An overview of the proposed Hyperspectral Locality-aware Image Transformer (HyLITE), block representation: The input image is patchified, linearly projected, and appended with a classifier token and a positional embedding. iii). Classification: At the end, the representation of the classifier token maps to a distinct category, such as [green, red], iv). Regularization: To further promote locality, we apply a 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 α attenuates the regularization strength.

\[ \mathcal{L} = CE(y, y^\prime) + \lambda \cdot Reg(X_p) \]

Regularization

To minimize the regularization loss, the global output token \( X_p \) 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 locality in the learning process.

\[ Reg(X_p) = \left\| X_p - \sum_{i \in \text{spectral tokens}} X_i \right\|_2^2 \]

In summary, our work makes three main contributions:

1. We propose Hyperspectral Locality-aware Image Transformer (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.

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 10.0% on Indian Pines, by 3.4% on Houston 2013, and by 6.6% on Pavia University.

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.