Seminar: Hot topics in Computer Vision

Matti J. Frind

DALL-E: “Generate a high-resolution photorealistic satellite image from one of the Sentinel-1 satellites with false colours.”

Motivation

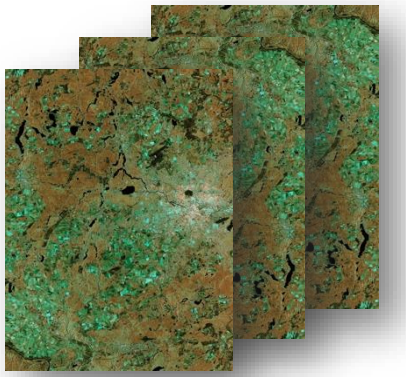
urban?

agriculture

water?

https://www.esa.int/ESA_Multimedia/Missions/Sentinel-1/(sortBy)/view_count/(result_type)/images

Motivation



Sentinel I/II

Landsat

ImageNet

...

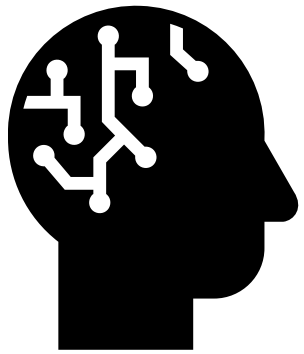
Trained using **contrastive learning** or **reconstruction learning**.

Foundational Model

https://www.esa.int/ESA_Multimedia/Missions/Sentinel-1/(sortBy)/view_count/(result_type)/images
https://eoimages.gsfc.nasa.gov/images/imagerecords/153000/153149/princecharles_oli2_20240715_lrg.jpg
https://cs.stanford.edu/people/karpathy/cnnembed/

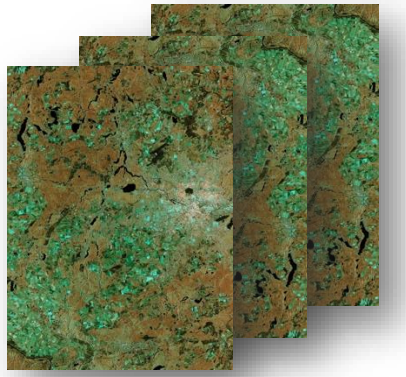
Goal

“Neural Plasticity-Inspired Foundation Model for Observing the Earth Crossing Modalities” - Zhitong Xiong et al. - 2024



DOFA Model

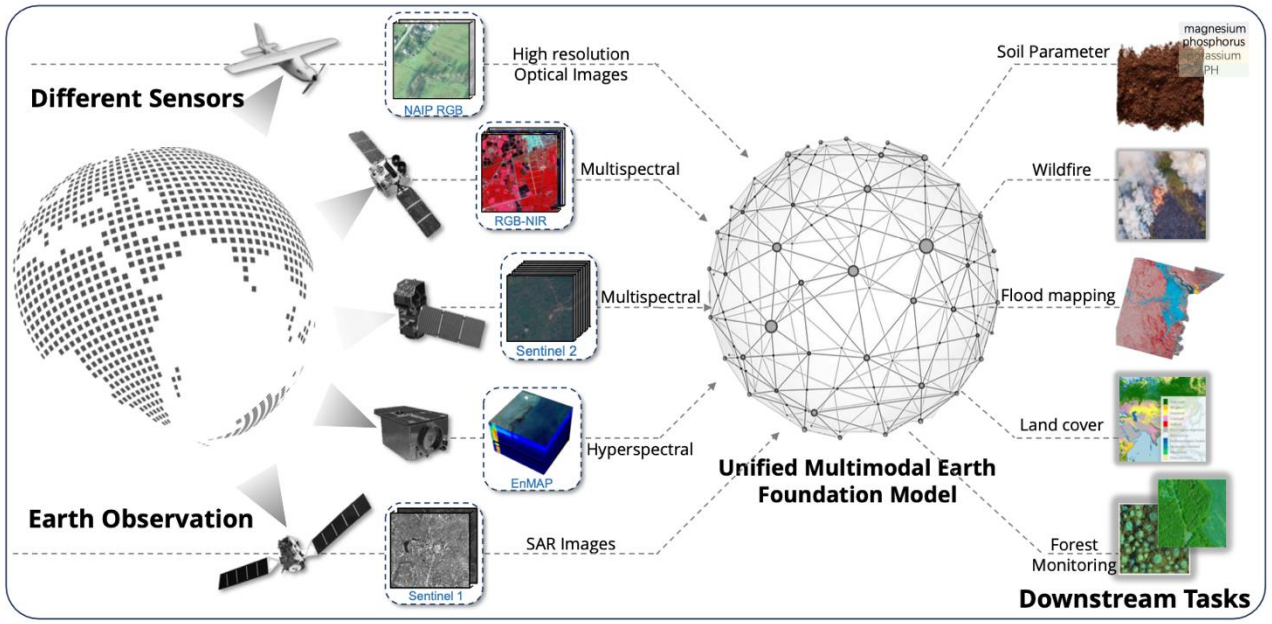
Test & Benchmark

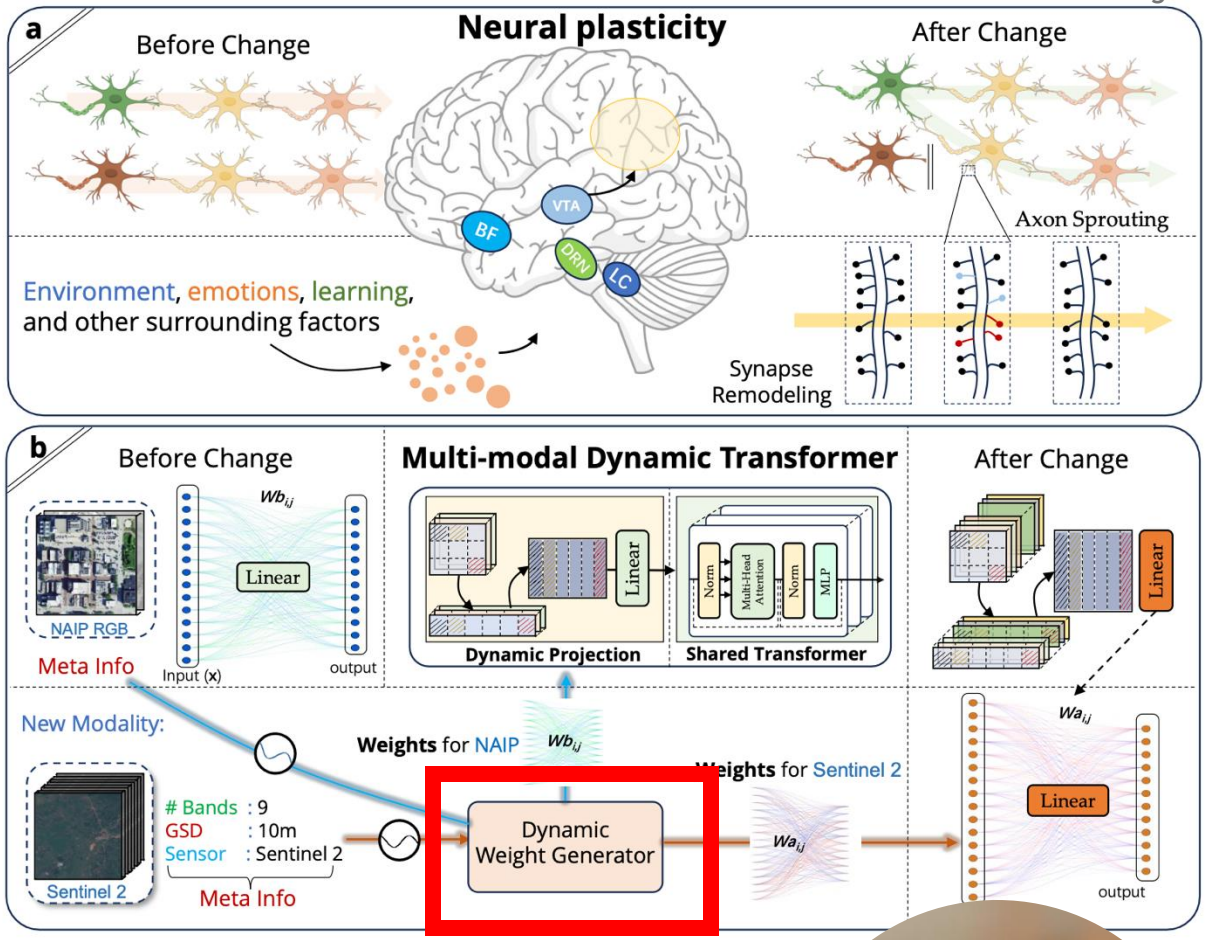


BigEarthNet

- Foundational model for remote sensing
- Multimodal

DOFA stands for ”Dynamic-One-For-All”





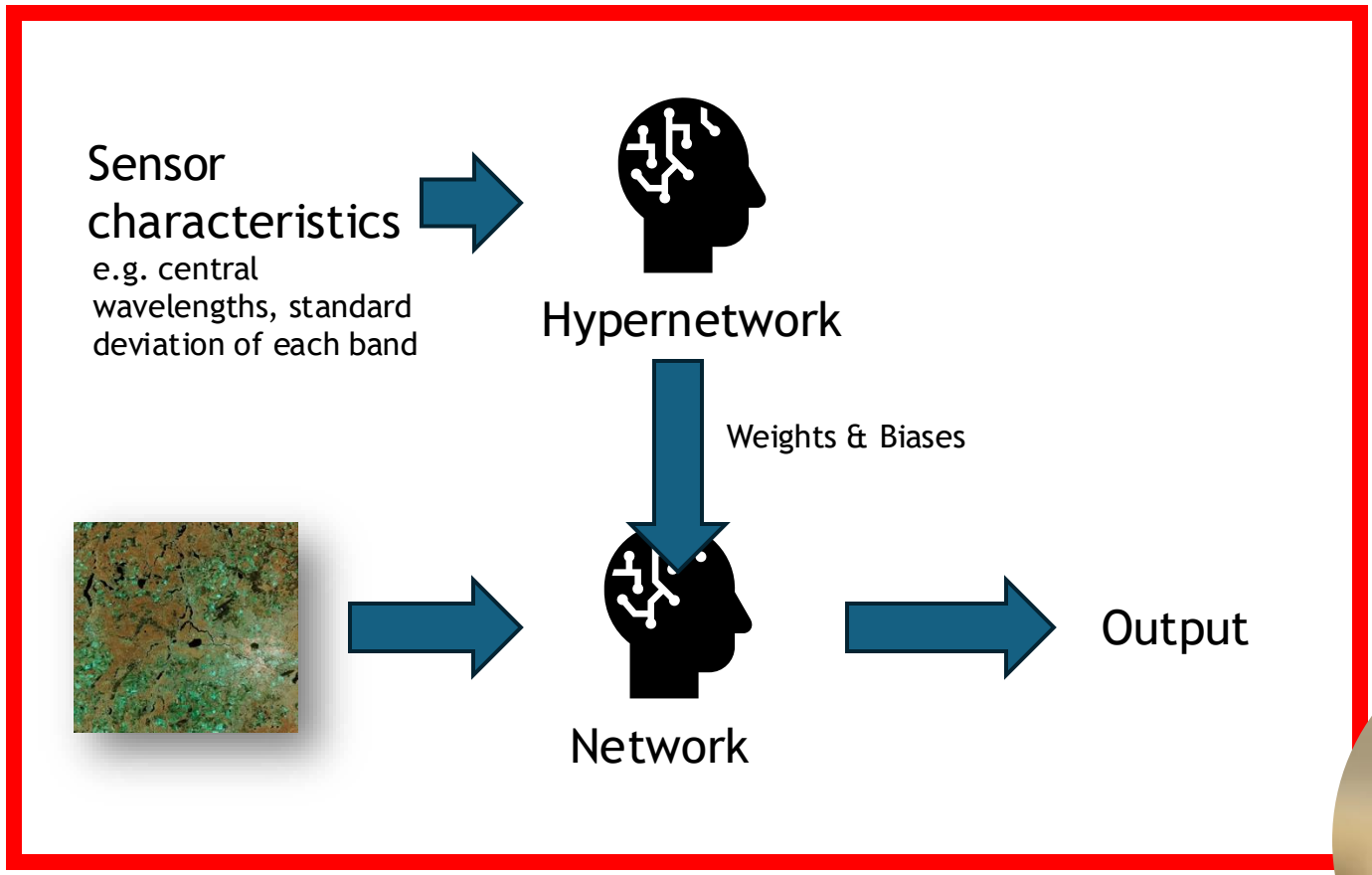
1st base concept:
Hypernetwork

Zhitong Xiong et al “Neural plasticity-inspired multimodal foundation model for earth observation” 2024

DOFA stands for ”Dynamic-One-For-All”

1st base concept:
Hypernetwork

Goal:
- “reduce computational overhead” and complexity
- Multimodality (also on unseen data)



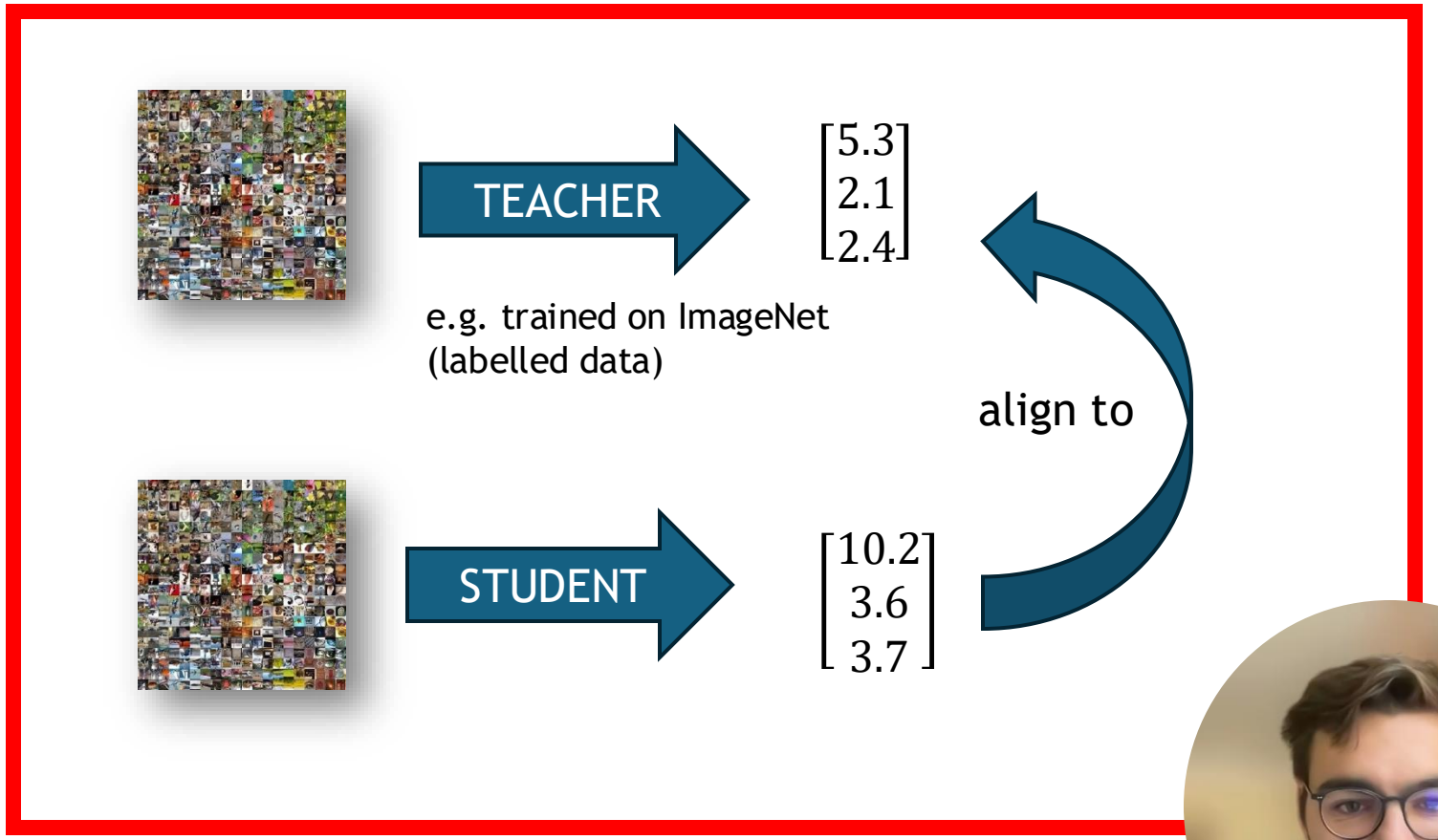
Zhitong Xiong et al “Neural plasticity-inspired multimodal foundation model for earth observation” 2024

DOFA stands for ”Dynamic-One-For-All”

2nd base concept:
Distillation loss

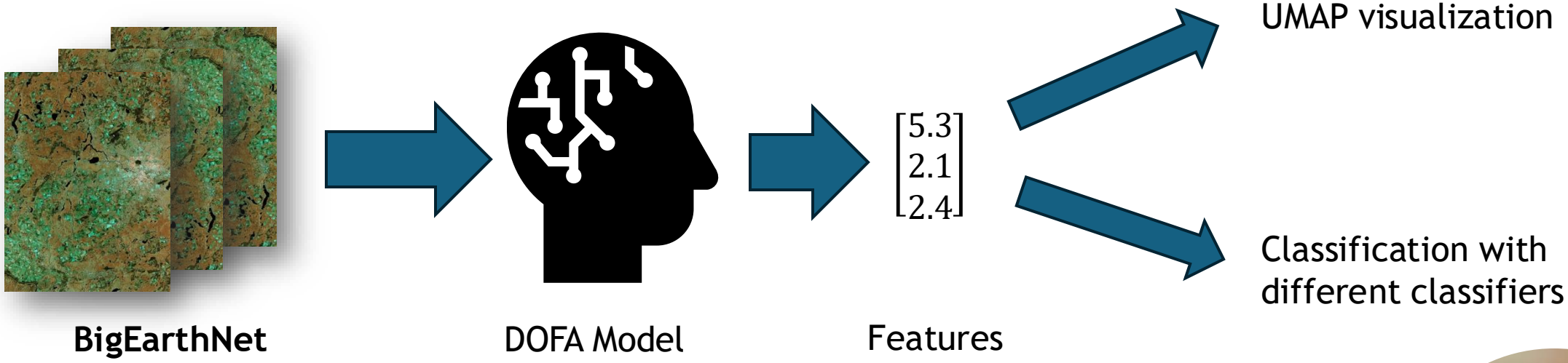
Goal:
- Accelerate training convergence
- Enhance overall performance

Combined with a reconstruction loss
→ Reproduce input data correctly + produce informed represantations



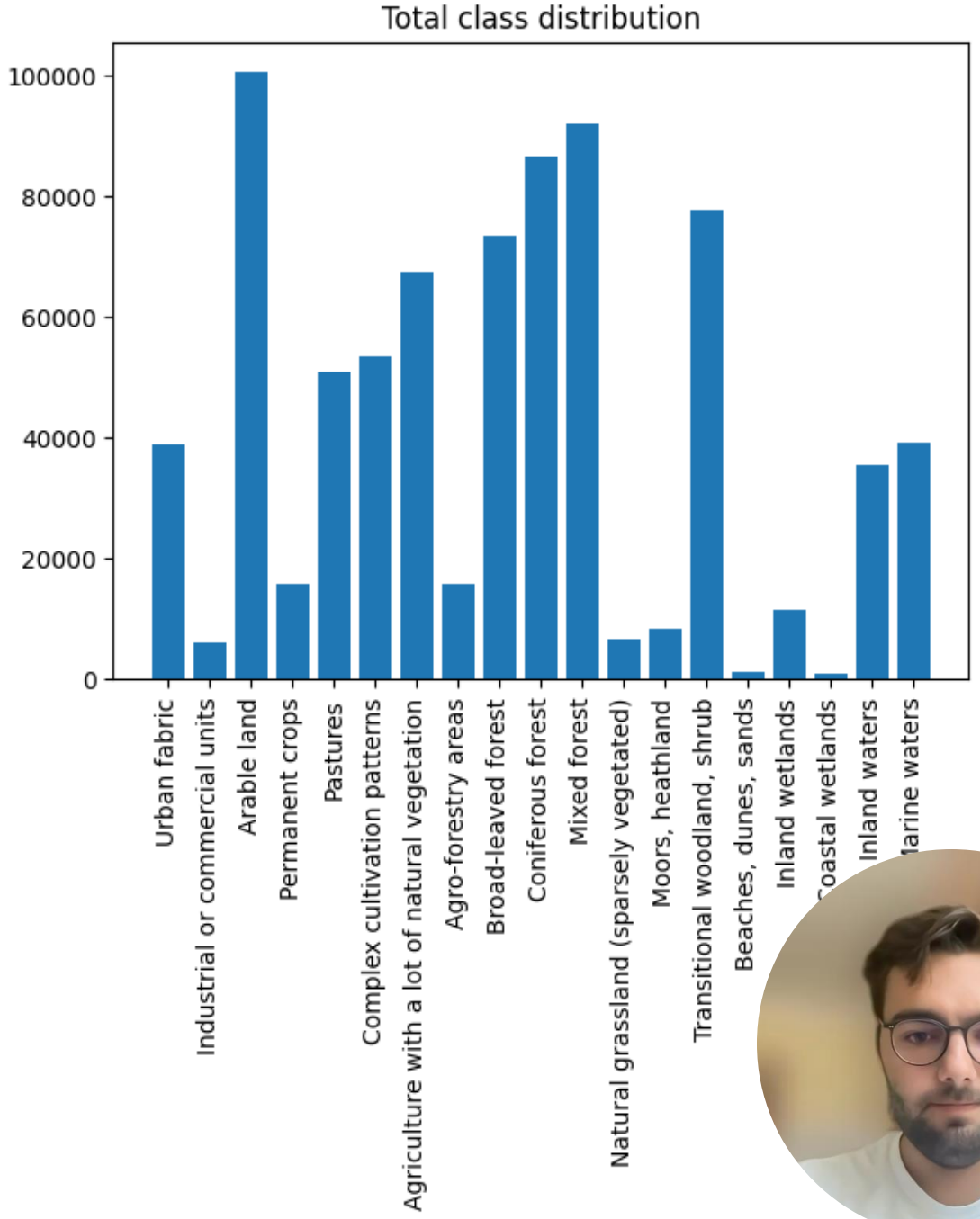
Zhitong Xiong et al “Neural plasticity-inspired multimodal foundation model for earth observation” 2024

Methodology



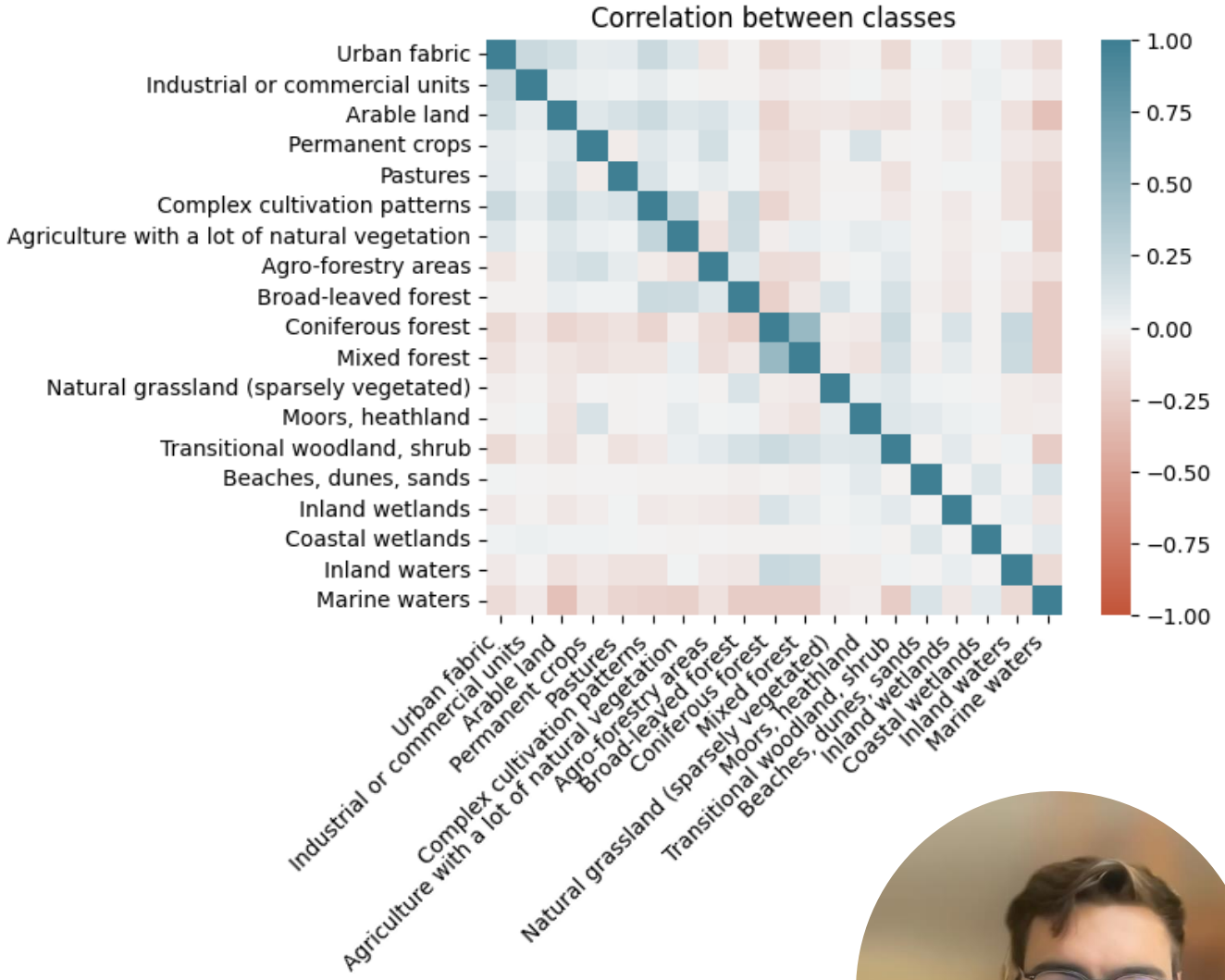
About the Dataset

- 550.000 labeled image patches from **Sentinel-I** and Sentinel-II
- 19 classes or 43 classes
- Split into train and test set by torchgeo repository



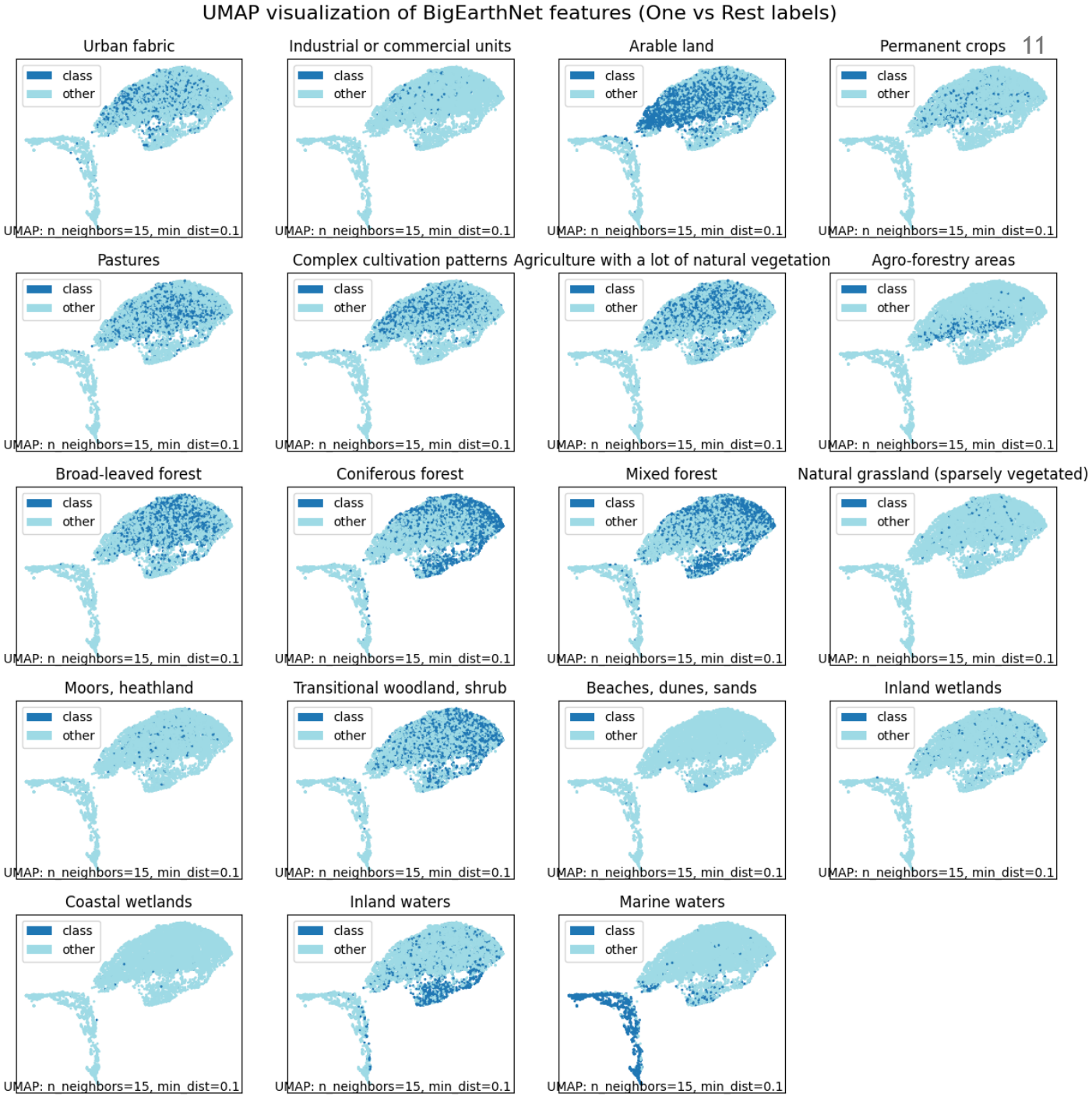
Dataset

- 550.000 labeled image patches from Sentinel-I and Sentinel-II
- 19 classes or 43 classes
- Split into train and test set by torchgeo repository



UMAP feature analysis

- UMAP transformation on DOFA features
- labelled with 19 classes

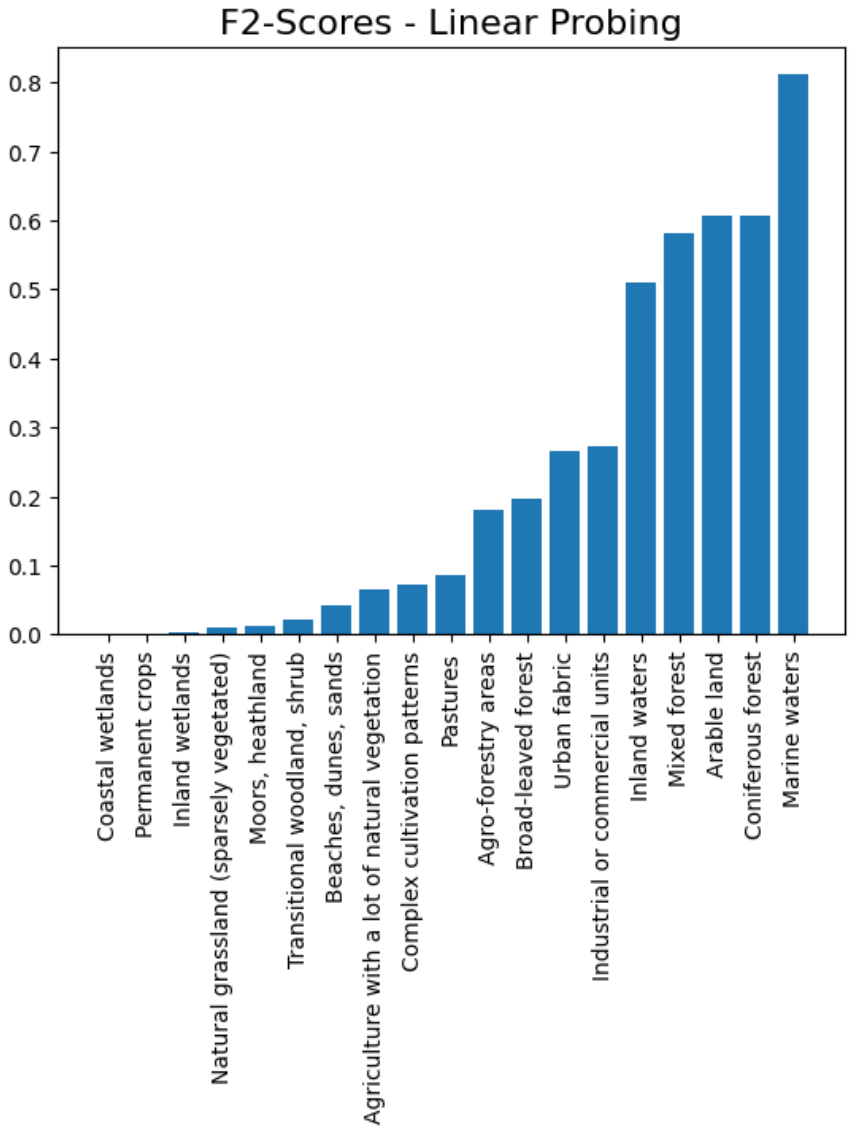


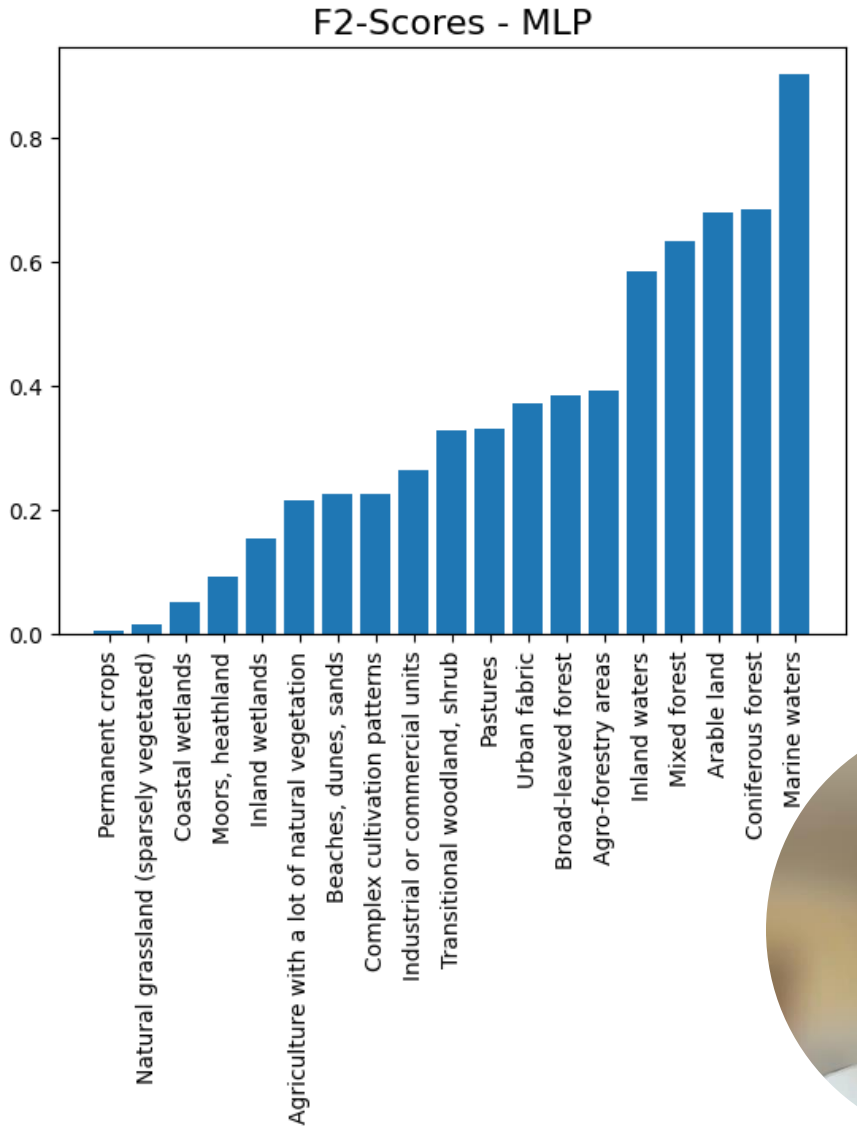
Classification results

Tests results on BigEarthNet using DOFA feature vectors and 19 classes:

	F^2_{macro} (%)	F^2_{micro} (%)	hamming loss	P_{macro} (%)	P_{micro} (%)
Random Forest	21.4	35.8	0.123	52	**72**
Linear Probing	22.9	35.3	0.124	49	70
MLP	**34.5**	**47.8**	**0.113**	**58**	71

Classification results





Key takeaway

- DOFA is able to produce meaningful features from the BigEarthNet
- DOFA has low performance for low-data-availability classes

References

- Zhitong Xiong, Yi Wang, Fahong Zhang, Adam J. Stewart, Jo¨elle Hanna, Damian Borth, Ioannis Papoutsis, Bertrand Le Saux, Gustau Camps-Valls, and Xiao Xiang Zhu. Neural plasticity-inspired multimodal foundation model for earth observation, 2024
- https://www.esa.int/ESA_Multimedia/Missions/Sentinel-1/(sortBy)/view_count/(result_type)/images
- https://eoimages.gsfc.nasa.gov/images/imagerecords/153000/153149/princecharles_oli2_20240715_lrg.jpg
- https://cs.stanford.edu/people/karpathy/cnnembed/
- Zhenda Xie, Zheng Zhang, Yue Cao, Yutong Lin, Jianmin Bao, Zhuliang Yao, Qi Dai, and Han Hu. Simmim: A simple framework for masked image modeling, 2022
- Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction,