AImage Lab

The color out of space: learning self-supervised representations for Earth Observation imagery

Stefano Vincenzi*, Angelo Porrello*, Pietro Buzzega*, Marco Cipriano*, Pietro Fronte‡, Roberto Cucco‡,

Carla Ippoliti†, Annamaria Conte†, Simone Calderara*

*University of Modena and Reggio Emilia, Italy

†Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise 'G.Caporale', Teramo, Italy

‡Progressive Systems Srl, Frascati – Rome, Italy

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

AImage Lab

Centro Interdipartimentale di Ricerca
Softech: ICT per le Imprese

Introduction

AImage Lab

Remote Sensing applications:

- Disaster prevention;
- Climate change;
- Vector-borne disease.

They require a higher number of satellitary images.

- ✓ Wide **increase in satellite missions**;
- ✗ **Lack of large annotated datasets,** acquiring ground truth data is expensive and requires expertise.

Current Approaches

AImage Lab

- models pre-trained on the **ImageNet** dataset
  - ✘ **limited** only on the tasks involving RGB images;
  - ✘ **satellitary images represent a different domain**.

Novel approaches tailored for satellitary images

- **In-domain representation learning**[1] involve pre-training on a different remote sensing dataset;
- **Tile2Vec**[2], rely on the assumption that spatially close tiles share similar information.

[1] M. Neumann et al. In-domain representation learning for remote sensing. arXiv preprint arXiv:1911.06721, 2019.
[2] N. Jean et al. Tile2vec: Unsupervised representation learning for spatially distributed data. In AAAI, 2019

Our Approach

AImage Lab

In this work, we propose**:**

- ✓ to learn meaningful representations from satellite images, **leveraging** its **high-dimensionality spectral bands to reconstruct the color channels (colorization);**
- ✓ to employ the learned representations on **two different tasks**: **land-cover classification and** prediction of the **presence/absence of the West Nile Virus**;
- ✓ an **ensemble model** between spectral and color channels.

Dataset

AImage Lab

- Images from satellites **Sentinel-2A** and **Sentinel-2B**, handled by ESA.

BigEarthNet

- Novel large-scale dataset[3] collecting **590 326 tiles**;
- Each example comprises of **12 bands (RGB included)** and multiple land-cover classes as ground truth;
- We adopt a class nomenclature involving **19 classes**[4];
- We **discards 70 987 patches** that are fully covered by clouds and snow.

	Bands	*Wavelength (μm)*	*Res (m)*
B_1	Coastal Aerosol	0.443	60
$B_{2,3,4}$	BGR channels	0.490	10
B_5	Vegetation Red Edge	0.705	20
B_6	Vegetation Red Edge	0.740	20
B_7	Vegetation Red Edge	0.783	20
B_8	NIR	0.842	10
B_{8A}	Vegetation Red Edge	0.865	20
B_9	Water Vapour	0.945	60
B_{10}	SWIR (Cirrus)	1.375	60
B_{11}	SWIR	1.610	20
B_{12}	SWIR	2.190	20

[3]G. Sumbul et al. Bigearthnet: A large-scale benchmark archive for remote sensing image understanding. In IGARSS, 2019.

[4] G. Sumbul et al. Bigearthnet deep learning models with a new class-nomenclature for remote sensing image understanding. arXiv preprint arXiv:2001.06372, 2020.

Dataset

AImage Lab

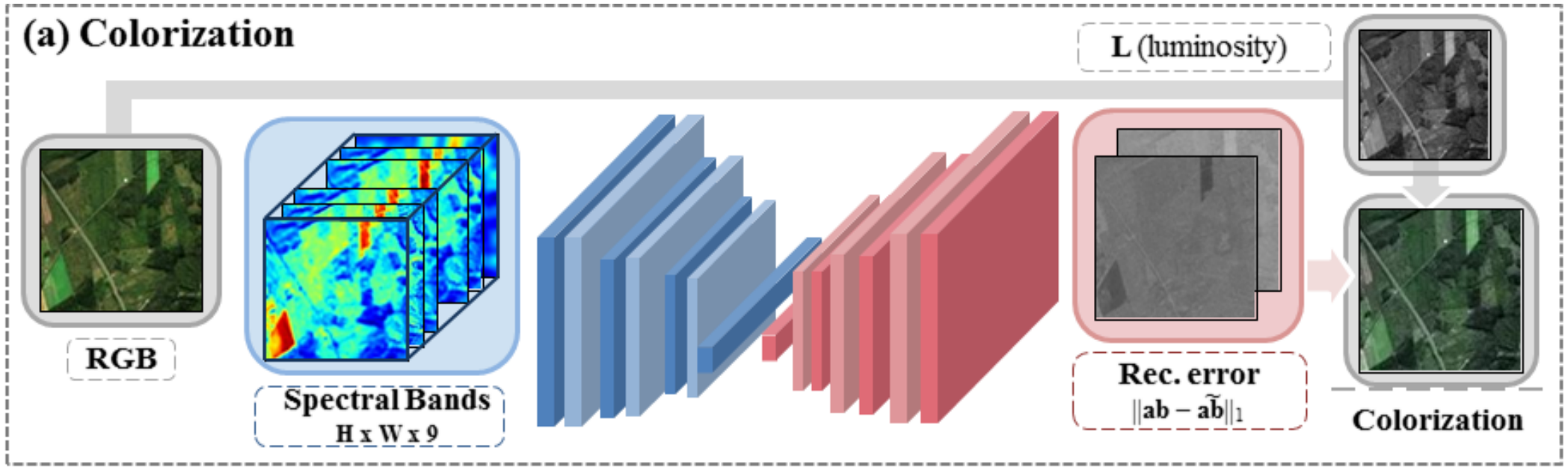
West Nile Disease Dataset (WND)

- a mosquito-borne disease caused by West Nile virus (WNV);
- **Sentinel 2A/2B** data paired with ground truth WND data;
- WND dataset for the **year 2018** comprises of **1 488 distinct cases**, divided into **962 negatives** and **526 positives**;
- Each case comes with a variable number of Sentinel-2 patches, thus leading to **18 684 spectral images in total**.

Model

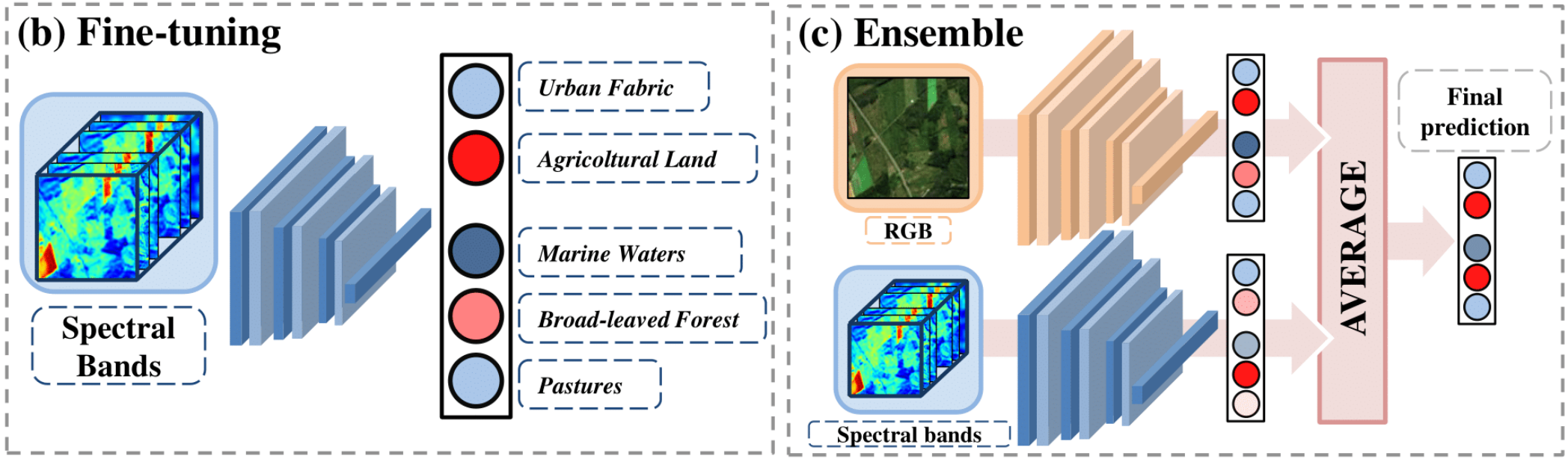
AImage Lab

- Our main goal consists in finding a **good initialization** for the network, such that it can later capture meaningful patterns even in presence of a **few labeled data**.
- We devises a two-stage procedure:
  a) **Self-supervised Colorization**;
  b) **Fine-tuning** on **Land-Cover** classification and **WND** prediction.



Model

- Once the encoder-decoder has been trained, we exploit **only the encoder as a pre-trained feature extractor**, adding a linear layer to maps the bottleneck features to the classification output space.
- The **ensemble** (c) is formed by **two independent branches taking the RGB channels and the spectral bands as input respectively**.



Land-Cover Classification

- The results are expresses in terms of **Mean-Average-Precision** (mAP);
- **We considered different dataset sizes**, ranging from 5,000 to 519,000 samples;
- Both ImageNet and colorization features lead to remarkable improvements;
- We assessed an **ensemble model, which brought to better results in all different ranges**;
- Finally, in the last table, we compare our best model with various baseline[4].

Input	pre-training	1k	5k	10k	50k	Full
RGB	from scratch	.486	.608	.645	.744	.851
RGB	ImageNet	.620	.695	.726	.786	**.879**
Spectral	from scratch	.555	.667	.711	.767	.866
Spectral	Color. (our)	**.622**	**.730**	**.760**	**.793**	.860

Input	pre-training	1k	5k	10k	50k	Full
RGB	ImageNet	.620	.695	.726	.786	.879
Spectral	Colorization	.622	.730	.760	.793	.860
Ensemble	Color.+ImageNet	**.656**	**.751**	**.778**	**.823**	**.896**

Method	pr.	rc.	F1
K-Branch CNN	.716	**.789**	.727
VGG19	.798	.767	.759
ResNet-50	.813	.774	.771
ResNet-101	.801	.774	.764
ResNet-152	.817	.762	.765
Ensemble (our)	**.843**	.781	**.811**

[4] G. Sumbul et al. Bigearthnet deep learning models with a new class-nomenclature for remote sensing image understanding. arXiv preprint arXiv:2001.06372, 2020.

West Nile Disease

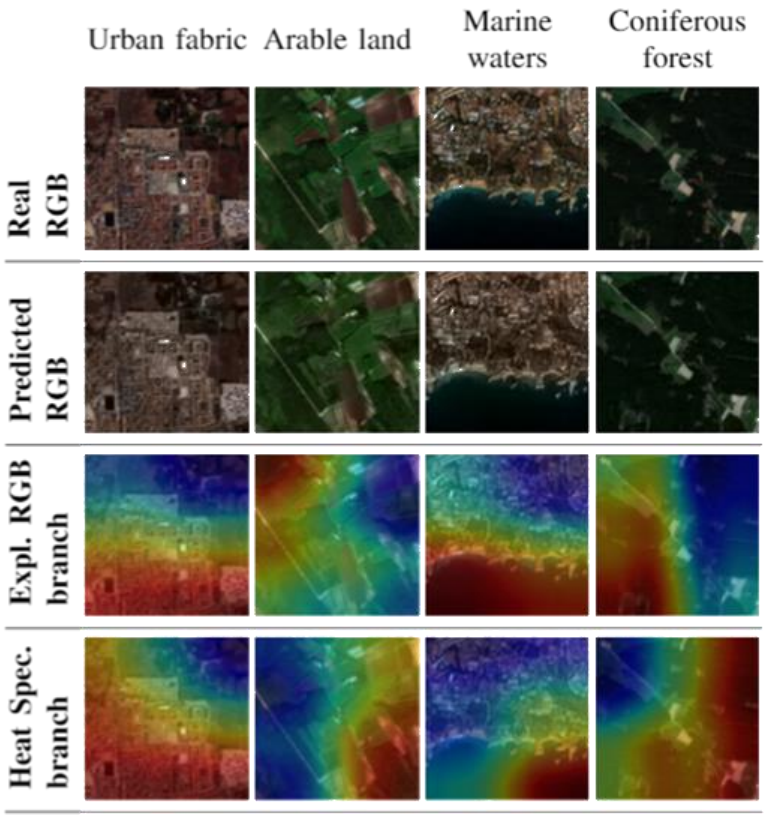
AImage Lab

- We define a simple baseline (**random classifier**) that computes predictions by randomly guessing from the class-prior distribution of the training set;
- **All the networks we trained exceed random guessing**, hence suggesting they effectively learned meaningful features;
- **The ensemble model** surpasses networks based on a single modality by a consistent margin.

Input	pre-training	acc.	pr.	rc.	F1
Random classifier	-	.503	.391	.395	.393
RGB	from scratch	.652	.542	.941	.688
RGB	ImageNet	.865	.819	.857	.838
$B_{1,8A,11,12}$	from scratch	.756	.662	.817	.732
$B_{1,8A,11,12}$	Colorization	.852	.823	.811	.817
Ensemble	Color.+ImageNet	**.880**	**.855**	**.850**	**.852**

Why does the ensemble work better?

- **diversity among individual learners**;
- **GradCam** to assess whether the two branches look for different portions within their inputs.
- The third and fourth rows of the figure highlight the input regions that have been considered important for predicting the target category.
- **The regions highlighted from the two branches visually diverge**, confirming the weak correlation between their representations.



Conclusion

AImage Lab

- In this work, we propose a self-supervised learning approach for satellitary images;
- We prepend a colorization phase to a fine-tuning one on downstream tasks;
- We observe that the initialization we devised leads to remarkable results, exceeding the baselines especially in presence of scarce labeled data;
- We qualitatively observe that representations learned through colorization are different from the ones driven by the RGB channels. Based on this finding, we set up an ensemble model that achieves the highest results.

AImage Lab

Thanks!

Source code: https://github.com/stevinc/TheColorOutOfSpace