Unraveling Metameric Dilemma for Spectral Reconstruction: A High-Fidelity Approach via Semi-Supervised Learning

Xingxing Yang
Department of Computer Science
Hong Kong Baptist University, Hong Kong
csxxyang@comp.hkbu.edu.hk

Jie Chen[*]
Department of Computer Science
Hong Kong Baptist University, Hong Kong
chenjie@comp.hkbu.edu.hk

Zaifeng Yang
Institute of High Performance Computing
Agency for Science Technology and Research, Singapore
Yang_zaifeng@a-star.edu.sg

Abstract

Spectral reconstruction from RGB images often suffers from a metameric dilemma, where distinct spectral distributions map to nearly identical RGB values, making them indistinguishable to current models and leading to unreliable reconstructions. In this paper, we present Diff-Spectra that integrates supervised physics-aware spectral estimation and unsupervised high-fidelity spectral regularization for HSI reconstruction. We first introduce an Adaptive illumiChroma Decoupling (AICD) module to decouple illumination and chrominance information, which learns intrinsic and distinctive feature distributions, thereby mitigating the metameric issue. Then, we incorporate the AICD into a learnable spectral response function (SRF) guided hyperspectral initial estimation mechanism to mimic the physical image formation and thus inject physics-aware reasoning into neural networks, turning an ill-posed problem into a constrained, interpretable task. We also introduce a metameric spectra augmentation method to synthesize comprehensive hyperspectral data to pre-train a Spectral Diffusion Module (SDM), which internalizes the statistical properties of real-world HSI data, enforcing unsupervised high-fidelity regularization on the spectral transitions via inner-loop optimization during inference. Extensive experimental evaluations demonstrate that our Diff-Spectra achieves competitive performance on both Spectral reconstruction and downstream HSI classification.

1 Introduction

Hyperspectral imaging captures hundreds of spectral bands, allowing for precise identification of materials and illumination conditions that are often indistinguishable from RGB imaging. It has been widely applied in remote sensing [1, 2], medical diagnosis [3, 4] and agriculture[5, 6].

Traditional hyperspectral imaging methods show limitations in time-consuming acquisition processes with limited spatial resolution. Recent advances in deep learning within the computer vision community have paved the way for hyperspectral image reconstruction from RGB inputs using data-driven methodologies [7, 8]. Early model-based methods, such as sparse coding [9] and low-rank

[*]Corresponding Author

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

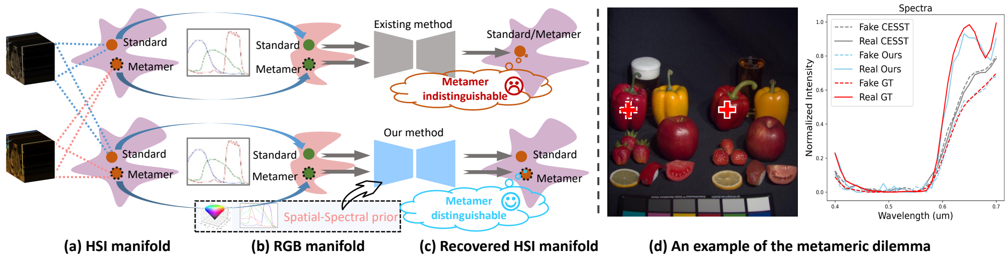

Figure 1: Motivation. (a) shows metamer and standard HSIs. In (b), standard and metamer RGB inputs are produced by the same SRF from corresponding HSIs. In (d), Left: the dashed and solid red highlighted points are two pixels from the fake and real peppers from the CAVE dataset [19], respectively; Right: Corresponding ground-truth spectral curves of the two pixels, and reconstructed spectral curves of CESST [8] (grey) and our method (blue). The fake and real peppers are in similar colors but have different spectra. Existing methods (*e.g.*, CESST) fail to distinguish either the two spectra from each other or from their true spectra, while our method can reconstruct faithfully.

representation [10], rely predominantly on manually crafted priors. In contrast, modern learning-based approaches [7, 11, 12, 13] leverage deep learning frameworks such as convolutional neural networks [14, 15, 16] and Vision Transformers [7, 8] to achieve superior reconstruction performance.

Limitations. However, we observe that most existing methods [7, 11, 12] suffer from a metameric dilemma [17, 18], where distinct spectral distributions map to nearly identical RGB values, making them indistinguishable to current models and leading to unreliable reconstructions, as shown in Fig.1. We argue that this stems from the direct nonlinear mapping from the sRGB input space to the hyperspectral output space, leading to three key limitations in terms of data, methodologies, and pipelines: (**i**) ***Data Poverty:*** Existing hyperspectral datasets lack diversity in rich metameric examples, forcing existing models to memorize a limited subset of spectra rather than internalizing the full spectral manifold; (**ii**) ***Spectral Manifold Blindness:*** Existing methods overlook the explicit modeling of the real-world spectral distribution, leading to hallucinated outputs that minimize pixel-wise losses (*e.g.*, MSE) but violate physical laws (*e.g.*, unnatural spectral intensity in Fig. 1); (**iii**) ***Architecture Myopia:*** Existing methods treat spectral radiance as a monolithic entity, conflating illumination and chrominance. It amplifies metameric failures that networks cannot distinguish whether RGB variations stem from lighting or material properties.

It raises a fundamental question: ***Can we introduce auxiliary information to address the above-mentioned metameric dilemma, thus achieving high-fidelity Spectral reconstruction? If so, how?***

To this end, we propose *Diff-Spectra*, a semi-supervised model for HSI reconstruction that integrates supervised physics-aware spectral estimation and unsupervised high-fidelity spectral regularization. Specifically, to deal with architecture myopia, we first introduce an Adaptive IllumiChroma Decoupling (AICD) module by factorizing RGB inputs into independent illumination and chrominance subspace via orthogonal decoupling. Then, the AICD is incorporated into a SRF-guided HSI initial estimation (SRF-guided HIE) mechanism to estimate the target spectral signal in a supervised manner. This process mimics the physical image formation and thus injects physics-aware reasoning into neural networks, turning an ill-posed problem into a constrained, interpretable task. To deal with the data poverty, we introduce a metameric spectra augmentation method to synthesize a comprehensive HSI dataset with diverse metamer samples and spectral perturbations, transforming sparse spectral data into a rich prior that can guide reconstruction beyond RGB ambiguities. To deal with the spectral manifold blindness, we introduce a spectral diffusion module (SDM) that learns to denoise corrupted spectra during pre-training on the comprehensive HSI dataset, which internalizes the statistical properties of real-world hyperspectral data. During reconstruction, the pre-trained SDM serves as a spectral prior that regularizes the coarsely estimated HSI signal from the SRF-guided HIE mechanism with high-fidelity real-world spectral distributions via our proposed unsupervised inner loop optimization. The main contributions are given as follows:

- We propose a semi-supervised paradigm, Diff-Spectra, to deal with the metameric dilemma in spectral reconstruction. It integrates supervised physics-aware spectral estimation and unsupervised high-fidelity spectral regularization.

- We introduce an Adaptive IllumiChroma Decoupling (AICD) module based on orthogonal decoupling to effectively factorize illumination and chrominance information, serving as an IllumiChroma prior. It learns distinctive image features, alleviating the metameric dilemma.
- We design a lightweight, learnable SRF-guided HIE mechanism to obtain an initial HSI estimation, which formulates the reconstruction model with physical constraints and enhances the interpretability.
- We introduce a metameric spectra augmentation method to synthesize a comprehensive HSI dataset to pre-train a Spectral Diffusion Module (SDM) to capture the real-world spectral distribution, serving as a spectral prior to improve spectral fidelity.
- Extensive experiments on both spectral reconstruction and HSI classification demonstrate that our Diff-Spectra framework significantly outperforms SOTA methods.

2 Related Work

2.1 Hyperspectral Image Reconstruction

Early efforts [9, 16] utilized model-based approaches that incorporated fidelity terms and physical priors to constrain the target solutions. For example, Arad et al. [9] improved this interpolation challenge using hyperspectral priors to forge a sparse dictionary of HSIs alongside their RGB counterparts. Despite their contributions, these model-based strategies depend on manually tailored priors, constricting their representational capability. Learning-based methods [11, 12, 20, 21], shifted the focus to data-driven approaches, learning implicit mappings from RGB to hyperspectral domains using specifically designed architectures. Notably, the HSCNN model [14] revolutionizes the field by mapping the input RGB image into hyperspectral feature space via a convolutional layer and harnessing deep residual convolutional blocks to approximate the enriched HSI. Cai et al. [7] proposed a transformer-based approach to capture the long-range channel-wise correlations that compute the self-attention map along the channel dimension, tailored for HSI reconstruction. Wang et al. [22] introduced an intrinsic image decomposition (IID) framework to decompose input images into reflectance and shading features and then reconstruct them in the spectral domain separately. However, all existing learning-based methods directly learn mappings between sRGB and HSI feature spaces, neglecting the intrinsic spectral distribution of HSIs and lacking physical constraints, thus encountering the challenges posed by the metameric dilemma.

2.2 Diffusion-based HSI Image Reconstruction

Diffusion models [23, 24] have witnessed an explosion of continuously growing capability and capacity architectures. In the context of HSI image reconstruction, Pang et al. [25] proposed HIR-Diff that leverages a powerful pre-trained diffusion prior and a product-of-experts guidance scheme to remove degradations and recover clean hyperspectral images in an unsupervised manner, demonstrating strong generalization across scenes and noise types. Wu et al. [26] proposed a conditional denoising transformer to fuse high-resolution multi-spectral images and low-resolution hyperspectral images to generate target high-resolution HSI images. Liu et al. [27] proposed incorporating the deep generative prior of diffusion models to constrain the high-resolution multi-spectral image and low-resolution hyperspectral image fusion process. While these works focus on restoration and reconstruction, diffusion's representational advantages have also benefited HSI classification, suggesting transferable priors for reconstruction. Chen et al. [28] proposed a spatial-spectral diffusion module to generate high-dimensional HSI signals for HSI classification, indicating the great potential of diffusion-based generative models in spectral distribution modeling. Beyond natural HSI classification, Sigger et al. [29] proposed a multistage unsupervised diffusion framework to extract complementary high- and low-level spectral features for challenging biomedical HSI classification, underscoring diffusion's capacity to model fine-grained spectral structure. However, applying diffusion models for practical HSI systems still faces trade-offs among cost, complexity, and acquisition speed. Moreover, directly reconstructing full hyperspectral cubes from RGB inputs with diffusion remains comparatively underexplored, presenting an opportunity to combine diffusion priors with cross-modal spectral constraints and measurement-aware conditioning for faithful RGB-to-HSI recovery grounded in both spectral accuracy and spatial details.

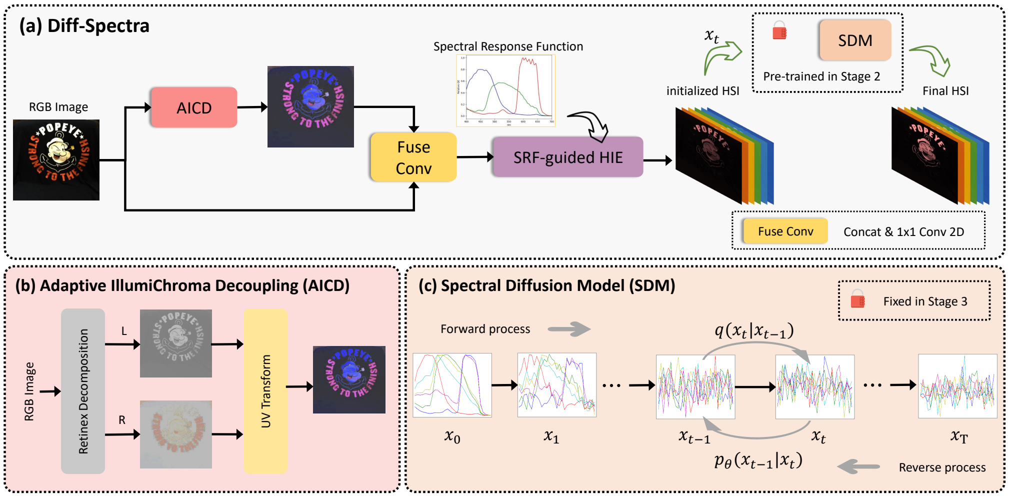

Figure 2: Overall framework of **Diff-Spectra**. Our model has three training stages. In Stage 1, the SRF-guided HIE and AICD are pre-trained using Eq. 10 to initialize a coarse-level HSI signal while freezing the SDM. In Stage 2, the SDM is pre-trained using Eq.13 to learn the spectral distribution of HSI signals while freezing the SRF-guided HIE and AICD. In Stage 3, we fine-tune the SRF-guided HIE and AICD using the objective function Eq. 14, which regularizes the initialized HSI signal in Stage 1 with the spectral prior of SDM in Stage 2. During inference, we further introduce an inner loop optimization as a test-time adaptation with Eq. 13 to generate refined HSI signals by following the spectral reverse generative sequence of SDM.

3 Methodology

3.1 Problem Definition

The RGB imaging process can be formulated as a sub-sampling process of the target HSI signal:

$$\mathbf{X} = \mathbf{Y}\mathbf{A} + \epsilon, \quad \mathbf{A} \in \mathbb{R}^{C\times c}, \tag{1}$$

where $\mathbf{X} \in \mathbb{R}^{H\times W\times c}$ denotes the observed RGB image, with H, W and c representing height, width, and channel ($c << C$) of $\mathbf{X}$, respectively. $\mathbf{Y} \in \mathbb{R}^{H\times W\times C}$ and $\mathbf{A} \in \mathbb{R}^{C\times c}$ are the target HSI signal and the spectral response function (SRF) of the RGB sensor, respectively. ϵ is the noise or residual terms that arise in the image processing pipeline. The HSI reconstruction task can be formulated as a maximum a posteriori problem: $\max_{\mathbf{Y}} p(\mathbf{Y} \mid \mathbf{X}, \mathbf{A})$. Applying Bayes' theorem, the posterior can be further reformulated as:

$$p(\mathbf{Y} \mid \mathbf{X}, \mathbf{A}) = \frac{p(\mathbf{X}, \mathbf{A} \mid \mathbf{Y})p(\mathbf{Y})}{p(\mathbf{X}, \mathbf{A})}, \tag{2}$$

Taking the negative logarithm and discarding irrelevant terms $\mathbf{X}$, $\mathbf{A}$, we obtain the objective function:

$$\min_{\mathbf{Y}}\{-\log p(\mathbf{X}, \mathbf{A} \mid \mathbf{Y}) - \log p(\mathbf{Y})\}, \tag{3}$$

where $\log p(\mathbf{X}, \mathbf{A} \mid \mathbf{Y})$ is the log-likelihood that depicts the degradation processes of RGB sampling from the HSI signal, and $\mathbf{Y}$ is the spectral prior that will contribute to the restoration of $\mathbf{Y}$.

3.2 Motivation and Solution

Existing HSI reconstruction methods [7, 8] only consider the first term in Eq. 3 (*i.e.*, $\log p(\mathbf{X}, \mathbf{A} \mid \mathbf{Y})$) and directly map RGB inputs to hyperspectral feature space, overlooking subtle variations in RGB feature distributions, thereby leading to the *metameric dilemma*, as shown in Fig. 1 and further demonstrated by the experimental results in Section 4.3.

In this paper, we would like to consider this challenging spectral reconstruction problem as supervised physics-aware spectral estimation and unsupervised high-fidelity spectral regularization. Thus, in terms of supervised physics-aware spectral estimation, which corresponds to to optimize the first term $-\log p(\mathbf{X}, \mathbf{A} \mid \mathbf{Y})$ in Eq. (3), we introduce an Adaptive IllumiChroma Decoupling (AICD) module by factorizing RGB inputs into independent illumination ($\mathbf{U}$) and chrominance ($\mathbf{V}$) subspace via orthogonal decoupling to explore intrinsic clues and distinguish image features, which effectively eliminates redundancy and enhances distinctive patterns among metameric hyperspectral data. Then, the AICD is incorporated into the SRF-guided HSI Initial Estimation (SRF-guided HIE) module to estimate the target spectral signal in a supervised manner. This process mimics the physical image formation and thus injects physics-aware reasoning into neural networks, turning an ill-posed problem into a constrained, interpretable task. Now, we express the first term $-\log p(\mathbf{X}, \mathbf{A} \mid \mathbf{Y})$ as the following according to Eq. (1):

$$\mathcal{L}_Y(\mathbf{Y}) = \|(\mathbf{X}, \mathbf{U}, \mathbf{V}) - \mathbf{YA} - \epsilon\|^2. \tag{4}$$

Then, we assume the second prior term, *i.e.*, $\log p(\mathbf{Y})$, as unsupervised high-fidelity spectral regularization for the estimated spectra, which further constrains the estimation via transferring distribution prior knowledge from a pre-trained model leveraging our metameric spectra dataset.

3.3 Supervised Physics-Aware Spectral Estimation

3.3.1 Adaptive IllumiChroma Decoupling

A spectral radiance $\mathbf{Y}(\lambda)$ can be represented as $\mathbf{Y}(\lambda) = \mathbf{E}(\lambda)\mathbf{S}(\lambda)$, where, $\lambda, \mathbf{E}(\lambda), \mathbf{S}(\lambda)$ represents the spectral entity, illumination and surface chrominance, respectively. Existing deep learning methods treat $\mathbf{Y}(\lambda)$ as a monolithic entity, failing to disentangle illumination and chrominance. This conflation amplifies metameric ambiguities, as networks struggle to distinguish whether RGB variations stem from lighting changes $\mathbf{E}(\lambda)$ or material properties $\mathbf{S}(\lambda)$. This dilemma is the same with RGB inputs. As such, we introduce AICD that aims to extract intrinsic and distinctive image features that are robust against different light variations. However, as shown in Fig. 3, we find that merely relying on reflectance $\mathbf{R}$ and illumination $\mathbf{L}$ decomposition based on the Retinex theory [30] cannot distinguish between metameric examples.

Considering that HSIs capturing often exhibit low-light characteristics that lead to potential information loss, we first customize an illumination sensitivity parameter S_k, which enables image-specific adjustment as: $\mathbf{S}_k = \sqrt[k]{\sin(\frac{\pi \mathbf{L}'}{2}) + \tau}$, where $k \in \mathbb{Q}^+, \tau = 1 \times 10^{-8}$. In specific, different from existing decomposition methods [31, 32, 33] that decompose images in a deterministic manner, which may not appropriately capture the diverse and complex imaging and lighting conditions that are specific to downstream tasks, we introduce trainable parameters to enable adaptive learning of intrinsic image features in an end-to-end manner, thereby enhancing compatibility with downstream tasks. Specifically, we employ convolutional layers to obtain embedded features: $R^{'} = Conv(\mathbf{R}), L^{'} = Conv(\mathbf{L})$.

Orthogonal Decoupling. As discussed, decomposed reflectance and illumination components cannot well distinguish metamer data. Inspired by the orthogonal decoupling method in [34, 35], which effectively eliminates redundancy and enhances distinctive patterns among visual features, we further deployed orthogonal UV transform to extract more distinctive features. We define the horizontal ($\mathbf{U}$) and vertical ($\mathbf{V}$) plane as:

$$\mathbf{U} = \mathbf{S}_k \odot \mathbf{R}^{'} \odot h, \qquad \mathbf{V} = \mathbf{S}_k \odot \mathbf{R}^{'} \odot v, \tag{5}$$

where $\odot$ denotes Hadamard production. Note that we orthogonalize the UV-planes using the two intermediate variables $h = \cos(2\pi \mathbf{R}^{'})$ and $v = \sin(2\pi \mathbf{R}^{'})$. Finally, the decomposed features ($\mathbf{U}$, and $\mathbf{V}$) will serve as the illumichroma prior for SRF-guided HIE.

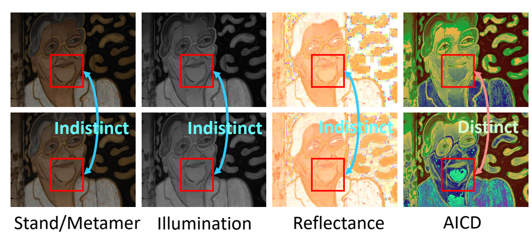

Figure 3: Difference between the decomposition methods of Retinex and AICD. The top row is based on standard data, while the bottom row is based on metamer data. AICD can transform metamer counterparts into more discriminative features than reflectance and illumination maps.

3.3.2 SRF-guided HSI Initial Estimation

Note that the SRF matrix $\mathbf{A}$ in Eq. 1 that projects the HSI to an RGB frame has a row-full rank, enabling a right-inverse solution based on the matrix inversion rule. Thus, there exists a transpose matrix $\mathbf{A}^T$ of $\mathbf{A}$ such that the following equation obtained from Eq. 1 holds

$$\mathbf{X}\mathbf{A}^T - \epsilon\mathbf{A}^T = \mathbf{Y}\mathbf{A}\mathbf{A}^T, \tag{6}$$

$\mathbf{A}\mathbf{A}^T$ is a square matrix, and there exists an inverse matrix of $\mathbf{A}\mathbf{A}^T$ such that the following equation holds:

$$\gamma\mathbf{X}\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} - \epsilon\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} = \mathbf{Y}\mathbf{A}\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}, \tag{7}$$

We can rewrite Eq. 7 as

$$\mathbf{Y} = \mathbf{X}\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} - \epsilon\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}, \tag{8}$$

Direct inversion of RGB images into spectral space via SRFs may be suboptimal due to variations in SRFs across different spectral cameras and disturbances arising from factors such as RGB format compression and noise, and can not distinguish metamers. To address these challenges, we incorporate the AICD into a spectral response function (SRF) guided HSI initial estimation (SRF-guided HIE) mechanism to estimate the target spectral signal in a supervised manner. This process mimics the physical image formation and thus injects physics-aware reasoning into neural networks, turning an ill-posed problem into a constrained, interpretable task. Specifically, we incorporate the decomposed illumichroma prior from the AICD module and then estimate the coarse HSI signal via the SRF-guided HIE function:

$$\tilde{\mathbf{Y}} = \mathcal{F}(\mathcal{H}(\mathbf{X}, \mathbf{U}, \mathbf{V})\mathbf{A}^T \left(\mathbf{A}\mathbf{A}^T\right)^{-1}, \tag{9}$$

where $\mathcal{F}$ denotes a neural network mapping function (a standard UNet-based network [7]), $\mathcal{H}$ denotes a fusion operator (via concatenation and convolution) and $\tilde{\mathbf{Y}}$ denotes the SRF-guided HIE output. Now, Eq. (4) is implemented as:

$$\mathcal{L}_Y(\mathbf{Y}) = \|\mathbf{Y} - \mathcal{F}(\mathcal{H}(\mathbf{X}, \mathbf{U}, \mathbf{V})\mathbf{A}^T \left(\mathbf{A}\mathbf{A}^T\right)^{-1})\|^2, \tag{10}$$

3.4 Unsupervised High-Fidelity Spectral Regularization

We introduce spectral diffusion models (SDMs) [27, 36] as spectral priors to refine the initial HSI estimated by the SRF-guided HIE and achieve accurate reconstruction. However, three key challenges must be addressed: (**i**) SDMs struggle to capture metameric feature distributions, limiting their ability to model subtle spectral variations. (**ii**) Directly modeling the spatial-spectral distribution of 3D hyperspectral data incurs a significant computational burden due to the high dimensionality of the data. (**iii**) A notable domain discrepancy exists between SRF-guided HIE results and real hyperspectral images (HSIs), suggesting that directly applying SDMs may result in suboptimal performance. To address these challenges, we introduce three key improvements for effectively integrating SDMs into spectral reconstruction: (1) metameric spectral augmentation, which creates comprehensive training data to enhance feature representation, (2) a lightweight spectral diffusion architecture to reduce computational complexity, and (3) gradient-based inner loop optimization to bridge the domain gap and improve reconstruction accuracy.

Metameric spectra augmentation. Metameric augmentation leverages an orthogonal subspace decomposition [37] of a spectrum $\mathbf{S}$ into a component that lies in the sensor-response range subspace and a residual in the null space, enabling different spectra to produce the same RGB response under a given spectral sensitivity [38]. With a spectral response matrix $\mathbf{A}$, a new metameric spectrum $\mathbf{S}^\dagger$ can be synthesized following the concept of metameric black [39]:

$$\mathbf{S}^\dagger = \mathcal{J} + \beta\mathcal{J}^\dagger, \tag{11}$$

where we project $\mathbf{S}$ onto the range space as $\mathcal{J} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{S}$ and define the residual null space as $\mathcal{J}^\dagger = \mathbf{S} - \mathcal{J}$. New metamer spectra are then synthesized by varying the scalar β. Because $\mathcal{J}^\dagger$ lies in the null space of the response, changing β does not alter the RGB tristimulus, so $\mathbf{S}$ and $\mathbf{S}^\dagger$ are colorimetric matches while differing spectrally. This property reflects the broader fact that rich spectra are reduced to three sensor channels in trichromatic systems, making metamers common and exploitable for augmentation.

We adopt this as a spectral-wise augmentation (rather than spatial flips or crops) to expand the diversity of training spectra while preserving color consistency seen by RGB sensors. Such augmentation can enhance robustness in spectral reconstruction pipelines, a concept established in color science as a way to explore spectra that map to the same color. Practically, we generate metamer data by sampling β uniformly from $[0, 1)$; note that setting $\beta = 1$ recovers the original spectrum, while other values yield alternative metamers that share the same RGB under $\mathbf{A}$. We produce metamers at a $1:1$ ratio with standard spectra, doubling the dataset and providing both the original and augmented hyperspectral inputs for pre-training.

Lightweight spectral diffusion architecture. HSI signals exhibit significant spatial sparsity [40, 41], suggesting that direct modeling of 3-D data cubes might be suboptimal. As such, we introduce a 1-D spectral diffusion model to capture the spectral distribution of HSI signals to address the second term in Eq. (3). Our SDM employs two iterative processes, following the standard DDPM [42], a forward diffusion process, and a reverse denoising process, as illustrated in Fig. 2. Unlike existing diffusion-based variations [33, 28], we adopt a 1-D MLP-based UNet denoising network to ensure that the diffusion model is compatible with 1-D spectral data. While it is feasible to capture the spectral distribution of HSI signals via our proposed SDM, the key question is: *how can the trained SDM be incorporated to solve the second prior term $-\log p(\mathbf{Y})$ of Eq. (3)?*

In our work, we assume that any spectral vector $\mathbf{y} \in \mathbb{R}^{\mathbb{C}}$ of the target HSI $\mathbf{Y}$ are i.i.d. (independently identically distributed), *i.e.*, $-\log p(\mathbf{Y}) = -\sum_{\mathbf{y}\in\mathbf{Y}} \log p(\mathbf{y})$, and each spectrum sample follows the spectral distribution learned by our proposed SDM, the deep generative prior, *i.e.*, $\mathbf{y} = \mathbf{y_0} \sim q(\mathbf{y}_0)$. Consequently, the optimization problem Eq. (3) can be rewritten as:

$$\min_{\mathbf{Y}} \mathcal{L}_Y(\mathbf{Y}) + \lambda \sum_{t,\mathbf{y}} \mathcal{L}_{kl}(q(\mathbf{y}_{t-1}|\mathbf{y}_t, \mathbf{y})||p_\theta(\mathbf{y}_{t-1}|\mathbf{y}_t)), \tag{12}$$

where λ and $\mathcal{L}_{kl}$ denote a balance hyperparameter and the KL divergence, respectively. Subsequently, we can use the parameterization trick to rewrite the second term in Eq. (3) as

(13)

With the above derivation, the final objective function of the proposed Diff-Spectra can be formulated as

$$\min_{\mathbf{Y}} \mathcal{L}_Y(\mathbf{Y}) + \lambda \sum_{t,\mathbf{y}} \mathcal{L}_\theta(\mathbf{y}, t). \tag{14}$$

Note that this objective function is adopted during the fine-tuning process in Stage 3, where λ is set to 0.1 empirically.

Inner loop optimization. It is impractical to optimize Eq. (14) simultaneously for all time steps. *This is because an inherent domain gap exists between the spectral distribution learned by the SDM and the spectral distribution of the coarse-level HSI learned by the SRF-guided HIE network.* Simply assuming these two spectral distributions are consistent without further adaptation can lead to suboptimal results. As such, we propose an inner loop optimization during each time step in the inference phase that performs gradient descent K times for each t, which serves as the test-time adaptation.

4 Experiments and Analysis

4.1 Implementation Details and Datasets

We implemented Diff-Spectra using Pytorch. In **Stage 1**, we pre-train the SRF-guided HIE and AICD on the training dataset for 300 epochs with the Adam optimizer following [7]. Empirically, we set the learning rate to 4×10^{-4} and the batch size is 20. In **Stage 2**, we pre-train the SDM on our generated metamer dataset for 300 epochs with the Adam optimizer. The learning rate is set to 1×10^{-2}. The batch size is 1024, and the total diffusion steps $\mathbf{T}$ is 5000. In **Stage 3**, we fine-tune the SRF-guided HIE and AICD, while freezing the SDM for 100 epochs with the Adam optimizer. Empirically, we set the learning rate to 1×10^{-4} and the batch size is 20. During inference, we input RGB images from the testing dataset and load the pre-trained SRF-guided HIE and AICD to obtain coarse-level HSIs. We then treat the coarse-level HSIs as learnable parameters and load the pre-trained SDM to update the parameters and generate refined HSIs, using a learning rate of 1×10^{-4} with diffusion steps of $\mathbf{S} = 50$ and inner loop $\mathbf{K} = 5$. To evaluate the generalization and fidelity of our method, we use two HSI reconstruction datasets (ARAD-1K [43] and ICVL [9]) and two classification datasets (Indian Pines [1] and Pavia University [44]). Further details for implementation and dataset descriptions are provided in the Appendix A.2.

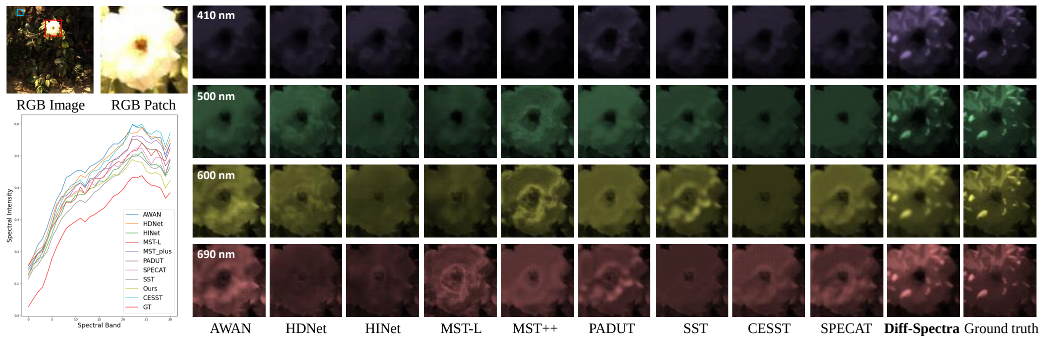

Figure 4: Visual comparisons on a randomly selected scene from the validation set of the ARAD-1K dataset with 4 spectral channels. The spectral curves (bottom-left) correspond to the selected blue Box of the RGB image. Please zoom in for a better comparison.

Table 1: Comparison with SOTA methods on ARAD-1K [43] and ICVL [9] datasets. The best and second are shown in red and blue, respectively. Our method achieves the best performance on most metrics with relatively fewer parameters.

Method	Venue	Params (M)	ARAD-1K Dataset				ICVL Dataset			
			ERGAS (↓)	SAM (↓)	SSIM (↑)	PSNR (↑)	ERGAS (↓)	SAM (↓)	SSIM (↑)	PSNR (↑)
AWAN *[45]*	CVPRW'20	4.04	17.45	9.83	0.909	31.22	8.32	4.59	0.918	31.92
HDNet *[46]*	CVPR'22	2.66	14.16	10.24	0.913	32.13	7.58	4.17	0.924	31.85
HINet *[47]*	CVPR'21	5.21	15.16	8.18	0.916	32.51	7.93	4.19	0.928	32.01
MST-L *[48]*	CVPR'22	2.45	12.53	7.47	0.922	33.90	5.44	3.51	0.935	32.87
MST++ *[7]*	CVPRW'22	1.62	9.18	6.05	0.928	34.32	4.82	3.04	0.941	32.44
PADUT *[49]*	ICCV'23	6.38	7.39	5.53	0.946	34.51	4.15	3.11	0.948	33.07
SST *[50]*	IJSWIS'24	12.74	8.29	6.01	0.933	33.95	4.67	3.71	0.935	32.71
CESST *[8]*	AAAI'24	1.54	7.85	5.88	0.931	34.74	4.09	3.27	0.939	32.96
SPECAT *[51]*	CVPR'24	0.37	8.62	6.10	0.930	33.48	4.92	3.81	0.944	32.54
Diff-Spectral ***(ours)***	–	2.49	4.63	3.96	0.940	35.47	2.84	2.39	0.941	34.71

4.2 Hyperspectral Image Reconstruction

We present quantitative and qualitative comparisons with 9 state-of-the-art methods including AWAN [45], HINet [47], HDNet [46], MST-L [48], MST++ [7],SST [50], PADUT [49], CESST [8] and SPECAT [51].

Qualitative Comparison. Visual Comparisons are given in Fig. 4 and Fig. 5. Fig. 4 compares the reconstructed HSIs with four randomly selected spectral channels using nine SOTA methods and our Diff-Spectra on the validation set of the ARAD-1K dataset. Fig. 5 shows the MSE error map between generated and ground-truth HSIs, calculating along the spectral dimension. It is observed that existing HSI reconstruction methods struggle with spectral intensity estimation and detail recovery, particularly in regions with high-frequency details such as the sky. In contrast, our approach excels at restoring intricate textures and achieving superior pixel-level smoothness. This improvement is attributed to the novel illumiChroma prior learned by the AICD module and the spectral prior learned by the SDM. The AICD module facilitates intrinsic image decomposition, such as illuminance and color information, guiding the initial estimation of the HSI signal with perceptually pleasing spatial features, while the SDM captures spectral distributions that refine the coarse-level HSI generated by the SRF-guided HIE, ensuring spectral consistency.

Quantitative Comparison. We evaluate the performance using metrics including ERGAS, SAM, SSIM, and PSNR. The first two metrics assess spectral quality, while the latter two evaluate spatial quality. Lower ERGAS and SAM values indicate better spectral quality, while higher SSIM and PSNR values signify better spatial quality. As shown in Table 1, our method achieves the best performance over most metrics on both the ARAD-1K dataset and the ICVL dataset. Our approach, Diff-Spectra, outperforms state-of-the-art methods by delivering the highest PSNR and lowest ERGAS (best spectral and spatial reconstruction quality) with significantly lower computational complexity.

4.3 Metameric Dilemma Evaluation

To demonstrate that existing methods suffer from the metameric dilemma and to validate the effectiveness of our proposed Diff-Spectra in mitigating this issue, we generate metamer HSI data from

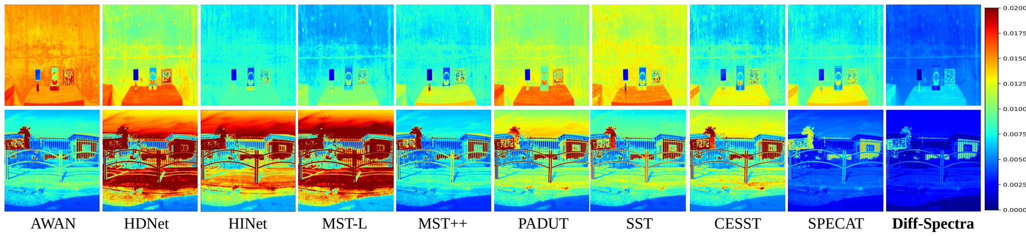

Figure 5: The MSE error maps obtained from the validation subset of the ICVL dataset, which are calculated along the spectral direction, showcasing the discrepancies between the reconstructed HSIs and the corresponding ground truths.

Table 2: Metameric dilemma Evaluation and HSI classification comparison.

Method	Data	ERGAS ↓	SAM ↓	SSIM ↑	PSNR ↑
MST++ [7]	std	9.18	6.05	0.928	34.32
	meta	84.77	50.92	0.8696	27.14
PADUT [49]	std	7.39	5.53	0.946	34.51
	meta	80.07	47.91	0.872	27.90
SST [50]	std	8.29	6.01	0.933	33.95
	meta	70.11	52.38	0.884	27.88
CESST [8]	std	7.85	5.88	0.931	34.74
	meta	59.82	40.74	0.885	28.59
SPECAT [51]	std	8.62	6.10	0.930	33.48
	meta	48.90	63.59	0.8701	26.63
Diff-Spectral ***(ours)***	std	4.63	3.96	0.940	35.47
	meta	21.98	24.11	0.906	31.61

(a) Comparison with SOTA methods on standard (std) and metamer (meta) data.

Class No.	*AWAN [45]*	*MST++ [7]*	*PADUT [49]*	*MST-L [48]*	***Ours***
1	82.71	81.45	87.49	90.47	92.72
2	59.36	84.27	92.35	86.56	91.45
3	73.65	60.33	62.54	88.47	89.97
4	80.04	95.14	91.87	89.58	89.71
5	99.46	99.17	100.00	99.47	100.00
6	95.19	90.62	84.28	47.59	90.68
7	78.42	76.04	75.61	89.55	81.83
8	81.47	98.15	93.46	70.43	69.39
9	94.83	90.52	94.88	82.64	98.44
OA (%)	75.26	82.37	85.48	81.45	89.02
AA (%)	83.35	85.19	86.93	82.97	88.15
κ	0.7124	0.7941	0.8039	0.7355	0.8349

(b) Quantitative comparison of different methods in terms of the accuracy for each class.

the original ARAD-1K dataset following [52] and use metamer HSI data to synthesize corresponding RGB images (*i.e.*, metamer). Next, we test several pre-trained models using both standard RGB images and metamer RGB images (Note that these models are pre-trained on standard data), including MST++ [7], CESST [8], PADUT [49], SST [50], SPECAT [51], and our proposed Diff-Spectra. The quantitative results are given in Table 2(a). As can be seen, all the existing methods experience catastrophic performance drops in terms of PSNR and SAM in the presence of metamers, which is also known as the metameric dilemma.

4.4 Evaluation on HSI Classification

To further evaluate the fidelity and verify the reliability of the hyperspectral images generated by our method, we conduct experiments on the hyperspectral image classification task based on a pre-trained HSI classification model, SpectralFormer [53]. We compare our approach with existing methods on two widely used HSI classification datasets: the Indian Pines dataset and the Pavia University dataset, conducting both quantitative and qualitative analyses. For our evaluation, we reconstruct HSI images using assorted pre-trained HSI reconstruction methods. Subsequently, these synthesized HSI images are employed as the input to a pre-trained HSI classification model, SpectralFormer [53], which serves as a benchmark for performance evaluation.

Evaluations. We evaluate the quantitative performance on the Pavia University dataset using three widely adopted metrics: Overall Accuracy (OA), Average Accuracy (AA), and the Kappa Coefficient (κ), as shown in Table 2(b). As can be seen, our method achieves the best OA (89.02%), AA (88.15%), and κ (0.8349). It ranks first in Classes 1, 3, and 9 and ties for first in Class 5, while remaining competitive in the remaining classes.

4.5 Ablation Study

Break-down Ablation. We perform bread-down ablation to investigate the effectiveness of each module in Table 3(a) and Figure 6. Comparing Variant 1 with SimDiff-Spectra (*i.e.*, a UNet-based HSI reconstruction network, which is similar to [7]), we find that the AICD primarily contributes to the spatial details recovery of the HSI signal, which aligns with our original design intention. Comparing Variant 1 with Variant 2, we observe that the SRF-guided HIE mechanism enhances

Table 3: Ablation studies of our proposed modules and inner loop mechanism.

Method	AICD	SRF	SDM	IP	PSNR	SAM
SimDiff-Spectral					32.57	7.52
Variant 1	✓				34.18	6.08
Variant 2	✓	✓			34.95	5.41
Variant 3	✓	✓	✓		34.19	5.84
Diff-Spectral	✓	✓	✓	✓	35.47	3.96

(a) Break-down ablations of our proposed Diff-Spectra, where IP denotes the inner loop.

Setting	K	PSNR	SAM
Diff-Spectral	1	34.19	5.84
Diff-Spectral	4	35.41	4.07
Diff-Spectral	5	35.47	3.96
Diff-Spectral	7	35.25	4.31
Diff-Spectral	10	35.01	4.33

(b) The right sub-table investigate the inner loop steps ($\mathbf{K}$) in the SDM.

both spatial and spectral performance due to the incorporation of physical constraints. Notably, by comparing Variant 2 with Variant 3, we find that simply incorporating SDM causes a severe performance drop. Finally, comparing Variant 3 with the full Diff-Spectra model, we find that the inner loop facilitates effective integration of SDM and primarily contributes to spectral recovery via the learned spectral distribution regularization.

Inner Loop Analysis. We analyze the impact of the inner loop step parameter $\mathbf{K}$ of the SDM in Table 3(b). The results indicate that performance improves significantly when $\mathbf{K}$ is greater than 1 as compared to when $\mathbf{K} = 1$. Note that when $\mathbf{K} = 1$, the performance is even worse than Variant 2. This is because a domain gap exists between the spectral distribution learned by the SDM and the spectral distribution of the coarse-level HSI learned by the SRF-guided HIE network. Assuming these two spectral distributions are consistent without further adaptation can lead to suboptimal results. Directly assuming these two spectral distributions are consistent will introduce an inferior influence.

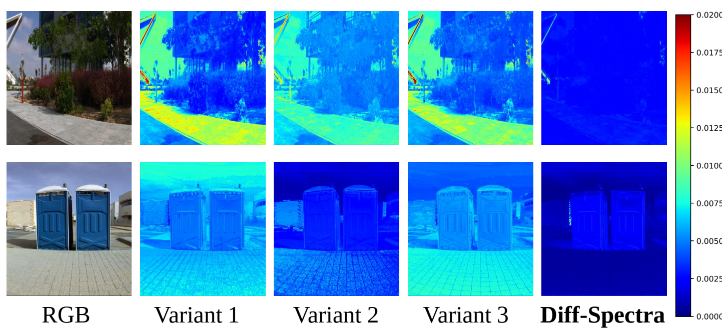

Figure 6: Break-down ablation study. The reconstruction MSE error map evaluation on the validation set of the ICVL dataset.

5 Conclusion

In this paper, we propose Diff-Spectra, which integrates supervised physics-aware spectral estimation and unsupervised high-fidelity spectral regularization for spectral reconstruction. Especially, the supervised physics-aware spectral estimation consists of an adaptive illumichroma decoupling (AICD) and a learnable SRF-guided HIE mechanism, mimicking the physical image formation, and thus injecting physics-aware reasoning into neural networks, turning an ill-posed problem into a constrained, interpretable task. We further introduce an unsupervised high-fidelity spectral regularization by incorporating a pre-trained spectral diffusion model (SDM) to regularize the coarsely estimated HSI signal from the SRF-guided HIE mechanism with high-fidelity real-world spectral distributions. Extensive experiments on both spectral reconstruction and HSI classification demonstrate that Diff-Spectra significantly outperforms SOTA methods. Future work will focus on the investigation of the spectral distribution gap between the HSI estimated by the SRF-guided HIE and the distribution learned by the SDM, such as quantizing each distribution into a dictionary and calculating their distance.

References

[1] Farid Melgani and Lorenzo Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8):1778–1790, 2004.

[2] Yuan Yuan, Xiangtao Zheng, and Xiaoqiang Lu. Hyperspectral image superresolution by transfer learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(5):1963–1974, 2017.

[3] William R Johnson, Daniel W Wilson, Wolfgang Fink, Mark S Humayun, and Gregory H Bearman. Snapshot hyperspectral imaging in ophthalmology. *Journal of biomedical optics*, 12(1):014036, 2007.

[4] Guolan Lu and Baowei Fei. Medical hyperspectral imaging: a review. *Journal of biomedical optics*, 19(1):010901, 2014.

[5] Telmo Adão, Jonáš Hruška, Luís Pádua, José Bessa, Emanuel Peres, Raul Morais, and Joaquim Joao Sousa. Hyperspectral imaging: A review on uav-based sensors, data processing and applications for agriculture and forestry. *Remote sensing*, 9(11):1110, 2017.

[6] Hui Huang, Li Liu, and Michael O Ngadi. Recent developments in hyperspectral imaging for assessment of food quality and safety. *Sensors*, 14(4):7248–7276, 2014.

[7] Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, Radu Timofte, and Luc Van Gool. Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 745–755, 2022.

[8] Xingxing Yang, Jie Chen, and Zaifeng Yang. Hyperspectral image reconstruction via combinatorial embedding of cross-channel spatio-spectral clues. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 6567–6575, 2024.

[9] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *European conference on computer vision*, 2016.

[10] Shipeng Zhang, Lizhi Wang, Lei Zhang, and Hua Huang. Learning tensor low-rank prior for hyperspectral image reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12006–12015, 2021.

[11] Lizhi Wang, Tao Zhang, Ying Fu, and Hua Huang. Hyperreconnet: Joint coded aperture optimization and image reconstruction for compressive hyperspectral imaging. *IEEE Transactions on Image Processing*, 28(5):2257–2270, 2018.

[12] Lizhi Wang, Chen Sun, Ying Fu, Min H Kim, and Hua Huang. Hyperspectral image reconstruction using a deep spatial-spectral prior. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8032–8041, 2019.

[13] Xingxing Yang, Jie Chen, and Zaifeng Yang. Cooperative colorization: Exploring latent cross-domain priors for nir image spectrum translation. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 2409–2417, 2023.

[14] Zhiwei Xiong, Zhan Shi, Huiqun Li, Lizhi Wang, Dong Liu, and Feng Wu. Hscnn: Cnn-based hyperspectral image recovery from spectrally undersampled projections. In *Proceedings of the IEEE/CVF international conference on computer vision workshops*, pages 518–525, 2017.

[15] Yigit Baran Can and Radu Timofte. An efficient cnn for spectral reconstruction from rgb images. *arXiv preprint arXiv:1804.04647*, 2018.

[16] Antonio Robles-Kelly. Single image spectral reconstruction for multimedia applications. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 251–260, 2015.

[17] Bernhard Hill. Color capture, color management, and the problem of metamerism: does multispectral imaging offer the solution? In *Color Imaging: Device-Independent Color, Color Hardcopy, and Graphic Arts V*, volume 3963, pages 2–14. SPIE, 1999.

[18] Samuel Ortega, Martin Halicek, Himar Fabelo, Gustavo M Callico, and Baowei Fei. Hyperspectral and multispectral imaging in digital and computational pathology: a systematic review. *Biomedical Optics Express*, 11(6):3195–3233, 2020.

[19] Fumihito Yasuma, Tomoo Mitsunaga, Daisuke Iso, and Shree K Nayar. Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum. *IEEE transactions on image processing*, 19(9):2241–2253, 2010.

[20] Zhan Shi, Chang Chen, Zhiwei Xiong, Dong Liu, and Feng Wu. Hscnn+: Advanced cnn-based hyperspectral recovery from rgb images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 939–947, 2018.

[21] Biebele Joslyn Fubara, Mohamed Sedky, and David Dyke. Rgb to spectral reconstruction via learned basis functions and weights. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 480–481, 2020.

[22] Nan Wang, Shaohui Mei, Yifan Zhang, Mingyang Ma, and Xiangqing Zhang. Hyperspectral image reconstruction from rgb input through highlighting intrinsic properties. *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[23] Florinel-Alin Croitoru, Vlad Hondru, Radu Tudor Ionescu, and Mubarak Shah. Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9):10850–10869, 2023.

[24] Yuchun Miao, Lefei Zhang, Liangpei Zhang, and Dacheng Tao. Dds2m: Self-supervised denoising diffusion spatio-spectral model for hyperspectral image restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12086–12096, 2023.

[25] Li Pang, Xiangyu Rui, Long Cui, Hongzhong Wang, Deyu Meng, and Xiangyong Cao. Hir-diff: Unsupervised hyperspectral image restoration via improved diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3005–3014, 2024.

[26] Chanyue Wu, Dong Wang, Yunpeng Bai, Hanyu Mao, Ying Li, and Qiang Shen. Hsr-diff: Hyperspectral image super-resolution via conditional diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7083–7093, 2023.

[27] Jianjun Liu, Zebin Wu, and Liang Xiao. A spectral diffusion prior for unsupervised hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[28] Ning Chen, Jun Yue, Leyuan Fang, and Shaobo Xia. Spectraldiff: A generative framework for hyperspectral image classification with diffusion models. *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[29] Neetu Sigger, Tuan T Nguyen, and Gianluca Tozzi. Brain tissue classification in hyperspectral images using multistage diffusion features and transformer. *Journal of Microscopy*, 2024.

[30] Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977.

[31] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12504–12513, 2023.

[32] Zhenqi Fu, Yan Yang, Xiaotong Tu, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Learning a simple low-light image enhancer from paired low-light instances. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22252–22261, 2023.

[33] Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, and Jiayi Ma. Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12302–12311, 2023.

[34] Zhen Zhao, Zhizhong Zhang, Xin Tan, Jun Liu, Yanyun Qu, Yuan Xie, and Lizhuang Ma. Rethinking gradient projection continual learning: Stability/plasticity feature space decoupling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3718–3727, 2023.

[35] Zhongze Wang, Haitao Zhao, Jingchao Peng, Lujian Yao, and Kaijie Zhao. Odcr: Orthogonal decoupling contrastive regularization for unpaired image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25479–25489, 2024.

[36] Keli Deng, Yuntao Qian, Jie Nie, and Jun Zhou. Diffusion model based hyperspectral unmixing using spectral prior distribution. *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[37] Virginia Klema and Alan Laub. The singular value decomposition: Its computation and some applications. *IEEE Transactions on automatic control*, 25(2):164–176, 1980.

[38] Carlos Eduardo Rodríguez-Pardo and Gaurav Sharma. Geometry of multiprimary display colors ii: Metameric control sets and gamut tiling color control functions. *IEEE Access*, 9:96912–96929, 2021.

[39] Ali Alsam and Reiner Lenz. Calibrating color cameras using metameric blacks. *Journal of the Optical Society of America A*, 24(1):11–17, 2006.

[40] Lizhi Wang, Zhiwei Xiong, Guangming Shi, Feng Wu, and Wenjun Zeng. Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(10):2104–2111, 2016.

[41] Jize Xue, Yong-Qiang Zhao, Yuanyang Bu, Wenzhi Liao, Jonathan Cheung-Wai Chan, and Wilfried Philips. Spatial-spectral structured sparse low-rank representation for hyperspectral image super-resolution. *IEEE Transactions on Image Processing*, 30:3084–3097, 2021.

[42] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[43] Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, et al. Ntire 2022 spectral recovery challenge and data set. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 863–881, 2022.

[44] Fabio Dell'Acqua, Paolo Gamba, Alessio Ferrari, Jon Aevar Palmason, Jón Atli Benediktsson, and Kolbeinn Árnason. Exploiting spectral and spatial information in hyperspectral urban data with high resolution. *IEEE Geoscience and Remote Sensing Letters*, 1(4):322–326, 2004.

[45] Jiaojiao Li, Chaoxiong Wu, Rui Song, Yunsong Li, and Fei Liu. Adaptive weighted attention network with camera spectral sensitivity prior for spectral reconstruction from rgb images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 462–463, 2020.

[46] Xiaowan Hu, Yuanhao Cai, Jing Lin, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Hdnet: High-resolution dual-domain learning for spectral compressive imaging. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17542–17551, 2022.

[47] Liangyu Chen, Xin Lu, Jie Zhang, Xiaojie Chu, and Chengpeng Chen. Hinet: Half instance normalization network for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 182–192, 2021.

[48] Yuanhao Cai, Jing Lin, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17502–17511, 2022.

[49] Miaoyu Li, Ying Fu, Ji Liu, and Yulun Zhang. Pixel adaptive deep unfolding transformer for hyperspectral image reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12959–12968, 2023.

[50] Zeyu Cai, Zheng Liu, Jian Yu, Ziyu Zhang, Feipeng Da, and Chengqian Jin. Reversible-prior-based spectral-spatial transformer for efficient hyperspectral image reconstruction. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 20(1):1–22, 2024.

[51] Zhiyang Yao, Shuyang Liu, Xiaoyun Yuan, and Lu Fang. Specat: Spatial-spectral cumulative-attention transformer for high-resolution hyperspectral image reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25368–25377, 2024.

[52] Graham D Finlayson and Peter Morovic. Metamer sets. *JOSA A*, 22(5):810–819, 2005.

[53] 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.

[54] Boaz Arad, Radu Timofte, Ohad Ben-Shahar, Yi-Tun Lin, and Graham D Finlayson. Ntire 2020 challenge on spectral reconstruction from an rgb image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 446–447, 2020.

[55] Magnus Magnusson, Jakob Sigurdsson, Sveinn Eirikur Armansson, Magnus O Ulfarsson, Hilda Deborah, and Johannes R Sveinsson. Creating rgb images from hyperspectral images using a color matching function. In *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*, pages 2045–2048, 2020.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction outline the key challenges addressed, our contributions, and a summary of the experimental findings.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We include the limitation of our work in Section 5.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

Justification:

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We have described the experimental details and submitted our code together with the submission. The code will be made publicly available upon acceptance.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We used the public dataset and submitted our code.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: See Section 4.1.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: Due to the limited computational resources, we conducted each experiment only once.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).

- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: See Section 4.1.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We fully adhere to the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: See Appendix **??**.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The paper properly credits the creators or original owners of the assets used, such as code, data, and models. We also have correctly cited the relevant literature.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: We have provided the source code of our proposed model.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.

A Implementation Details and Datasets.

A.1 Implementation Details

Three-stage Training Scheme. We implement our proposed Diff-Spectra in the PyTorch framework. Specifically, our proposed Diff-Spectra adopts a three-stage training strategy:

- **Stage 1**, we follow [7] that utilizes RGB-HSI pairs from the training dataset to pre-train the SRF-guided HIE network $\mathcal{F}(\cdot)$, fuse-and-conv operation $\mathcal{H}(\cdot)$, and AICD module $\mathcal{G}(\cdot)$ jointly, optimized with the corresponding loss (*i.e.*, Eq. 10 in the main content), while freezing the parameters of the SDM;
- **Stage 2**, we follow the standard diffusion model DDPM [42] that utilizes HSI from the training dataset to pre-train the SDM, optimized with the loss Eq. 13, while freezing the parameters of the SRF-guided HIE;
- **Stage 3**, to align the distribution gap between the estimated HSI from the SRF-guided HIE network and the learned spectral prior from SDM, we use Eq. 14 in the main content to fine-tune the SRF-guided HIE network and AICD module for 100 epochs, while freezing the SDM. Note that the SDM is treated as a regularizer in this stage. Empirically, we set the learning rate to 1×10^{-4} and the batch size is 20.

Test-time Adaptation. The pre-trained SRF-guided HIE is applied to RGB images from the testing dataset to generate an initial coarse-level HSI signal $\tilde{\mathbf{Y}}$. Next, we treat $\tilde{\mathbf{Y}}$ as trainable parameters, and a spectrum $\tilde{\mathbf{y}}$ is sampled from $\tilde{\mathbf{Y}}$. We assume each sampled spectrum satisfies the spectral distribution learned by the SDM, *i.e.*, $\tilde{\mathbf{y}} = \mathbf{y}_s \sim q(\mathbf{y}_0)$, where $s \in (0, T)$ and T is the time step trained in **Stage 2**. Now, we can use Eq. 13 (in the main content) to refine the coarse-level HSI signal $\tilde{\mathbf{Y}}$ during the sampling process of SDM. However, it is impractical to optimize Eq. 13 for all t due to the inherent distribution difference between the HSI signal generated by SRF-guided HIE and the spectral distribution learned by SDM. Thus, we perform the gradient update $\mathbf{K}$ times in each time step t, which we named as the inner loop optimization.

A.2 Datasets.

ARAD-1K Dataset [43]. The ARAD-1K dataset includes 950 RGB-HSI pairs, with 900 for training and 50 for validation, at a 482×512 resolution across 31 spectral channels (400–700nm). This dataset not only stands as the largest collection available for HSI reconstruction tasks but also integrates content from preceding compilations, notably the NTIRE 2020 HSI dataset [54]. Each HSI in this collection is captured with a spatial resolution of 482×512, spanning 31 spectral channels ranging from 400nm to 700nm.

ICVL Dataset [9]. The ICVL dataset contains 201 HSIs with a resolution of 1300×1392. As it lacks the provision of aligned RGB images, we use the spectral sampling method proposed by Magnusson et al. [55] to generate the corresponding RGB images. Given that 18 of these images have different resolutions, we leverage the remaining 183 image pairs that maintain resolution consistency, allocating 147 pairs for training and 36 for testing.

Indian Pine Dataset [1]. The Indian Pine dataset records the landscape over an area in North-Western Indiana, USA, using the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor. It comprises 145×145 pixels, each with a ground sampling distance (GSD) of 20 meters, and encompasses 220 spectral bands that span the wavelength range from 400 nanometers to 2500 nanometers, achieving a spectral resolution of 10 meters. Following the elimination of 20 bands characterized by noise and water absorption, 200 spectral bands were kept, specifically bands 1-103, 109-149, and 164-219. The scene under investigation features 16 primary categories, including corn, oats, buildings, etc.

Pavia University Dataset [44]. It is collected by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor, which surveyed the area surrounding Pavia University in Pavia, Italy. Capable of capturing 103 spectral bands that range from 430 nanometers to 860 nanometers, the resulting image is composed of 610×340 pixels, each with a ground sampling distance (GSD) of 1.3 meters. This particular scene encompasses 9 distinct land cover classes, including asphalt, grass, trees, etc.

Table 4: Quantitative comparison of different methods in terms of the accuracy for each class, as well as the overall performance using metrics - Overall Accuracy (OA), Average Accuracy (AA), and Kappa coefficient (κ) on the Indian Pines dataset. The best one is shown in bold.

Class No.	*AWAN [45]*	*HDNet [46]*	*HINet [47]*	*SPECAT [51]*	*MST-L [48]*	*MST++ [7]*	*PADUT [49]*	***Diff-Spectra (Ours)***
1	54.16	55.69	**72.74**	61.71	66.48	68.32	70.15	67.41
2	40.21	57.38	71.47	74.24	72.34	74.35	75.18	**78.72**
3	73.82	82.14	91.74	91.38	95.54	90.16	94.27	**96.46**
4	85.68	84.17	90.67	90.05	96.24	85.48	83.74	**96.72**
5	81.11	78.63	93.45	83.48	85.21	91.52	92.45	**94.01**
6	96.75	96.03	**98.42**	94.17	96.72	95.92	97.22	94.94
7	66.31	76.54	72.81	73.88	76.18	77.51	**80.92**	73.48
8	48.38	59.44	65.59	64.02	59.18	60.17	62.30	**67.44**
9	44.69	63.19	71.32	72.42	**80.03**	68.27	65.27	70.02
10	96.70	95.88	**100.00**	95.32	**100.00**	**100.00**	**100.00**	98.87
11	73.14	89.57	85.11	90.54	**91.26**	84.58	85.52	85.77
12	17.25	55.42	82.47	89.27	89.81	85.92	84.88	**90.25**
13	90.44	98.31	**100.00**	98.93	94.42	**100.00**	**100.00**	**100.00**
14	32.27	57.45	40.25	87.34	76.54	62.51	67.24	**91.91**
15	81.82	81.82	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
16	40.00	**100.00**	80.00	**100.00**	**100.00**	80.00	**100.00**	**100.00**
OA (%)	58.26	70.41	73.02	77.32	76.36	74.15	75.85	**78.14**
AA (%)	63.92	76.98	82.25	85.42	86.25	82.79	84.95	**87.88**
κ	0.5161	0.5918	0.6844	0.7088	0.7304	0.7008	0.7344	**0.7571**

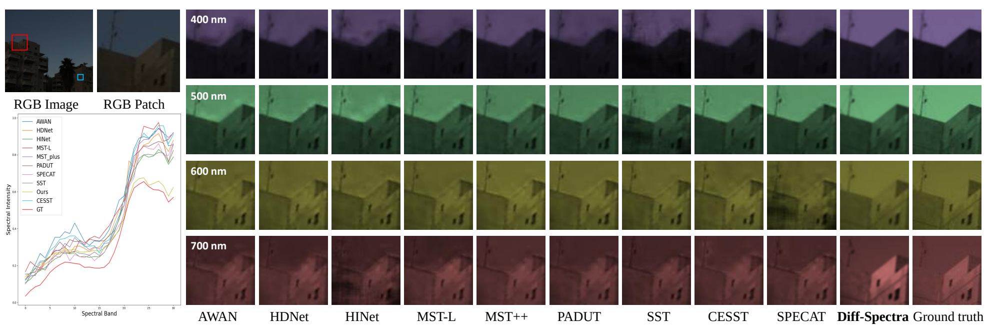

Figure 7: Visual comparisons on a randomly selected scene from the validation set of the ARAD-1K dataset with 4 spectral channels. The RGB patch corresponds to the selected red box of the RGB image. The spectral curves (bottom-left) correspond to the selected blue Box of the RGB image. Please zoom in for a better comparison.

B Additional Experiment Results

Hyperspectral Image Classification. We further provide more quantitative comparisons on the Indian Pines dataset in Table 4. In this table, we further add three methods for more generalized comparisons, including HDNet [46], HINet [47], and SPECAT [51]. As can be seen, our method outperforms existing methods on most classes, and enjoys the best performance over the three metrics: Overall Accuracy (OA), Average Accuracy (AA), and Kappa coefficient (κ).

Hyperspectral Image Reconstruction. We provide more visual comparisons in Fig. 7 and Fig. 8. Specifically, Fig. 7 illustrates the false color results on four spectrum bands, 400nm, 500nm, 600nm and 700nm, respectively, which is the *"ARAD-1K-0901"* image chosen from the validation set of ARAD-1K dataset. In the top-left corner of Fig. 7, the RGB patch corresponds to the selected red box of the RGB image. The spectral curves (bottom left) correspond to the selected blue Box of the RGB image. Please zoom in for a better comparison. Fig. 8 illustrates the false color images on four spectrum bands, 440nm, 500nm, 600nm and 700nm, respectively, which is the *"nachal-0823-1147"* image chosen from the testing set of ICVL dataset. In the top-left corner of Fig. 8, the RGB patch corresponds to the selected red box of the RGB image. The spectral curves (bottom left) correspond to the selected blue Box of the RGB image. Please zoom in for a better comparison. As can be seen, our method can recover more precise texture information and better pixel-level smoothness over other SOTA methods. In addition, both spectral intensity curves in Fig. 7 and Fig. 8 show that our method can recover more precise spectral values and the spectral distribution of our method is closer to the ground truth compared with existing methods, especially in the long-wavelength spectrum (*e.g.*, from

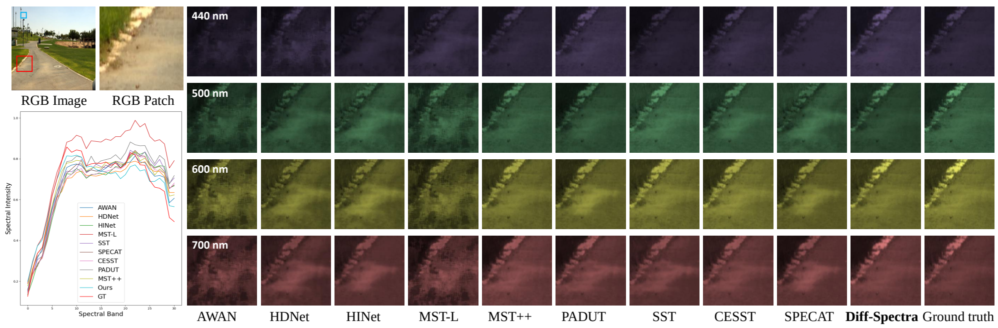

Figure 8: Visual comparisons on a randomly selected scene from the validation set of the ICVL HSI dataset with 4 spectral channels. The RGB patch corresponds to the selected red box of the RGB image. The spectral curves (bottom left) correspond to the selected blue Box of the RGB image. Please zoom in for a better comparison.

$600nm - 700nm$), which is benefited by both image-level prior and spectral-level prior introduced in our model.