

Supplementary Materials: *HINER: Neural Representation for Hyperspectral Image*

Anonymous Authors

1 CLASSIFICATION ON COMPRESSED HSI

1.1 Adaptive Spectral Weighting (ASW)

Architecture. ASW consists of two modules: WeightMLP \mathcal{W} and ConvMLP \mathcal{M} , in which the output $I_c \in \mathbb{R}^{N \times H \times W}$ maintains the original shape. Next, let's ignore the residual for simplicity, and this process can be abbreviated as:

$$I_c = \mathcal{M}(\mathcal{W}(\hat{I})). \quad (1)$$

WeightMLP generates an n -dimensional vector $W \in \mathbb{R}^{N \times 1}$ using a small MLP. The vector W is then utilized to weight the HSI spectral-wisely. The purpose of this step is to adaptively emphasize or de-emphasize certain spectral bands. Assuming $\hat{I} \in \mathbb{R}^{N \times H \times W}$, the output P of WeightMLP can be written as:

$$P = \hat{I} \odot W = \begin{bmatrix} \hat{I}_1 \\ \hat{I}_2 \\ \dots \\ \hat{I}_n \end{bmatrix} \odot \begin{bmatrix} W_1 \\ W_2 \\ \dots \\ W_n \end{bmatrix} = \begin{bmatrix} W_1 \hat{I}_1 \\ W_2 \hat{I}_2 \\ \dots \\ W_n \hat{I}_n \end{bmatrix} = \begin{bmatrix} P_1 \\ P_2 \\ \dots \\ P_n \end{bmatrix} \in \mathbb{R}^{N \times H \times W}. \quad (2)$$

Then P is passed to ConvMLP \mathcal{M} comprising 1×1 conv to aggregate cross-spectral information. Let $A \in \mathbb{R}^{N \times M \times 1 \times 1}$ and $B \in \mathbb{R}^{M \times N \times 1 \times 1}$ represent the two convolution layers used in the \mathcal{M} , respectively:

$$A = [\alpha_1, \alpha_2, \dots, \alpha_n]^T, \forall \alpha \in \mathbb{R}^{M \times 1}; \\ B = [\beta_1, \beta_2, \dots, \beta_n], \forall \beta \in \mathbb{R}^{1 \times M}. \quad (3)$$

Considering the λ -th band of output I_c , it can be written as

$$I_c[\lambda, :, :] = \alpha_1^T \beta_\lambda^T P_1 + \alpha_2^T \beta_\lambda^T P_2 + \dots + \alpha_n^T \beta_\lambda^T P_n \quad (4)$$

Combining Eq. 1, it can be found that ASW first spectral-wisely re-weight the reconstructed HSI by multiplying learned vector W , and then aggregate cross-spectral information.

Optimization. By employing ASW, the optimization of \hat{I} is converted into the optimization of network parameters. This conversion can be readily accomplished through gradient descent techniques. Then the input of classification network becomes the output of ASW I_c , by which the $u(\hat{I}) = ||I - \hat{I}||$ is translated to $u(\hat{I}) = ||I - I_c||$ to constrain classifier's input. We relax the constraint to prevent the necessity of introducing ground truth,

$$u(I) = ||I - I_c|| \approx ||\hat{I} - I_c|| = ||\hat{I} - \mathcal{SAW}(\hat{I})||. \quad (5)$$

Given the condition $||I - \hat{I}|| < 10^{-3}$, this relaxation holds valid. Finally, our optimization objective can be expressed as the amalgamation of the classification loss \mathcal{L}_C and the reconstruction loss \mathcal{L}_R , as described in Eq. (8) of the main paper:

$$\arg \min \mathcal{L}_C + \beta \cdot \mathcal{L}_R(\hat{I}, I_c) \quad (6)$$

1.2 Implicit Spectral Interpolation (ISI)

Data Augmentation has shown promise for training robust deep neural networks against unforeseen data bias or corruptions [1, 2]. Intuitively, augmented samples encourage perturbing the underlying source distribution to enlarge predictive uncertainty of the current model, so that the generated perturbations can improve the model generalization during training. One intuitive manifestation of generalization is the flatness of the loss landscape. As described in the main paper, a flatter loss landscape, indicative of better generalization, exhibits relatively small loss changes under parameter perturbations, whereas a sharp loss landscape indicates otherwise.

We propose a simple yet effective strategy, Implicit Spectral Interpolation, to augment training samples, thereby facilitating improved performance on compressed HSI,

$$S = \sum \mathcal{HINER}(\lambda + U(-\eta, \eta)), \quad (7)$$

where $U(-\eta, \eta)$ represents a uniform distribution that adds random variables to λ to generate diverse reconstructed (perturbed) samples. In addition, we randomly disable and enable the spectral interpolation of the wavelengths in each forward pass, like [3]:

$$\eta = \begin{cases} 0 & \text{with probability } p \\ 0.1 & \text{with probability } 1 - p \end{cases}. \quad (8)$$

Here we use $\eta = 0.1$ and $p = 0.5$.

2 EXPERIMENTS

2.1 Experimental Setup

2.1.1 Datasets.

- **Indian Pines** is collected by the AVIRIS sensor over the Indian Pines Proving Ground in northwestern Indiana, used for compression and classification purposes. It consists a scene of 145×145 pixels with 224 spectral bands spanning the wavelength range of 400-2500 nm. This scene is a subset of a larger scene. The Indian Pines scene predominantly consists of two-thirds agriculture and one-third forest or other perennial natural vegetation. Additionally, there are two major two-lane highways, a railroad line, and some low-density housing areas, along with other buildings and smaller roads. Sixteen classes are labeled (e.g., corn, grass, soybean, woods, and so on), with some classes being very rare (fewer than 100 samples for alfalfa or oats). After removing noisy bands, the number of bands is reduced to 200: [104-108], [150-163], 220. Despite its limited size, this dataset serves as one of the main reference datasets in the community. A graphical representation of a sample from this dataset is presented in Fig. 1(a).
- **Pavia University** is captured by the ROSIS sensor in Pavia, Northern Italy, with the purpose of compression and classification. The image dimensions are 610×340 pixels, and it comprises 103 spectral bands. The image has been segmented



(a) Indian Pines



(b) Pavia University



(c) Pavia Centre



(d) CHILD

Figure 1: Dataset visualization

into 9 distinct classes, including asphalt, meadows, gravel, trees, metal sheet, bare soil, bitumen, brick, and shadow.

- **Pavia Centre** is a 1096×715 pixels image, where the number of spectral bands is 102. The geometric resolution is 1.3 meters. Image differentiates 9 classes each, including water, trees, asphalt, self-blocking bricks, bitumen, tiles, shadows, meadows, and bare soil.
- **CHILD** [4] comprises 141 HSI images captured by the PMVIS system, which measures 145 spectral samples ranging from 450 nm to 950 nm. The spatial resolution of each image is 960 × 1056 pixels. In this paper, we selected one HSI image from the dataset, named 20210803172558, for our experiment. Figure 1(d) shows its sample image.

Here Table 1 displays the training and testing datasets distribution in classification.

Table 1: Land-cover classes of used Indian Pine and Pavia University datasets, with the standard training and testing distribution.

	classes	training	testing	spatial resolution
Indian	16	695 (3.3%)	9671	145x145
PaviaU	9	3921 (1.9%)	40002	610x340

2.1.2 Evaluation Metrics.

- **PSNR** (Peak Signal-to-Noise Ratio) quantifies the ratio of a signal to its noise, which calculates the ratio of the square of the maximum possible amplitude of the signal to the mean square error (MSE) in the signal. PSNR is employed as a measure of distortion in compression, where higher values correspond to better quality. The PSNR for an HSI with N spectral bands can be formulated as:

$$PSNR(I, \hat{I}) = \frac{1}{N} \sum_{i=1}^N 10 \log_{10} \left(\frac{\max^2(I_i)}{MSE(I_i, \hat{I}_i)} \right) \quad (9)$$

- **bpppb** (bits per pixel per band) is used to evaluate the consumption of compressed bitrate. For $I \in \mathbb{R}^{N \times H \times W}$, the bpppb is calculated as follows:

$$bpppb = \frac{\theta(embeddings) \cdot b_e + \theta(decoder) \cdot b_d}{H \times W \times N} \quad (10)$$

where θ measures the parameters quantities and b denotes the corresponding bit-width.

- **CR** (Compression Ratio) serves as a metric to quantify the compression effect, and it is defined as:

$$CR = \frac{bpppb_{gt}}{bpppb_{compressed}}. \quad (11)$$

- **OA** (Overall Accuracy) is employed to measure the overall classification accuracy. OA is calculated as the number of correctly categorized samples divided by the total sample size.
- **AA** (Average Accuracy) refers to the mean value of classification accuracy across all classes. It involves calculating the accuracy of each individual category and then averaging the accuracies of all categories.
- κ (kappa coefficient) serves as a statistical measure of consistency between the classification maps and the ground truth. The κ ranges from -1 to 1, where 1 signifies perfect consistency, 0 indicates stochastic consistency, and -1 implies complete inconsistency, where a higher κ signifies better performance of the model.

2.1.3 Implementation.

- **HINER**. In addition to the specifications outlined in the main paper, we employ a quantization bit-width of 8 bits for our experiments. Furthermore, for positional encoding, we set $b = 1.25$ and $l = 80$.
- **FHNeRF and Rezasoltani**. FHNeRF and Rezasoltani are two state-of-the-art methods in the implicit neural representation of HSI, which take the original pixel coordinates as input and use *sine* activations. Given that there are no publicly accessible source codes, we faithfully reproduce them. We use a 5-layer/15-layer perceptron and change the hidden

dimension to build models of different sizes for FHNeRF [5]/ Rezasoltani [6], respectively. Both methods are trained for 15000 iterations with Adam optimizer [7] using a learning rate cosine descent strategy.

- **JPEG2000.** For JPEG2000 compression, we utilize OpenJPEG to independently encode each spectral band. Initially, we transform the original HSI into individual raw files, with each file corresponding to a spectral band. Subsequently, we compress and decompress each raw file using OpenJPEG. After the decompression process, we convert the reconstructed raw files back into the MAT (matlab) format. This facilitates the comparison between the reconstructed data and the original data, enabling the computation of PSNR.
- **VVC.** For VVC compression, we initially convert a MAT file into individual PNG files, with each PNG file corresponding to a spectral band. These PNG files are then merged into a YUV file, comprising a sequence of 'frames' at consecutive wavelengths. Subsequently, we perform compression using the VTM tool on the YUV file. However, due to VTM's lack of support for compressing 16-bit YUV files, we utilize 8-bit YUV files instead. After compression and subsequent decompression, we obtain the reconstructed YUV file. Next, we employ ffmpeg to convert the YUV file back into PNG files. The subsequent steps are akin to the JPEG2000 process, where we combine the individual PNG files back into a single MAT data format and compare the results with the original data to compute the PSNR.

2.2 Encoding Complexity

In Sec. 4.2 of the main paper, we have shown that HINER is faster than pixel-wise FHNeRF and Rezasoltan in encoding. Here, we further evaluate the image encoding speed compared to HNeRV, as shown in Table 2. As observed, HINER achieves a higher speed compared to HNeRV, partly due to our encoder having fewer parameters. Additionally, after positional encoding, only a small input vector $\in \mathbb{R}^{1 \times 160}$ needs to be processed by the MLP. This is smaller than the image matrix, e.g., $\in \mathbb{R}^{720 \times 360}$ in Pavia University, requiring multiple down-sampling operations with convolution. Consequently, our encoder has lower encoding complexity and better compression performance.

Table 2: Encoding time comparison.

Method	Encoder Size	Model Size (MB)		
		0.2	0.5	1.5
HINER	0.12 MB	480s	500s	790s
HNeRV	0.22 MB	570s	620s	850s

2.3 Classification on Compressed HSI

Here, we present additional results regarding classification on compressed HSI samples. In Fig.2, we visualize the spatial distribution of the training and testing sets, along with the classification map. Additionally, Table3 and Table 4 exhibit quantitative performance at various compression ratios (CRs). For our method, all compressed

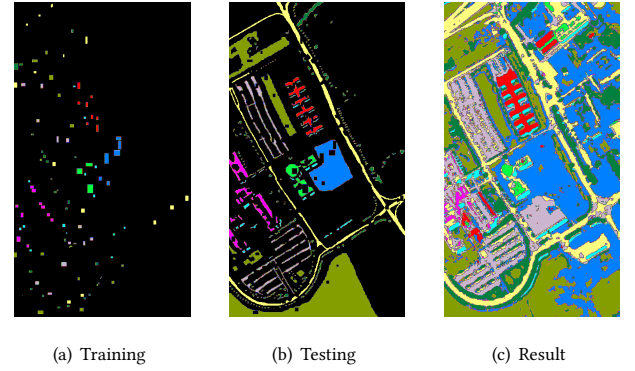


Figure 2: Classification map obtained by our model on the Pavia University dataset

Table 3: Quantitative performance of the Indian Pines.

Method	CR	OA (%)	AA (%)	κ
SF	$\times 1$	81.86	87.81	0.7919
SF [*]	$\times 71$	80.61	85.90	0.7784
Ours[*]	$\times 71$	86.54	91.17	0.8465
SF [*]	$\times 28$	79.15	84.27	0.7633
Ours[*]	$\times 28$	87.03	90.99	0.8519
SF [*]	$\times 13$	79.23	86.73	0.7651
Ours[*]	$\times 13$	86.45	90.94	0.8457

Table 4: Quantitative performance of the Pavia University.

Method	CR	OA (%)	AA (%)	κ
SF	$\times 1$	91.07	90.20	0.8805
SF [*]	$\times 109$	86.29	87.89	0.8203
Ours[*]	$\times 109$	88.93	88.96	0.8529
SF [*]	$\times 54$	86.75	88.77	0.8249
Ours[*]	$\times 54$	88.55	89.36	0.8484
SF [*]	$\times 35$	86.13	89.09	0.8189
Ours[*]	$\times 35$	88.20	89.27	0.8438

HSIs with different CRs are evaluated using the same classification model trained at a CR of 28/109 for Indian Pine/ Pavia University datasets. In contrast, the SF method is re-trained for each CR to achieve the best performance. It is evident from the results that our method demonstrates superior performance in all cases, showcasing high robustness across various compression ratios.

An interesting observation is that a lower compression ratio may not result in better accuracy. This phenomenon is consistent with previous works [8–10] that for certain compression techniques, a higher CR may not significantly degrade the performance of pixel-based classification as the homogenization effect increases

the similarity among pixels of the same area. In addition, land-cover type is also believed to be one of the factors as compression also has different effects on classification results of different land-cover types [11].

2.4 Ablation Studies

2.4.1 Positional Encoding. As discussed in Section 3.2, MLPs are susceptible to the well-known spectral bias [12, 13], wherein they tend to learn low frequency components of the signal. Thus, directly inputting the wavelength λ into the encoder without positional encoding would lead to the network's incapacity to adequately capture high-frequency variation [14, 15]. We illustrate this phenomenon with the regression curve shown in Figure 3. Initially, during the earlier epochs, the performance of HINER w/o PE exhibits a similar regression performance with HINER w/ PE, indicating comparable capability in learning low-frequency components of the signal. However, as the epochs progress, the gap widens, highlighting the superior efficiency of positional encoding in capturing high-frequency information.

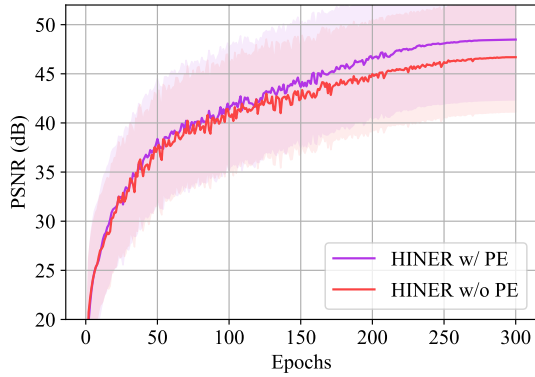


Figure 3: Regression curve of HINER w/ & w/o PE.

2.4.2 Reconstruction Loss. Here, we present an ablation study concerning the γ in Eq. (4) of the main paper:

$$\mathcal{L}_R = \underbrace{\sum_{n=1}^N \|\hat{I}_n - I_n\|}_{L1 \text{ loss}} + \underbrace{\gamma \cdot \sum_{n=1}^N \frac{180}{\pi} \arccos\left(\frac{\vec{\hat{I}}_n^T \cdot \vec{I}_n}{\|\vec{\hat{I}}_n\|_2 \|\vec{I}_n\|_2}\right)}_{CAM}, \quad (12)$$

As depicted in Table 5, we set $\gamma = 0.01$ in our experiments.

Table 5: Ablations on coefficient γ between L1 loss and CAM.

γ	PSNR
0.005	44.11
0.001	43.93
0.01	44.14
0.1	43.87

2.4.3 Adaptive Spectral Weighting. We conduct thorough experiments from next two aspects.

Optimization Objective. As described in Sec. 3.3 of the main paper and Sec. 1.1 of the supplementary material, the optimization objective of ASW is formulated as:

$$\arg \min \mathcal{L}_C + \beta \cdot \mathcal{L}_R, \quad (13)$$

where \mathcal{L}_R is introduced to constrain the input of the classifier (also the output of ASW) in the neighborhood of the ground truth. As illustrated in Table 6, when \mathcal{L}_R is omitted (i.e., $\beta = 0$), there is a notable decrease in accuracy. This phenomenon also corroborates the validity of our theoretical analysis, i.e., *for downstream classification on compressed HSI, task accuracy is not only related to the classification loss but also to the reconstruction fidelity*. Ultimately, we set $\beta = 2.5$ to achieve a balance between these two losses.

Table 6: Ablations on reconstruction loss.

β	OA(%)	AA(%)	κ
5	82.3	87.85	0.7979
2.5	87.03	90.99	0.8519
1.4	84.88	90.57	0.8282
0.5	83.88	88.6	0.8166
0	81.5	85.09	0.789

Classification-Oriented Reconstruction. Additionally, we conduct ablation experiments to examine the effect of adding ASW before the classifier, as shown in Table 7. The inclusion of ASW results in PSNR decrease of the inputted reconstructed HSI of the classifier but an obvious improvement in classification accuracy. This suggests that ASW is able to adaptively weight HSI under the supervision of classification loss, thereby facilitating the translation of reconstruction from perceived visual quality to classification accuracy.

Table 7: Ablations on ASW.

	PSNR	OA (%)	AA (%)	κ
w/o ASW	44.25	79.15	84.27	0.7633
w/ ASW	34.71	84.06	88.24	0.8187

2.4.4 Random uniform variables in ISI. In Sec. 3.3 of the main paper, we implement Implicit Spectral Interpolation (ISI) by introducing random variables on wavelengths:

$$\mathcal{S} = \sum \mathcal{HINER}(\lambda + U(-\eta, \eta)), \quad (14)$$

where $U(-\eta, \eta)$ represents a uniform distribution that adds random variables to λ . When trained with \mathcal{S} , the classification network exhibits improved generalization and reduced accuracy degradation on compressed HSI. It is crucial to note that ground truth HSI is not introduced during training for ISI. Table 8 shows the impact of different η (here the η is not normalized). ISI proves to be a robust method across various η values. For consistency, we set $\eta = 0.1$ as the default setting.

Table 8: Ablations on uniform perturbation η .

η	OA(%)	AA(%)	κ
0.05	87.13	90.06	0.8526
0.1	87.03	90.99	0.8519
0.15	85.68	90.9	0.8369
0.2	85.93	91.29	0.8399
0.4	86.88	91.37	0.8502

REFERENCES

- [1] Riccardo Volpi, Hongseok Namkoong, Ozan Sener, John C Duchi, Vittorio Murino, and Silvio Savarese. Generalizing to unseen domains via adversarial data augmentation. *Advances in neural information processing systems*, 31, 2018.
- [2] Long Zhao, Ting Liu, Xi Peng, and Dimitris Metaxas. Maximum-entropy adversarial data augmentation for improved generalization and robustness. *Advances in Neural Information Processing Systems*, 33:14435–14447, 2020.
- [3] Xiuying Wei, Ruihao Gong, Yuhang Li, Xianglong Liu, and Fengwei Yu. Qdrop: Randomly dropping quantization for extremely low-bit post-training quantization. In *International Conference on Learning Representations*, 2021.
- [4] Erqi Huang, Maoqi Zhang, Zhan Ma, Linsen Chen, Yiyu Zhuang, and Xun Cao. High-fidelity hyperspectral snapshot of physical world: System architecture, dataset and model. *IEEE Journal of Selected Topics in Signal Processing*, 16(4):608–621, 2022.
- [5] Lili Zhang, Tianpeng Pan, Jiahui Liu, and Lin Han. Compressing hyperspectral images into multilayer perceptrons using fast-time hyperspectral neural radiance fields. *IEEE Geoscience and Remote Sensing Letters*, 2024.
- [6] Shima Rezasoltani and Faisal Z Qureshi. Hyperspectral image compression using implicit neural representations. In *2023 20th Conference on Robots and Vision (CRV)*, pages 248–255. IEEE, 2023.
- [7] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [8] Jingru Wei, Li Mi, Ye Hu, Jing Ling, Yawen Li, and Zhenzhong Chen. Effects of lossy compression on remote sensing image classification based on convolutional sparse coding. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2021.
- [9] Fernando Garcia-Vilchez, Jordi Muñoz-Mari, Maciel Zortea, Ian Blanes, Vicente González-Ruiz, Gustavo Camps-Valls, Antonio Plaza, and Joan Serra-Sagristà. On the impact of lossy compression on hyperspectral image classification and unmixing. *IEEE Geoscience and remote sensing letters*, 8(2):253–257, 2010.
- [10] Alaitz Zabala and Xavier Pons. Effects of lossy compression on remote sensing image classification of forest areas. *International Journal of Applied Earth Observation and Geoinformation*, 13(1):43–51, 2011.
- [11] Liang Zhai, Xinming Tang, and Guo Zhang. A new quality assessment index for compressed remote sensing image. In *Mathematics of Data/Image Pattern Recognition, Compression, and Encryption with Applications XI*, volume 7075, pages 175–182. SPIE, 2008.
- [12] Shaowen Xie, Hao Zhu, Zhen Liu, Qi Zhang, You Zhou, Xun Cao, and Zhan Ma. Diner: Disorder-invariant implicit neural representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6143–6152, 2023.
- [13] Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *International Conference on Machine Learning*, pages 5301–5310. PMLR, 2019.
- [14] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.
- [15] Matthew Tancik, Pratul P Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in Neural Information Processing Systems*, 33:7537–7547, 2020.