Feb 21, 2024

Multimodal Earth Observation Modeling using AI

RemoteSensing@UniMiB

Mirko Paolo Barbato, Simone Zini, Massimiliano Clemenza, Flavio Piccoli, **Paolo Napoletano**

http://www.ivl.disco.unimib.it
https://www.pignolettomibinfn.it
In collaboration with National Institute for Nuclear Physics

Introduction to our work - Multimodal approach

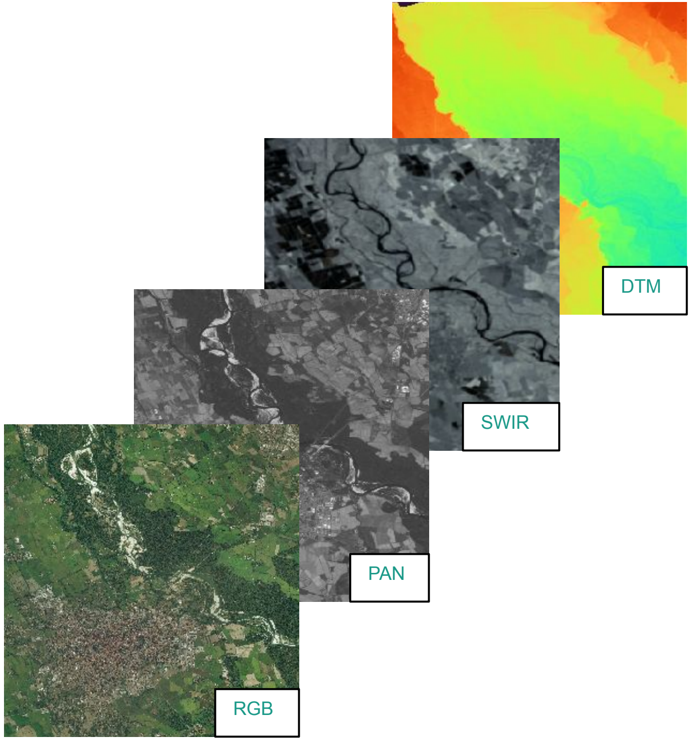
- Earth Observation is used to **monitor the environment** with consequences in the study of climate change, management of resources, agriculture of precision procedures, etc...

- The **environment complexity poses hard challenges** in its monitoring and investigation and **cannot be expressed using only one kind of sensors**

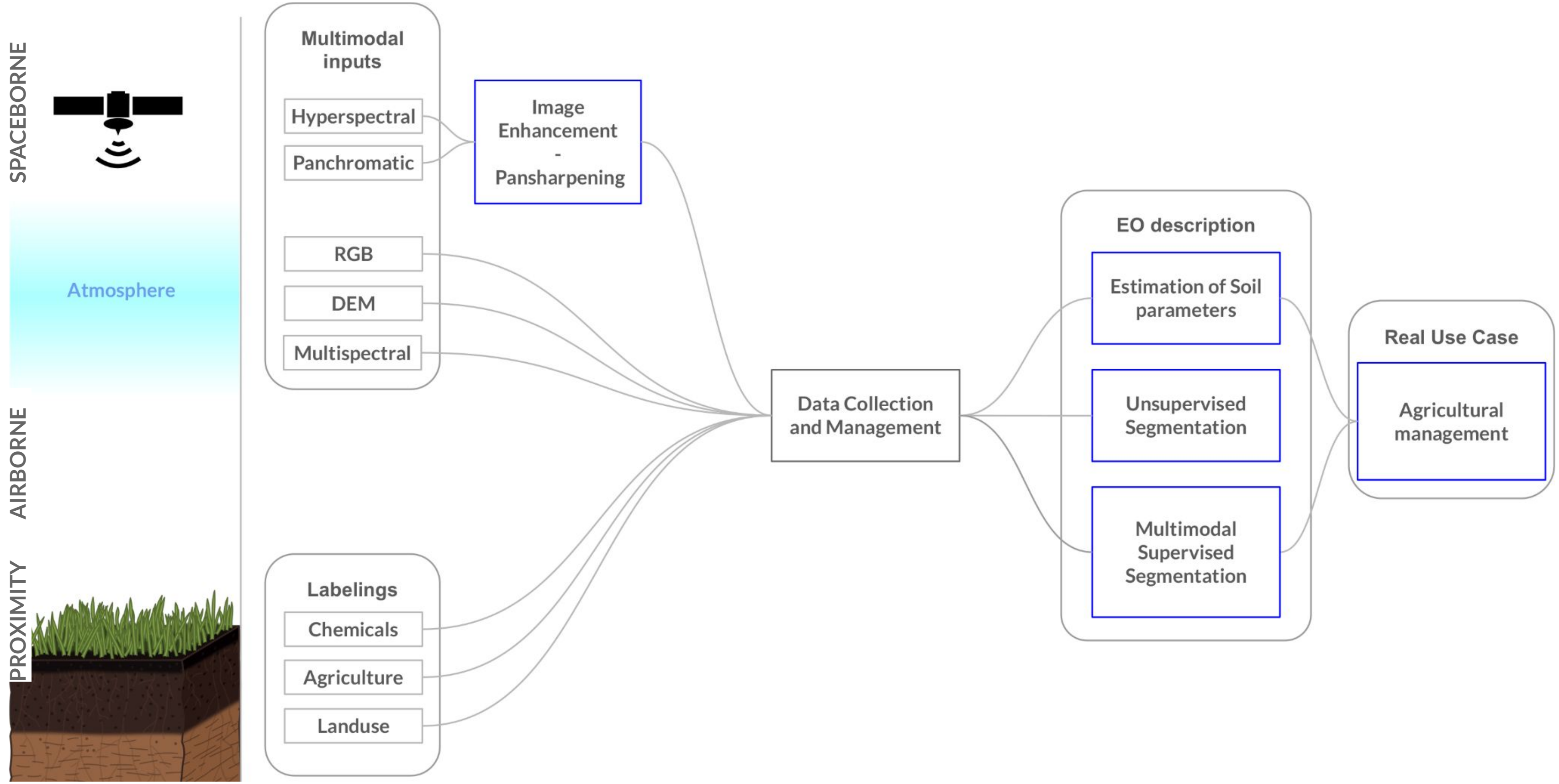
- **Multimodal approach**:

  - Multimodal remote sensing **combines data** from **multiple sensors** to overcome limitations and enhance the analysis.

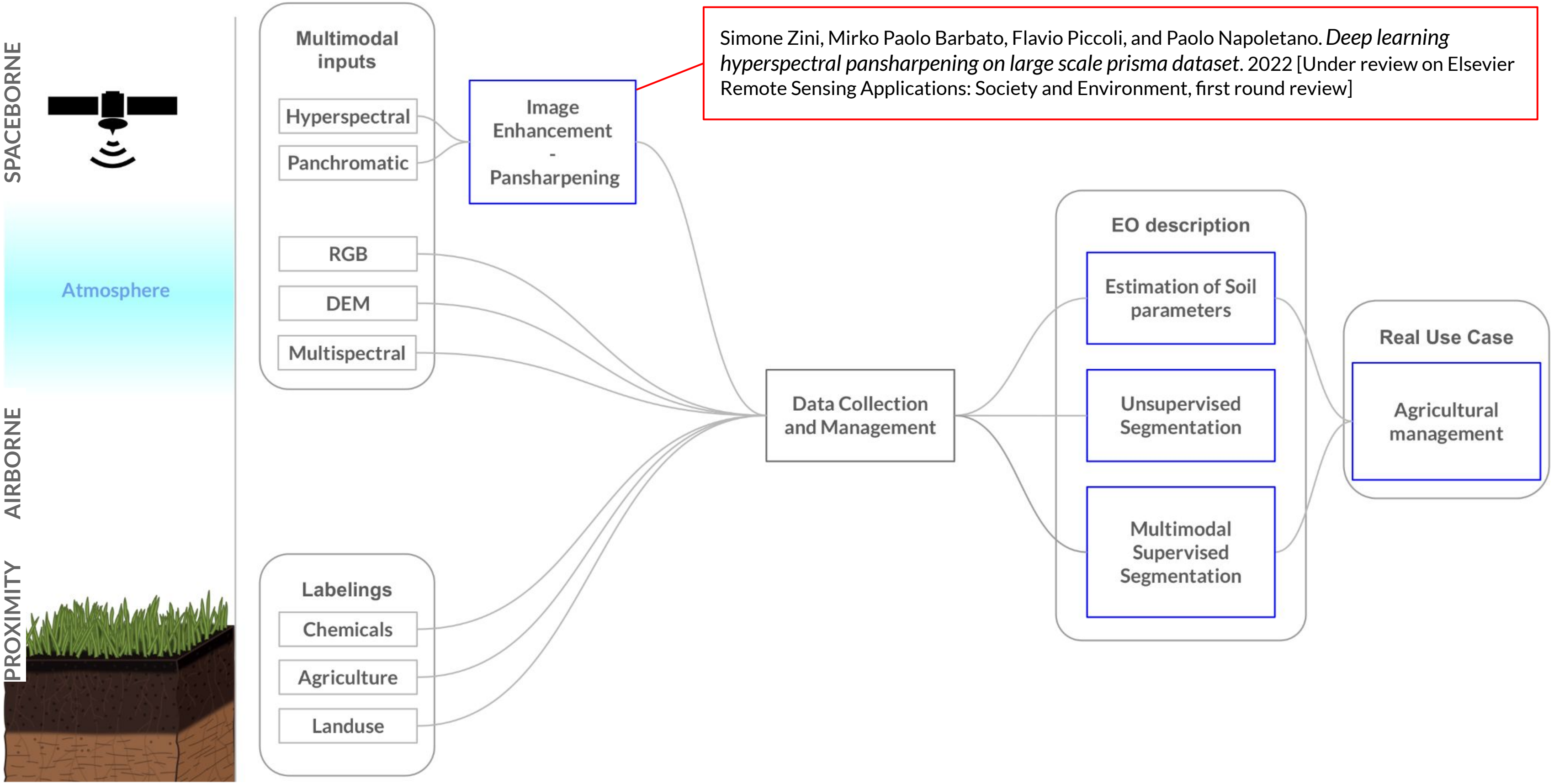
  - The **fusion** of different **modalities** can provide **complementary information** and **improve the performance** of different tasks.



Overview of our topics (selected) in Earth Observation



Overview of our topics (selected)



Simone Zini, Mirko Paolo Barbato, Flavio Piccoli, and Paolo Napoletano. *Deep learning hyperspectral pansharpening on large scale prisma dataset*. 2022 [Under review on Elsevier Remote Sensing Applications: Society and Environment, first round review]

Dataset pre-processing - Pansharpening

The **hyperspectral pansharpening** consists of **fusing the PAN and the correspondent HS images** to enhance the spatial resolution of the spectral cube

→ **HS image** reaches the **same spatial resolution as the correspondent PAN image**



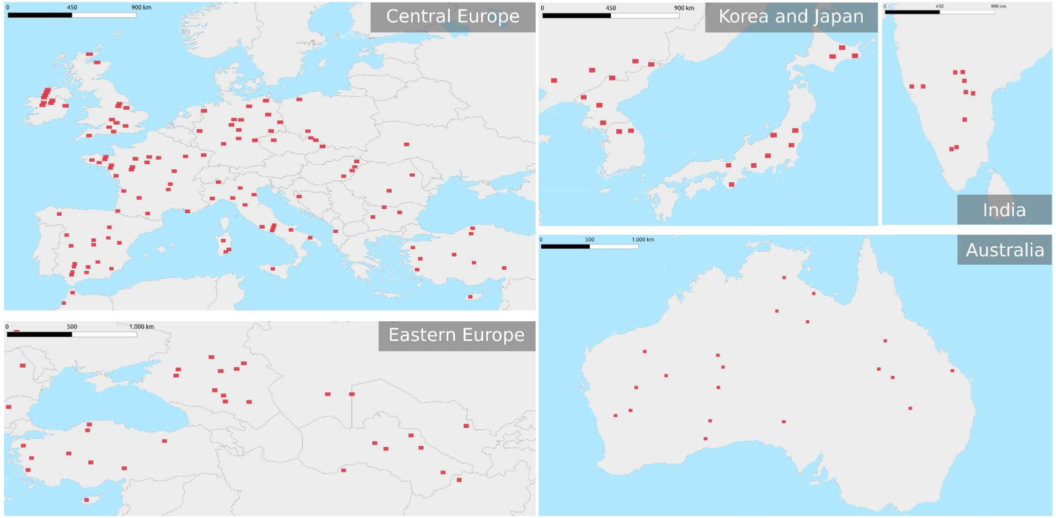

Issues:

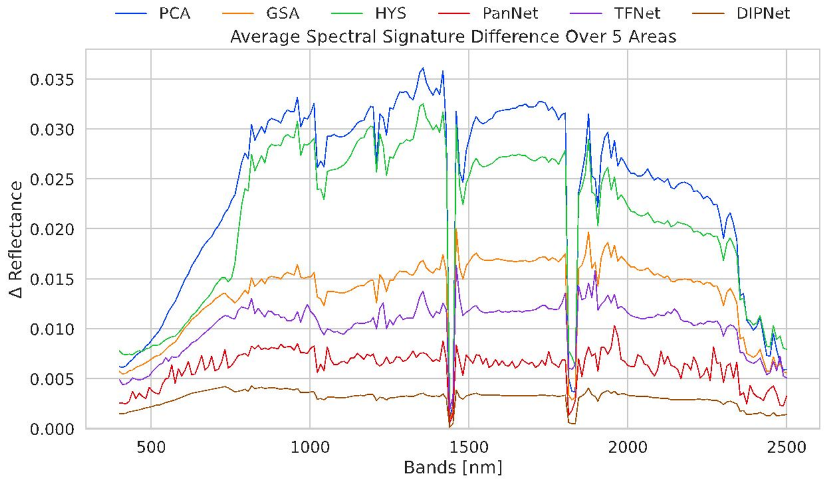
- hyperspectral datasets for pansharpening are small and not suitable for classic deep learning techniques (same as segmentation)
- models are not generalizable
- pansharpening techniques are usually built upon multispectral datasets

Simone Zini, Mirko Paolo Barbato, Flavio Piccoli, and Paolo Napoletano. Deep learning hyperspectral pansharpening on large scale prisma dataset, 2022 [Under review on Elsevier Remote Sensing Applications: Society and Environment, first round review]; Images at full resolution: https://thezino.github.io/HSbenchmarkPRISMA/

Investigation of pansharpening

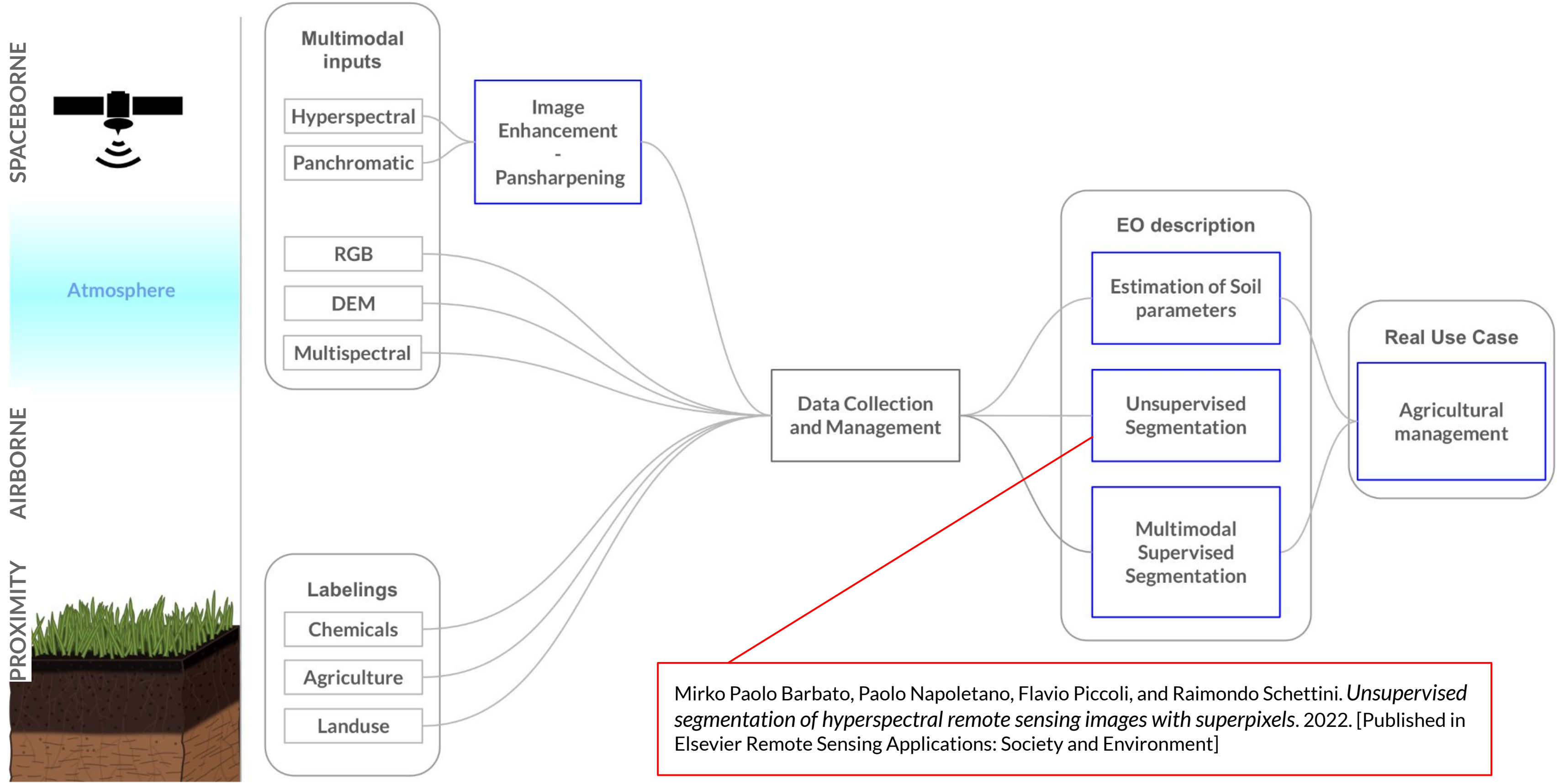
- Built a **dataset of 190 images from ASI Prisma** (262200 km^2)
  - first hyperspectral dataset statistically relevant for the task
  - different areas from all the world
- Adaptation of several **Deep Learning methods** and **comparison with traditional machine learning methods** (no-training).





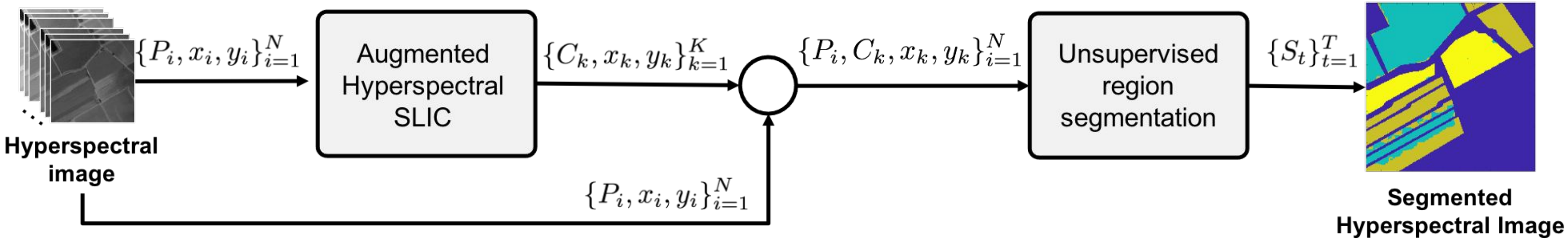
Simone Zini, Mirko Paolo Barbato, Flavio Piccoli, and Paolo Napoletano. Deep learning hyperspectral pansharpening on large scale prisma dataset, 2022 [Under review on Elsevier Remote Sensing Applications: Society and Environment, first round review]; Images at full resolution: https://thezino.github.io/HSbenchmarkPRISMA/

Overview of our topics (selected)



Mirko Paolo Barbato, Paolo Napoletano, Flavio Piccoli, and Raimondo Schettini. *Unsupervised segmentation of hyperspectral remote sensing images with superpixels*. 2022. [Published in Elsevier Remote Sensing Applications: Society and Environment]

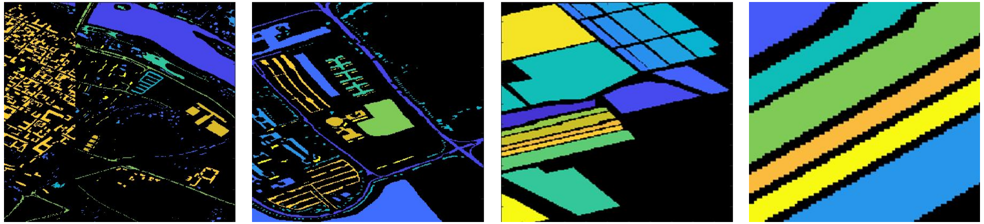
Unsupervised Segmentation - Pipeline

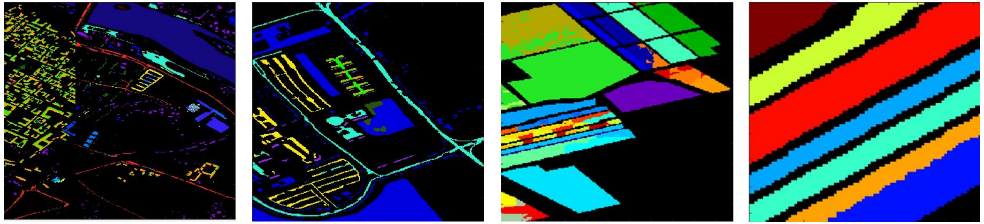


- Does **not need preliminary information**
- Does **not need to be trained**
- **Easily adaptable to different kinds of data and features**
- Robust to noise

Unsupervised Segmentation - Results

Ground truth

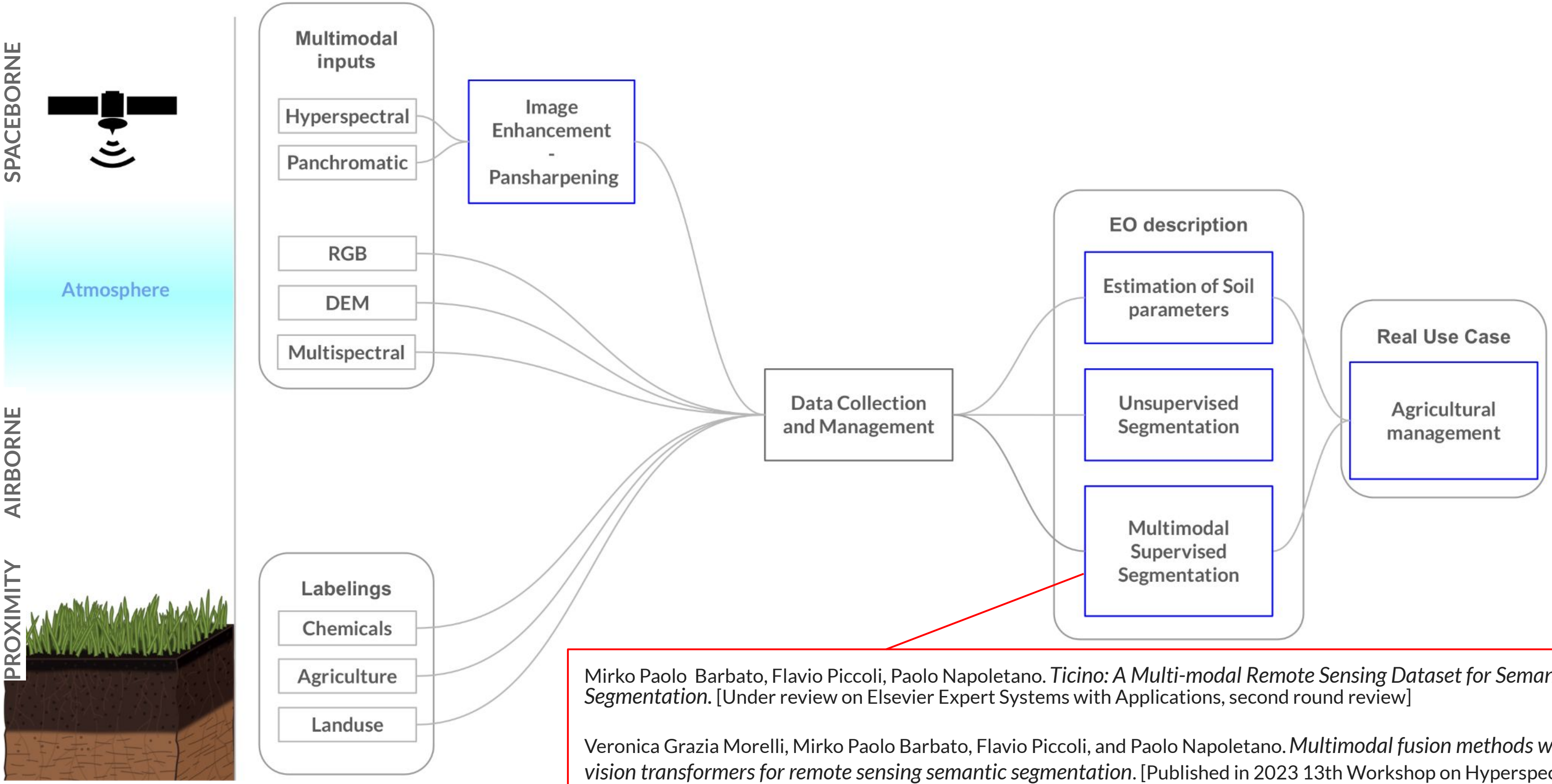




Unsupervised segmentation

	Pavia Center ARI - NMI	Pavia Univ. ARI - NMI	Salinas ARI - NMI	SalinasA ARI - NMI	**Average** ARI - NMI	Require number of classes
K-means (Obeid et al. (2021))	0.85 - 0.82	0.40 - 0.63	0.63 - 0.83	0.67 - 0.78	0.64 - 0.77	Yes
GMM (Obeid et al. (2021))	0.77 - 0.74	0.29 - 0.53	0.53 - 0.79	0.78 - 0.87	0.59 - 0.73	Yes
HNMF (Gillis et al. (2014))	0.85 - 0.77	0.38 - 0.57	0.53 - 0.79	0.78 - 0.87	0.64 - 0.75	Yes
SMCE (Elhamifar and Vidal (2013))	0.80 - 0.77	0.31 - 0.56	0.57 - 0.78	0.76 - 0.81	0.61 - 0.73	Yes
DLSS (Murphy and Maggioni (2018))	0.52 - 0.42	0.49 - 0.57	0.37 - 0.39	0.63 - 0.81	0.55 - 0.50	Yes
3D-CAE (Nalepa et al. (2020))	0.96 - 0.86	0.36 - 0.59	0.67 - 0.85	0.77 - 0.87	0.69 - 0.79	Yes
DEC (Xie et al. (2016))	0.83 - 0.80	0.41 - 0.67	0.57 - 0.80	0.78 - 0.87	0.65 - 0.79	Yes
BDEC (Obeid et al. (2021))	**0.97** - **0.91**	**0.60** - 0.70	0.68 - 0.87	0.81 - 0.87	0.77 - 0.84	Yes
OUR_BW	0.81 - 0.80	0.53 - 0.70	0.67 - 0.87	0.82 - 0.92	0.71 - 0.82	**No**
OUR	0.88 - 0.87	0.59 - **0.72**	**0.85** - **0.91**	**0.90** - **0.95**	**0.81** - **0.86**	**No**

Overview of our topics (selected)



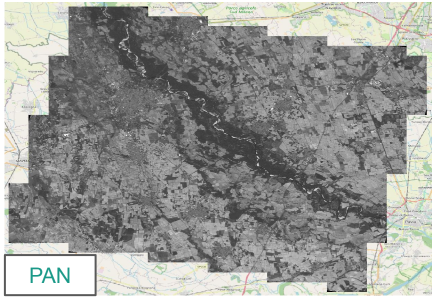
Mirko Paolo Barbato, Flavio Piccoli, Paolo Napoletano. *Ticino: A Multi-modal Remote Sensing Dataset for Semantic Segmentation.* [Under review on Elsevier Expert Systems with Applications, second round review]

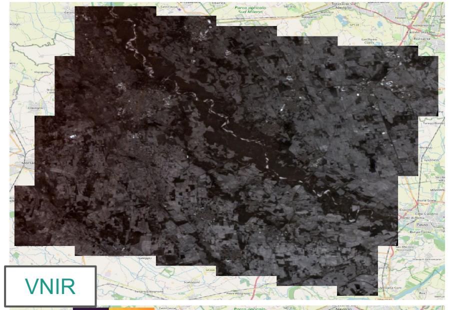
Veronica Grazia Morelli, Mirko Paolo Barbato, Flavio Piccoli, and Paolo Napoletano. *Multimodal fusion methods with vision transformers for remote sensing semantic segmentation*. [Published in 2023 13th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS). IEEE]

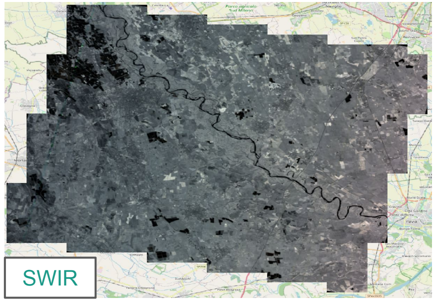
Proposed dataset - sources

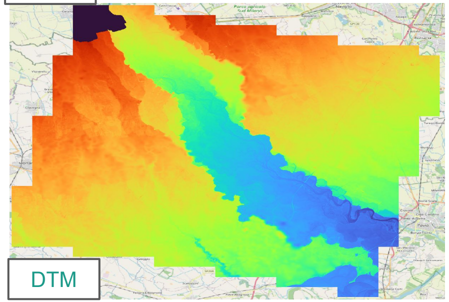
- **RGB** data (Microsoft Bing Maps) ~2.25 m per pixel
- **Hyperspectral*** (Prisma- 30 m per pixel)
- **Panchromatic**** data (Prisma- 5 m per pixel)
- **Digital Terrain Model** of the area considered (Geoportal of Lombardia Region – 5 m per pixel)

* **Hyperspectral is** Visual and Near-Infrared (VNIR 400-1000 nm) and Short-wave Infrared (SWIR 900-2500 nm)
** **Panchromatic** uses a single band that combines Red, Green, and Blue bands





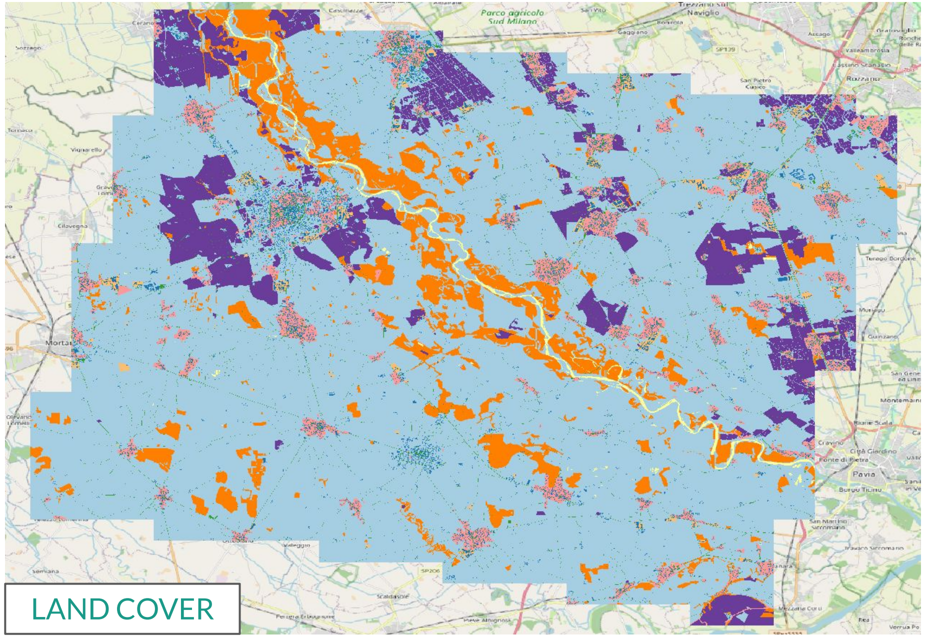


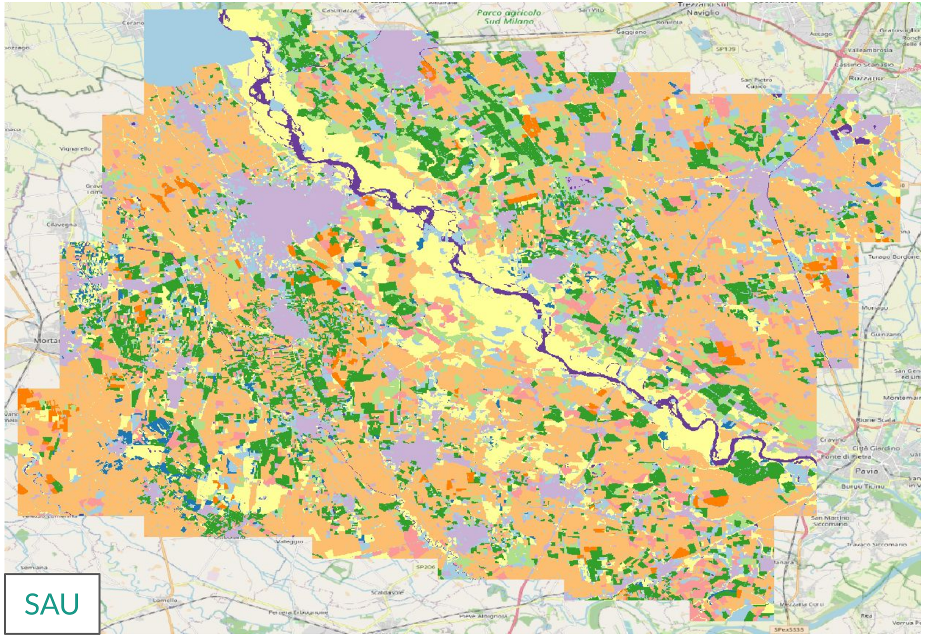


- **RGB: 3 channels**
- **PAN: 1 channel**
- **Vnir: 63 bands**
- **Swir: 171 bands**
- **DTM: 1 channels**

Microsoft Bing Maps https://www.bing.com/maps/?cp=45.470725%7E9.214096&lvl=11.0
Prisma satellites data. https://www.asi.it/scienze-della-terra/prisma/ Geoportale Regione Lombardia https://www.geoportale.regione.lombardia.it/

Proposed dataset - labelings





Land Cover (OpenStreetMaps and Italian Agenzie delle Entrate) with 8 classes:

> *Background, Building, Road, Residential, Industrial, Forest, Farmland, Water*

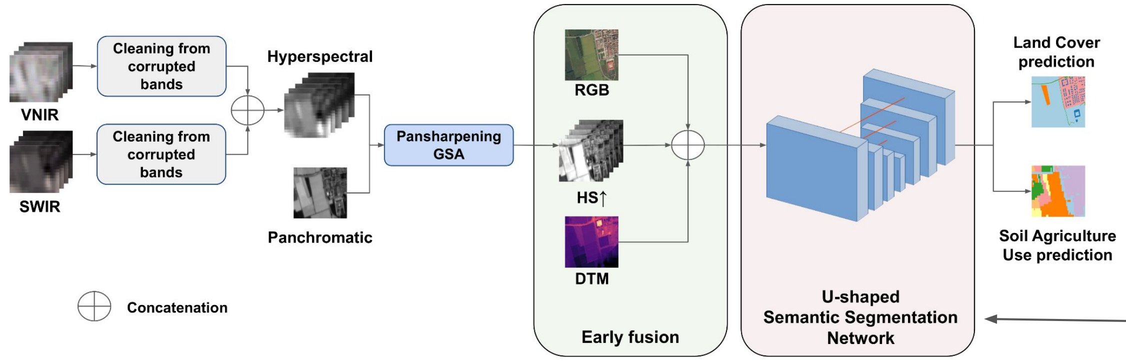
Soil Agriculture Use *(Geoportal of Lombardia Region) with 10 classes:*

> *Background, Other agricultural crops, Forage crops, Corn, Industrial plants, Rice, Seeds, Man-made areas, Water bodies, and Natural vegetation.*

OpenStreetMap. OSM https://www.openstreetmap.org/#map=6/42.088/12.564 Agenzia delle Entrate - Consultazione cartografia catastale https://www.agenziaentrate.gov.it/portale/web/guest/schede/fabbricatiterreni/consultazione-cartografia-catastale/servizio-consultazione-cartografia

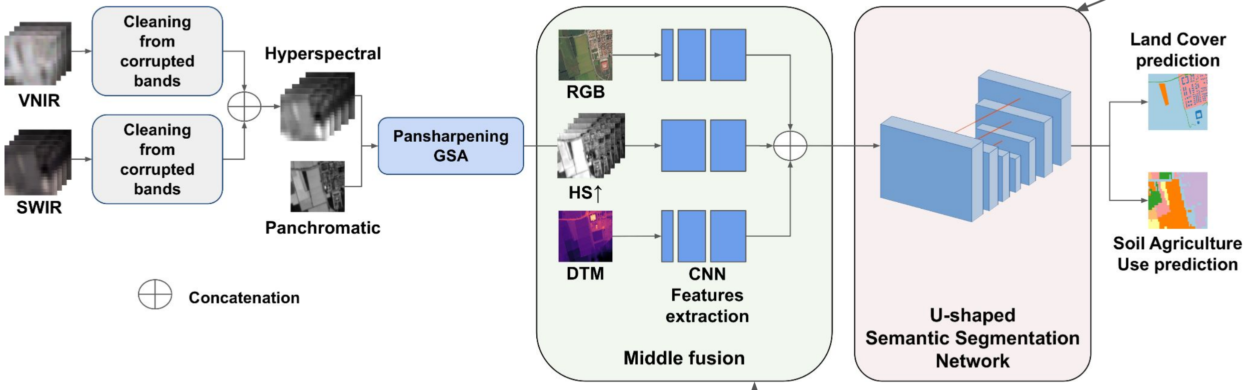
CNN-based fusion pipeline

Early fusion



Unet with a Resnet18 backbone

Middle fusion



This module extracts **high-level features** from each modality and concatenates them. It consists of **convolution** and **ReLU layers** for each modality that use **padding** to adapt the width and height of the inputs.

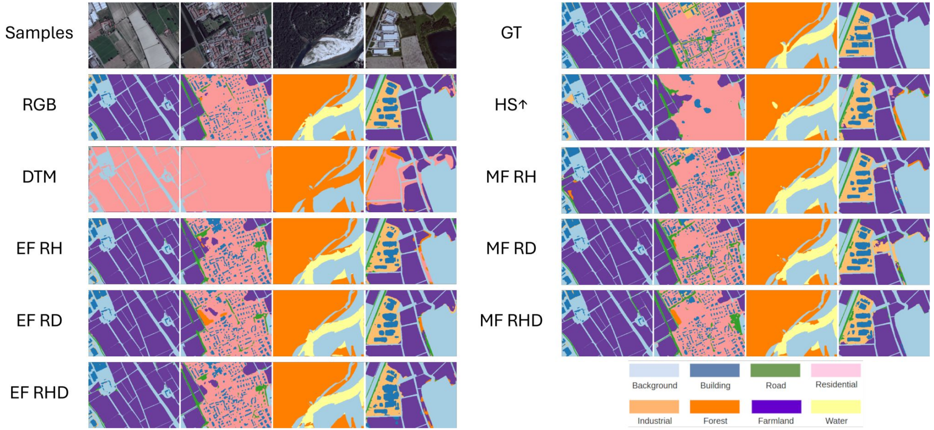
Mirko Paolo Barbato, Flavio Piccoli, Paolo Napoletano. Ticino: A Multi-modal Remote Sensing Dataset for Semantic Segmentation. [Under review on Elsevier Expert Systems with Applications, second round review]

Results on Land Cover

Results demonstrate the **usefulness of multimodality** in the Land Cover scenario with an **Overall increment** of 3% Accuracy, 4% mIoU and 1% Precision.

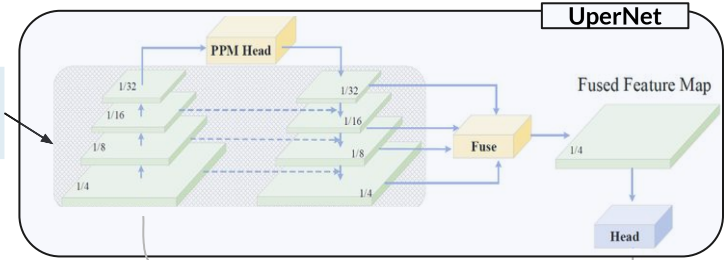
Land Cover										
		No fusion			Early fusion			Middle fusion		
Class	Metric	RGB	HS↑	DTM	RGB HS↑	RGB DTM	RGB HS↑ DTM	RGB HS↑	RGB DTM	RGB HS↑ DTM
Overall	Acc	0.75	0.68	0.26	0.75	0.73	0.72	**0.79**	0.78	0.78
	IoU	0.63	0.56	0.17	0.63	0.61	0.61	0.66	0.66	**0.67**
	Prec.	0.78	0.72	0.30	0.76	0.78	0.77	0.78	**0.79**	**0.79**

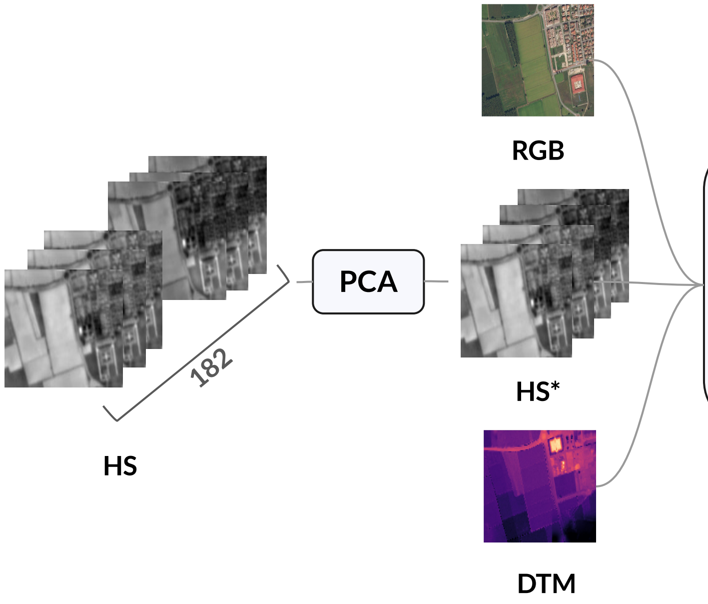
Land Cover										
		No fusion			Early fusion			Middle fusion		
Class	Metric	RGB	HS↑	DTM	RGB HS↑	RGB DTM	RGB HS↑ DTM	RGB HS↑	RGB DTM	RGB HS↑ DTM
Building	Acc	0.62	0.39	0.00	0.64	0.62	0.63	**0.75**	0.74	0.69
	IoU	0.50	0.31	0.00	0.49	0.49	0.48	**0.54**	**0.54**	0.53
	Prec.	**0.71**	0.60	0.36	0.68	0.70	0.67	0.66	0.68	0.69
	Acc	0.52	0.29	0.03	0.45	0.45	0.41	0.57	**0.61**	0.55
Farmland	Acc	0.93	0.91	0.63	0.93	0.94	**0.95**	0.93	**0.95**	**0.95**
	IoU	0.85	0.82	0.39	0.86	0.83	0.88	0.87	0.89	**0.90**
	Prec.	0.91	0.89	0.51	0.91	0.87	0.92	0.94	0.93	**0.95**
Water	Acc	0.79	0.86	0.02	0.87	0.75	0.85	**0.89**	0.76	0.88
	IoU	0.65	0.72	0.02	**0.74**	0.66	0.73	**0.74**	0.65	0.73
	Prec.	0.79	0.82	0.06	0.83	0.83	**0.85**	0.81	0.82	0.81
Overall	Acc	0.75	0.68	0.26	0.75	0.73	0.72	**0.79**	0.78	0.78
	IoU	0.63	0.56	0.17	0.63	0.61	0.61	0.66	0.66	**0.67**
	Prec.	0.78	0.72	0.30	0.76	0.78	0.77	0.78	**0.79**	**0.79**

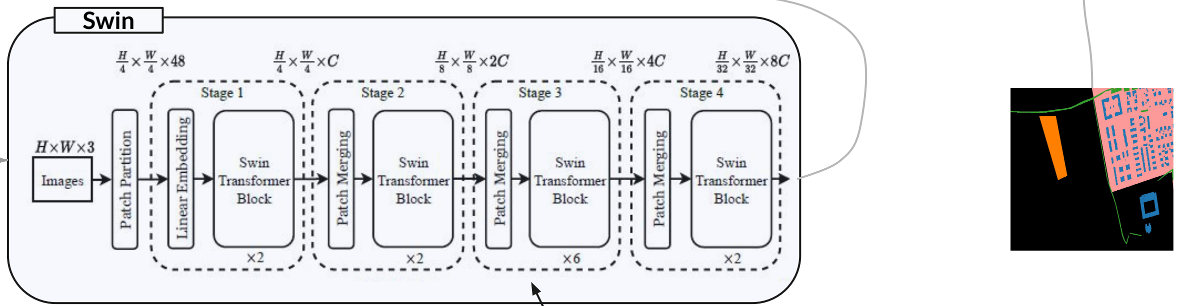


Transformer-based fusion techniques - General pipeline

UperNet is used as **decoder**







The **Swin architecture** has been used as **backbone** of the U-shaped architecture considered and as the **base for all the fusion methods** tested, modifying it accordingly.

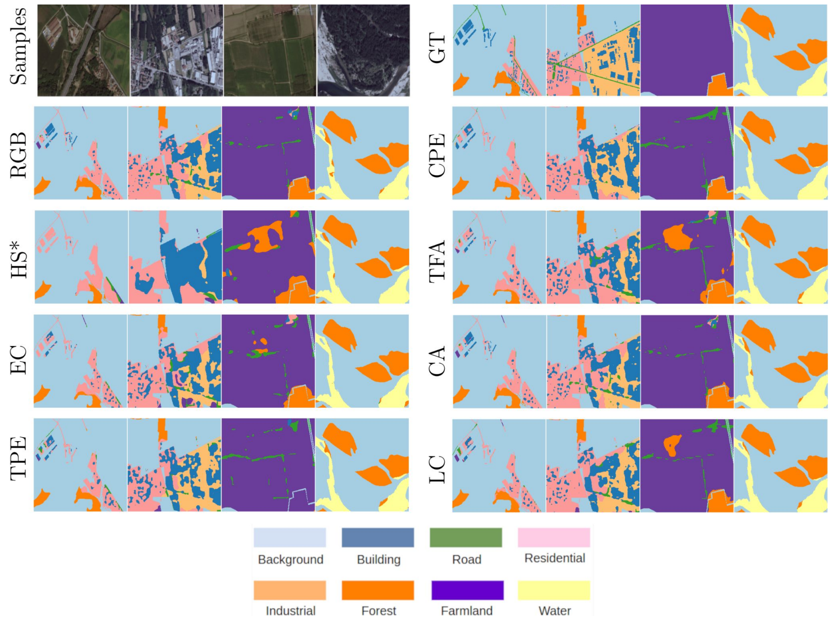
LIU, Ze, et al. Swin transformer: Hierarchical vision transformer using shifted windows. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021. p. 10012-10022.
MA, Xiaoxuan, et al. Tnseg: adversarial networks with multi-scale joint loss for thyroid nodule segmentation. *The Journal of Supercomputing*, 2023, 1-26.

Transformer-based analysis Results

The **Late concatenation is the best compromise** in terms of performance and complexity of the model.

Token Fusion as Attention Level is able to achieve better results than RGB **without incrementing the complexity** of the model in terms of parameters.

	Method	Acc	Pr	mIoU	Macs	Pars
Single	RGB	67.22	72.75	55.71	9.65	39.28
	HS*	58.10	62.71	45.51	9.65	39.28
Multi	Early Conc. (EC)	67.57	73.15	56.06	9.68	39.29
	Tok. Pat. Emb. (TPE)	68.89	64.71	73.95	16.40	60.60
	Cha. Pat. Emb. (CPE)	65.01	71.18	53.85	65.43	241.96
	Tok. Fus. Att. (TFA)	69.13	74.27	57.51	16.14	38.74
	Cross-Att. (CA)	**71.85**	74.72	59.42	37.86	111.61
	Late Conc. (LC)	71.84	**75.31**	**59.69**	16.14	63.29

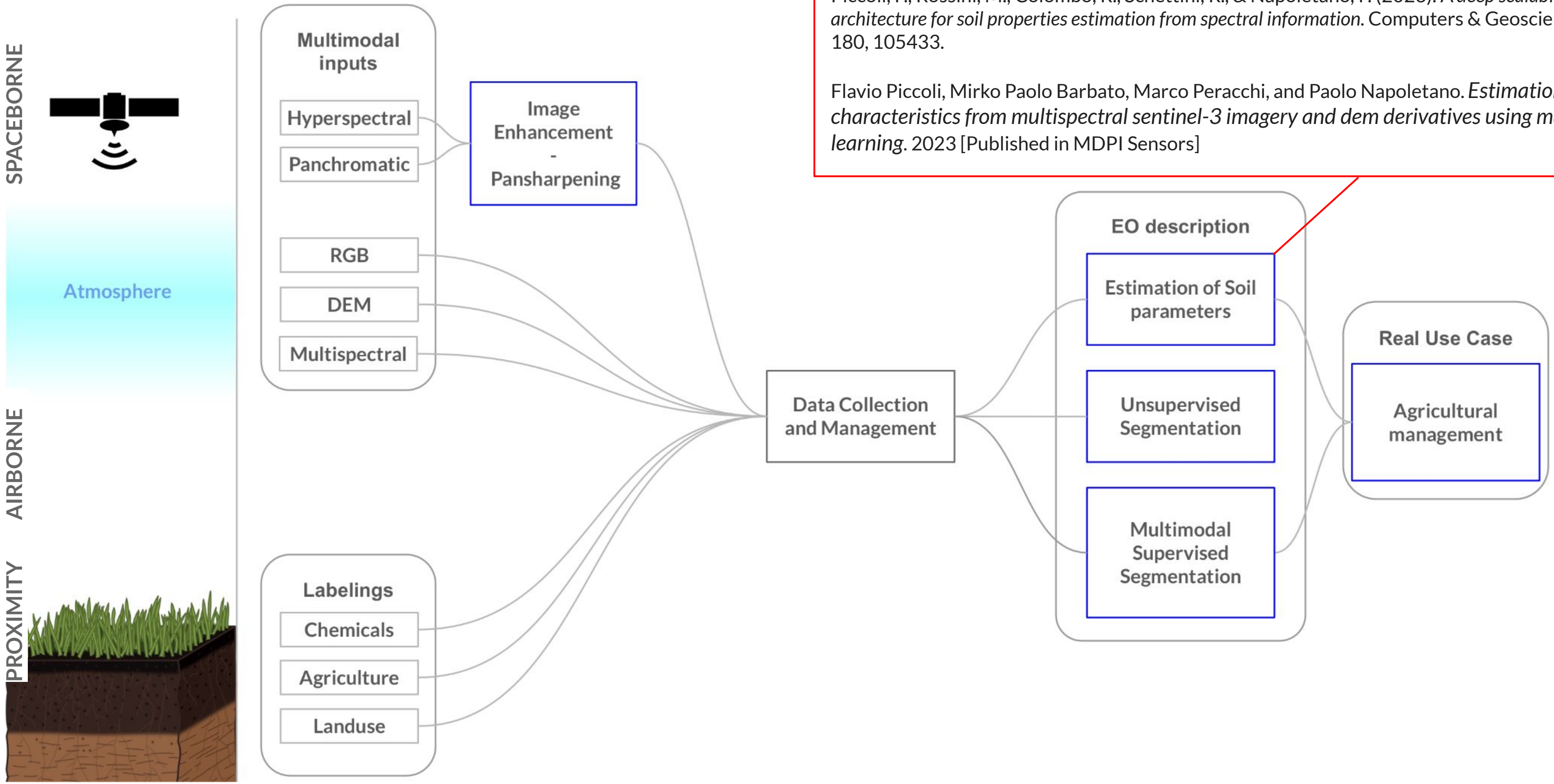


Veronica Grazia Morelli, Mirko Paolo Barbato, Flavio Piccoli, and Paolo Napoletano. Multimodal fusion methods with vision transformers for remote sensing semantic segmentation. In 2023 13th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS). IEEE, 2023 (in press)

Overview of our topics (selected)

Piccoli, F., Rossini, M., Colombo, R., Schettini, R., & Napoletano, P. (2023). A *deep scalable neural architecture for soil properties estimation from spectral information*. Computers & Geosciences, 180, 105433.

Flavio Piccoli, Mirko Paolo Barbato, Marco Peracchi, and Paolo Napoletano. *Estimation of soil characteristics from multispectral sentinel-3 imagery and dem derivatives using machine learning*. 2023 [Published in MDPI Sensors]



Dataset 1

LUCAS:

- **19,036 topsoil** observations from Europe;

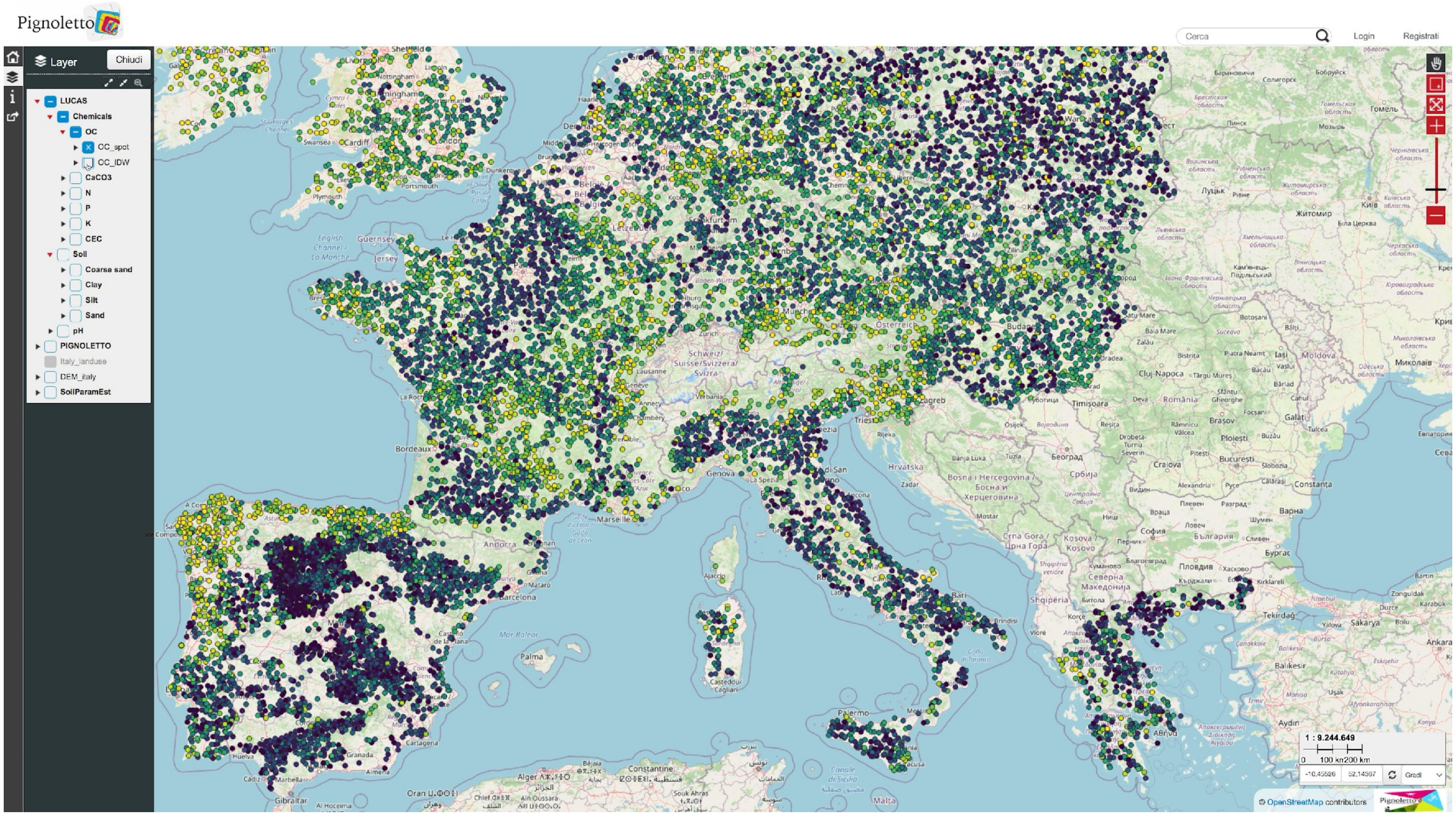
- **Spectral measurement** of air-dried samples in the range **400** to **2500 nm** with a spectral sampling interval of 0.5 nm.

- **12 chemical and physical soil properties**: the percentage of *coarse fragments*, particle size distribution (% clay, silt and sand content), *pH* (in $CaCl_2$ and H_2O), *organic carbon* (g/kg), *carbonate* content (g/kg), *phosphorous* content (mg/kg), total *nitrogen* content (g/kg), extractable *potassium* content (mg/kg) and the *cation exchange capacity* (cmol(+)/kg).

Tóth, G., Jones, A., Montanarella, L., 2013. The lucas topsoil database and derived information on the regional variability of cropland topsoil properties in the european union. Environmental monitoring and assessment 185. doi:10.1007/s10661-013-3109-3.

Dataset 1

LUCAS



Tóth, G., Jones, A., Montanarella, L., 2013. The lucas topsoil database and derived information on the regional variability of cropland topsoil properties in the european union. Environmental monitoring and assessment 185. doi:10.1007/s10661-013-3109-3.

Proposed method

- A **flexible** deep convolutional neural networks (**CNNs**) for soil characteristics estimation from hyperspectral signal;
- A **fast architecture search** and a scalable hyper parameter search: *input resolution*, *layer types/size*, *loss functions*, *training hyperparameters*
- A **parametric network architecture** as a composition of N **building blocks** and two **fully connected layers**. The output are V estimations of soil variables

CNN structure: N blocks. R is the spatial resolution of the input signal, p and p_{max} are two parameters controlling the number of filters of each block and V is the number of output variables.

Stage	Operation	Output size
Pre-processing	Input	$1 \times R$
Building blocks	Block 1	$2^p \times \frac{R}{2}$
	Block 2	$2^{1+p} \times \frac{R}{4}$
	$\vdots$	
	Block N	$2^{N-1+p} \times \frac{R}{2^N}$
Projection	Conv 1×1	$V \times 1 \times 1$
	Flatten	V

Piccoli, F., Rossini, M., Colombo, R., Schettini, R., & Napoletano, P. (2022). A deep scalable neural architecture for soil properties estimation from spectral information. *arXiv preprint arXiv:2210.17314*.

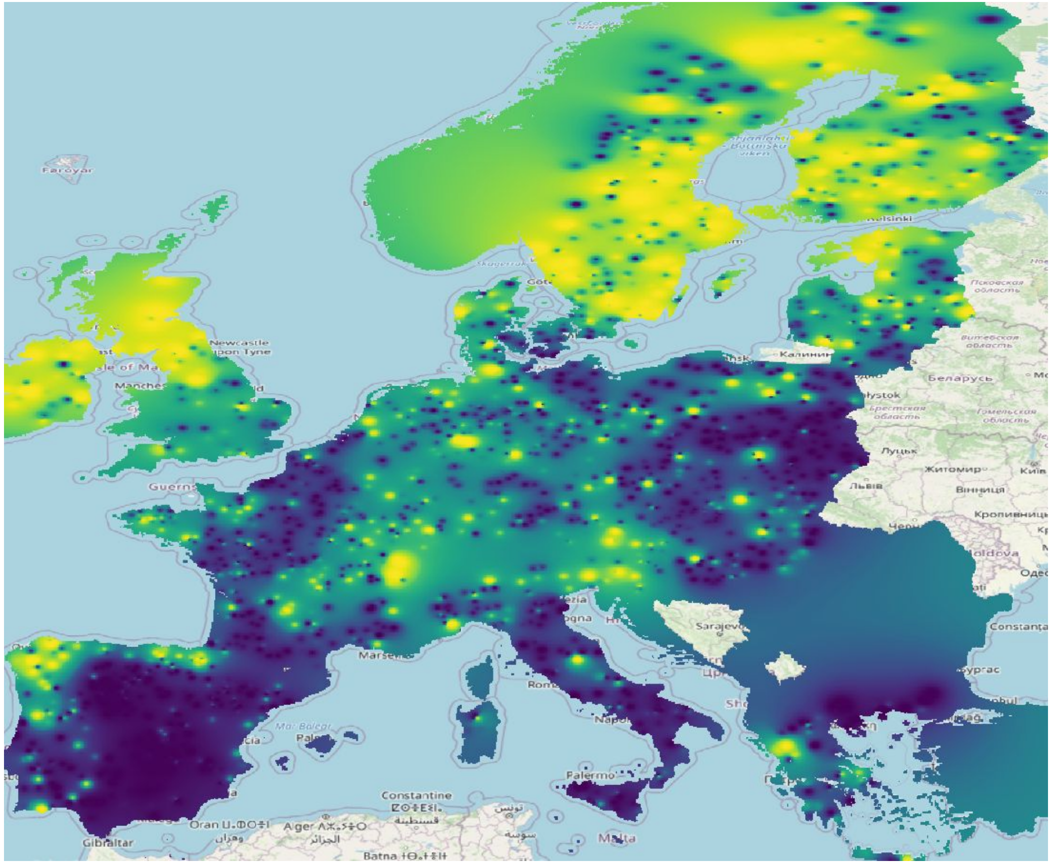
Comparison with the state of the art

- Comparison on Dataset 1 of the best architecture with Random Forests (RF), Support Vector Machines (SVR), Boosting Regression Trees (BRT) and their multi-output counterparts.

method	clay	silt	sand	pH_{CaCl2}	pH_{H2O}	OC	CaCO3	N	P	K	CEC	avg
RF [sv]	0.4780	0.3961	0.3996	0.453	0.4737	0.5304	0.5454	0.4408	0.0902	0.1422	0.4069	0.3960
SVR [sv]	0.7402	0.5222	0.5824	0.8494	0.8441	0.7219	0.7591	0.7198	0.1455	0.1386	0.6305	0.6049
BRT [sv]	0.5054	0.3998	0.4132	0.5556	0.5753	0.6001	0.6669	0.4728	0.096	0.1133	0.4480	0.4406
RF [mv]	0.5519	0.4385	0.4546	0.5899	0.6003	0.5771	0.6340	0.4894	0.0974	0.1583	0.4821	0.4612
SVR [mv]	0.7453	0.5280	0.5910	0.8537	0.8469	0.7232	0.7567	0.7219	0.1575	0.1401	0.6346	0.6090
BRT [mv]	0.5028	0.4069	0.4163	0.5522	0.5703	0.6022	0.6676	0.4779	0.0915	0.1266	0.4401	0.4413
ours [mv]	**0.7897**	**0.7048**	**0.7522**	0.9026	0.8983	**0.8303**	**0.9372**	**0.7924**	**0.3665**	**0.5441**	**0.7313**	**0.7499**
ours [sv]	0.7626	0.5243	0.6592	**0.9032**	**0.8999**	0.7637	0.9213	0.6904	0.2315	0.3433	0.3433	0.6402

Ref...

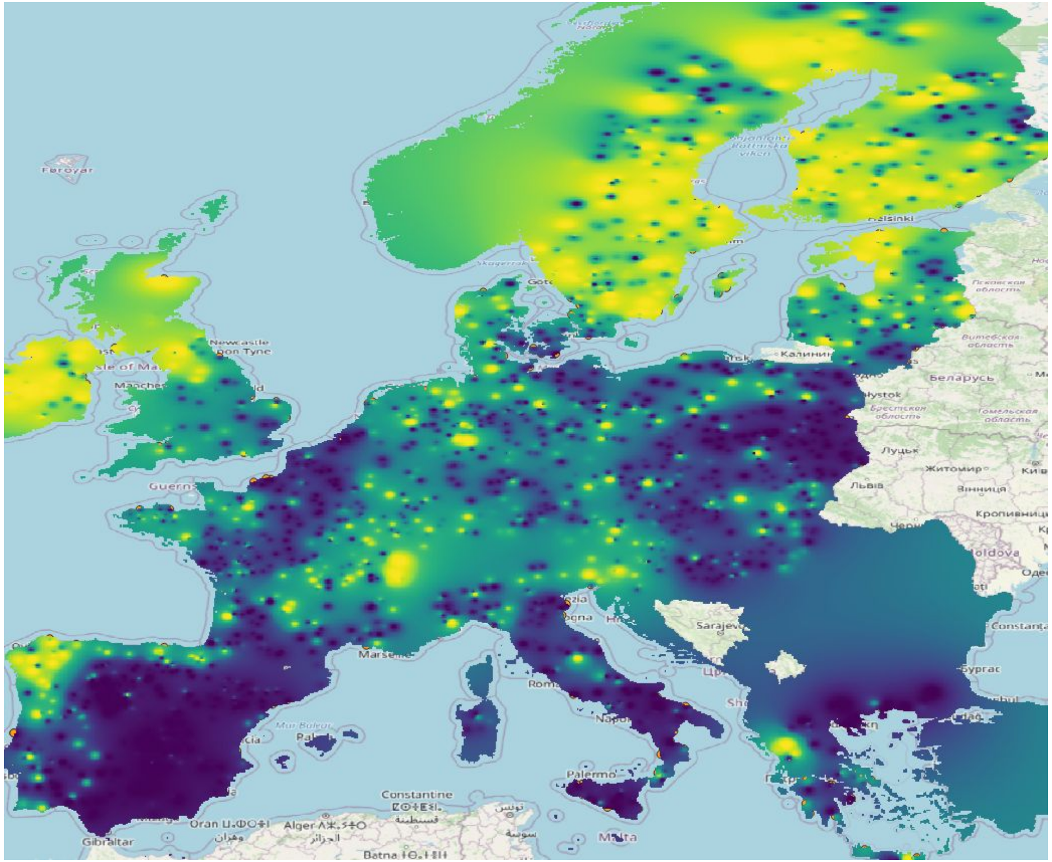
Visual results – Maps



Ref…

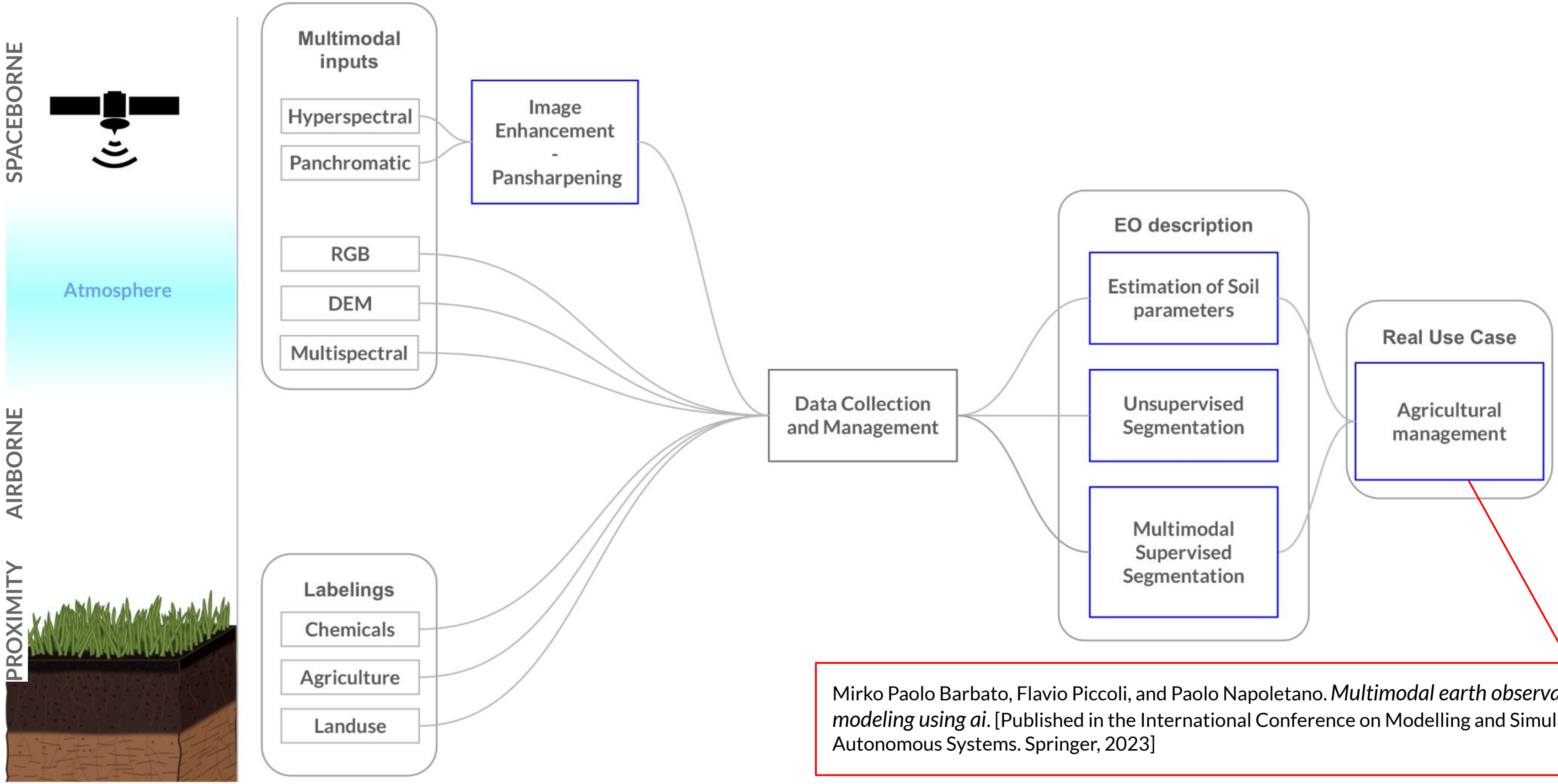
GT – OC

Visual results – Maps



Estimated – OC

Overview of our topics (selected)

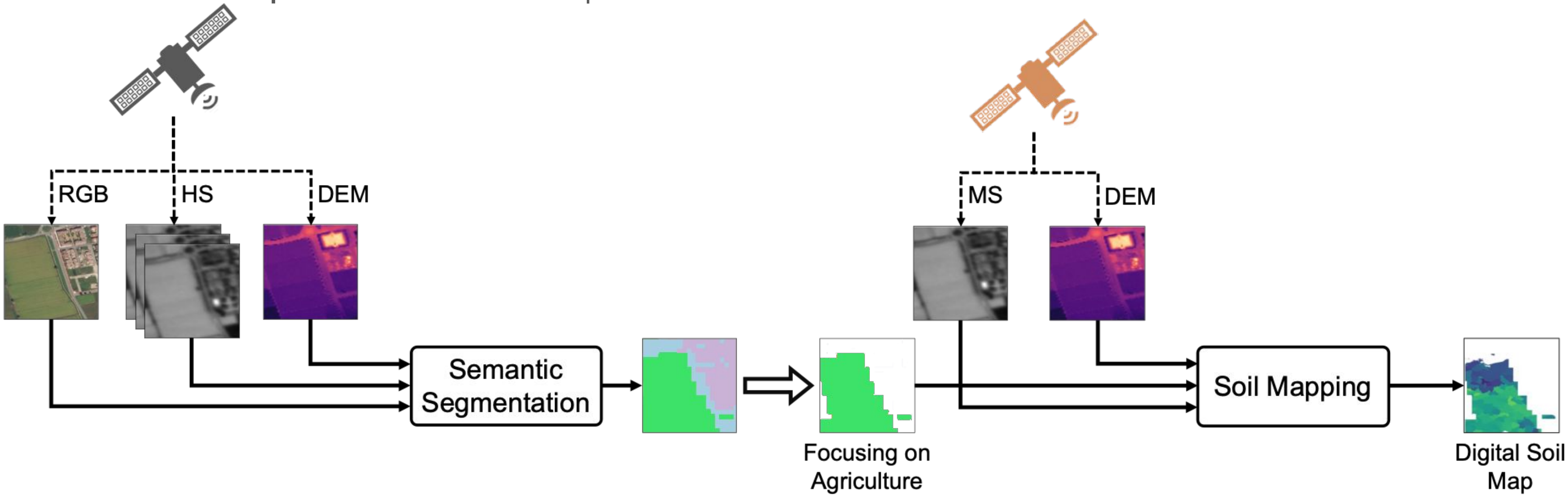


Mirko Paolo Barbato, Flavio Piccoli, and Paolo Napoletano. *Multimodal earth observation modeling using ai*. [Published in the International Conference on Modelling and Simulation for Autonomous Systems. Springer, 2023]

Estimation parameters of Agricultural areas

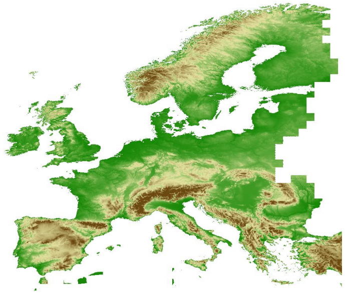
Soil textures play a determining role in water-holding capacity, drainage characteristics, nutrient retention, and susceptibility to erosion, influencing plant growth and agricultural productivity. The pHs affect nutrient availability and microbial activity in the soil, giving important information on the soil health.

1. **Semantic Segmentation** to identify agricultural areas
2. **Estimation of soil parameters**: textures and pHs



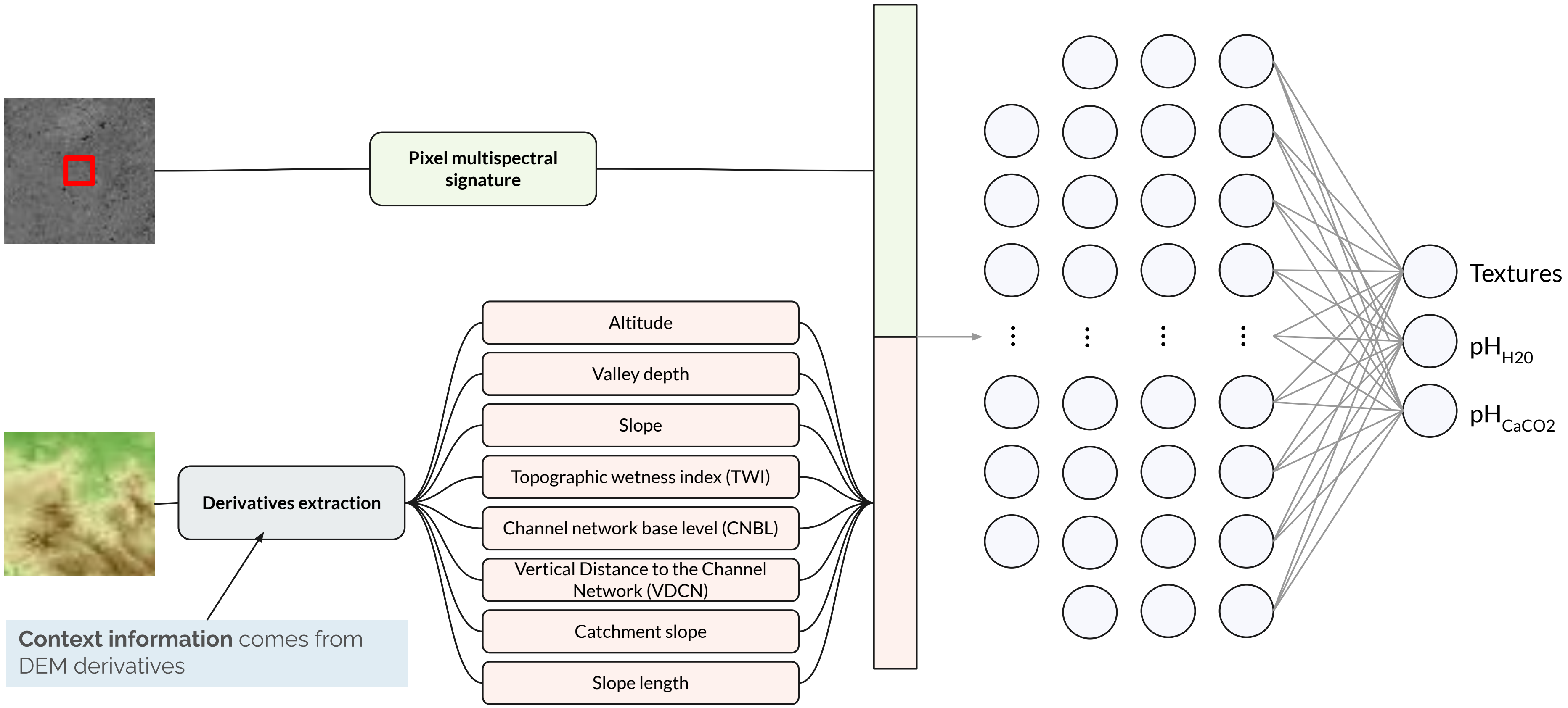
Multimodal dataset for Soil parameters estimation

1. **Soil parameters labeling**:
   - **Lucas** → 20,000 manually collected samples of textures and chemicals soil parameters
     - Textures → percentage of clay, silt, coarse, and sand
     - pH → CaCl2 and H2O)
2. **Multispectral information**:
   - **Sentinel-3** → 14 images covering Europe with a resolution of 300 m/px and 21 bands (400 nm to 1020 nm)
3. **Digital Elevation Model**:
   - **Copernicus project** → cover the Europe with a resolution of 8 m/px



Tóth, G.; Jones, A.; Montanarella, L. The LUCAS topsoil database and derived information on the regional variability of cropland
Sentinel 3 - [Sentinel-3 - Sentinel Online (esa.int)]

Multi estimation using Artificial Neural Network



Context information comes from DEM derivatives

Flavio Piccoli, Mirko Paolo Barbato, Marco Peracchi, and Paolo Napoletano. Estimation of soil characteristics from multispectral sentinel-3 imagery and dem derivatives using machine learning. Sensors, 23(18):7876, 2023

Results on the estimation of soil parameters

Results demonstrate the usefulness of **MS** and **DEM** information in the Land Cover scenario with an improvement of 0.07 R^2, 0.04 RMSE and 0.03 MAE

Results demonstrate the advantages of multimodal approaches compared with MS single modality

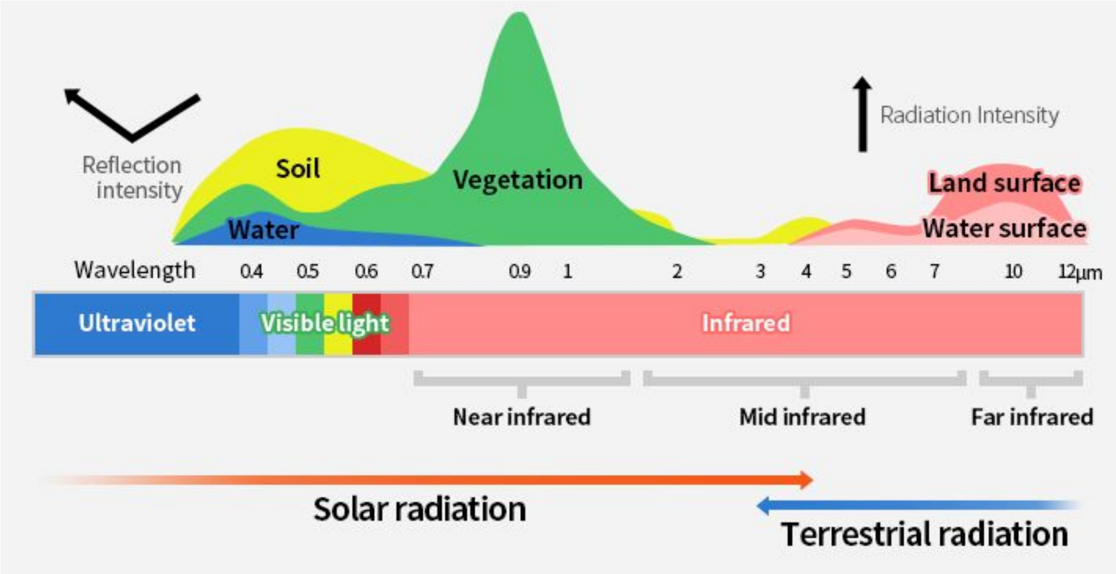
Parameter	Metric	Single modality (MS)	Multimodality (MS+DEM)
Textures	R^2 (↑)	0.27	**0.37**
	RMSE (↓)	0.86	**0.81**
	MAE (↓)	0.68	**0.64**
pH_{H_2O}	R^2 (↑)	0.51	**0.56**
	RMSE (↓)	0.70	**0.67**
	MAE (↓)	0.56	**0.53**
pH_{CaCl_2}	R^2 (↑)	0.50	**0.56**
	RMSE (↓)	0.71	**0.67**
	MAE (↓)	0.57	**0.53**
	R^2 (↑)	0.43	**0.50**
Overall	RMSE (↓)	0.76	**0.72**
	MAE (↓)	0.60	**0.57**

Questions?

Spectral response (wavelengths) to different materials

A **single sensor** may **not capture relevant characteristics** of a scene or object

Every **material** on earth shows its own strength of **reflection** in each **wavelength** when it is exposed to the **Electromagnetic** waves



Introduction to Remote Sensing, Lal Samarakoon Director, Geoinformatics Center, AIT Senior Scientist, JAX
https://earth.jaxa.jp/en/eo-knowledge/remote-sensing/index.html