

# FLOW-MATCHING GUIDED DEEP UNFOLDING FOR HYPERSPPECTRAL IMAGE RECONSTRUCTION

Anonymous authors

Paper under double-blind review

## ABSTRACT

Hyperspectral imaging (HSI) provides rich spatial-spectral information but remains costly to acquire due to hardware limitations and the difficulty of reconstructing three-dimensional data from compressed measurements. Although compressive sensing systems such as CASSI improve efficiency, accurate reconstruction is still challenged by severe degradation and loss of fine spectral details. We propose the *Flow-Matching-guided Unfolding network* (FMU), which, to our knowledge, is the first to integrate flow matching into HSI reconstruction by embedding its generative prior within a deep unfolding framework. To further strengthen the learned dynamics, we introduce a mean velocity loss that enforces global consistency of the flow, leading to a more robust and accurate reconstruction. This hybrid design leverages the interpretability of optimization-based methods and the generative capacity of flow matching. Extensive experiments on both simulated and real datasets show that FMU significantly outperforms existing approaches in reconstruction quality. Code and models will be available.

## 1 INTRODUCTION

Hyperspectral Imaging (HSI) captures information across numerous spectral bands, providing richer spectral and spatial details compared to traditional RGB images. This enhanced information is valuable for diverse applications, including material characterization, environmental monitoring, remote sensing, and medical imaging. (Li et al., 2019; 2020; Uzkent et al., 2017; Van Nguyen et al., 2010; Rao et al., 2022; Goetz et al., 1985; Lu et al., 2020; Stuart et al., 2019; Rajabi et al., 2024; Khan et al., 2022; Lu & Fei, 2014; ul Rehman & Qureshi, 2021).

Despite the benefits, acquiring hyperspectral images is challenging due to the need for high-performance sensors and time-intensive data collection across many spectral bands. The resulting high acquisition costs limit the scalability of hyperspectral imaging. Recent advancements, such as Compressive Sensing-based Coded Aperture Snapshot Spectral Imaging (CASSI) systems (Gehm et al., 2007), allow for the acquisition of compressed hyperspectral images in a single snapshot, improving collecting efficiency. However, reconstructing accurate 3D hyperspectral images from compressed measurements remains a significant challenge.

HSI reconstruction is an ill-posed inverse problem. Various techniques, including classical model-based methods (Bioucas-Dias & Figueiredo., 2007; Yuan, 2016) and modern learning-based methods (Charles et al., 2011; Meng et al., 2020b; Miao et al., 2019), have been proposed. One promising approach is the Deep Unfolding Network (DUN), which combines convex optimization and powerful neural network priors, providing both interpretability and the power of modern learning-based methods, thus achieving state-of-the-art (SOTA) performance (Cai et al., 2022c; Wang et al., 2022; Dong et al., 2023; Li et al., 2023; Meng et al., 2023).

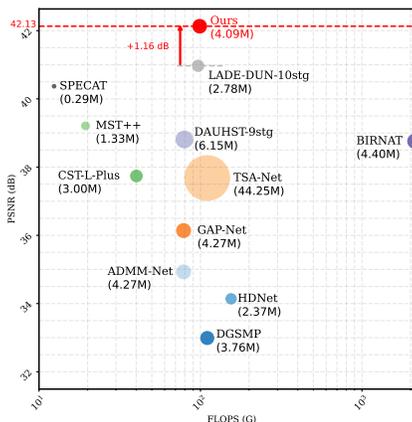


Figure 1: PSNR-FLOPS comparison with recent HSI reconstruction methods. Input size is  $256 \times 256$  for FLOPs computation.

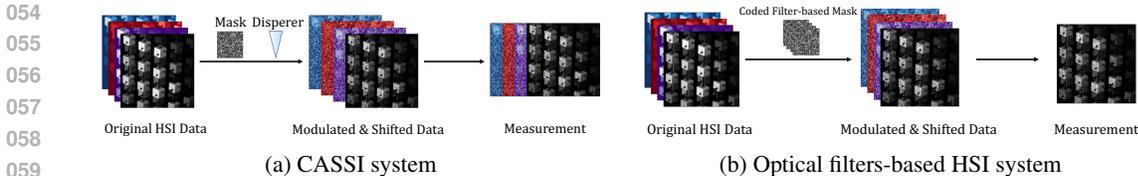


Figure 2: System Overview: Classical CASSI vs. Optical Filters-Based HSI Systems

Unlike typical restoration tasks (e.g., super-resolution or deblurring), CASSI reconstruction involves additional complexities such as spectral compression, physical modulation, and noise. These factors not only increase the dimensionality gap between measurements and the target signal but also amplify the reconstruction difficulty. The denoising module in unfolding frameworks is critical, but the ill-posed nature of the problem limits performance, thus requiring stronger denoisers.

Another limitation of regression-based methods lies in their inability to recover fine details. Standard regression losses often suppress high-frequency components, producing overly conservative reconstructions. Hence, enhancing detail recovery in CASSI remains an open challenge, which requires approaches that can better preserve high-frequency information while mitigating degradation.

To address these challenges, we introduce flow matching prior in this paper to guide reconstruction in an unfolding framework. During the training phase, we train flow matching to extract prior information conditioned on the compressed measurements with a pretrained encoder which learned clean hyperspectral images knowledge. To incorporate this knowledge into the reconstruction process, the prior is injected into the denoising modules of the unfolding network via a prior-guided Transformer architecture. Flow matching prior allows FMU to benefit simultaneously from external prior knowledge derived from clean HSIs and the strong generative capabilities of Flow Matching, ultimately boosting reconstruction quality. The main contributions of this work are summarized as follows:

- We propose a flow-matching guided unfolding network (FMU) for hyperspectral image reconstruction, where priors are generated by flow-matching conditioned on compressed measurements to facilitate high-quality reconstruction. To the best of our knowledge, this is the first attempt to investigate physics-driven deep unfolding with flow matching in HSI reconstruction.
- We introduce a mean velocity constraint to ensure flow consistency during reconstruction. This key contribution strengthens the generative prior from flow matching, addressing potential instability in the flow prediction and leading to more robust hyperspectral image reconstruction.
- Relative to previous state-of-the-art, our method substantially improves reconstruction quality, reaching 42.13 dB PSNR in simulation datasets. It enhances overall fidelity and also better recovers high-frequency details and fine textures, enabling more reliable hyperspectral reconstructions.
- We further benchmark previous works and our model on optical filter-based hyperspectral imaging systems (e.g., liquid crystal tunable filter-based HSI (Marois et al., 2023) and Fabry-Perot filters-based HSI (Xiong et al., 2022)) and CASSI system. The results demonstrate that our algorithm provides robust and efficient reconstruction performance, suggesting its strong potential for deployment in future compact chip-integrated HSI devices with different imaging systems.

## 2 RELATED WORK

### 2.1 HYPERSPECTRAL IMAGING SYSTEM

Since its inception, Hyperspectral Imaging (HSI) technology has undergone significant and rapid development, leading to various practical modalities such as pushbroom, whiskbroom, and snapshot systems (Ortega et al., 2019; Uto et al., 2016; Baek et al., 2017). Notably, snapshot-based HSI, which employs a coded mask for compressed image acquisition in the spatial-spectral domain, offers high resolution across temporal, spatial, and spectral dimensions. As illustrated in Fig. 2a, traditional CASSI systems separate the encoding of the spatial and spectral domains, which results in larger system sizes and lower resilience to shocks and vibrations. Filter-based HSI systems (see Fig. 2b) utilize broadband optical filters to encode both the spatial and the spectral domains with a single mask. These systems not only provide high light throughput and fine resolution, but also simplify the optical path of the imaging system, making it possible to integrate them onto a chip (Monakhova et al., 2020; Yako et al., 2023). Therefore, investigating hyperspectral reconstruction techniques in optical filter-based HSI systems is essential to advance the miniaturization and practical deployment of next-generation spectrometers (Yang et al., 2021).

## 2.2 HYPERSPECTRAL IMAGE RECONSTRUCTION

Before the application of deep learning, traditional hyperspectral image reconstruction methods mainly relied on physics-based optimization frameworks. These methods modeled the reconstruction process as an inverse problem and solved it using convex optimization algorithms (Wagadarikar et al., 2008; Wang et al., 2016). They utilized priors such as sparsity (Kittle et al., 2010) and low-rankness (Liu et al., 2018) to regularize the problem and improve reconstruction quality. While these methods are interpretable and robust in some cases, they are highly sensitive to parameter selection and computationally expensive, limiting their scalability for large-scale data.

With the advent of deep learning, researchers began exploring convolutional neural networks (CNNs) and Transformer-based models for hyperspectral image reconstruction. Early deep learning approaches (Chan et al., 2016; Chen et al., 2023; 2024; Qiu et al., 2021) incorporated pre-trained denoising networks into optimization frameworks (Plug-and-Play, PnP), which provided strong feature representation but lacked adaptability for complex noise and compressed measurements, limiting their effectiveness in noisy or low-quality data.

End-to-end deep learning methods, particularly CNNs, have become popular for directly mapping compressed measurements to complete hyperspectral images, leveraging global feature learning instead of relying on handcrafted features (Cheng et al., 2022; Hu et al., 2022). However, CNNs struggle with capturing long-range dependencies, which are important in hyperspectral data because of the complex relationships between spectral bands.

To overcome CNN limitations, Cai et al. (2022a;b) introduced transformer-based models, which excel at modeling non-local dependencies and capturing complex spatial-spectral relationships. However, these methods often lack physical interpretability and are sensitive to system configuration, leading to challenges in generalization and robustness, especially under degradation.

More recently, Deep Unfolding Networks (DUN) have emerged, combining physics-driven optimization with data-driven deep learning (Cai et al., 2022c; Wang et al., 2022). DUN models map iterative optimization steps to neural network layers, replacing prior components with learnable denoisers, offering the advantages of both interpretability and deep learning flexibility. Despite their success, DUNs still face challenges, particularly in recovering high-frequency details and efficiently aggregating features from compressed measurements. These limitations motivate the exploration of stronger priors and more expressive architectures to further enhance reconstruction performance.

## 2.3 FLOW MATCHING

Flow Matching (FM) has recently emerged as a promising paradigm for generative modeling. Its core idea is to directly fit continuous flow paths between probability densities, thereby learning a continuous-time dynamical system. Unlike diffusion models (Ho et al., 2020) that rely on a fixed forward process, flow matching offers greater flexibility and interpretability, supporting efficient density modeling and high-quality sample generation. This makes flow matching a natural alternative to diffusion when both efficiency and theoretical clarity are desired.

Lipman et al. (2022) first proposed FM by fitting optimal transport maps along random paths, constructing a path-independent vector field to approximate the target distribution. Subsequent work extended FM to practical applications: Rectified Flow (Liu et al., 2022) introduced path correction for stable image generation, while FM-OT (Kornilov et al., 2024) combined FM with optimal transport to improve efficiency and generalization. These developments demonstrate the versatility of FM in bridging theoretical advances with real-world generative tasks.

Beyond generative tasks, FM has also been successfully applied to trajectory modeling and probabilistic differential equations, demonstrating strong adaptability across diverse domains. These applications highlight its ability to capture complex temporal dependencies and stochastic behaviors, further broadening its utility beyond image or signal generation. Compared to diffusion models such as DDPM (Ho et al., 2020) and Score-based Models (Song et al., 2020), FM achieves faster inference and provides more explicit control over path evolution.

Overall, FM pushes generative modeling toward greater efficiency and interpretability. With its flexible path modeling, end-to-end optimization, and strong generative performance, FM shows broad potential in areas including image synthesis, speech, and physical modeling. This broad applicability highlights its promise as a foundation for next-generation generative frameworks.

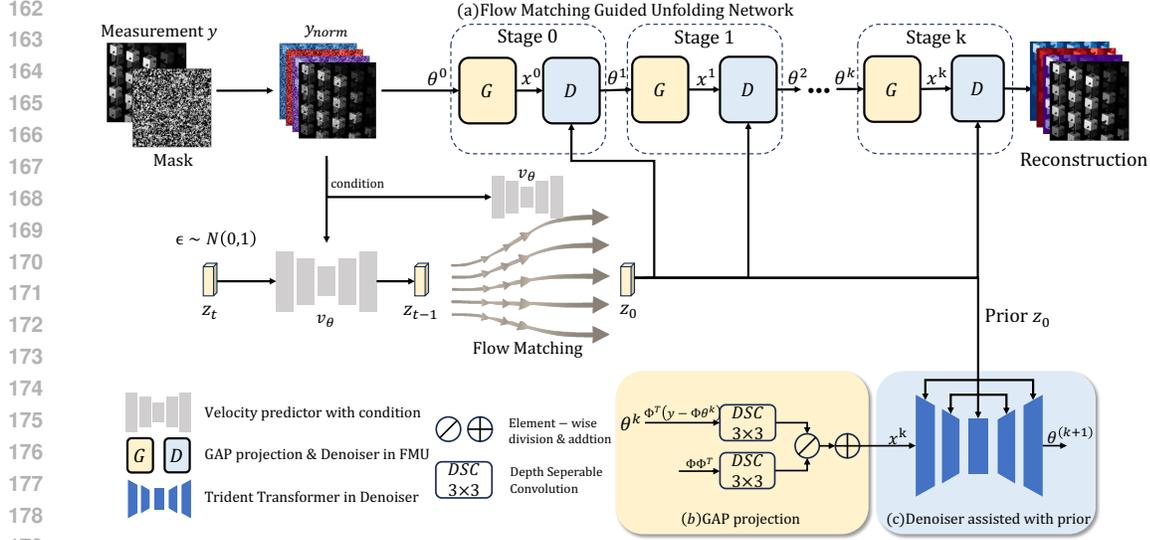


Figure 3: Overall pipeline of our method. The measurement  $\mathbf{y}$  passes through FMU with  $N$  stages and finally get the output reconstruction. In each stage, there is a GAP projection and a U-shaped denoiser consists of Trident Transformers, which is assisted with the prior from flow matching.

### 3 METHOD

#### 3.1 MODELING OF THE HSI SYSTEM

The CASSI architecture acquires hyperspectral information by encoding spectral slices and projecting them onto a 2D detector. Let the hyperspectral cube be  $\mathbf{X} \in \mathbb{R}^{W \times H \times N_\lambda}$ , where  $W$ ,  $H$ , and  $N_\lambda$  denote spatial width, height, and the number of spectral channels. A mask  $\mathbf{m} \in \mathbb{R}^{W \times H}$  modulates each slice, and after shifting with step  $d$ , we obtain the spatial-spectral mask  $\mathbf{M}_{\text{CASSI}} \in \mathbb{R}^{W \times (H + (N_\lambda - 1)d) \times N_\lambda}$ . The measurement is then calculated as follows

$$\mathbf{Y}_{\text{CASSI}} = \sum_{n_\lambda=1}^{N_\lambda} \text{shift}(\mathbf{X})(:, :, n_\lambda) \odot \mathbf{M}_{\text{CASSI}}(:, :, n_\lambda) + \mathbf{N}_{\text{CASSI}}, \quad (1)$$

where  $\odot$  denotes the Hadamard product,  $\mathbf{N}$  is measurement noise, and the shifted mask is  $\mathbf{M}_{\text{CASSI}}(i, j, n_\lambda) = \mathbf{m}(i, j + (n_\lambda - 1)d)$ .

For optical filter-based systems, a 2D filter mask directly encodes all spectral bands with distinct transmittance rates, yielding

$$\mathbf{Y} = \sum_{n_\lambda=1}^{N_\lambda} \mathbf{X}(:, :, n_\lambda) \odot \mathbf{M}(:, :, n_\lambda) + \mathbf{N}, \quad (2)$$

where  $\mathbf{Y} \in \mathbb{R}^{W \times H}$ ,  $\mathbf{M} \in \mathbb{R}^{W \times H \times N_\lambda}$  is the 3D mask, and  $\mathbf{N}$  denotes noise. Compared with CASSI, this design simplifies the optical path and reduces correlation among spectral responses, thus increasing the burden on reconstruction algorithms.

For both systems, vectorizing the cube and measurements as  $\mathbf{x} = \text{vec}(\mathbf{X}) \in \mathbb{R}^{W \times H \times N_\lambda}$  and  $\mathbf{y} = \text{vec}(\mathbf{Y}) \in \mathbb{R}^{W \times H}$ , the forward model becomes

$$\mathbf{y} = \Phi \mathbf{x} + \boldsymbol{\eta}, \quad (3)$$

where  $\boldsymbol{\eta} \in \mathbb{R}^n$ ,  $n = W \times H$  is the noise and  $\Phi \in \mathbb{R}^{n \times n N_\lambda}$  is the sensing matrix, represented as  $\Phi = [\mathbf{D}_1, \dots, \mathbf{D}_{N_\lambda}]$ ,  $\mathbf{D}_{n_\lambda} = \text{Diag}(\text{vec}(\mathbf{M}(:, :, n_\lambda)))$ . Thus, hyperspectral image reconstruction can be viewed as solving the ill-posed linear inverse problem of reconstructing  $\mathbf{x}$  from  $\mathbf{y}$ .

#### 3.2 PROPOSED MODEL

We propose a flow matching-guided unfolding network (FMU) (Fig. 3). The measurement is processed by an  $N$ -stage deep unfolding network (DUN), comprising a projection and a denoiser in each stage. The denoiser is a U-shaped transformer in which the flow-matching prior is injected to assist denoising; Trident Transformer module from Wu et al. (2024) is used to aggregate high-quality, degradation-free prior features for compensation.

### 3.2.1 FLOW-MATCHING GUIDED UNFOLDING FRAMEWORK

We adopt the generalized alternating projection framework (Liao et al., 2014) as the backbone of our reconstruction method. We can solve Eq. (3) by the objective:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \Phi \mathbf{x}\|^2 + \tau R(\mathbf{x}) \quad (4)$$

The objective enforces consistency between the reconstruction and compressed measurement  $\mathbf{y}$ , while  $\tau R(\mathbf{x})$  serves as the regularization. GAP extends the classical alternating projection framework by enabling systematic projections between convex sets that change in a structured manner, thereby allowing for flexible interruption and efficient resumption. To decouple the measurement-consistency constraint from the prior-regularization step, we introduce an auxiliary variable  $\theta$ , where  $\mathbf{x}$  represents the measurement-consistent solution and  $\theta$  incorporates the prior-driven refinement:

$$(\hat{\mathbf{x}}, \hat{\theta}) = \arg \min_{\mathbf{x}, \theta} \frac{1}{2} \|\mathbf{x} - \theta\|^2 + \tau R(\theta), \quad \text{s.t. } \mathbf{y} = \Phi \mathbf{x}. \quad (5)$$

The update scheme alternates between two operations. Euclidean projection enforces measurement consistency and data fidelity by projecting the intermediate solution onto the feasible set. The prior-guided denoising step then leverages learned priors to refine the reconstruction:

$$\mathbf{x}^{(k+1)} = \theta^{(k)} + \Phi^\dagger(\mathbf{y} - \Phi \theta^{(k)}), \quad \theta^{(k+1)} = \mathcal{F}_{k+1}(\mathbf{x}^{(k+1)}; \mathbf{z}_{FM}). \quad (6)$$

Here,  $\mathcal{F}_{k+1}$  denotes a flow-matching guided denoiser in the (k+1)-th stage, and  $\mathbf{z}_{FM}$  is the flow-matching prior learned from clean HSIs (see Sec. 3.2.2). This prior provides strong generative regularization and enables the model to better capture spectral-spatial correlations, leading to improved recovery even with a limited number of unfolding stages. By combining the projection consistency of GAP with the expressive modeling capability of the flow-matching prior, our method effectively stabilizes the optimization process and accelerates convergence. In contrast to conventional iterative solvers, which require hundreds of iterations, our framework achieves comparable or superior performance within only a few unfolding stages.

Furthermore, following Wu et al. (2024), we incorporate gradient correction into each stage. This learnable adjustment rectifies the imperfect gradient updates caused by the limited number of iterations, thereby stabilizing optimization and yielding faster and more accurate reconstructions.

### 3.2.2 TRAINING STRATEGY OF FMU

During reconstruction in FMU, a flow matching prior is incorporated in denoiser to mitigate the effects of severely degraded measurements. We train the FMU in two phases (see Fig. 4).

**In the first phase**, our goal is to extract prior knowledge from HSIs without degradation. We train an encoder that jointly takes the ground-truth HSIs  $\mathbf{x}$  and the compressed measurement  $\mathbf{y}$  as input. Before encoding, we normalize both  $\mathbf{y}$  and the sensing matrix  $\Phi$ . Specifically,  $\mathbf{y}_{norm} \in \mathbb{R}^{W \times H \times N_\lambda} = \Phi^\dagger \mathbf{y}$ . The input to the encoder is then defined as  $I_E \in \mathbb{R}^{W \times H \times 2N_\lambda} = \text{concat}(\mathbf{y}_{norm}, \mathbf{x})$ , and the latent prior is obtained as  $\mathbf{z}_{LE} \in \mathbb{R}^{n \times c} = \text{LE}(\text{concat}(\mathbf{y}_{norm}, \mathbf{x}))$ , where  $n, c$  denote the feature dimension and channel number, and LE refers to the latent encoder. The encoder first applies a pixel unshuffle to reduce spatial resolution while increasing channel dimensionality, followed by MobileBlocks (Howard et al., 2019) for efficient local feature learning, and an MLP-Mixer module (Tolstikhin et al., 2021) for effective channel-wise and spatial mixing. The detailed architecture of the latent encoder is provided in the supplementary material. The output  $\mathbf{z}_{LE}$  is then used as the prior to guide denoising, leading to the reconstruction  $\hat{\mathbf{x}} = \text{FMU}(\mathbf{y}, \Phi, \mathbf{z}_{LE})$ . In this phase, we adopt the L1 norm as the reconstruction loss:  $\mathcal{L}_1 = \mathcal{L}_{rec} = \|\hat{\mathbf{x}} - \mathbf{x}\|$ .

**In the second phase**, our aim is to train a flow matching to generate a prior conditioned on  $\mathbf{y}_{norm}$ . We fix the encoder LE trained in the first phase and use the output  $\mathbf{z}_{LE}$  as the generative object of flow matching. Specifically, the goal of our flow matching framework is to estimate a coupling  $\pi(p_0, p_1)$ , where  $x_0 \sim p_0$  is the latent feature distribution of  $\mathbf{z}_{LE}$  and  $x_1 \sim p_1$  is a Gaussian noise distribution. The flow is formulated as solving an ordinary differential equation:

$$d\mathbf{x} = v(\mathbf{x}_t, t, \mathbf{y}) dt, \quad (7)$$

on  $t \in [0, 1]$ , where the velocity field  $v : \mathbb{R}^D \times [0, 1] \rightarrow \mathbb{R}^D$  drives the transformation from  $p_0$  to  $p_1$ . We parameterize the velocity as  $v_\theta(\mathbf{x}_t, t, \mathbf{y})$  and estimate the parameters  $\theta$  via a standard

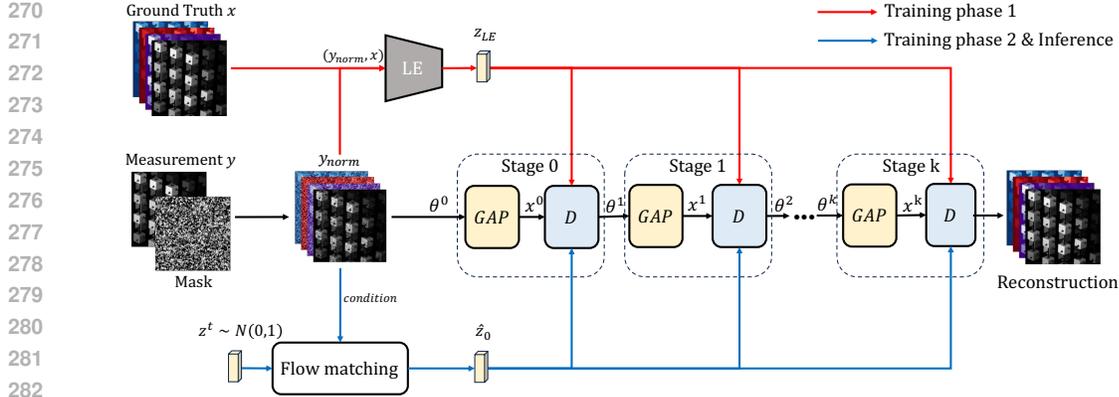


Figure 4: Two-phase training procedure of our method. In first phase, we train a latent encoder to learn knowledge from clean HSIs; and in second phase we fix the latent encoder and train flow matching with the prior from clean HSI to generate prior conditioned on measurement.

least-squares regression (Hastie et al., 2009) approach:

$$\hat{\theta} = \arg \min_{\theta} \mathbb{E}_{t, \mathbf{x}_t} [\|v(\mathbf{x}_t, t, \mathbf{y}) - v_{\theta}(\mathbf{x}_t, t, \mathbf{y})\|_2^2]. \quad (8)$$

Inspired by Dao et al. (2023), we adopt a constant velocity ODE (Ordinary Differential Equation) where the trajectory is defined as  $\mathbf{x}_t = (1-t)\mathbf{x}_0 + t\mathbf{x}_1$ , i.e., a linear interpolation between  $\mathbf{x}_0$  and  $\mathbf{x}_1$ . This yields the simplified flow matching loss:

$$\hat{\theta} = \arg \min_{\theta} \mathbb{E}_{t, \mathbf{x}_t} [\|\mathbf{x}_1 - \mathbf{x}_0 - v_{\theta}(\mathbf{x}_t, t, \mathbf{y})\|_2^2]. \quad (9)$$

The flow matching module integrates over the learned velocity field to generate  $\hat{z}_0$ , which is then employed as a prior to guide the unfolding network for HSI reconstruction. The overall second-stage training objective is given by  $\mathcal{L}_2 = \mathcal{L}_{FM} + \mathcal{L}_{rec}$ , where  $\mathcal{L}_{rec}$  is the reconstruction loss and  $\mathcal{L}_{FM}$  is the flow matching loss. We also introduce a **mean velocity constraint** in  $\mathcal{L}_{FM}$  to improve the consistency of flow matching.  $\mathcal{L}_{FM}$  and details will be described in the next subsection.

### 3.2.3 MEAN VELOCITY CONSTRAINT

Similar to prior works that explored stability in flow matching, such as Rectified Flow (Liu et al., 2022) and FM-OT (Kornilov et al., 2024), we also introduce a regularization mechanism. In our case, we propose a **mean velocity constraint**, which enforces consistency of the predicted flow in expectation, providing global stability complementary to pointwise regression.

Formally, the mean velocity loss is defined as

$$\mathcal{L}_{mean} = \|\mathbb{E}_{t, \mathbf{z}} [v_{\theta}(t, \mathbf{z})] - \mathbb{E}_{t, \mathbf{z}} [v^*(t, \mathbf{z})]\|_2^2, \quad (10)$$

where  $v_{\theta}(t, \mathbf{z})$  denotes the predicted velocity field,  $v^*(t, \mathbf{z})$  is the target velocity field, and  $\mathbb{E}_{t, \mathbf{z}}[\cdot]$  denotes expectation over timesteps and samples.

By incorporating this constraint, the final flow matching objective becomes

$$\mathcal{L}_{FM} = \|\hat{z}_0 - z_{LE}\|_1 + \lambda_{mean} \mathcal{L}_{mean}, \quad (11)$$

where  $\lambda_{mean}$  controls the trade-off between flow matching loss and mean velocity regularization.

## 4 EXPERIMENT

### 4.1 EXPERIMENT SETTINGS

Consistent with Cai et al. (2022b), we employed 28 spectral channels with wavelengths ranging from 450 to 650 nm. Our experiments were conducted on both simulated and real datasets.

**Simulated Optical Filter-based HSI System Data.** We utilized two benchmark datasets, namely CAVE (Park et al., 2007) and KAIST (Choi et al., 2017). The CAVE dataset consists of 32 hyperspectral images, each with a spatial resolution of  $512 \times 512$  pixels. The KAIST dataset provides 30 hyperspectral images at a higher resolution of  $2,704 \times 3,376$  pixels. We employed the CAVE dataset for training. For evaluation, 10 representative scenes from the KAIST dataset were selected. In line

| Method         | Params | GFLOPS  | Scene1          | Scene2          | Scene3          | Scene4          | Scene5          | Scene6          | Scene7          | Scene8          | Scene9          | Scene10         | Avg             |
|----------------|--------|---------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| $\lambda$ -Net | 62.64M | 117.98  | 32.62<br>0.9247 | 31.43<br>0.9003 | 30.49<br>0.9109 | 32.34<br>0.9185 | 31.57<br>0.9391 | 33.44<br>0.9410 | 32.98<br>0.9092 | 29.22<br>0.9075 | 31.33<br>0.9218 | 30.90<br>0.9178 | 31.63<br>0.9191 |
| DGSMP          | 3.76M  | 110.06  | 34.09<br>0.9495 | 30.83<br>0.9188 | 30.41<br>0.9182 | 38.05<br>0.9654 | 32.74<br>0.9584 | 35.20<br>0.9719 | 33.96<br>0.9349 | 31.20<br>0.9510 | 31.39<br>0.9371 | 31.99<br>0.9666 | 32.99<br>0.9472 |
| HDNet          | 2.37M  | 154.76  | 33.73<br>0.9523 | 33.46<br>0.9416 | 34.14<br>0.9555 | 40.18<br>0.9736 | 33.15<br>0.9649 | 35.70<br>0.9713 | 34.71<br>0.9445 | 31.13<br>0.9463 | 32.66<br>0.9495 | 32.53<br>0.9659 | 34.14<br>0.9565 |
| ADMM-Net       | 4.27M  | 78.58   | 35.57<br>0.9622 | 35.22<br>0.9539 | 34.18<br>0.9473 | 39.80<br>0.9716 | 34.67<br>0.9642 | 36.11<br>0.9675 | 34.46<br>0.9353 | 31.82<br>0.9475 | 33.76<br>0.9506 | 33.66<br>0.9698 | 34.93<br>0.9570 |
| BiSRNet        | 0.036M | 1.18    | 36.23<br>0.9628 | 35.79<br>0.9526 | 34.11<br>0.9289 | 40.05<br>0.9563 | 35.55<br>0.9632 | 37.34<br>0.9653 | 35.27<br>0.9381 | 34.44<br>0.9512 | 34.71<br>0.9392 | 34.95<br>0.9660 | 35.84<br>0.9524 |
| GAP-Net        | 4.27M  | 78.58   | 36.35<br>0.9674 | 37.16<br>0.9615 | 35.85<br>0.9568 | 41.17<br>0.9799 | 35.89<br>0.9701 | 36.45<br>0.9747 | 35.90<br>0.9523 | 32.26<br>0.9537 | 35.98<br>0.9609 | 34.39<br>0.9754 | 36.14<br>0.9653 |
| TSA-Net        | 44.25M | 110.06  | 37.87<br>0.9745 | 39.51<br>0.9809 | 36.08<br>0.9641 | 43.72<br>0.9852 | 36.43<br>0.9799 | 38.06<br>0.9821 | 36.71<br>0.9596 | 35.18<br>0.9738 | 37.30<br>0.9762 | 35.89<br>0.9830 | 37.68<br>0.9759 |
| CST-L-Plus     | 3.00M  | 40.10   | 37.42<br>0.9736 | 38.99<br>0.9801 | 37.88<br>0.9697 | 42.41<br>0.9800 | 36.94<br>0.9819 | 38.08<br>0.9828 | 36.88<br>0.9681 | 35.39<br>0.9718 | 36.18<br>0.9720 | 37.19<br>0.9820 | 37.74<br>0.9762 |
| MST++          | 1.33M  | 19.42   | 39.19<br>0.9815 | 41.58<br>0.9871 | 39.40<br>0.9759 | 44.36<br>0.9893 | 38.63<br>0.9865 | 39.15<br>0.9870 | 38.27<br>0.9727 | 35.49<br>0.9756 | 39.33<br>0.9832 | 36.66<br>0.9859 | 39.21<br>0.9825 |
| BIRNAT         | 4.40M  | 2122.66 | 38.72<br>0.9787 | 40.83<br>0.9839 | 39.25<br>0.9757 | 43.46<br>0.9854 | 37.92<br>0.9832 | 38.83<br>0.9838 | 37.30<br>0.9632 | 35.90<br>0.9747 | 38.88<br>0.9799 | 36.50<br>0.9840 | 38.76<br>0.9792 |
| DAUHST-9stg    | 6.15M  | 79.50   | 38.26<br>0.9785 | 40.28<br>0.9850 | 38.26<br>0.9748 | 44.04<br>0.9859 | 37.87<br>0.9869 | 39.23<br>0.9852 | 38.44<br>0.9750 | 35.66<br>0.9746 | 38.62<br>0.9814 | 37.40<br>0.9881 | 38.81<br>0.9815 |
| SPECAT         | 0.29M  | 12.4    | 40.24<br>0.9820 | 42.40<br>0.9860 | 41.43<br>0.9780 | 44.90<br>0.9820 | 39.62<br>0.9870 | 39.90<br>0.9840 | 39.41<br>0.9770 | 37.49<br>0.9770 | 40.45<br>0.9820 | 37.90<br>0.9830 | 40.37<br>0.9860 |
| LADE-DUN-10stg | 2.78M  | 96.69   | 40.10<br>0.9853 | 43.27<br>0.9919 | 41.76<br>0.9849 | 47.55<br>0.9949 | 39.76<br>0.9906 | 40.43<br>0.9910 | 39.94<br>0.9809 | 37.72<br>0.9833 | 41.21<br>0.9897 | 37.94<br>0.9895 | 40.97<br>0.9882 |
| FMU-S (ours)   | 2.78M  | 99.87   | 41.42<br>0.9877 | 44.28<br>0.9927 | 43.29<br>0.9860 | 46.75<br>0.9949 | 40.77<br>0.9918 | 41.00<br>0.9917 | 41.69<br>0.9850 | 37.83<br>0.9828 | 43.11<br>0.9906 | 39.21<br>0.9905 | 41.94<br>0.9894 |
| FMU (ours)     | 4.09M  | 98.84   | 41.84<br>0.9883 | 44.29<br>0.9930 | 43.54<br>0.9869 | 47.14<br>0.9955 | 41.15<br>0.9926 | 41.15<br>0.9922 | 41.65<br>0.9851 | 38.17<br>0.9840 | 42.85<br>0.9910 | 39.49<br>0.9913 | 42.13<br>0.9900 |

Table 1: Quantitative results (PSNR/SSIM) of hyperspectral image reconstruction across ten representative scenes from simulated dataset and the overall average. A variety of state-of-the-art methods are compared to demonstrate the effectiveness of our approach.

with previous work by Yao et al. (2024), both the training and the testing data were processed using an optical filter-based mask to simulate measurement acquisition. In particular, we adopted the Fabry–Perot filter-based mask proposed in Yako et al. (2023), as it serves as a representative design for optical filter-based HSI systems in terms of both structure and performance.

**Real CASSI System Data.** For real hyperspectral data, we adopted a dataset collected using the CASSI system, which was previously introduced in TSA-Net (Meng et al., 2020b). To align with real acquisition conditions, all training samples were generated as simulated CASSI measurements, with additional noise incorporated to approximate the sensor projection process.

**Metrics and Implementation Specifics.** To evaluate the reconstruction performance, we employed peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) as quantitative metrics. The proposed model was implemented in PyTorch (Paszke et al., 2019) and trained on a single NVIDIA A6000 GPU with 48 GB memory. The training process was conducted over 300 epochs with the Adam optimizer ( $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ). A Cosine Annealing schedule was applied to the learning rate, which started from  $4 \times 10^{-4}$  and decreased progressively to a minimum of  $\gamma = 1 \times 10^{-6}$ .

## 4.2 COMPARISONS WITH STATE-OF-THE-ART METHODS

### 4.2.1 SIMULATION EXPERIMENTS RESULTS

To provide a comprehensive evaluation, we compare our method against a wide spectrum of state-of-the-art approaches in hyperspectral image reconstruction. These include CNN-based networks such as TSA-Net (Meng et al., 2020b), DGSMP (Huang et al., 2021), HDNet (Hu et al., 2022), and BiSRNet (Cai et al., 2023) that emphasize local feature extraction, the RNN-based BIRNAT (Cheng et al., 2022) that captures sequential spectral dependencies, Transformer-based architectures like MST++ (Cai et al., 2022b), CST-L-Plus (Cai et al., 2022a), and SPECAT (Yao et al., 2024) that exploit long-range spatial-spectral modeling, and deep unfolding frameworks such as ADMM-Net (Ma et al., 2019), GAP-Net (Meng et al., 2020a), DAUHST (Cai et al., 2022c) and LADE-DUN (Wu et al., 2024). This diverse selection covers the major methodological paradigms and enables a fair comparison with our proposed approach, under identical experimental settings to ensure fairness.

As presented in Tab. 1, our method achieves an average PSNR of 42.13 dB and SSIM of 0.9900, setting a new state-of-the-art in HSI reconstruction. Beyond numerical gains, it also produces reconstructions with sharper structures and more consistent spectral responses, highlighting superior

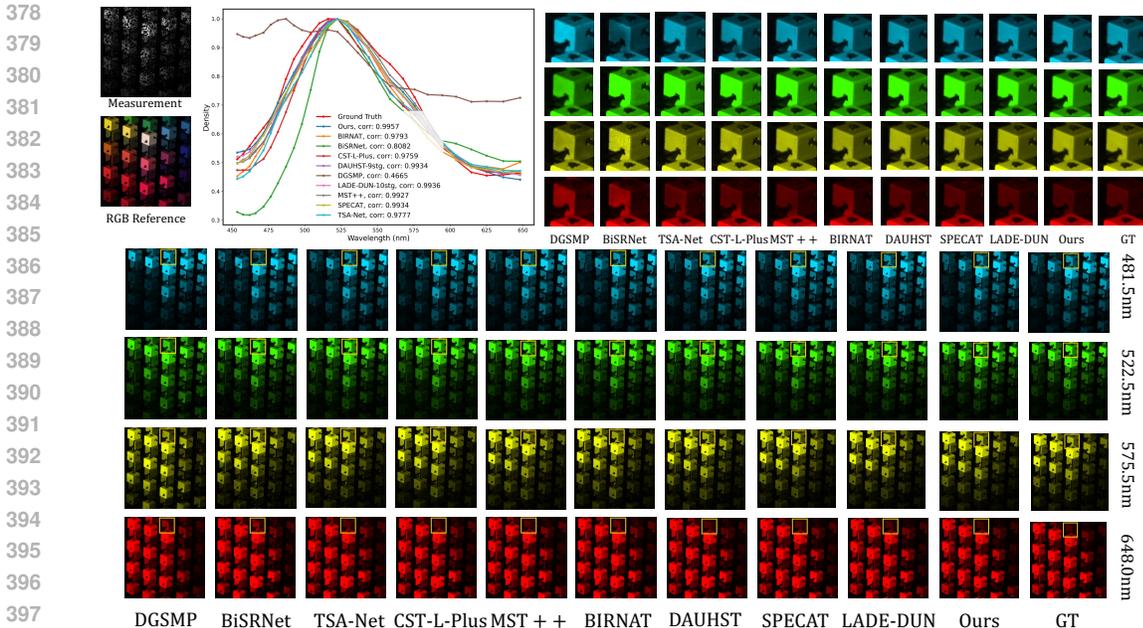


Figure 5: Qualitative comparison on simulated data. From top to bottom, each row visualizes the reconstructed channels at wavelengths of 481.5 nm, 522.5 nm, 575.5 nm, and 648.0 nm in scene 2. The top left part shows the measurement, the RGB reference image of the original HSI and the spectral density curves within the yellow region of interest.

spatial detail and spectral fidelity. In Fig. 5 we compare the reconstructions across four representative spectral channels with color-coded visualizations. Our method yields the closest visual results to the ground truth, with the spectral density curve correlation reaching the highest value of 0.9417. To provide a more comprehensive comparison, we also conducted experiments on the CASSI system. Our method outperforms the previous state-of-the-art (LADE-DUN) by +0.93 dB PSNR and +0.009 SSIM, as shown in the supplementary materials.

In particular, compared to the recent SOTA unfolding-based method LADE-DUN (Wu et al., 2024), which attains a PSNR of 40.97 dB and SSIM of 0.9882 with 2.78M parameters and 96.69 G FLOPS, our method achieves both higher accuracy (42.13 dB / 0.9900) and better spectral fidelity while maintaining a modest parameter scale (4.09 M). Furthermore, we propose a compact variant, FMU-S, whose parameter count is aligned with that of LADE-DUN for a fair comparison. Remarkably, FMU-S still surpasses LADE-DUN, achieving 41.94 dB PSNR and 0.9894 SSIM, thereby demonstrating the effectiveness of our design even under the same parameter budget.

#### 4.2.2 REAL CASSI SYSTEM RESULTS

To assess real-world performance, we apply our method to CASSI-captured measurements without ground-truth HSIs. As shown in Fig. 6, our reconstructions exhibit clearer spatial details, sharper edges, and better cross-channel spectral consistency than competing methods, with fewer artifacts. These results demonstrate the robustness and practicality of our approach for real CASSI data. The improved fidelity and stability suggest strong potential for deployment in real imaging systems, paving the way toward reliable applications of hyperspectral reconstruction in practical scenarios.

#### 4.3 ABLATION STUDY

To further validate the effectiveness of our proposed method, we conduct a series of ablation studies. The results of these studies are summarized in Table 2.

**Effect of Flow Matching.** To evaluate the impact of our proposed flow matching mechanism, we conduct an ablation study comparing the model with and without flow matching. The results, shown in Table 2a, demonstrate a significant performance boost with flow matching. Specifically, the model incorporating flow matching achieves a PSNR of 42.13 dB and SSIM of 0.9900, surpassing the baseline model without flow matching (40.58 dB / 0.9878). And the small FLOPS overhead (98.84 G vs. 96.40 G) highlights that performance gains come at a very modest computational cost.

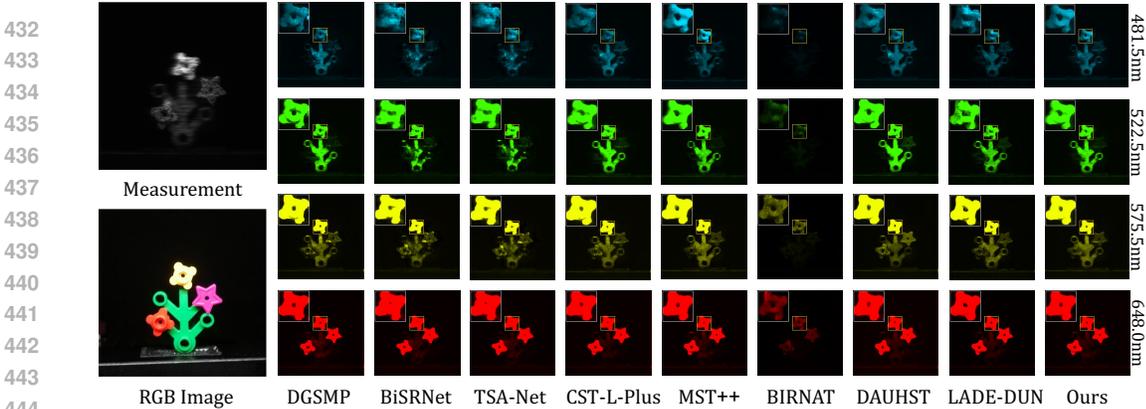


Figure 6: Qualitative comparison on real CASSI data. From top to bottom, each row visualizes the reconstructed channels at wavelengths of 481.5 nm, 522.5 nm, 575.5 nm, and 648.0 nm in scene 1.

| Method               | PSNR (dB)    | SSIM          | FLOPS (G)    |
|----------------------|--------------|---------------|--------------|
| Baseline (no prior)  | 40.58        | 0.9878        | <b>96.40</b> |
| +Latent Diffusion    | 40.97        | 0.9882        | 96.69        |
| +FlowMatching (ours) | <b>42.13</b> | <b>0.9900</b> | 98.84        |

(a) Effect of prior used in the denoiser.

| Variant          | PSNR(dB)     | Params(M)   | FLOPS(G)     |
|------------------|--------------|-------------|--------------|
| MLP              | 41.95        | <b>2.78</b> | 99.87        |
| gMLP             | 42.05        | 4.36        | 103.27       |
| Tiny Transformer | 41.90        | 9.09        | 109.45       |
| SimpleCNN (ours) | <b>42.13</b> | 4.09        | <b>98.84</b> |

(c) Effect of various type of denoisers.

| $\lambda_{mean}$ | 0      | 0.1    | 1      | 5             | 10     | 100    |
|------------------|--------|--------|--------|---------------|--------|--------|
| PSNR (dB)        | 41.90  | 41.97  | 41.99  | <b>42.13</b>  | 42.02  | 41.84  |
| SSIM             | 0.9887 | 0.9898 | 0.9898 | <b>0.9900</b> | 0.9900 | 0.9897 |

(b) Effect of  $\lambda_{mean}$  loss weight.

| Methods    | ADMM         | GAP          | DAUHST | LADE-DUN | <b>FMU</b>    |
|------------|--------------|--------------|--------|----------|---------------|
| PSNR (dB)  | 34.93        | 36.14        | 38.81  | 40.97    | <b>42.13</b>  |
| SSIM       | 0.9570       | 0.9653       | 0.9815 | 0.9882   | <b>0.9900</b> |
| Params (M) | 4.27         | 4.27         | 6.15   | 2.78     | <b>4.09</b>   |
| FLOPS (G)  | <b>78.58</b> | <b>78.58</b> | 79.50  | 96.69    | 98.84         |

(d) Comparison with series of unfolding framework.

Table 2: Ablation studies on simulation datasets with PSNR, SSIM, Params, and FLOPS reported.

**Impact of Mean Velocity Constraint.** We also explore the effect of introducing the mean velocity constraint, controlled by the weight parameter  $\lambda_{mean}$ . The results, as presented in Table 2b, reveal that the model performs best when  $\lambda_{mean} = 5$ , achieving 42.13 dB PSNR and 0.9900 SSIM. Larger or smaller values degrade the results, indicating that proper adjustment of  $\lambda_{mean}$  is crucial for balancing reconstruction accuracy and regularization, ensuring robustness across different settings.

**Denoiser Choice for Flow Matching.** We investigate the effect of different denoisers for velocity prediction. Table 2c compares MLP (Rosenblatt, 1958), gMLP (Liu et al., 2021), Tiny Transformer (Vaswani et al., 2017), and our SimpleCNN. While all variants achieve competitive results, SimpleCNN delivers the best PSNR of 42.13 dB with 4.09 M parameters and the lowest FLOPS (98.84 G), demonstrating a favorable trade-off between efficacy, and reconstruction quality.

**Comparison with Different Unfolding Frameworks.** We further compare our framework with representative unfolding-based approaches, including ADMM-Net (Ma et al., 2019), GAP-Net (Meng et al., 2020a), DAUHST (Cai et al., 2022c), and LADE-DUN (Wu et al., 2024). As shown in Table 2d, our method achieves the best results with 42.13 dB PSNR and 0.9900 SSIM, outperforming LADE-DUN (40.97 dB / 0.9882) and DAUHST (38.81 dB / 0.9815). Although ADMM and GAP show slightly lower FLOPS (78.58G), their accuracy is far inferior, highlighting the superior accuracy–efficiency trade-off achieved by our approach.

## 5 CONCLUSION

In this paper, we introduce flow matching into hyperspectral image reconstruction for the first time, embedding its generative prior into a deep unfolding framework and further enforcing global consistency via a mean-velocity loss. This integration enhances the recovery of fine spectral details and better handles heavily degraded measurements, yielding a stronger and more robust prior. Extensive experiments on both simulated and real datasets demonstrate state-of-the-art performance, particularly in scenarios involving optical filter–based systems and CASSI acquisitions. Beyond quantitative improvements, our reconstructions show sharper structures, smoother spectral responses. Overall, these results highlight flow matching’s promise for computational imaging and its suitability for future compact chip-integrated HSI systems across diverse application domains.

## REFERENCES

- 486  
487  
488 Seung-Hwan Baek, Incheol Kim, Diego Gutierrez, and Min H Kim. Compact single-shot hyper-  
489 spectral imaging using a prism. *ACM TOG*, 2017. 2
- 490  
491 J.M. Bioucas-Dias and M.A.T. Figueiredo. A new twist: Two-step iterative shrinkage/thresholding  
492 algorithms for image restoration. *IEEE TIP*, 2007. 1
- 493  
494 Yuanhao Cai, Jing Lin, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and  
495 Luc Van Gool. Coarse-to-fine sparse transformer for hyperspectral image reconstruction. In  
496 *ECCV*, 2022a. 3, 7
- 497  
498 Yuanhao Cai, Jing Lin, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and  
499 Luc Van Gool. Mask-guided spectral-wise transformer for efficient hyperspectral image recon-  
500 struction. In *CVPR*, 2022b. 3, 6, 7
- 501  
502 Yuanhao Cai, Jing Lin, Haoqian Wang, Xin Yuan, Henghui Ding, Yulun Zhang, Radu Timofte,  
503 and Luc V Gool. Degradation-aware unfolding half-shuffle transformer for spectral compressive  
504 imaging. In *NeurIPS*, 2022c. 1, 3, 7, 9
- 505  
506 Yuanhao Cai, Yuxin Zheng, Jing Lin, Xin Yuan, Yulun Zhang, and Haoqian Wang. Binarized  
507 spectral compressive imaging. In *NeurIPS*, 2023. 7
- 508  
509 Stanley H Chan, Xiran Wang, and Omar A Elgendy. Plug-and-play admm for image restoration:  
510 Fixed-point convergence and applications. *IEEE TCI*, 2016. 3
- 511  
512 Adam S Charles, Bruno A Olshausen, and Christopher J Rozell. Learning sparse codes for hyper-  
513 spectral imagery. *IEEE Journal of Selected Topics in Signal Processing*, 2011. 1
- 514  
515 Yong Chen, Xinfeng Gui, Jinshan Zeng, Xi-Le Zhao, and Wei He. Combining low-rank and deep  
516 plug-and-play priors for snapshot compressive imaging. *IEEE TNNLS*, 2023. 3
- 517  
518 Yong Chen, Wenzhen Lai, Wei He, Xi-Le Zhao, and Jinshan Zeng. Hyperspectral compressive  
519 snapshot reconstruction via coupled low-rank subspace representation and self-supervised deep  
520 network. *IEEE TIP*, 2024. 3
- 521  
522 Ziheng Cheng, Bo Chen, Ruiying Lu, Zhengjue Wang, Hao Zhang, Ziyi Meng, and Xin Yuan.  
523 Recurrent neural networks for snapshot compressive imaging. *IEEE TPAMI*, 2022. 3, 7
- 524  
525 Inchang Choi, Dong Seon Jeon, Giljoo Nam, Diego Gutierrez, and Min H. Kim. High-quality  
526 hyperspectral reconstruction using a spectral prior. *ACM TOG*, 2017. 6
- 527  
528 Quan Dao, Hao Phung, Binh Nguyen, and Anh Tran. Flow matching in latent space. *arXiv preprint*  
529 *arXiv:2307.08698*, 2023. 6
- 530  
531 Yubo Dong, Dahua Gao, Tian Qiu, Yuyan Li, Minxi Yang, and Guangming Shi. Residual degrada-  
532 tion learning unfolding framework with mixing priors across spectral and spatial for compressive  
533 spectral imaging. In *CVPR*, 2023. 1
- 534  
535 Michael E Gehm, Renu John, David J Brady, Rebecca M Willett, and Timothy J Schulz. Single-shot  
536 compressive spectral imaging with a dual-disperser architecture. *Optics Express*, 2007. 1
- 537  
538 Alexander FH Goetz, Gregg Vane, Jerry E Solomon, and Barrett N Rock. Imaging spectrometry for  
539 earth remote sensing. *Science*, 1985. 1
- Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2009. 6
- Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, 2020. 3
- Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. Searching for mobilenetv3. In *ICCV*, 2019. 5

- 540 Xiaowan Hu, Yuanhao Cai, Jing Lin, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and  
541 Luc Van Gool. Hdnet: High-resolution dual-domain learning for spectral compressive imaging.  
542 In *CVPR*, 2022. 3, 7
- 543 Tao Huang, Weisheng Dong, Xin Yuan, Jinjian Wu, and Guangming Shi. Deep gaussian scale  
544 mixture prior for spectral compressive imaging. In *CVPR*, 2021. 7
- 546 Atiya Khan, Amol D Vibhute, Shankar Mali, and Chandrashekhar H Patil. A systematic review on  
547 hyperspectral imaging technology with a machine and deep learning methodology for agricultural  
548 applications. *Ecological Informatics*, 2022. 1
- 549 David Kittle, Kerkil Choi, Ashwin Wagadarikar, and David J Brady. Multiframe image estimation  
550 for coded aperture snapshot spectral imagers. *Applied optics*, 2010. 3
- 552 Nikita Kornilov, Petr Mokrov, Alexander Gasnikov, and Aleksandr Korotin. Optimal flow matching:  
553 Learning straight trajectories in just one step. In *NeurIPS*, 2024. 3, 6
- 554 Lu Li, Wei Li, Ying Qu, Chunhui Zhao, Ran Tao, and Qian Du. Prior-based tensor approximation  
555 for anomaly detection in hyperspectral imagery. *IEEE TNNLS*, 2020. 1
- 557 Miaoyu Li, Ying Fu, Ji Liu, and Yulun Zhang. Pixel adaptive deep unfolding transformer for hyper-  
558 spectral image reconstruction. In *ICCV*, 2023. 1
- 559 Shutao Li, Weiwei Song, Leyuan Fang, Yushi Chen, Pedram Ghamisi, and Jon Atli Benediktsson.  
560 Deep learning for hyperspectral image classification: An overview. *IEEE TGRS*, 2019. 1
- 562 Xuejun Liao, Hui Li, and Lawrence Carin. Generalized alternating projection for weighted-2,1  
563 minimization with applications to model-based compressive sensing. *SIAM Journal on Imaging  
564 Sciences*, 2014. 5
- 565 Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching  
566 for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022. 3
- 568 Hanxiao Liu, Zihang Dai, David So, and Quoc V Le. Pay attention to mlps. In *NeurIPS*, 2021. 9
- 569 Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and  
570 transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022. 3, 6
- 572 Yang Liu, Xin Yuan, Jinli Suo, David J Brady, and Qionghai Dai. Rank minimization for snapshot  
573 compressive imaging. *IEEE TPAMI*, 2018. 3
- 574 Guolan Lu and Baowei Fei. Medical hyperspectral imaging: a review. *Journal of biomedical optics*,  
575 2014. 1
- 577 Ruiying Lu, Bo Chen, Ziheng Cheng, and Penghui Wang. Rafnet: Recurrent attention fusion net-  
578 work of hyperspectral and multispectral images. *Signal Processing*, 2020. 1
- 579 Jiawei Ma, Xiao-Yang Liu, Zheng Shou, and Xin Yuan. Deep tensor admm-net for snapshot com-  
580 pressive imaging. In *ICCV*, 2019. 7, 9
- 582 Mikael Marois, Jonathan D Olson, Dennis J Wirth, Jonathan T Elliott, Xiaoyao Fan, Scott C Davis,  
583 Keith D Paulsen, and David W Roberts. A birefringent spectral demultiplexer enables fast hyper-  
584 spectral imaging of protoporphyrin ix during neurosurgery. *Communications Biology*, 2023. 2
- 585 Ziyi Meng, Shirin Jalali, and Xin Yuan. Gap-net for snapshot compressive imaging. *arXiv preprint  
586 arXiv:2012.08364*, 2020a. 7, 9
- 587 Ziyi Meng, Jiawei Ma, and Xin Yuan. End-to-end low cost compressive spectral imaging with  
588 spatial-spectral self-attention. In *ECCV*, 2020b. 1, 7
- 590 Ziyi Meng, Xin Yuan, and Shirin Jalali. Deep unfolding for snapshot compressive imaging. *IJCV*,  
591 2023. 1
- 592 Xin Miao, Xin Yuan, Yunchen Pu, and Vassilis Athitsos. l-net: Reconstruct hyperspectral images  
593 from a snapshot measurement. In *ICCV*, 2019. 1

- 594 Kristina Monakhova, Kyrollos Yanny, Neerja Aggarwal, and Laura Waller. Spectral diffusercam:  
595 lensless snapshot hyperspectral imaging with a spectral filter array. *Optica*, 2020. 2  
596
- 597 Samuel Ortega, Raúl Guerra, Maria Diaz, Himar Fabelo, Sebastián López, Gustavo M Callico, and  
598 Roberto Sarmiento. Hyperspectral push-broom microscope development and characterization.  
599 *IEEE Access*, 2019. 2
- 600 Jong-Il Park, Moon-Hyun Lee, Michael D Grossberg, and Shree K Nayar. Multispectral imaging  
601 using multiplexed illumination. In *2007 IEEE 11th International Conference on Computer Vision*,  
602 2007. 6  
603
- 604 Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor  
605 Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-  
606 performance deep learning library. In *NeurIPS*, 2019. 7
- 607 Haiquan Qiu, Yao Wang, and Deyu Meng. Effective snapshot compressive-spectral imaging via  
608 deep denoising and total variation priors. In *CVPR*, 2021. 3  
609
- 610 Roozbeh Rajabi, Amin Zehtabian, Keshav D Singh, Alireza Tabatabaeenejad, Pedram Ghamisi, and  
611 Saeid Homayouni. Hyperspectral imaging in environmental monitoring and analysis, 2024. 1  
612
- 613 Weiqiang Rao, Lianru Gao, Ying Qu, Xu Sun, Bing Zhang, and Jocelyn Chanussot. Siamese trans-  
614 former network for hyperspectral image target detection. *IEEE Transactions on Geoscience and*  
615 *Remote Sensing*, 2022. 1
- 616 Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization  
617 in the brain. *Psychological review*, 1958. 9
- 618 Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben  
619 Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint*  
620 *arXiv:2011.13456*, 2020. 3  
621
- 622 Mary B Stuart, Andrew JS McGonigle, and Jon R Willmott. Hyperspectral imaging in environmen-  
623 tal monitoring: A review of recent developments and technological advances in compact field  
624 deployable systems. *Sensors*, 2019. 1
- 625 Ilya O Tolstikhin, Neil Houlsby, Alexander Kolesnikov, Lucas Beyer, Xiaohua Zhai, Thomas Un-  
626 terthiner, Jessica Yung, Andreas Steiner, Daniel Keysers, Jakob Uszkoreit, et al. Mlp-mixer: An  
627 all-mlp architecture for vision. In *NeurIPS*, 2021. 5  
628
- 629 Aziz ul Rehman and Shahzad Ahmad Qureshi. A review of the medical hyperspectral imaging sys-  
630 tems and unmixing algorithms’ in biological tissues. *Photodiagnosis and Photodynamic Therapy*,  
631 2021. 1
- 632 Kuniaki Uto, Haruyuki Seki, Genya Saito, Yukio Kosugi, and Teruhisa Komatsu. Development  
633 of a low-cost hyperspectral whiskbroom imager using an optical fiber bundle, a swing mirror,  
634 and compact spectrometers. *IEEE Journal of Selected Topics in Applied Earth Observations and*  
635 *Remote Sensing*, 2016. 2  
636
- 637 Burak Uzket, Aneesh Rangnekar, and Matthew Hoffman. Aerial vehicle tracking by adaptive  
638 fusion of hyperspectral likelihood maps. In *CVPRW*, 2017. 1
- 639 Hien Van Nguyen, Amit Banerjee, and Rama Chellappa. Tracking via object reflectance using a  
640 hyperspectral video camera. In *CVPRW*, 2010. 1  
641
- 642 Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez,  
643 Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NeurIPS*, 2017. 9
- 644 Ashwin Wagadarikar, Renu John, Rebecca Willett, and David Brady. Single disperser design for  
645 coded aperture snapshot spectral imaging. *Applied optics*, 2008. 3  
646
- 647 Lishun Wang, Zongliang Wu, Yong Zhong, and Xin Yuan. Snapshot spectral compressive imaging  
reconstruction using convolution and contextual transformer. *Photonics Research*, 2022. 1, 3

648 Lizhi Wang, Zhiwei Xiong, Guangming Shi, Feng Wu, and Wenjun Zeng. Adaptive nonlocal sparse  
649 representation for dual-camera compressive hyperspectral imaging. *IEEE transactions on pattern*  
650 *analysis and machine intelligence*, 2016. 3  
651  
652 Zongliang Wu, Ruiying Lu, Ying Fu, and Xin Yuan. Latent diffusion prior enhanced deep unfolding  
653 for snapshot spectral compressive imaging. In *ECCV*, 2024. 4, 5, 7, 8, 9  
654  
655 Jian Xiong, Xusheng Cai, Kaiyu Cui, Yidong Huang, Jiawei Yang, Hongbo Zhu, Wenzheng Li,  
656 Bo Hong, Shijie Rao, Zekun Zheng, et al. Dynamic brain spectrum acquired by a real-time  
657 ultraspectral imaging chip with reconfigurable metasurfaces. *Optica*, 2022. 2  
658  
659 Motoki Yako, Yoshikazu Yamaoka, Takayuki Kiyohara, Chikai Hosokawa, Akihiro Noda, Klaas  
660 Tack, Nick Spooren, Taku Hirasawa, and Atsushi Ishikawa. Video-rate hyperspectral camera  
661 based on a cmos-compatible random array of fabry-pérot filters. *Nature Photonics*, 2023. 2, 7  
662  
663 Zongyin Yang, Tom Albrow-Owen, Weiwei Cai, and Tawfique Hasan. Miniaturization of optical  
664 spectrometers. *Science*, 2021. 2  
665  
666 Zhiyang Yao, Shuyang Liu, Xiaoyun Yuan, and Lu Fang. Specat: Spatial-spectral cumulative-  
667 attention transformer for high-resolution hyperspectral image reconstruction. In *CVPR*, 2024.  
668 7  
669  
670 Xin Yuan. Generalized alternating projection based total variation minimization for compressive  
671 sensing. In *IEEE ICIP*, 2016. 1  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701