Locality-Aware Hyperspectral Classification

Fangqin Zhou [1] Mert Kilickaya [1] Joaquin Vanschoren [1]

[1]Automated Machine Learning Group of Eindhoven University of Technology

Motivation

Limitations of existing methods:

- Hyperspectral Imaging (HSI) provides fine-grained information that is not typically available in conventional RGB images, which has led to breakthroughs in various industries.
- Classic machine learning methods and deep learning models have shown limitations in modeling long-range dependencies across spatial-spectral dimensions of HSI.
- Recent research based on vision transformers for HSI focuses solely on spectral information and lacks attention to the spatial locality.

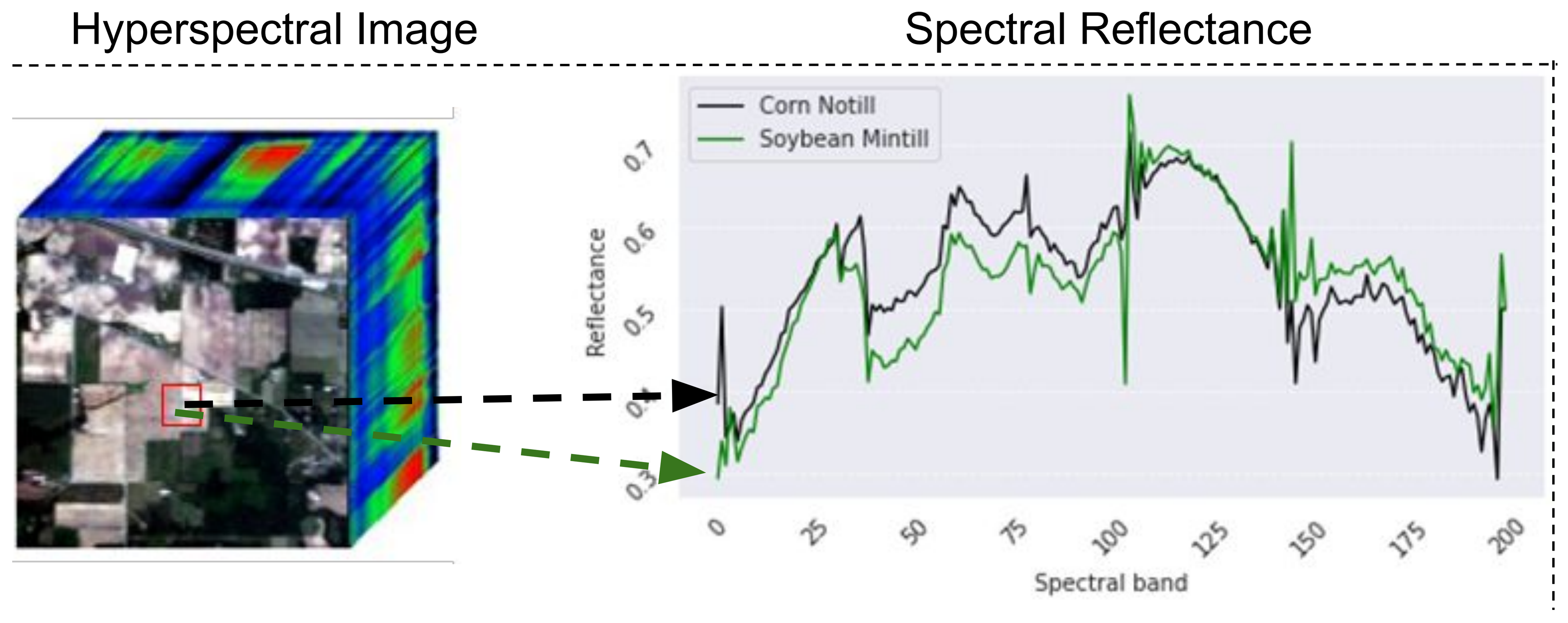

Figure 1. An example of Indian Pines dataset. A Hyperspectral image cube, with the spectral signals of two pixels from separate categories ('notill' and 'mintill').

Main contributions

In summary, our work makes three main contributions:

1. We propose **Hy**perspectral **L**ocality-aware **I**mage **T**ransform**E**r (HyLITE), a novel architecture that can model the local-spectral relationships in Hyperspectral data.
2. We equip HyLITE with a novel local-global regularization objective, to balance global and local spectral information.
3. We conduct experiments on three well-established benchmarks, and show that HyLITE significantly improves over the competitive SpectralFormer [1] baseline, across *all* benchmarks and metrics.

References

[1] Danfeng Hong, Zhu Han, Jing Yao, Lianru Gao, Bing Zhang, Antonio Plaza, and Jocelyn Chanussot. Spectralformer: Rethinking hyperspectral image classification with transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2021.

[2] Damian Ibanez, Ruben Fernandez-Beltran, Filiberto Pla, and Naoto Yokoya. Masked auto-encoding spectral–spatial transformer for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022.

Hyperspectral Locality-aware Image TransformEr

An overview of the proposed Hyperspectral Locality-aware Image TransformEr (HyLITE). *i). Preprocessing*: The input image is patchified, linearly projected, and appended with a classifier token and a positional embedding. *ii). Representation*: The input is processed by identical spectral and local multi-head attention (MHA) blocks. *iii). Classification*: At the end, the representation of the classifier token is mapped to a distinct category, such as $\{grass, road\}$. *iv). Regularization*: To further promote locality, we apply our novel regularization on top of the learned token representations.

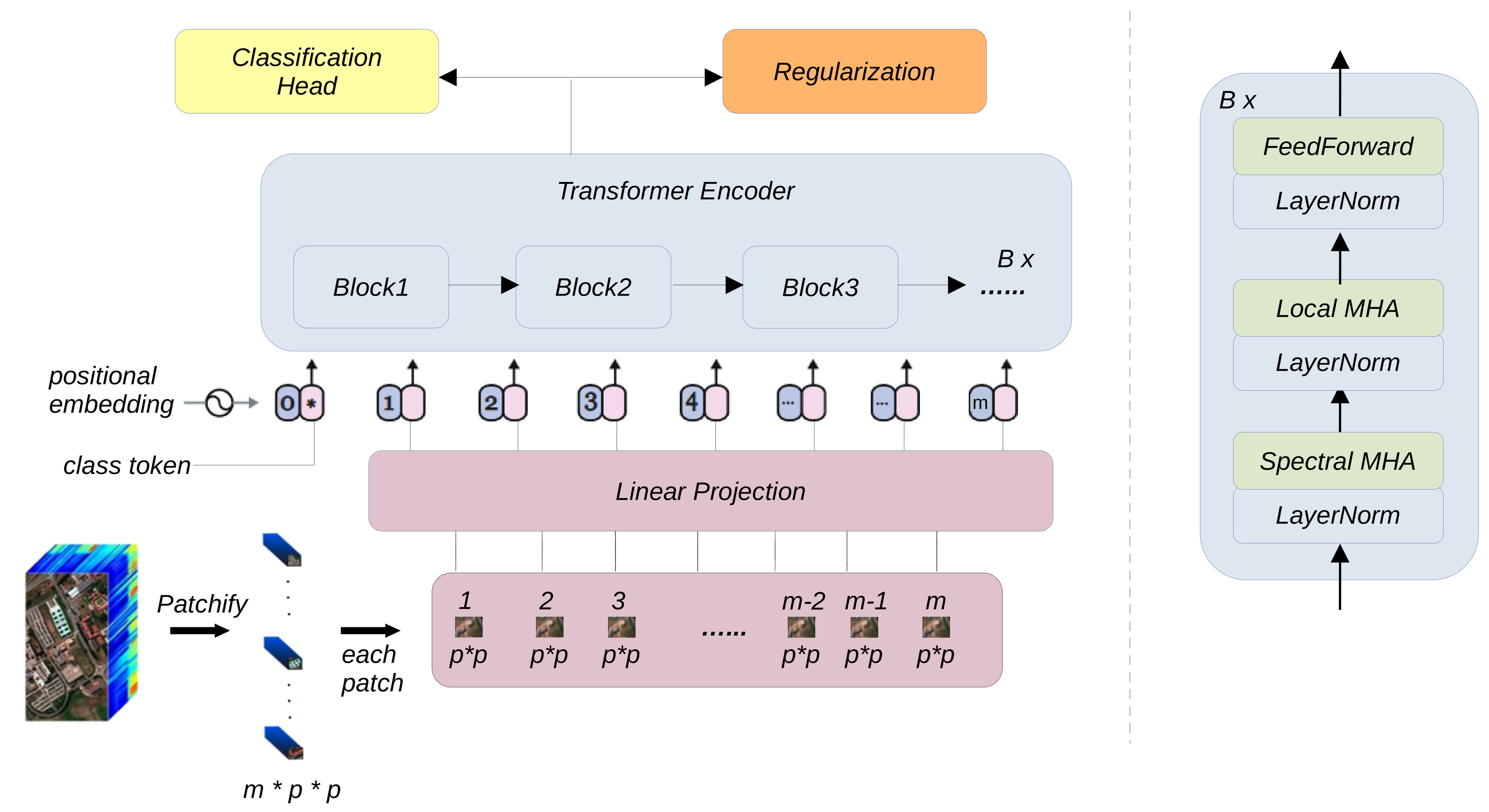

Figure 2. An overview of the proposed Hyperspectral Locality-aware Image TransformEr

The Objective

Our model aims to optimize the following loss function, in which CE(·) is the standard cross-entropy loss, Reg(·) is our novel local-global regularization objective, and λ attenuates the regularization strength.

$$\mathcal{O} = \mathrm{CE}(y, y\prime) + \lambda \cdot \mathrm{Reg}(X_B)$$

Regularization

To minimize the regularization loss, the global output token X_B^0 should be close to the center of the spectral tokens. Hence, the gradients will nudge the representations of the global and spectral tokens closer together, causing them to converge rather than diverge, and aggregating information from each other, thus incorporating globality in the learning process.

$$\mathrm{Reg}(X_B) = \left\| X_B^0 - \frac{1}{m}\sum_{i=1}^{m} X_B^i \right\|_2^2$$

Experimental Results

Our model consistently outperforms all techniques by a wide margin across all datasets and evaluation metrics. For example, in comparison to SpectralFormer, we improve the overall accuracy by 10.83% on Indian Pines, by 3.41% on Houston2013, and by 6.64% on Pavia University.

	IndianPines			*Houston2013*			*PaviaUniversity*		
	OA	*AA*	*Kappa*	*OA*	*AA*	*Kappa*	*OA*	*AA*	*Kappa*
kNN	59.17	63.90	0.54	77.30	78.28	0.75	70.53	79.68	0.62
RF	69.80	76.78	0.65	77.48	80.35	0.75	69.67	80.18	0.62
SVM	72.36	83.16	0.68	76.91	78.99	0.79	70.82	84.44	0.64
$1-DCNN$	70.43	79.60	0.66	80.04	82.74	0.78	75.50	86.26	0.69
$2-DCNN$	75.89	86.64	0.72	83.72	84.35	0.82	86.05	88.99	0.81
RNN	70.66	76.37	0.66	82.23	85.04	0.81	77.13	84.29	0.71
miniGCN	75.11	78.03	0.71	81.71	83.09	0.80	79.79	85.07	0.73
ViT	71.86	78.97	0.68	80.41	82.50	0.78	76.99	80.22	0.70
SpectralFormer	78.97	85.39	0.76	85.08	86.39	0.83	84.64	86.75	0.79
HyLITE(Ours)	89.80	94.69	0.88	88.49	89.74	0.87	91.28	92.25	0.88
Δ	10.83	9.30	0.12	3.41	3.35	0.03	6.64	5.50	0.08
MAEST[2]	82.12	87.63	0.79	83.61	84.89	0.82	87.20	89.91	0.83

Table 1. Comparison against the State-of-the-Art

Incorporating locality not only improves accuracy, but also improved sample efficiency of Hyperspectral imaging, which is promising for low-shot learning applications.

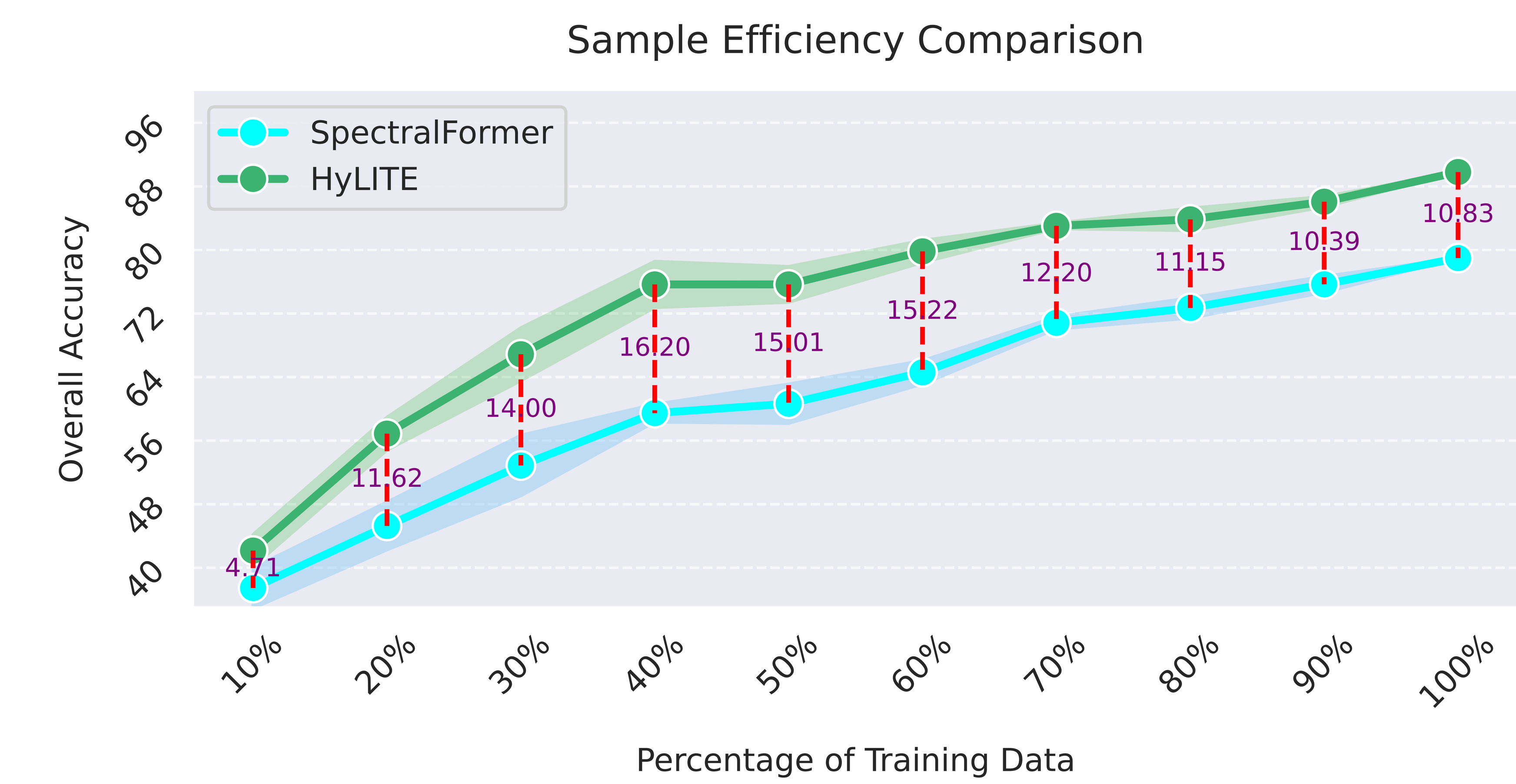

Figure 3. Comparing the sample efficiency of HyLITE and SpectralFormer [1] on Indian Pines.