# ADAPT: General-to-Specific ADAptive Prior Transfer for Hyperspectral Image Reconstruction

Zhaolu Chen * and Xian-Hua Han

Graduate School of Artificial Intelligence and Science, Rikkyo University, Tokyo, Japan
Email: mark.zhaolu.chen@rikkyo.ac.jp (Z.C.); hanxhua@rikkyo.ac.jp (X.-H.H.)
*Corresponding author

***Abstract*—Hyperspectral Imaging (HSI) captures rich spectral details far beyond RGB cameras. To enable real-time acquisition, contemporary HSI systems often rely on compressive sensing theory, making the reconstruction of hyperspectral data from compressed measurements a fundamental and challenging task in computational imaging. While supervised deep learning has achieved remarkable success in Hyperspectral (HS) image reconstruction tasks, the generalization capability of these existing models is often hindered by limited training data, resulting in poor performance on unseen scenes. To address this issue, we proposed ADAPT (ADAptive Prior Transfer), a test-time prior transfer framework that bridges the gap between the general priors learned by a pre-trained model and the specific information of a target scene. Our method leverages a arbitrary pre-trained deep unfolding or end-to-end network and test-time fine-tuning using a loss function derived from the physical Coded Aperture Snapshot Spectral Imaging (CASSI) imaging forward model and the emprical prior of HSI. Experiments on the KAIST dataset demonstrated that our method improved the reconstruction quality for both transformer- and deep unfolding-based models.**

***Keywords*—hyperspectral image reconstruction, test-time adaptation, self-supervised learning**

## I. Introduction

Hyperspectral Imaging (HSI) provides rich spectral information for a wide range of applications including remote sensing [1–3], medical imaging [4, 5], and agriculture [6]. However, conventional HSI systems based on Whiskbroom (e.g. AVIRIS [7, 8]) or Pushbroom (e.g. Hyperion [9]) scanning systems are limited by their temporal resolution, rendering them incapable of capturing dynamic scenes in real time [10–13]. This issue can be addressed by Snapshot Compressive Imaging (SCI) [14, 15], which offers a solution by capturing and compressing high-dimensional spectral data into a single 2-dimensional measurement [10]. One of the most representative SCI architectures is the Coded Aperture Snapshot Spectral Imaging (CASSI) system [14, 16, 17], which optically encodes and compresses a 3-dimensional HSI data cube $X$ into an encoded 2-dimensional snapshot $Y$, and a software decoder is required to reconstruct the compressive snapshot into a high-quality 3D HSI cube.

HSI reconstruction has been extensively investigated through a wide range of strategies, ranging from traditional model-based techniques [18–20] to modern deep learning—driven frameworks [21–25]. Conventional model-based methods, grounded in physical priors and mathematical regularization principles, offer clear interpretability and theoretical tractability [18–20]. However, their reconstruction performance is often limited by the reliance on handcrafted priors, simplified assumptions about spectral–spatial correlations, and the high computational burden associated with iterative optimization to solve the inverse problem [10, 26, 27]. In contrast, deep learning-based approaches have achieved remarkable advances by leveraging data-driven prior modeling and end-to-end optimization, resulting in substantial improvements in both reconstruction accuracy and computational efficiency [21–25]. Among these, Deep Unfolding Networks (DUNs) [28–30] have emerged as a particularly promising paradigm that bridges the gap between interpretability and learning flexibility. By embedding priors that are automatically learned through neural network architectures into the underlying mathematical imaging model, DUNs effectively unify the complementary strengths of model-based optimization and deep learning. Specifically, DUNs unfold iterative optimization algorithms into trainable neural networks, thus preserving the interpretability and physical constraints of model-based methods while simultaneously leveraging the powerful representation and generalization capabilities of neural networks. This hybrid paradigm facilitates the principled integration of domain-specific knowledge such as physical constraints and degradation models, within a learnable and data-adaptive framework. Consequently, DUNs achieve a favorable trade-off between theoretical soundness and empirical performance, making them a compelling choice for complex inverse imaging problems such as HSI reconstruction. Despite these advances, a key limitation of many existing deep learning methods lies in their reliance on dataset-driven priors, which restricts their generalization capability across diverse imaging conditions and application domains.

Manuscript received October 13, 2025; revised November 12, 2025; accepted December 29, 2025; published May 27, 2026.

This generalization gap presents a primary obstacle to the real-world application of supervised HSI reconstruction models [10, 24, 31].

To bridge this generalization gap that often arises when models trained on limited datasets are deployed in real-world scenarios, we propose ADAPT, a self-supervised test-time adaptation framework designed to enhance model robustness and adaptability under unseen conditions. Unlike purely self-supervised frameworks that learn priors from scratch using only measurements, we leverage the rich spectral-spatial representations learned by supervised models from labeled data by introducing a dynamic self-supervised learning mechanism during the inference stage to align these general priors with the specific statistics of the target scene. This "General-to-Specific" transfer strategy enables the model to learn scene-specific priors directly from the target measurement, thereby adapting to distributional shifts and unmodeled variations that were not captured during the initial training phase.

Specifically, the self-supervised objective in the inference stage integrates two complementary components: (1) A measurement consistency constraint, which enforces fidelity between the reconstructed hyperspectral signal and the observed compressed measurement, ensuring physical plausibility; and (2) a signal prior regularization, which encodes the inherent spectral–spatial smoothness and structural correlations of hyperspectral data. These jointly formulated objectives steer a lightweight fine-tuning process with the available test measurement only, thereby allowing the model to adjust its parameters in response to the specific spectral and statistical characteristics of the novel scene. Quantitative experiments on benchmark HS image dataset confirm the efficacy of the proposed ADAPT framework over both transformer-based End-to-End (E2E) and deep unfolding-based models.

In summary, our main contributions are as follows.

- We propose ADAPT, a test-time adaptation framework for HSI reconstruction that enables a supervised pre-trained deep unfolding network or E2E network to dynamically adapt to unseen scenes during inference. By introducing a self-supervised learning objective, the framework allows the model to capture scene-specific spectral-spatial characteristics without requiring additional labeled data, thereby addressing the generalization gap inherent in conventional supervised methods.
- Our self-supervised objective combines a measurement consistency constraint, ensuring physical fidelity between reconstructed signals and observed compressed measurements, with a signal prior regularization, which encodes intrinsic spectral-spatial correlations and structural smoothness.
- Extensive experiments on benchmark HSI datasets validate the effectiveness of ADAPT, achieving significant gains in reconstruction metrics for both deep unfolding and E2E networks.

## II. Related Works

### A. Hyperspectral Image Reconstruction

Mathematically, the reconstruction of 3D hyperspectral images from 2D compressed measurements represents an ill-posed linear inverse problem [10]. Early solutions primarily relied on model-based optimization methods, which leveraged hand-crafted priors such as Total Variation (TV) [26] or low-rank matrix decomposition [20, 27] to regularize the solution space. While such methods offer strong mathematical interpretability, they suffer from high computational burdens and limited reconstruction quality, due to the iterative optimization process and the simplicity of hand-crafted process.

With the growing success of deep neural networks across a wide range of vision tasks, deep learning-based HSI reconstruction has emerged as a dominant paradigm, enabling data-driven learning of spectral-spatial correlations that are difficult to model analytically. Early approaches predominantly employed end-to-end networks that learn a direct mapping from two-dimensional compressed measurements to three-dimensional hyperspectral cubes. For instance, Miao *et al.* [23] proposed a dual-stage generative model that captures representative features in both the spatial and spectral domains, while Meng *et al.* [32] enhanced a U-Net architecture with self-attention modules to achieve accurate and real-time reconstruction. Yorimoto *et al.* [33] developed HyperMixNet, which jointly learns spatial context and spectral dependencies using multiple supervision signals, and Hu *et al.* [22] designed a dual-domain learning framework to better preserve high-frequency spectral details in compressive imaging tasks. More recently, the field has witnessed a paradigm shift from Convolutional Neural Networks (CNNs) to Transformer-based architectures, driven by their superior capability to capture long-range dependencies. Transformer models such as Mask-guided Spectral-wise Transformer (MST) [34] and Coarse-to-fine Sparse Transformer (CST) [35] exploit self-attention mechanisms to model global spectral relationships while maintaining computational efficiency by operating primarily in the spectral dimension.

### B. General Prior Learning in HSI Reconstruction

Despite the remarkable success and architectural diversity ranging from deep unfolding networks to transformer-based networks, these networks share a common paradigm of general prior learning. In this paradigm, models are trained in a supervised or self-supervised manner on a fixed dataset, such as Columbia Imaging and Vision Laboratory dataset (CAVE dataset) [36] or Korea Advanced Institute of Science and Technology dataset of Hyperspectral Reflectance Images (KAIST dataset) [37], to learn a mapping from measurements to hyperspectral images. During the model training process, the network parameters encapsulate a general prior representing the statistical characteristics of the training data. However, such a paradigm would cause

a generation gap that arises when the pre-trained model is deployed to reconstruct unseen scenes. Due to the high dimensionality of hyperspectral data, the spectral characteristics and spatial structures in real-world testing scenarios often diverge significantly from that in the training data, thus the general prior learned by the network may fail to capture unique prior of the target scene, resulting in artifacts, spectral distortion, or performance degradation. Consequently, while general prior learning provides strong reconstruction capability, it remains essential to integrate physically grounded modeling or adaptive learning mechanisms to enhance robustness and generalization in real-world HSI reconstruction.

### C. Test-Time Adaptation

To overcome distribution shifts and improve robustness in real-world deployments, Test-Time Adaptation (TTA) techniques emerged as a paradigm to address distribution shifts by adapting pre-trained models to test data during inference. TTA approaches typically begin with a pre-trained network and perform lightweight parameter fine-tuning during inference, guided by a self-supervised objective defined directly on the test measurements [38, 39]. Through this mechanism, TTA frameworks effectively adapt the broadly learned representations from the training phase to the scene-specific characteristics of the target domain without additional ground-truth annotations. In high-level computer vision tasks such as classification, methods such as Test Entropy Minimization (TENT) [40] and Source Hypothesis Transfer (SHOT) [41] typically minimize the entropy of model predictions to align source and target distributions, assuming that low-entropy predictions on target data correlate with high accuracy. However, these classification-oriented TTA methods are not suitable for direct application to low-level reconstruction problems, such as HSI reconstruction, because the latter requires pixel-wise fidelity rather than classification accuracy. In addition, the effectiveness of TTA largely depends on the quality of the pre-trained model, which determines the richness and generality of the prior knowledge available for transfer [39].

Motivated by such insight, our proposed ADAPT framework starts from high-performance pretrained deep unfolding- and transformer-based models; rather than the entropy minimization approach in high-level computer vision tasks, we employ a self-supervised, physics-driven adaptation loss to guide fine-tuning at test time. This design enables effective knowledge transfer from general training distributions to scene-specific domains, thereby maximizing reconstruction fidelity and robustness in real-world hyperspectral imaging scenarios.

## III. Methods

This section first introduces the compressive imaging process of the CASSI system, which provides the fundamental mathematical formulation underlying our self-supervised learning objective. We then present the overall architecture of the proposed ADAPT framework, highlighting its integration of physical measurement constraints and adaptive learning mechanisms. Subsequently, we elaborate on the general prior learning strategy based on the Generalized Alternating Projection (GAP)-net model, followed by a detailed description of the scene-specific prior transfer that occurs during the test-time adaptation phase.

### A. Physical Imaging Process of CASSI Imaging Model

To ensure the fidelity of the reconstructed measurement, we integrated the physical imaging process of the CASSI system into the forward model of our ADAPT framework. The CASSI system enables the acquisition of a three-Dimensional (3D) HS cube within a single two-Dimensional (2D) measurement, thereby achieving real-time compressive HS imaging. In CASSI, each spectral channel of the target scene $X \in \mathbb{R}^{H\times W\times B_\lambda}$, where $H$, $W$ and $B_\lambda$ are the height, width, and number of wavelengths, respectively, is modulated by a coded aperture mask $M \in \mathbb{R}^{H\times W}$ and subsequently dispersed by a prism according to its wavelength. The detector records the superimposed measurement $Y \in \mathbb{R}^{H\times W}$, which can be described as:

$$Y = \sum_{\lambda=1}^{B_\lambda} S_\lambda (M \odot X_\lambda) + N, \tag{1}$$

where $S_\lambda(\cdot)$ denotes the wavelength-dependent shift operator, $\odot$ represents element-wise multiplication, and $N$ is the system noise.

For simplicity, the imaging model in Eq. (1) can be written in the vectorized linear form by noting three facts. First, the element-wise modulation by the coded aperture satisfies $\mathrm{vec}(M \odot X_\lambda) = \mathrm{diag}(\mathrm{vec}(M))\, x_\lambda$, which turns the masking operation into a diagonal matrix. Second, the wavelength-dependent shift operator $S_\lambda(\cdot)$ is linear and can therefore be represented by a permutation-type matrix $P_\lambda$ acting on the vectorized image. Finally, by stacking the contributions of all spectral bands, the overall sensing matrix is obtained as $\Phi = [\,P_1\mathrm{diag}(m)\;\; P_2\mathrm{diag}(m)\;\cdots\; P_{B_\lambda}\mathrm{diag}(m)\,]$, leading to the compact vectorized representation as:

$$y = \Phi x + n, \tag{2}$$

where $y$ and $x$ are the vectorized forms of $Y$ and $X$, respectively, and $\Phi$ refers to the system sensing matrix encoding both optical modulation and spectral dispersion. The primary task of HSI construction is to recover HSI images $x$ given a known snapshot $y$ and sensing matrix $\Phi$. This forward model provides the physical foundation of our proposed ADAPT framework, ensuring measurement consistency during self-supervised test-time adaptation.

### B. Overall Architecture

Building upon the CASSI forward model, our proposed ADAPT framework introduces a self-supervised test-time adaptation strategy to enhance the generalization capability of HSI reconstruction networks under unseen conditions. As illustrated in Fig. 1, ADAPT operates on a supervised pre-trained GAP-net, which serves as the backbone network. Through supervised learning on

diverse training samples with paired input–output data, GAP-net effectively captures a general spectral–spatial prior that represents the underlying correlations across multiple scenes.

During inference, ADAPT augments this general prior with a self-supervised learning objective that adapts the model to the unique characteristics of the target measurement. Specifically, the adaptation objective integrates two complementary components: (1) a measurement consistency loss and (2) a signal prior regularization. These objectives drive a lightweight fine-tuning process performed directly on the test measurement, without requiring additional labeled data. Through the adaptive optimization, ADAPT effectively transfers scene-specific priors into the pre-trained model, achieving improved reconstruction fidelity and robustness.

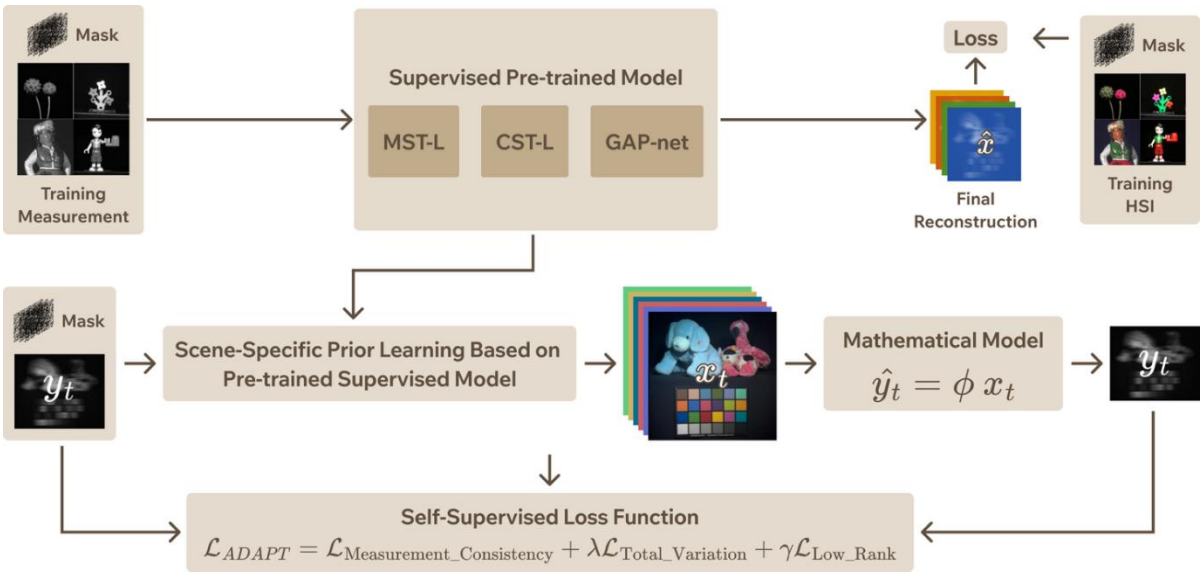


Fig. 1. Our ADAptive prior transfer framework for CASSI HSI reconstruction.

### C. General Prior Learning with GAP-net

To achieve accurate and interpretable HSI reconstruction, our framework employs the Generalized Alternating Projection Network (GAP-net) [42] as the backbone model for learning general spectral–spatial priors from data. GAP-net is a representative DUN method, which bridges traditional optimization-based reconstruction and deep learning by unrolling iterative algorithms into a trainable neural network architecture [30, 34]. This design allows each network stage to correspond to one iteration of an optimization procedure, thereby maintaining the interpretability and convergence properties of the underlying physical model.

Concretely, GAP-net formulates HSI reconstruction as an optimization problem to estimate the target HS signal $x$ by minimizing a combination of measurement fidelity and learned prior regularization:

$$\hat{x} = \arg\min_{x} \frac{1}{2} \| y - \Phi x \|_2^2 + \eta f_{\text{CNN}}^{\theta}(x), \tag{3}$$

where $f_{\text{CNN}}^{\theta}(\cdot)$ represents a CNN parameterized by $\theta$, which implicitly encodes data-driven spectral–spatial priors. The first term enforces consistency with the physical measurement process, while the second term introduces a learnable regularization that captures the intrinsic structures and smoothness properties of hyperspectral data.

To efficiently solve the optimization problem in Eq. (3), GAP-net unfolds the iterative reconstruction procedure into a sequence of learnable network stages, each corresponding to one iteration of the underlying optimization algorithm. The reconstruction alternates between two complementary modules: a data-consistency module, which enforces fidelity between the reconstructed hyperspectral estimate and the observed measurement, and a prior-update module, which employs a CNN to adaptively refine the spectral–spatial representation of the signal. The conceptual scheme of the GAP-net is illustrated in the top part of Fig. 1. Through supervised training with paired measurement–ground-truth datasets, the network learns optimal parameters $\theta^*$ that effectively encode a general prior representation reflecting the statistical regularities and correlations inherent in hyperspectral imagery.

After supervised pre-training, the learned model can generate an initial reconstruction for an unseen compressive measurement $\mathrm{y}_t$ as:

$$x_t = \text{GAPNet}^{\theta^*}(y_t, \Phi), \tag{4}$$

where the pre-trained parameters $\theta^*$ embody the general prior knowledge acquired from the training dataset [42].

### D. Specific Prior Learning with Test-Time Adaptation

While the supervised pre-trained GAP-net successfully captures a general spectral-spatial prior from the training dataset, its reconstruction performance may deteriorate

when confronted with domain shifts in real-world test scenarios. Such discrepancies often arise from variations in illumination conditions, material properties, or sensor noise, leading to distributional mismatches between the training and deployment environments. To alleviate this limitation, we introduce a Specific Prior Learning strategy based on Test-Time Adaptation (TTA), which enables the model to refine its learned priors during inference without requiring ground-truth supervision.

In the proposed ADAPT framework, test-time adaptation is formulated as a self-supervised optimization process that fine-tunes the parameters of the pre-trained GAP-net using only the observed measurement $y_t$. This adaptation is guided by a hybrid objective that integrates two complementary constraints: measurement consistency and signal prior regularization. The overall test-time objective can be expressed as:

$$\mathcal{L}_{\text{ADAPT}} = \mathcal{L}_{\text{MC}} + \lambda \mathcal{L}_{\text{SP}}, \tag{5}$$

where $\mathcal{L}_{\text{MC}}$ enforces physical fidelity between the reconstructed hyperspectral estimate and the corresponding compressed measurement [43], while $\mathcal{L}_{\text{SP}}$ regularizes the reconstruction by imposing implicit signal priors derived from the model's own predictions.

In this study, we employ two complementary priors to formulate $\mathcal{L}_{\text{SP}}$: the Total Variation (TV) prior and the Spectral Low-Rank (LR) prior. Consequently, $\mathcal{L}_{\text{SP}}$ is defined as:

$$\mathcal{L}_{\text{SP}} = \mathcal{L}_{\text{TV}} + \gamma \mathcal{L}_{\text{LR}}, \tag{6}$$

where the hyperparameter $\gamma$ modulates the relative contribution of the two regularization terms, thereby controlling the balance between $\mathcal{L}_{\text{TV}}$ and $\mathcal{L}_{\text{LR}}$ in the loss item $\mathcal{L}_{\text{SP}}$.

Next, we will elaborate on the specific formulation of each loss component employed in the proposed adaptation framework.

### E. Measurement Consistency Loss

Denoted as $\mathcal{L}_{mc}$, the measurement consistency loss acts as a data filedity term which ensures that the reconstructed HSI $\hat{x}$ will reproduce a similar measurement $\hat{y}$ compared to input measurement $y$ when passed through the forward model, neglecting the noise. In our experiment, the measurement consistency is defined as the root mean square error between the input and simulated measurements as inspired by [43], written as:

$$\mathcal{L}_{mc} = \text{RMSE}(y, \Phi\hat{x}), \tag{7}$$

### F. Total Variation Loss

The total variation loss is denoted $\mathcal{L}_{tv}$. The total variation regularizer applies a linear penalty by minimizing the L1-norm of the image gradient, thereby preserving sharp edges while removing noise [44, 45]. We use the total variation to achieve flat regions without over-penalizing large gradients that possibly define object boundaries. The total variation loss we use in our settings can be written as:

$$\mathcal{L}_{tv} = \sum_{k=1}^{B_\lambda} (\| \mathcal{D}_h \hat{x}_k \|_1 + \| \mathcal{D}_v \hat{x}_k \|_1), \tag{8}$$

where $B_\lambda$ is the number of spectral bands, $\hat{x}_k$ is the k-th 2D spectral band of the reconstructed HSI $\hat{x}$, and $\mathcal{D}_h$ and $\mathcal{D}_v$ are the horizontal and vertical gradients, respectively [26]. The total variation loss in our setting is critical because the measurement consistency alone may be insufficient to prevent the foundation model from producing noisy solutions by guiding the model to produce spatially consistent images while preserving clean edges.

### G. Spectral Low-Rank Loss

The spectral low-rank loss, denoted by $\mathcal{L}_{lr}$, is a key component of our signal prior which exploits the high spectral correlation inherent in HSI data.

It is well established that hyperspectral images exhibit significant redundancy across their spectral dimensions, meaning that spectral signatures within a local spatial patch can be represented by a low-rank matrix [43, 46–48]. The primary goal of low-rank approximation is to recover the structure, which often involves rank minimization; however, minimizing the rank of a matrix is a non-convex, NP-hard problem, which is often computationally intractable [46, 49, 50] . Therefore, the standard approach is to replace the rank minimization problem with its tightest convex surrogate, Nuclear Norm Minimization (NNM) [46, 51], which minimizes $||X||_* = \sum \sigma_i(X)$, the sum of the singular values.

As pointed out in [46], one of the key limitations of standard NNM is that NNM penalizes all singular values equally. This can be suboptimal in HSI reconstruction, as larger singular values typically correspond to the principal signal components, while smaller singular values are more likely to be associated with noise [20, 43]. Advanced methods, such as Weighted Nuclear Norm Minimization (WNNM) [46], would address this issue by assigning non-uniform weights to the singular values.

Inspired by Gu *et al.* [46] and He *et al.* [43] and with reference to [52, 53], our framework adopts a logarithm-based penalty in the low-rank loss calculation, which achieves a similar non-uniform penalization. By taking the logarithm, we penalize smaller singular values, which often correspond to noise, more heavily, whereas larger singular values containing the main spectral information receive smaller weights in the loss. To compute the loss, we first divide the reconstructed HSI $\hat{x}$ into $N_p$ overlapping 3D spatial-spectral patches with a spatial size of $s \times s$ and overlap length of $\frac{s}{2} - 1$. We then unfold each patch into a 2D matrix $Z_i \in \mathbb{R}^{s^2 \times B_\lambda}$, where $B_\lambda$ is the number of spectral bands. We ensure $s^2 > B_\lambda$ so that the rank is constrained by the spectral dimension.

Finally, we define the spectral loss as the mean of the log-sum of the singular values of these matrices, written as:

$$\mathcal{L}_{\text{lr}} = \frac{1}{N_p} \sum_{i=1}^{N_p} \sum_{r=1}^{B_\lambda} \log(\sigma_r(Z_i) + \epsilon), \tag{9}$$

where $\sigma_r(Z_i)$ is the $r$-th singular value of the matrix $Z_i$, and $\epsilon$ is a small positive constant (e.g., $10^{-8}$) used to ensure numerical stability and prevent $\log(0)$.

Finally, we present our self-supervise loss function as a weighted sum of all priors mentioned above, as:

$$\mathcal{L}_{total} = \mathcal{L}_{mc} + \lambda_{tv}\mathcal{L}_{tv} + \lambda_{lr}\mathcal{L}_{lr} \quad (10)$$

## IV. Experiments and Results

We conducted experiments on the KAIST dataset [37] to validate the effectiveness of our framework in improving generalization performance compared with a pre-trained GAP-net model using 205 HS images in CAVE dataset [36]. We followed the experimental settings of CASSI-SSL [43] and reconstructed our HSIs with 28 bands ranging from 450–650 nm.

### A. Experiment Settings

**Datasets:** In our experimental study, we utilized two widely recognized hyperspectral image datasets. The CAVE dataset [36] comprises 32 HSIs with a spatial resolution of 512×512 pixels. In contrast, the KAIST dataset [37] contains 30 high-resolution hyperspectral scenes, each with dimensions of 2704×3376 pixels. In accordance with established experimental protocols [32], an augmented version of the CAVE dataset, consisting of 205 samples, was employed to train the baseline GAP-net, MST and CST model. For the subsequent ADAPT transfer and evaluation phase, a subset of ten representative scenes from the KAIST dataset was selected in accordance to [34, 35, 42], and [43]. To ensure consistency across both datasets, the spectral range was constrained to 28 bands, corresponding to wavelengths between 450 nm and 650 nm.

**Evaluation Metrics:** We quantitatively evaluated our reconstruction performance using the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM).

**Implementation Details:** We implemented the ADAPT network using PyTorch on a system running Ubuntu 22.04.4 LTS, equipped with a Nvidia L40S GPU. The pre-trained models include one deep unfolding model (GAP-net) and two transformer-based models i.e. CST-L and MST-L. All pre-trained models are supervised, and each model is tested separately. Since the domain shift in hyperspectral data often involves complex spectral-spatial variations [54, 55], we fine-tune all the learnable parameters of the reconstruction network (e.g. the CNN denoisers in all $K$ stages of GAP-net, and all transformer blocks in MST-L and CST-L model) using the self-supervised loss. The physical sensing matrix $\Phi$ is kept frozen. For each test scene, the model was fine-tuned for 400 iterations using the Adam optimizer with parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and initial learning rate at $10^{-4}$, as outlined in [56], with a cosine annealing scheme. $\lambda_{tv}$ was set to 0.0012 and $\lambda_{lr}$ was set to 0.0001.

### B. Simulation Results

In our study, we applied the ADAPT framework to the pre-trained GAP-net model at test time. Table I provides detailed results of the reconstruction performance before (baseline) and after (ADAPT) fine-tuning.

The results in Table I demonstrate that our ADAPT framework can guide GAP-net to generate better reconstruction results for most scenes. Across the 10 test scenes, our results improved the performance in seven cases, achieving an average PSNR gain of 1.43 dB. Scenes 3, 4, and 9 benefited the most from our ADAPT framework, with PSNR increases of 3.96, 3.65, and 2.97 dB, respectively. This suggests that the specific priors of the scenes diverge significantly from the general prior learned by the model, in which case our framework steps out to bridge the generation gap between the prior knowledge learned by the model and the priors in specific scenes. To illustrate the improvements afforded by ADAPT, we present Fig. 2 as a qualitative comparison for Scene 3, which shows a PSNR gain of 3.96 dB. As observed in Fig. 2, the reconstruction after applying our approach achieved better visual results than the baseline model, successfully suppressing noise and artifact color banding while preserving the edges of the objects.

TABLE I. PSNR / SSIM Comparison of GAP-net before and after Applying the ADAPT Framework on the KAIST Dataset

| Scene ID | GAP-net (baseline) | ADAPT (ours) | PSNR Gain (dB) |
|---|---|---|---|
| 1 | 33.70 / 0.9175 | 34.84 / 0.9263 | +1.14 |
| 2 | 33.23 / 0.9021 | 33.75 / 0.9024 | +0.52 |
| 3 | 34.26 / 0.9333 | 38.22 / 0.9545 | +3.96 |
| 4 | 41.13 / 0.9726 | 44.78 / 0.9814 | +3.65 |
| 5 | 31.41 / 0.9236 | 32.39 / 0.9286 | +0.98 |
| 6 | 32.39 / 0.9303 | 32.39 / 0.9303 | 0.00 |
| 7 | 32.23 / 0.9060 | 33.23 / 0.9273 | +1.00 |
| 8 | 30.41 / 0.9119 | 30.42 / 0.9049 | +0.01 |
| 9 | 33.47 / 0.9204 | 36.44 / 0.9436 | +2.97 |
| 10 | 30.20 / 0.9037 | 30.20 / 0.9037 | 0.00 |
| Average | 33.24 / 0.9221 | 34.67 / 0.9303 | +1.43 |

TABLE II. PSNR / SSIM Comparison of CST-L before and after Applying the ADAPT Framework on the 10 KAIST Test Scenes

| Scene ID | CST-L (baseline) | ADAPT (ours) | PSNR Gain (dB) |
|---|---|---|---|
| 1 | 35.87 / 0.9540 | 36.91 / 0.9572 | +1.04 |
| 2 | 36.88 / 0.9579 | 38.47 / 0.9618 | +1.59 |
| 3 | 38.25 / 0.9663 | 41.99 / 0.9767 | +3.74 |
| 4 | 42.41 / 0.9817 | 47.72 / 0.9911 | +5.31 |
| 5 | 33.20 / 0.9584 | 36.36 / 0.9700 | +3.16 |
| 6 | 35.79 / 0.9673 | 36.41 / 0.9698 | +0.62 |
| 7 | 34.72 / 0.9468 | 38.12 / 0.9700 | +3.40 |
| 8 | 34.26 / 0.9667 | 35.13 / 0.9699 | +0.87 |
| 9 | 36.39 / 0.9616 | 41.54 / 0.9790 | +5.15 |
| 10 | 33.00 / 0.9516 | 34.26 / 0.9574 | +1.26 |
| Average | 36.08 / 0.9612 | 38.69 / 0.9703 | +2.61 |

However, we also noticed that the performance stagnated for scenes 6, 8, and 10. We hypothesized that for these scenes, the image statistics were already well-aligned with the general prior learned by the pre-trained GAP-net model. This leaves little room for improvement via self-supervised adaptation; however, our method does not degrade performance in such cases.

### C. Generalization Capability

To validate the generalization capability of our framework, we further applied ADAPT to two other pre-trained Transformer-based HSI reconstruction models, CST [35] and MST [34]. The results are presented in Tables II and III.

The results for CST-L and MST-L provided further evidence that despite the architectural differences in the HSI reconstruction models, our ADAPT framework delivered consistent and significant improvements across all three models. In addition, the performance gain of the transformer-based CST model was even more prominent than GAP-net baseline model. ADAPT improved reconstruction results for CST-L in all 10 scenes, achieving an average PSNR gain of 2.61 dB and per-scene gains as high as 5.31 dB (Scene 4) and 5.15 dB (Scene 9). We also observed an average gain of 0.86 dB on the MST-L model, with improvements in 9 out of 10 scenes. These experiments show that our ADAPT framework is a general strategy capable of bridging the generalization gap for various pre-trained models, regardless of whether they are based on DUN or Transformer architectures.

TABLE III. PSNR / SSIM COMPARISON OF MST-L BEFORE AND AFTER APPLYING THE ADAPT FRAMEWORK ON THE 10 KAIST TEST SCENES

| Scene ID | MST-L (baseline) | ADAPT (ours) | PSNR Gain (dB) |
|---|---|---|---|
| 1 | 35.44 / 0.9464 | 35.81 / 0.9446 | +0.37 |
| 2 | 36.12 / 0.9486 | 36.12 / 0.9486 | 0.00 |
| 3 | 36.39 / 0.9546 | 38.43 / 0.9632 | +2.04 |
| 4 | 42.06 / 0.9773 | 44.33 / 0.9812 | +2.27 |
| 5 | 32.94 / 0.9503 | 33.79 / 0.9440 | +0.85 |
| 6 | 34.71 / 0.9573 | 34.92 / 0.9547 | +0.21 |
| 7 | 34.08 / 0.9320 | 34.79 / 0.9427 | +0.71 |
| 8 | 32.88 / 0.9529 | 33.14 / 0.9513 | +0.26 |
| 9 | 35.04 / 0.9466 | 36.67 / 0.9540 | +1.63 |
| 10 | 32.74 / 0.9462 | 33.04 / 0.9447 | +0.30 |
| Average | 35.24 / 0.9512 | 36.10 / 0.9529 | +0.86 |

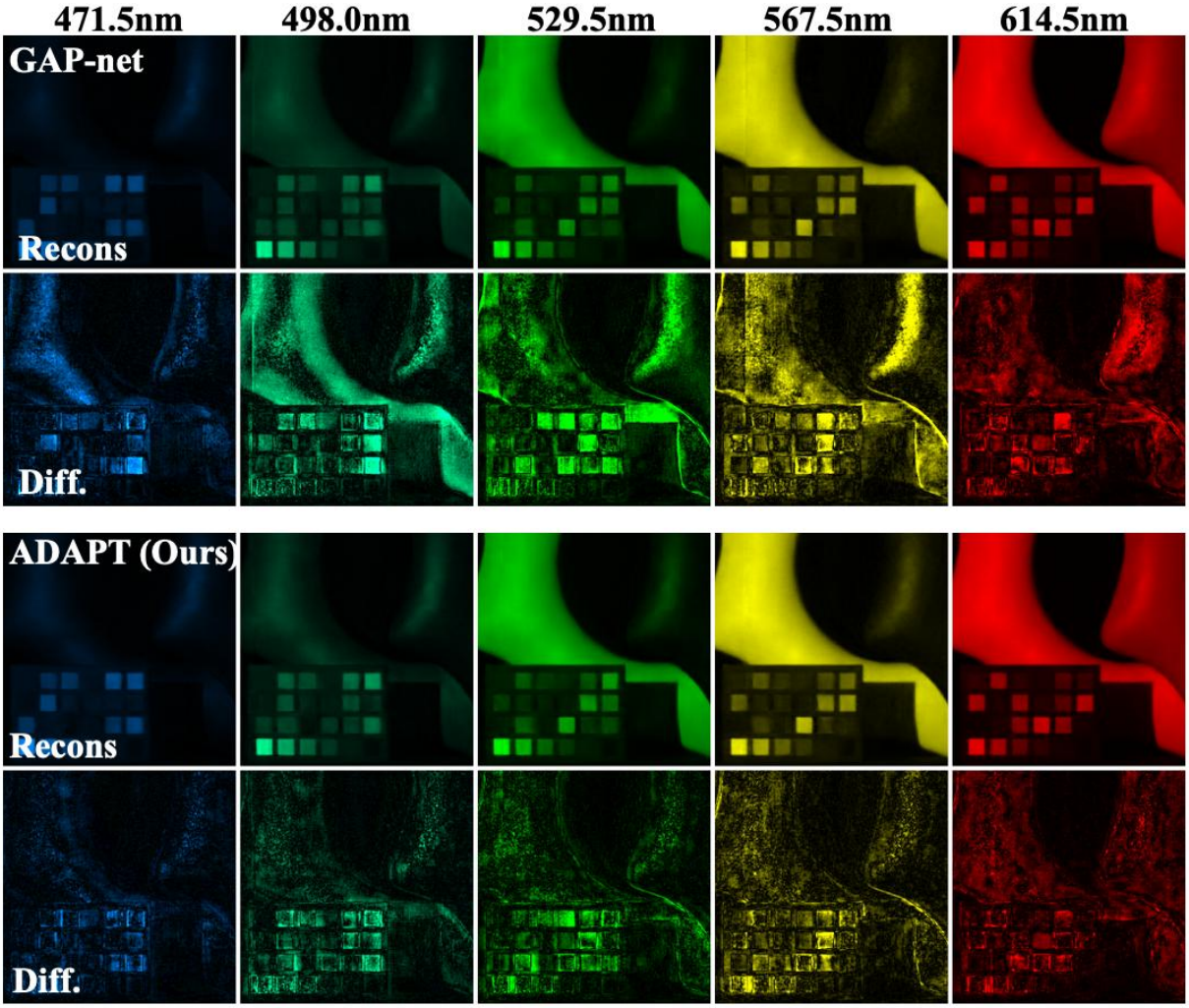


Fig. 2. Visual comparison of reconstruction results and difference images (Diff.) between the reconstructed and ground-truth data for 3 representative spectral bands (out of 28) of Scene 3 from the KAIST dataset.

### D. Ablation Study

To break down the impact of each component in our self-supervised loss function, we conducted an ablation study on the ADAPT framework by adding measurement consistency loss ($\mathcal{L}_{mc}$), total variation loss ($\mathcal{L}_{tv}$), and spectral low-rank loss ($\mathcal{L}_{lr}$) in each run. The results of the ablation study are listed in Table IV.

TABLE IV. ABLATION STUDY ON THE COMPONENTS OF THE SELF-SUPERVISED LOSS FUNCTION. AVERAGE PSNR AND SSIM ARE REPORTED ON THE KAIST DATASET

| Method | Avg. PSNR (dB) | Avg. SSIM |
|---|---|---|
| GAP-net (No fine-tuning) | 33.24 | 0.9221 |
| ADAPT (GAP-net, $\mathcal{L}_{mc}$ only) | 34.68 | 0.9282 |
| ADAPT (GAP-net, $\mathcal{L}_{mc} + \mathcal{L}_{tv}$) | 34.69 | 0.9281 |
| ADAPT (GAP-net, $\mathcal{L}_{mc} + \mathcal{L}_{lr}$) | 34.68 | 0.9306 |
| ADAPT (GAP-net, $\mathcal{L}_{mc} + \mathcal{L}_{tv} + \mathcal{L}_{lr}$) | 34.67 | 0.9303 |
| CST-L (No fine-tuning) | 36.03 | 0.9611 |
| ADAPT (CST-L, $\mathcal{L}_{mc}$ only) | 38.66 | 0.9695 |
| ADAPT (CST-L, $\mathcal{L}_{mc} + \mathcal{L}_{tv}$) | 38.65 | 0.9697 |
| ADAPT (CST-L, $\mathcal{L}_{mc} + \mathcal{L}_{lr}$) | 38.68 | 0.9700 |
| ADAPT (CST-L, $\mathcal{L}_{mc} + \mathcal{L}_{tv} + \mathcal{L}_{lr}$) | 38.69 | 0.9703 |
| MST-L (No fine-tuning) | 35.24 | 0.9512 |
| ADAPT (MST-L, $\mathcal{L}_{mc}$ only) | 36.10 | 0.9520 |
| ADAPT (MST-L, $\mathcal{L}_{mc} + \mathcal{L}_{tv}$) | 36.09 | 0.9520 |
| ADAPT (MST-L, $\mathcal{L}_{mc} + \mathcal{L}_{lr}$) | 36.11 | 0.9524 |
| ADAPT (MST-L, $\mathcal{L}_{mc} + \mathcal{L}_{tv} + \mathcal{L}_{lr}$) | 36.10 | 0.9529 |

The results indicate that the measurement consistency loss is the main contributor to the performance improvement, boosting the average PSNR by approximately 1.46 dB. This is expected because $\mathcal{L}_{mc}$ directly enforces that the reconstruction must conform to the physical measurement. The spectral low-rank loss further improves a slight improment in the structural and spectral accuracies, thus improving the SSIM. Notably, we observed a slight degradation in SSIM but stable PSNR when total variation loss ($\mathcal{L}_{tv}$) is added to the loss function, compared to using the measurement consistency loss ($\mathcal{L}_{mc}$) alone, which can be attributed to the inherent characteristics of total variation regularization called the staircase effect discussed in [44, 45]. While total variation loss effectively suppresses noise by enforcing piecewise smoothness, fine spatial textures may also be over-smoothed, causing loss of high frequency details in reconstructed image. While removing high-frequency noise improves the pixel-wise accuracy calculated by mean square error, the loss of fine textures penalizes the SSIM, which highly sensitive to structural information. However, the overall image quality can benefit from less frequent noise, given that PSNR is maintained.

However, the staircase effect is alleviated by the spectral low-rank loss ($\mathcal{L}_{lr}$). Unlike total variation, $\mathcal{L}_{lr}$ exploits the high correlation across spectral bands to separate signal from noise without blurring spatial details, thus preserving detailed spatial information. As shown in Table IV, introducing $\mathcal{L}_{lr}$ yielded the best SSIM performance for all networks, maintaining the balance between less noise and more high-frequency spatial textures in final reconstruction.

Fig. 3 shows the visualized loss curve during the fine-tuning of Scene 3, which achieved one of the highest PSNR gains. The total loss converged rapidly within the first 50 steps and then stabilized, demonstrating the efficiency and stability of the adaptation process. It is also evident that the total loss is dominated by the measurement consistency loss, denoted as MC Loss, showing the important role of the physics-informed forward model in guiding the prior transfer process.

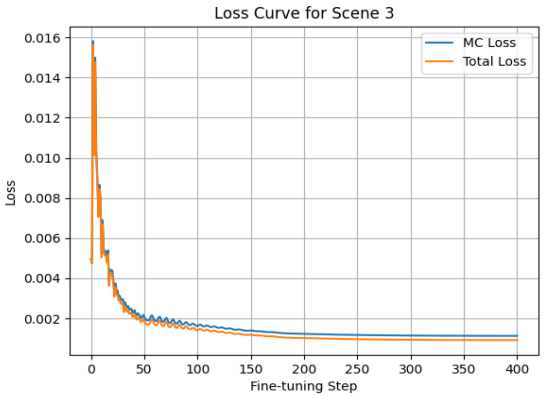


Fig. 3. Total loss curve against measurement consistency loss curve during prior transfer phase of Scene 3.

### E. Computational Analysis

To assess the computational costs introduced in our ADAPT framework, we quantified the additional time cost in Table V.

TABLE V. ADDITIONAL COMPUTATIONAL TIME INTRODUCED BY ADAPT FRAMEWORK

| Method | Relative time per scene (seconds) |
|---|---|
| GAP-net (No fine-tuning) | 0.00 |
| ADAPT (GAP-net, $\mathcal{L}_{mc}$ only) | +31.50 |
| ADAPT (GAP-net, $\mathcal{L}_{mc} + \mathcal{L}_{tv}$) | +31.14 |
| ADAPT (GAP-net, $\mathcal{L}_{mc} + \mathcal{L}_{lr}$) | +172.15 |
| ADAPT (GAP-net, $\mathcal{L}_{mc} + \mathcal{L}_{tv} + \mathcal{L}_{lr}$) | +171.66 |
| CST-L (No fine-tuning) | 0.00 |
| ADAPT (CST-L, $\mathcal{L}_{mc}$ only) | +26.93 |
| ADAPT (CST-L, $\mathcal{L}_{mc} + \mathcal{L}_{tv}$) | +27.14 |
| ADAPT (CST-L, $\mathcal{L}_{mc} + \mathcal{L}_{lr}$) | +173.35 |
| ADAPT (CST-L, $\mathcal{L}_{mc} + \mathcal{L}_{tv} + \mathcal{L}_{lr}$) | +172.79 |
| MST-L (No fine-tuning) | 0.00 |
| ADAPT (MST-L, $\mathcal{L}_{mc}$ only) | +79.51 |
| ADAPT (MST-L, $\mathcal{L}_{mc} + \mathcal{L}_{tv}$) | +97.23 |
| ADAPT (MST-L, $\mathcal{L}_{mc} + \mathcal{L}_{lr}$) | +240.38 |
| ADAPT (MST-L, $\mathcal{L}_{mc} + \mathcal{L}_{tv} + \mathcal{L}_{lr}$) | +221.72 |

The results indicate that the Measurement Consistency ($\mathcal{L}_{mc}$) and Total Variation ($\mathcal{L}_{tv}$) losses are computational lightweight. In GAP-net, adding $\mathcal{L}_{tv}$ to $\mathcal{L}_{mc}$ results in negligible time variations (from +31.50 s to +31.14 s), which can be attributed to system fluctuations rather than algorithmic complexity. In contrast, the spectral low-rank loss $\mathcal{L}_{lr}$ significantly incresed the computational burden, adding an average of 143.46 s to testing time. This is expected, as the singular value decomposition required for $\mathcal{L}_{lr}$ is a computationally intensive operation with a complexity of $O(\min(mn^2, m^2n))$. However, although ADAPT introduces an average latency of 188.72 s per scene across different models, we recognize the trade-off acceptable for high-stakes scientific and medical

imaging applications, where reconstruction fidelity is prioritized over real-time processing.

## V. Conclusion and Limitation

In this study, we addressed the issue of poor generalization in supervised HSI reconstruction models. We proposed ADAPT, a test-time prior transfer framework that enhances the performance of a pre-trained model on unseen scenes by fine-tuning it using a self-supervised loss function. Our experiments demonstrated the efficacy of this method. By applying the ADAPT framework to a pre-trained GAP-net model, we achieved an average of 1.46 dB of PSNR improvement on the KAIST dataset. Our work suggests that such a test-time adaptation strategy is highly compatible with deep unfolding architectures, such as GAP-net, offering insights into more robust and adaptive hyperspectral imaging systems.

However, while our proposed ADAPT framework demonstrates remarkable performance, this approach introduces additional computational costs. Although significantly faster than retraining, the required fine-tuning process for each new measurement may be a constraint for applications requiring real-time inference. In addition, we used fixed values for $\lambda_{tv}$ and $\lambda_{lr}$, which may slightly hinder the reconstruction performance. We expect future studies to address these drawbacks.

## Conflict of Interest

The authors declare no conflict of interest.

## Author Contributions

Conceptualization and methodology, Z.C. and X.H.; software, validation, formal analysis and investigation and writing—original draft preparation, review and editing, Z.C.; writing—review and editing, Z.C. and X.H.; project administration and funding acquisition, X.H. All authors had approved the final version.

## References


[1] M. Borengasser, W. S. Hungate, and R. Watkins, *Hyperspectral Remote Sensing: Principles and Applications*, CRC Press, 2007.

[2] J. Solomon and B. Rock, "Imaging spectrometry for earth remote sensing," *Science*, 1985.

[3] Y. Yuan, X. Zheng, and X. Lu, "Hyperspectral image superresolution by transfer learning," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 10, no. 5, pp. 1963–1974, 2017. doi: 10.1109/JSTARS.2017.2655112

[4] L. Zhi, D. Zhang, J. Yan, Q.-L. Li, and Q. Tang, "Classification of hyperspectral medical tongue images for tongue diagnosis," *Comput. Med. Imaging Graph.*, vol. 31, no. 8, pp. 672–678, Dec. 2007. doi: 10.1016/j.compmedimag.2007.07.008

[5] B. Fei, "Hyperspectral imaging in medical applications," *Data Handl. Sci. Technol.*, vol. 32, pp. 523–565, 2019. doi: 10.1016/B978-0-444-63977-6.00021-3

[6] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas, "Modern trends in hyperspectral image analysis: A review," *IEEE Access*, vol. 6, pp. 14118–14129, 2018. doi: 10.1109/ACCESS.2018.2812999

[7] R. O. Green *et al.*, "Imaging spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 227–248, 1998. doi: 10.1016/S0034-4257(98)00064-9

[8] G. Vane. (1987). Airborne Visible/Infrared Imaging Spectrometer (AVIRIS): A description of the sensor, ground data processing facility, laboratory calibration, and first results. *Jet Propulsion Laboratory, California Institute of Technology, JPL-PUBL-87-38*. [Online]. Available: https://ntrs.nasa.gov/citations/19880004943

[9] J. S. Pearlman, P. S. Barry, C. C. Segal, J. Shepanski, D. Beiso, and S. L. Carman, "Hyperion, a space-based imaging spectrometer," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1160–1173, 2003. doi: 10.1109/TGRS.2003.815018

[10] X. Yuan, D. J. Brady, and A. K. Katsaggelos, "Snapshot compressive imaging: Theory, algorithms, and applications," *IEEE Signal Process. Mag.*, vol. 38, no. 2, pp. 65–88, Mar. 2021. doi: 10.1109/MSP.2020.3023869

[11] S. Jalali and X. Yuan, "Snapshot compressed sensing: Performance bounds and algorithms," *IEEE Trans. Inf. Theory*, vol. 65, no. 12, pp. 8005–8024, Dec. 2019. doi: 10.1109/TIT.2019.2940666

[12] X.-H. Han, J. Wang, and H. Jiang, "Recent advancements in hyperspectral image reconstruction from a compressive measurement," *Sensors*, vol. 25, no. 11, 3286, May 2025. doi: 10.3390/s25113286

[13] A. F. H. Goetz, "Three decades of hyperspectral remote sensing of the earth: A personal view," *Remote Sens. Environ.*, vol. 113, pp. S5–S16, 2009. doi: 10.1016/j.rse.2007.12.014

[14] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Appl. Opt.*, vol. 47, no. 10, pp. B44–B51, Apr. 2008. doi: 10.1364/AO.47.000B44

[15] P. Llull *et al.*, "Coded aperture compressive temporal imaging," *Opt. Express*, vol. 21, no. 9, pp. 10526–10545, May 2013. doi: 10.1364/OE.21.010526

[16] A. A. Wagadarikar, N. P. Pitsianis, X. Sun, and D. J. Brady, "Video rate spectral imaging using a coded aperture snapshot spectral imager," *Opt. Express*, vol. 17, no. 8, pp. 6368–6388, Apr. 2009. doi: 10.1364/OE.17.006368

[17] M. E. Gehm, R. John, D. J. Brady, R. M. Willett, and T. J. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Opt. Express*, vol. 15, no. 21, pp. 14013–14027, Oct. 2007. doi: 10.1364/OE.15.014013

[18] X. Lin, Y. Liu, J. Wu, and Q. Dai, "Spatial-spectral encoded compressive hyperspectral imaging," *ACM Trans. Graph*, vol. 33, no. 6, Nov. 2014. doi: 10.1145/2661229.2661262

[19] Y. Wang, J. Peng, Q. Zhao, Y. Leung, X.-L. Zhao, and D. Meng, "Hyperspectral image restoration via total variation regularized low-rank tensor decomposition," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 4, pp. 1227–1243, 2017.

[20] S. Zhang, L. Wang, Y. Fu, X. Zhong, and H. Huang, "Computational hyperspectral imaging based on dimension-discriminative low-rank tensor recovery," in *Proc. 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, 2019, pp. 10182–10191. doi: 10.1109/ICCV.2019.01028

[21] J. Ma, X.-Y. Liu, Z. Shou, and X. Yuan, "Deep tensor ADMM-Net for snapshot compressive imaging," in *Proc. 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 10223–10232. doi: 10.1109/ICCV.2019.01032

[22] X. Hu *et al.*, "HDNet: High-resolution dual-domain learning for spectral compressive imaging," arXiv Preprint, arXiv:2203.02149, 2022.

[23] X. Miao, X. Yuan, Y. Pu, and V. Athitsos, "l-Net: Reconstruct hyperspectral images from a snapshot measurement," in *Proc. the IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2019.

[24] L. Wang, C. Sun, Y. Fu, M. H. Kim, and H. Huang, "Hyperspectral image reconstruction using a deep spatial-spectral prior," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[25] X. Zhang, Y. Zhang, R. Xiong, Q. Sun, and J. Zhang, "HerosNet: Hyperspectral explicable reconstruction and optimal sampling deep network for snapshot compressive imaging," in *Proc. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2022, pp. 17511–17520. doi: 10.1109/CVPR52688.2022.01701

[26] X. Yuan, "Generalized alternating projection based total variation minimization for compressive sensing," in *Proc. 2016 IEEE International Conference on Image Processing (ICIP)*, Sept. 2016, pp. 2539–2543. doi: 10.1109/ICIP.2016.7532817

[27] Y. Liu, X. Yuan, J. Suo, D. J. Brady, and Q. Dai, "Rank minimization for snapshot compressive imaging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 12, pp. 2990–3006, 2019. doi: 10.1109/TPAMI.2018.2873587

[28] Z. Wu, C. Yang, X. Su, and X. Yuan, "Adaptive deep PnP algorithm for video snapshot compressive imaging," *Int. J. Comput. Vis.*, vol. 131, no. 7, pp. 1662–1679, July 2023. doi: 10.1007/s11263-023-01777-y

[29] Y. Cai *et al.*, "Degradation-aware unfolding half-shuffle transformer for spectral compressive imaging," *Adv. Neural Inf. Process. Syst.*, vol. 35, pp. 37749–37761, Dec. 2022.

[30] Z. Meng, X. Yuan, and S. Jalali, "Deep unfolding for snapshot compressive imaging," *Int. J. Comput. Vis.*, vol. 131, no. 11, pp. 2933–2958, Nov. 2023. doi: 10.1007/s11263-023-01844-4

[31] Z. Meng, Z. Yu, K. Xu, and X. Yuan, "Self-supervised neural networks for spectral snapshot compressive imaging," in *Proc. the Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 2622–2631.

[32] Z. Meng, J. Ma, and X. Yuan, "End-to-end low cost compressive spectral imaging with spatial-spectral self-attention," in *Proc. Computer Vision—ECCV 2020*, 2020, pp. 187–204. doi: 10.1007/978-3-030-58592-1_12

[33] K. Yorimoto and X.-H. Han, "HyperMixNet: Hyperspectral image reconstruction with deep mixed network from a snapshot measurement," in *Proc. 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, IEEE, 2021, pp. 1184–1193. doi: 10.1109/ICCVW54120.2021.00138

[34] Y. Cai *et al.*, "Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction," arXiv Preprint, arXiv:2111.07910, 2022.

[35] Y. Cai *et al.*, "Coarse-to-fine sparse transformer for hyperspectral image reconstruction," arXiv Preprint, arXiv:2203.04845, 2022

[36] J.-I. Park, M.-H. Lee, M. D. Grossberg, and S. K. Nayar, "Multispectral imaging using multiplexed illumination," in *Proc. 2007 IEEE 11th International Conference on Computer Vision*, Oct. 2007, pp. 1–8. doi: 10.1109/ICCV.2007.4409090

[37] I. Choi, M. H. Kim, D. Gutierrez, D. S. Jeon, and G. Nam, "High-quality hyperspectral reconstruction using a spectral prior," *ACM Trans. Graph.*, 2017. doi: 10.1145/3130800.3130810

[38] H. Qu, H. Rahmani, L. Xu, B. Williams, and J. Liu, "Recent advances of continual learning in computer vision: An overview," arXiv Preprint, arXiv:2109.11369, 2024.

[39] Y. Zhao, T. Zhang, and H. Ji, "Test-time model adaptation for image reconstruction using self-supervised adaptive layers," in *Proc. Computer Vision—ECCV 2024*, Cham: Springer Nature Switzerland, 2025, pp. 111–128. https://doi.org/10.1007/978-3-031-72913-3_7

[40] D. Wang, E. Shelhamer, S. Liu, B. Olshausen, and T. Darrell, "Tent: Fully test-time adaptation by entropy minimization," arXiv Preprint, arXiv:2006.10726, 2021.

[41] J. Liang, D. Hu, and J. Feng, "Do we really need to access the source data? Source hypothesis transfer for unsupervised domain adaptation," arXiv Preprint, arXiv:2002.08546, 2021.

[42] Z. Meng, S. Jalali, and X. Yuan, "GAP-net for snapshot compressive imaging," arXiv Preprint, arXiv:2012.08364, 2020.

[43] Z. He, L. Wang, Z. Meng, and X. Yuan, "Self-supervised learning with spectral low-rank prior for hyperspectral image reconstruction," in *Proc. 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Tucson, AZ, USA: IEEE, Feb. 2025, pp. 9136–9145. doi: 10.1109/WACV61041.2025.00885

[44] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. Nonlinear Phenom.*, vol. 60, no. 1, pp. 259–268, Nov. 1992. doi: 10.1016/0167-2789(92)90242-F

[45] D. Strong and T. Chan, "Edge-preserving and scale-dependent properties of total variation regularization," *Inverse Probl.*, vol. 19, no. 6, S165, Nov. 2003. doi: 10.1088/0266-5611/19/6/059

[46] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2862–2869.

[47] Z. Zha, B. Wen, X. Yuan, J. Zhou, and C. Zhu, "Hyperspectral image denoising via nonlocal rank residual modeling," in *Proc. ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, June 2023, pp. 1–5. doi: 10.1109/ICASSP49357.2023.10096242

[48] Z. Zha, B. Wen, X. Yuan, J. Zhou, C. Zhu, and A. C. Kot, "Low-rankness guided group sparse representation for image restoration," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 10, pp. 7593–7607, Oct. 2023. doi: 10.1109/TNNLS.2022.3144630

[49] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, Dec. 2009. doi: 10.1007/s10208-009-9045-5

[50] M. Fazel, H. Hindi, and S. P. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *Proc. the 2001 American Control Conference. (Cat. No.01CH37148)*, 2001, pp. 4734–4739. doi: 10.1109/ACC.2001.945730

[51] J.-F. Cai, E. J. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," arXiv Preprint, arXiv:0810.3286, 2008. doi: 10.48550/arXiv.0810.3286

[52] S. Soleymanbaigi, A. Seyedi, F. A. Tab, and F. Daneshfar, "Encoder-decoder nonnegative matrix factorization with β-divergence for data clustering," *Pattern Recognit.*, vol. 171, 112211, Mar. 2026. doi: 10.1016/j.patcog.2025.112211

[53] F. Daneshfar, B. S. Saifee, S. Soleymanbaigi, and M. Aeini, "Elastic deep multi-view autoencoder with diversity embedding," *Inf. Sci.*, vol. 689, 121482, Jan. 2025. doi: 10.1016/j.ins.2024.121482

[54] L. Zhang, J. Nie, W. Wei, and Y. Zhang, "Unsupervised test-time adaptation learning for effective hyperspectral image super-resolution with unknown degeneration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 7, pp. 5008–5025, July 2024. doi: 10.1109/TPAMI.2024.3361894

[55] M. Li, Y. Fu, T. Zhang, and G. Wen, "Supervise-assisted self-supervised deep-learning method for hyperspectral image restoration," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 4, pp. 7331–7344, Apr. 2025. doi: 10.1109/TNNLS.2024.3386809

[56] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv Preprint, arXiv:1412.6980, 2014.