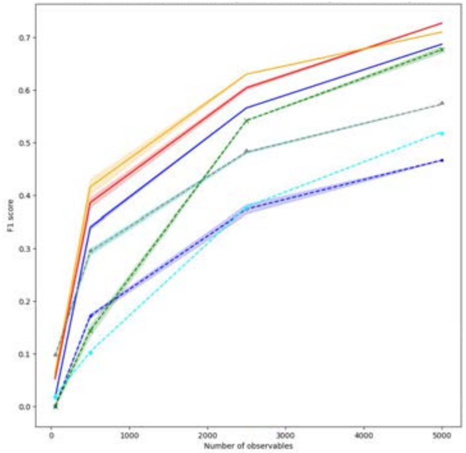
- Hundreds of trainings on large-scale datasets
- Various architectures and sizes

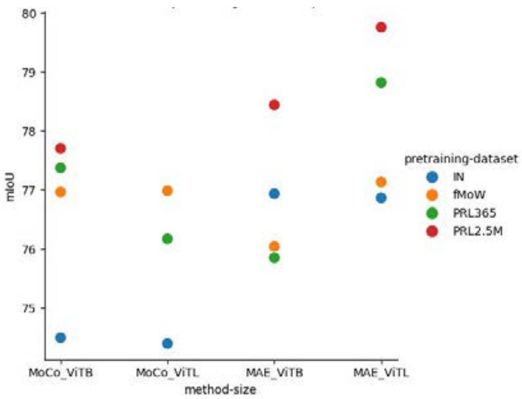


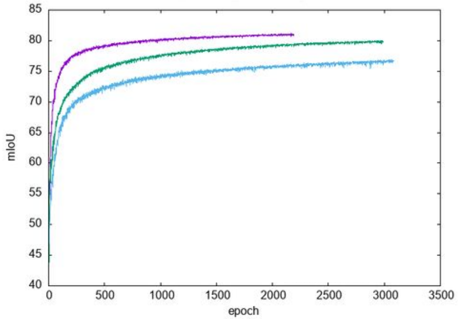
Comparing Experiments

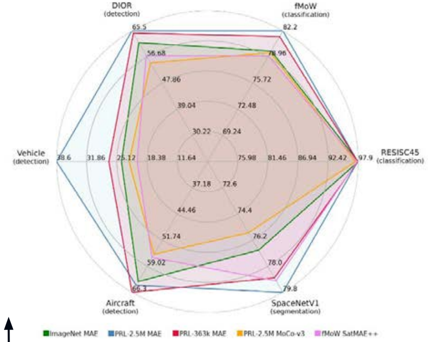
Results

► Compare all trainings by visualizing performance









Developing custom graphs and monitoring routines for Jean-Zay

SAFRAN

Agenda

SAFRAN

Summary

Experiments with SSL methods to build a Geo-Foundation Model

Processing of remote sensing imagery at scale
(up to 17M images)

Hundreds of trainings of neural networks with different scales

MAE is the best SSL method in our benchmarks

Can we build a Foundation Model for Remote Sensing applications ?

Future Directions

Try other SSL methods (e.g. DINOv2)

Knowledge Distillation to produce smaller models

Multimodal Learning as a promising improvement

POWERED BY TRUST