Contrastive Learning for Regression on Hyperspectral Data

Mohamad DHAINI [1,2], Maxime BERAR [1], Paul HONEINE [1], Antonin VAN EXEM [2]

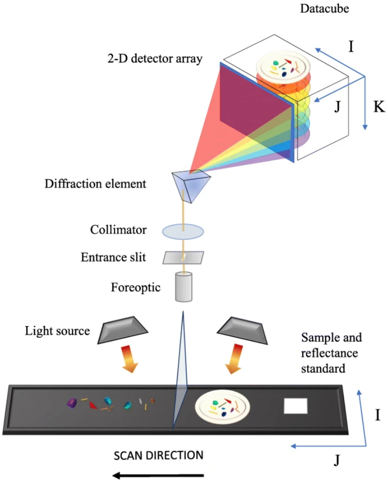
[1] *LITIS Lab, Université de Rouen Normandie, Saint-Etienne-du-Rouvray, France*
[2] *TELLUX Company, Maromme, France*

April 2024

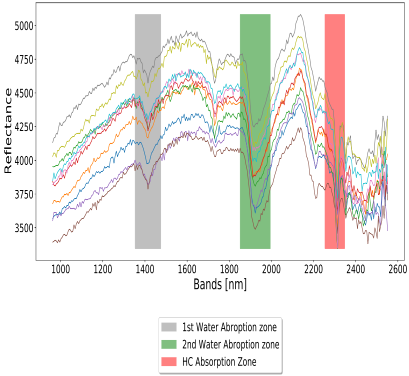
Outline

Hyperspectral Data

Hyperspectral Sensor:

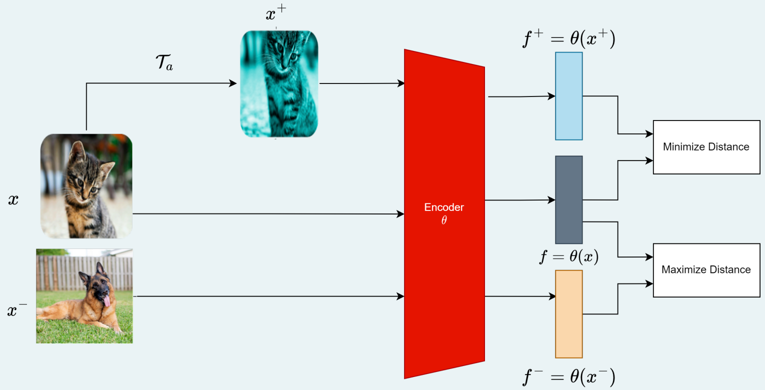


Spectral Data:



Contrastive Learning

How does Contrastive Learning Work ?



Objectives

- Contrastive Learning Framework for regression on Hyperspectral Data
- Augmentation techniques adequate for Hyperspectral Data
- Validation on Synthetic and Real Hyperspectral Datasets

Proposed Architecture

Proposed Method

- Hyperspectral image $X = [x^1(\lambda), x^2(\lambda), \ldots, x^N(\lambda)]^\top \in \mathbb{R}^{N \times b}$, where λ is the wavelength with b total number of wavelengths.
- $\widetilde{X} = \Phi_{transform}(X)$
- $\widetilde{F} = \Phi_w\left(\widetilde{X}\right)$ and $F = \Phi_w(X)$
- $\hat{Y} = g_\theta(F) \in \mathbb{R}^{N \times s}$

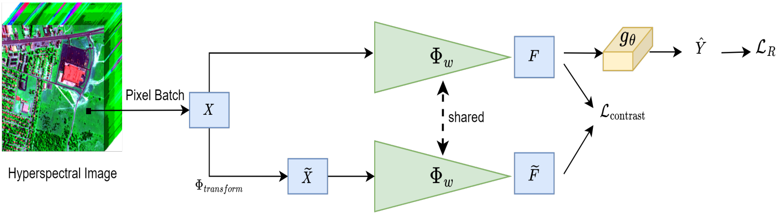


Figure: Architecture of the proposed method.

Spectral Augmentation

Challenges:

- Incompatibility of geometric transformations, noise injection, and color distortions with spectral domain
- Used transformations should not create strong deformations to the original spectrums

$\longrightarrow$ We propose a set of 8 augmentations technique adequate for spectral domain.

Spectral Augmentation

1. The Spectral Shift involves shifting the spectrum in the wavelength, such as

$$\widetilde{x}(\lambda) = x(\lambda - \Delta) \tag{1}$$

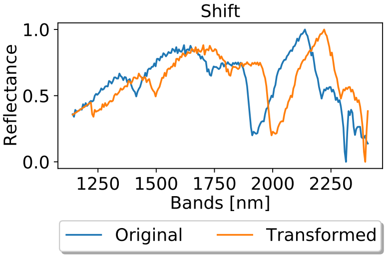


Figure: Spectral Shift Transformation.

Spectral Augmentation

2. The Spectral Flipping involves reversing the order of spectral bands in a spectrum according to the following:

$$\widetilde{x}(\lambda) = x(\lambda_{min} + \lambda_{max} - \lambda). \quad (2)$$

λ_{min} and λ_{max}: minimum and maximum wavelengths in the spectrum, respectively.

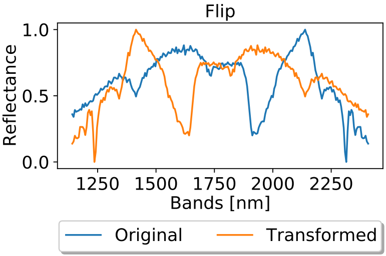


Figure: Spectral Flipping.

Spectral Augmentation

3. The Scattering Hapke's Model (Hapke 1981) :

$$\widetilde{x} = \frac{\omega}{(1 + 2\mu_1\sqrt{1-\omega})(1 + 2\mu_2\sqrt{1-\omega})}, \tag{3}$$

$$\omega = 1 - \left(\frac{\sqrt{\mu_0^4 x^2 + (1 + 4\mu_0^2 x)(1 - x)} - 2\mu_0 x}{1 + 4\mu_0^2 x} \right)^2 \tag{4}$$

ω: single scattering albedo of the material
μ_1: cosine of the angle between incoming radiation and the normal to the surface
μ_0: initial cosine angle of the incoming radiation.

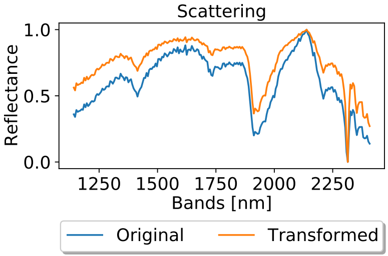


Figure: Spectral Scattering.

Spectral Augmentation

4. The Atmospheric Compensation model (Uezato et al. 2016)

$$\widetilde{x} = x \frac{E_{\text{sun-gr}}\, \mu_1 + E_{\text{sky}}}{E_{\text{sun-gr}}\, \mu_2 + E_{\text{sky}}}, \tag{5}$$

$E_{\mathrm{sun-gr}}$: solar radiance observed at the ground level
E_{sky}: denotes the skylight.
μ_1 and μ_2: cosines of the angles between the surface normal and the direction of the sun at each pixel and at the calibration panel, respectively.

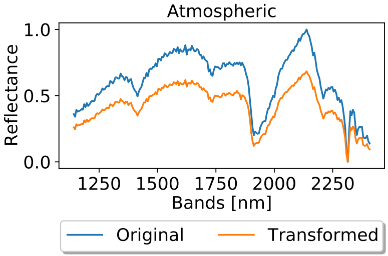


Figure: Example of Atmospheric Transformation.

Spectral Augmentation

5. The Elastic Distortion consists in a displacement grid on the wavelength axis, such as

$$\widetilde{x}(\lambda) = x(\lambda + \epsilon(\lambda)) = x\left(\lambda + \sum_{i=1}^{N_G} A_i e^{-\frac{(\lambda-\lambda_i)^2}{2\sigma^2}}\right) \tag{6}$$

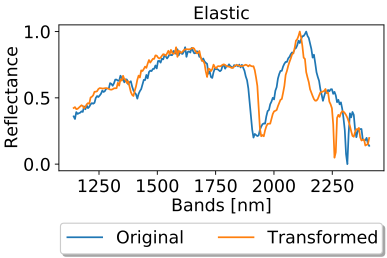


Figure: Example of Elastic Transformation.

Spectral Augmentation

6. The Band Erasure (Hu et al. 2021) randomly removing certain wavelength from the spectral data.
7. The Band Permutation (Hu et al. 2021) involves randomly permuting the order of the spectral bands.
8. The Nearest Neighbor (Wang et al. 2023) involves creating new synthetic samples based on the average of the closest samples.

Training Loss

Defining Positive Pairs:

Define for the i-th sample a ball $\mathbf{B}^i$ of radius r where the positive pair j is selected as the following:

$$r \geq \left\| y^i - y^j \right\|_2 . \tag{10}$$

Cross Entropy Based Loss:

The common contrastive loss used in most recent work is based on the cross entropy, which can be written as

$$\mathcal{L}_{Contrastive} = -\frac{1}{N} \sum_{i=1}^{2N} \sum_{j \in \mathbf{B}^i} \log \frac{\exp\left(\mathrm{sim}\left(f^i, f^j\right)/\tau\right)}{\sum_{k \notin \mathbf{B}^i} \exp\left(\mathrm{sim}\left(f^i, f^k\right)/\tau\right)} \tag{11}$$

$\mathrm{sim}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u}^T \boldsymbol{v}/(\|\boldsymbol{u}\|\|\boldsymbol{v}\|)$: cosine similarity between two vectors, and τ is a temperature scalar.

Training Loss

Regression Loss:

For training, the contrastive loss is combined with a standard mean squared error regression loss according to the following:

$$\mathcal{L}_{\mathrm{R}} = \frac{1}{N} \sum_{i=1}^{N} \left\| y^i - g_\theta \left(f^i \right) \right\|^2 \tag{12}$$

$$\mathcal{L}_{total} = \mathcal{L}_{\mathrm{R}} + \alpha \, \mathcal{L}_{\mathrm{Contrastive}} \tag{13}$$

Results on Synthetic Data

Synthetic Data:

- Four random spectrums from USGS digital spectral library (Swayze et al. 1993)
- Each spectrum is composed of 224 contiguous bands
- A total of 100×100 mixed pixels were generated with abundances following a Dirichlet distribution.

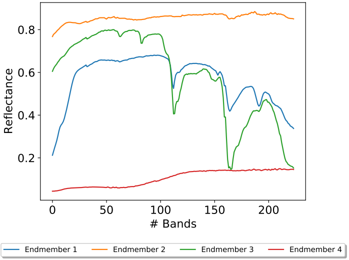


Figure: USGS Spectrums.

Results on Synthetic Data

Synthetic Data:

- Four random spectrums from USGS digital spectral library (Swayze et al. 1993)
- Each spectrum is composed of 224 contiguous bands
- A total of 100 $\times$ 100 mixed pixels were generated with abundances following a Dirichlet distribution.

	R^2	MAE
Baseline (No Contrastive)	0.55 ± 0.004	0.073 ± 0.07
Band erasure (Hu et al. 2021)	0.62 ± 0.003	0.064 ± 0.05
Band Permutation (Hu et al. 2021)	0.63 ± 0.003	0.063 ± 0.05
Nearest Neighbor (Wang et al. 2023)	0.61 ± 0.004	0.065 ± 0.05
Scattering	0.64 ± 0.004	0.061 ± 0.06
Atmospheric	0.65 ± 0.004	0.059 ± 0.06
Flipping	0.62 ± 0.005	0.062 ± 0.07
Elastic	0.66 ± 0.003	0.058 ± 0.05
Shift	0.75 ± 0.003	0.053 ± 0.05

Table: Regression Results on Synthetic Data.

Results on Synthetic Data

Synthetic Data:

- Four random spectrums from USGS digital spectral library (Swayze et al. 1993)
- Each spectrum is composed of 224 contiguous bands
- A total of 100×100 mixed pixels were generated with abundances following a Dirichlet distribution.

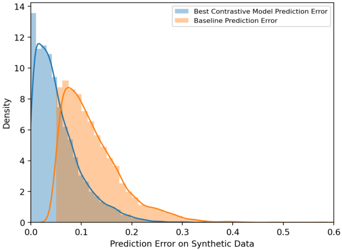


Figure: Error Distribution on Synthetic Dataset.

Results on Real Data

Real Data

- Datasets provided by Tellux company, specialized in soil pollution analysis using hyperspectral imaging (Dhaini et al. 2021)
- 10000 spectra with a spectral range [1130-2450 nm] containing hydrocarbon pollution concentration ranging [0-20000 mg/kg]

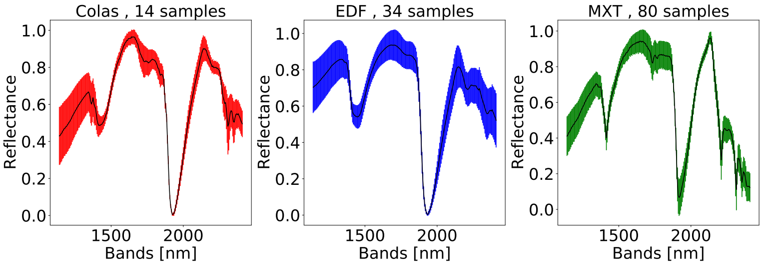


Figure: Tellux Datasets.

Results on Real Data

Real Data

- Datasets provided by Tellux company, specialized in soil pollution analysis using hyperspectral imaging (Dhaini et al. 2021)
- 10000 spectra with a spectral range [1130-2450 nm] containing hydrocarbon pollution concentration ranging [0-20000 mg/kg]

	R^2	MAE
Baseline (No Contrastive)	0.45 ± 0.002	2274.04 ± 21.2
Band erasure (Hu et al. 2021)	0.54 ± 0.003	1620.22 ± 18.1
Band Permutation (Hu et al. 2021)	0.53 ± 0.003	1700.04 ± 20.5
Nearest Neighbor (Wang et al. 2023)	0.54 ± 0.003	1850.40 ± 34.5
Scattering	0.56 ± 0.003	1737.26 ± 18.2
Atmospheric	0.55 ± 0.002	1796.63 ± 16.1
Elastic	0.58 ± 0.002	1709.33 ± 18.3
Flip	0.59 ± 0.002	1708.52 ± 18.4
Shift	0.59 ± 0.002	1380.37 ± 16.5

Table: Regression Results on Real Data.

Results on Real Data

Real Data

- Datasets provided by Tellux company, specialized in soil pollution analysis using hyperspectral imaging (Dhaini et al. 2021)
- 10000 spectra with a spectral range [1130-2450 nm] containing hydrocarbon pollution concentration ranging [0-20000 mg/kg]

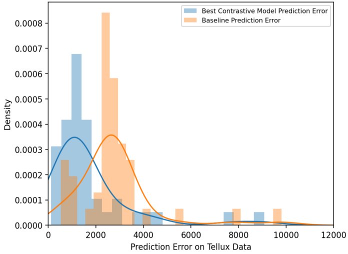


Figure: Error Distribution on Tellux Data.

Combination Study

Methodology

- Combine several transformations to increase robustness of feature extractor
- To reduce the number of possible combinations of proposed transformations, we propose an incremental evaluation technique
- We start by taking the transformation that provided the best result (shift transformation) and then we do all the 2-element combinations.

Combination Study

	R^2	ΔR^2
Shift	0.7522	–
Shift + Atmospheric	0.7774	0.0252
Shift + Atmospheric + Scattering	0.7921	0.0147
Shift + Atmospheric + Scattering + Elastic	0.7922	0.0001

Table: Combination Study Results on Synthetic Data.

	R^2	ΔR^2
Shift	0.59000	–
Shift + Atmospheric	0.60639	0.01639
Shift + Atmospheric + Elastic	0.61791	0.01152
Shift + Atmospheric + Elastic + Scattering	0.61793	0.00002

Table: Combination Study Results on Real Data.

Conclusion

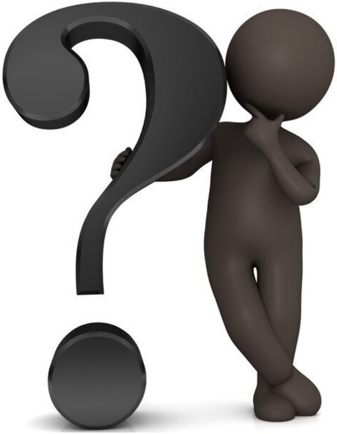
Conclusions and Future Work

- ★ We highlighted the ability of using contrastive learning for regression tasks on hyperspectral data
- ★ We proposed a set of spectral transformations adequate for hyperspectral data

Future Work

- ★ Combining the presented framework with domain adaptation frameworks to generalize knowledge to unseen domains.

THANK YOU FOR LISTENING !



References

Dhaini, Mohamad et al. (2021). "Hyperspectral imaging for the evaluation of lithology and the monitoring of hydrocarbons in environmental samples". In: *RemTech (International event on Remediation, Coasts, Floods, Climate, Seismic, Regeneration Industry)*.

Hapke, Bruce (1981). "Bidirectional reflectance spectroscopy: 1. Theory". In: *Journal of Geophysical Research: Solid Earth* 86.B4, pp. 3039–3054.

Hu, Xiang et al. (2021). "Deep spatial-spectral subspace clustering for hyperspectral images based on contrastive learning". In: *Remote Sensing* 13.21, p. 4418.

Swayze, GA et al. (1993). "The US Geological Survey, Digital Spectral Library: Version 1: 0.2 to 3.0 mum". In: *Bulletin of the American astronomical society*. Vol. 25, p. 1033.

Uezato, Tatsumi et al. (2016). "A novel endmember bundle extraction and clustering approach for capturing spectral variability within endmember classes". In: *IEEE Transactions on Geoscience and Remote Sensing* 54.11, pp. 6712–6731.

Wang, Meng et al. (2023). "Nearest Neighbor-Based Contrastive Learning for Hyperspectral and LiDAR Data Classification". In: *IEEE Transactions on Geoscience and Remote Sensing*.

6. The Band Erasure (Hu et al. 2021) randomly removing certain wavelength from the spectral data.

$$\tilde{x}(\lambda) = \begin{cases} 0, & \text{with probability } p \\ x(\lambda), & \text{with probability } 1 - p \end{cases} \quad (7)$$

where p is the probability of erasing a band.

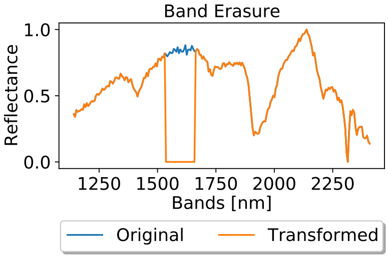


Figure: Example of Band Erasure.

7. The Band Permutation (Hu et al. 2021) involves randomly permuting the order of the spectral bands.

$$\widetilde{x}(\lambda) = x(\pi(\lambda)), \tag{8}$$

where $\pi(\lambda)$ is a random permutation function.

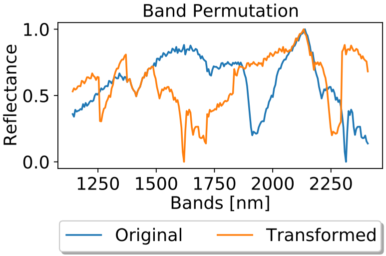


Figure: Example of Band Permutation.

8. The Nearest Neighbor (Wang et al. 2023) involves creating new synthetic samples based on the average of the closest samples.

$$\widetilde{x} = \frac{1}{k} \sum_{x^i \in \mathcal{B}(x,r)} x^i, \tag{9}$$

where $\mathcal{B}(x, r)$ is the set of k closest points to x within the radius r.

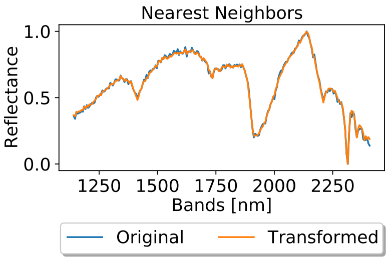


Figure: Example of Nearest Neighbor Transformation.