In2SET: Intra-Inter Similarity Exploiting Transformer for Dual-Camera Compressive Hyperspectral Imaging

Xin Wang[1]   Lizhi Wang[1*]   Xiangtian Ma[1]   Maoqing Zhang[1]   Lin Zhu[1]   Hua Huang[2]

[1]Beijing Institute of Technology   [2] Beijing Normal University

Abstract

Dual-camera compressive hyperspectral imaging (DC-CHI) offers the capability to reconstruct 3D hyperspectral image (HSI) by fusing compressive and panchromatic (PAN) image, which has shown great potential for snapshot hyperspectral imaging in practice. In this paper, we introduce a novel DCCHI reconstruction network, intra-inter similarity exploiting Transformer (In2SET). Our key insight is to make full use of the PAN image to assist the reconstruction. To this end, we propose to use the intra-similarity within the PAN image as a proxy for approximating the intra-similarity in the original HSI, thereby offering an enhanced content prior for more accurate HSI reconstruction. Furthermore, we propose to use the inter-similarity to align the features between HSI and PAN images, thereby maintaining semantic consistency between the two modalities during the reconstruction process. By integrating In2SET into a PAN-guided deep unrolling (PGDU) framework, our method substantially enhances the spatial-spectral fidelity and detail of the reconstructed images, providing a more comprehensive and accurate depiction of the scene. Experiments conducted on both real and simulated datasets demonstrate that our approach consistently outperforms existing state-of-the-art methods in terms of reconstruction quality and computational complexity. The code is available at `https://github.com/2JONAS/In2SET`.

1. Introduction

Hyperspectral imaging collects and processes scene information by dividing the whole spectrum into tens or hundreds of bands [15, 29, 57]. Owing to the capability of detailed scene representation, this technique has been widely adopted in many fields, e.g., medical diagnosis, health care, remote sensing [5, 11, 28, 37], and various computer vision tasks, e.g., recognition, classification, segmentation [26, 35, 49].

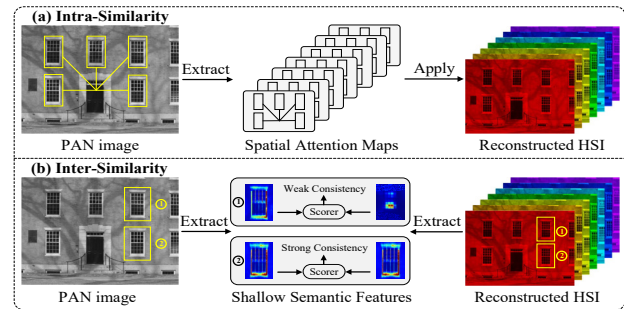The hyperspectral image (HSI) inherently possesses a

Figure 1. Illustration of the proposed In2SET method for hyperspectral image reconstruction. (a) Intra-Similarity: extraction and application of spatial attention maps from the PAN image to enhance the spatial resolution of the reconstructed HSI. (b) Inter-Similarity: utilization of semantic features from the HSI and PAN image, scored by their consistency, to inform and refine the reconstruction of HSI.

three-dimensional (3D) structure, comprising two dimensions of spatial information and one dimension of spectral information. Consequently, existing 2D imaging sensors lack the capability to directly capture this complete 3D signal. With the rapid progress in computational photography, snapshot spectral imagers have been developed to enable hyperspectral imaging in dynamic scenes [12, 25, 53]. The coded aperture snapshot spectral imaging (CASSI) [2, 19, 38], serving as a representative prototype of snapshot spectral imagers, captures one compressive image that encodes the spectral information by leveraging a decorated optical system. Inspired by multi-modal image fusion techniques, dual-camera compressive hyperspectral imaging (DCCHI) [39, 40, 43] upgrades the original CASSI by adding a panchromatic (PAN) camera to supply spatial information lost in CASSI. As a result, DCCHI can achieve more accurate reconstructions of HSI by fusing the compressive and PAN images compared to CASSI, which has shown great potential for snapshot hyperspectral imaging in practice.

The significant challenge of DCCHI lies in the HSI reconstruction method, where the key factor that determines

---

*Corresponding Author: Lizhi Wang (wanglizhi@bit.edu.cn)

the reconstruction performance is the ability to model HSI priors. Early methods mainly rely on general image priors such as smoothness[4, 39, 51], sparsity[17, 41, 52], low-rank[18, 30, 56], and non-local similarity[20, 42], which ignore the continuous and complete spatial characteristics of PAN images, a feature that could provide stronger image priors.

In recent years, some PAN-guided methods like ANSR [42, 45], PFusion [21], and PIDS [13] have utilized PAN image as prior information to guide the reconstruction of HSI. While these approaches have demonstrated merit, their performance is limited by the following two reasons: firstly, current PAN-guided methods rely on hand-crafted prior constraints, which lack a precise depiction of the intrinsic structures of HSI and often require manual parameter tuning. Consequently, when dealing with complex scenes, these methods tend to produce degraded reconstruction quality. Secondly, current PAN-guided methods typically model inter-modality correlations between PAN and HSI images from a single perspective, such as using edge information from the PAN images to constrain the HSI edges [13] or employing PAN image content to enrich HSI spatial detail [21]. However, HSI exhibits multi-faceted consistency with PAN images across various aspects such as structure, texture, luminance, contrast, and semantic content. Relying on one constraint may not fully exploit the comprehensive spatial-semantic guidance offered by the PAN image. In summary, accurately leveraging PAN image-guided prior modeling for HSI reconstruction is the key to further improving reconstruction quality.

In this paper, we propose an intra-inter similarity exploiting Transformer (In2SET) for DCCHI. The proposed In2SET module is integrated into a PAN-guided deep unrolling (PGDU) framework for reconstruction. The PGDU makes use of a multi-scale feature pyramid extracted from the PAN image as guidance. This feature pyramid provides hierarchical representations of the PAN at different resolutions, allowing the network to capture both global context and fine details, resulting in improved spatial fidelity and detail in the reconstructed images.

Our key insight is to make full use of the PAN image to assist the reconstruction. The In2SET is based on two observations: (1) given the large redundancy in both the spatial and the spectral dimensions, intra-similarity exist ubiquitously within HSI. Since the original HSI is not available in practice, intra-similarity needs to be computed based on intermediate reconstruction results. However, relying on these intermediate results for intra-similarity calculations introduces unreliability. Fortunately, the PAN image can be regarded as the integral of all spectral bands and provides a natural reference for estimating the intra-similarity in the original HSI. Thus, we propose to employ the intra-similarity in the PAN image as an approximation proxy for

the intra-similarity in the original HSI, which provides an enhanced content prior for HSI reconstruction. (2) Since the HSI and the PAN image describe the same scenes, inter-similarity should exist between the HSI and the PAN image, considering semantic information. Inter-similarity, serving as a scorer, assigns higher weights to features in areas of the HSI that are more similar to the corresponding areas in the PAN image, thereby improving feature representation in those areas and minimizing uncertainty and risk in the reconstruction process. This capability enables the reconstruction of the highly ill-posed inverse problem in a more confidently supervised manner, providing new contextual information for HSI reconstruction.

In a nutshell, this work integrates intra-similarity and inter-similarity with the advanced Transformer mechanism, specifically for DCCHI, resulting in high-fidelity HSI reconstruction. The main contributions of this work are as follows:

- We propose the PGDU, which employs the feature pyramid from the PAN image to guide the reconstruction of the HSI.
- We propose the In2SET, which employs a novel attention mechanism, specifically designed to concurrently capture both intra-similarity and inter-similarity between spectral and PAN images.
- Experiments on real and simulated data demonstrate that our method consistently outperforms state-of-the-art approaches.

2. Related Work

This section provides an overview of the advancements in HSI reconstruction, focusing on three pivotal areas: hand-crafted HSI priors modeling, data-driven HSI priors modeling, and PAN-guided HSI reconstruction.

2.1. Hand-Crafted HSI Priors Modeling

Early HSI reconstruction relied heavily on hand-crafted priors. Total Variation (TV) regularization [39, 54, 55] utilized the piecewise smoothness of images to constrain TV, albeit at the risk of suppressing important high-frequency structural details. Sparse representation techniques [3, 41, 52] have been employed, focusing on constructing dictionaries to model image sparsity. Low-rank priors [30, 47, 56] have been applied, designed to capture contextual information in high-dimensional data. However, these hand-crafted methods often required extensive manual parameter tuning and struggled to represent complex data scenarios effectively.

2.2. Data-Driven HSI Priors Modeling

The advent of deep learning has revolutionized the field of HSI reconstruction. Models such as HyperReconNet [44], λ-net [33], TSA-Net [31], DNU [46], DGSMP [23] and

PnP-DIP-HSI [32] have markedly improved reconstruction quality by modeling HSI data prior knowledge. Furthermore, the introduction of Transformer-based methods like MST [7], CST [6], DAU [9], RDLUF [16], and PADUT [27] has led to even more significant advances in reconstruction performance. Benefiting from the attention mechanism, Transformers can efficiently model both spatial and spectral similarities within HSI data, enabling a more flexible handling of input data and providing a richer feature representation for HSI reconstruction.

2.3. PAN-Guided HSI Reconstruction

The integration of PAN guidance in HSI reconstruction represents a significant stride in addressing the inherent limitations of traditional CASSI. ANSR[42, 45] proposed an adaptive non-local sparse representation model, guided by the PAN image. PFusion[21] leveraged RGB measurements for spatial coefficient estimation and CASSI measurements for spectral basis acquisition. It features a patch processing strategy to enhance the spectral low-rank property, optimizing the model efficiently without iterative methods or a spectral sensing matrix. PIDS[13] utilized the RGB measurement as a prior image to enhance semantic correspondence between HSI and the PAN image. These methods have marked significant advancements in the field. However, they often require manual parameter tuning and fail to leverage the rich spatial context and detailed features embedded within the PAN image, suggesting potential areas for further improvements in reconstruction.

Our motivation is to utilize the spatial semantic information provided by the PAN image to improve the HSI reconstruction result in DCCHI. We aim to develop a method that combines the strengths of PAN guidance to model the intra-inter similarity for more accurate reconstruction.

3. PGDU

3.1. DCCHI Forward Model

DCCHI consists of a beam splitter, a CASSI branch, and a PAN camera branch, as shown in Figure 2. The beam splitter divides the incident light from the scene into two equal-intensity parts. One part of the light is captured by CASSI, which undergoes spatial modulation and spectral modulation to form the CASSI measurement. The other part of the light is directly captured by the PAN camera.

Let $\mathcal{X} \in \mathbb{R}^{H \times W \times C}$ represent the spectral image of the target scene, where H, W, and C are the height, width, and number of spectral bands, respectively. The spatial and spectral modulation in CASSI is given by

$$\mathbf{Y^c} = \sum_{c=1}^{C} \text{shift}(\phi \odot \mathcal{X}[:, :, c]) + \mathbf{V^c}, \quad (1)$$
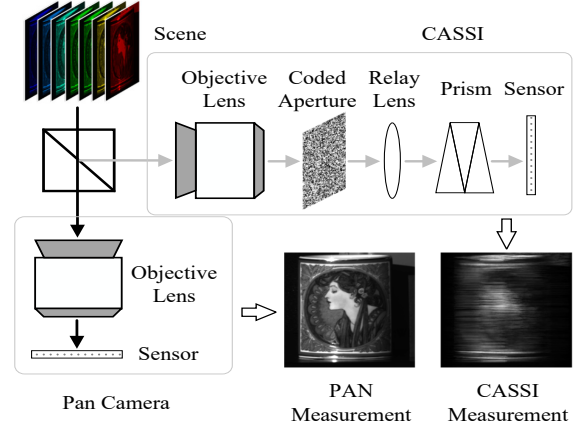

Figure 2. The dual-camera compressive hyperspectral imaging system.

where ϕ is the transmission intensity of the coded aperture, represented as $\phi \in \mathbb{R}^{H \times W}$, $\text{shift}(\cdot)$ indicates the spectral modulation caused by the dispersion prism, $\mathbf{Y^c}$ is the image captured by the CASSI branch, and $\mathbf{V^c}$ is regarded as Gaussian noise in CASSI.

To simplify to a matrix-vector form, the imaging model of the CASSI branch is represented as

$$\mathbf{y^c} = \mathbf{\Phi^c x} + \mathbf{v^c}, \quad (2)$$

where $\mathbf{\Phi^c}$ denotes the imaging model of the CASSI branch sensing matrix, $\mathbf{y^c}$ is the vector form of $\mathbf{Y^c}$, $\mathbf{x}$ is the vector form of $\mathcal{X}$, and $\mathbf{v^c}$ is the vector form of $\mathbf{V^c}$.

Correspondingly, the imaging model of the PAN camera branch is described by

$$\mathbf{y^p} = \mathbf{\Phi^p x} + \mathbf{v^p}, \quad (3)$$

where $\mathbf{\Phi^p}$ denotes the imaging model of the PAN camera branch sensing matrix, determined by the spectral response of the sensor, $\mathbf{y^c}$ is the vector form of PAN measurement, and $\mathbf{v^p}$ is regarded as Gaussian noise in the PAN measurement. To connect the imaging models to the subsequent discussion, we define $\mathbf{y}$, $\mathbf{\Phi}$, and $\mathbf{v}$ as

$$\mathbf{y} = \begin{bmatrix} \mathbf{y^c} \\ \mathbf{y^p} \end{bmatrix}, \quad \mathbf{\Phi} = \begin{bmatrix} \mathbf{\Phi^c} \\ \mathbf{\Phi^p} \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} \mathbf{v^c} \\ \mathbf{v^p} \end{bmatrix}. \quad (4)$$

Using the above definitions, the imaging model can be expressed as

$$\mathbf{y} = \mathbf{\Phi x} + \mathbf{v}, \quad (5)$$

where $\mathbf{\Phi}$ is a sensing matrix in $\mathbb{R}^{M \times N}$ that encodes the system mapping from the original scene $\mathbf{x}$ to the measurements $\mathbf{y}$. $\mathbf{y} \in \mathbb{R}^{M \times 1}$, $\mathbf{x} \in \mathbb{R}^{N \times 1}$, and $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$, The dimensions M and N are defined as follows: M is calculated as $H(W + d(C - 1)) + HW$ and N is calculated as HWC, where d represents the dispersion step size. It is important to note that in the context of HSI reconstruction, M is typically much smaller than N, highlighting that the HSI reconstruction task in DCCHI is an ill-posed problem.
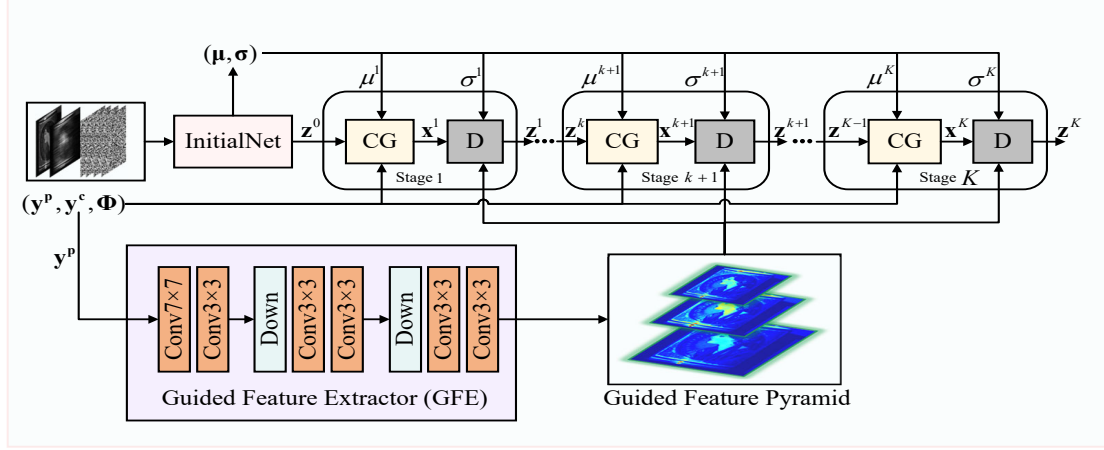
Figure 3. **Overview of PGDU.** The InitialNet initiates the process with compressive measurements and sensing matrix, followed by a series of stages each containing a conjugate gradient (CG) block and an In2SET denoiser.

3.2. Reconstruction Framework

Compared to learn a brute-force mapping between HSI $\mathbf{x}$ and measurement $\mathbf{y}$, model-based methods have potential in compressive imaging due to fully considering the imaging model. Model-based methods usually fomulate HSI reconstruction as a Bayesian inference challenge, solving Eq. (5) under a unified Maximum A Posteriori (MAP) framework. Mathematically, the optimization problem for HSI reconstruction could be expressed as

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{\Phi x}\|_2^2 + \eta J(\mathbf{x}), \quad (6)$$

where $J(\mathbf{x})$ represents regularization. The parameter η is a regularization coefficient, controlling the trade-off between data fidelity and the regularization term imposed by $J(\mathbf{x})$.

To solve the optimization problem shown in Eq. (6), we design the PGDU based on the physical structure of DC-CHI, as shown in Figure 3. Utilizing the half quadratic splitting (HQS) method, we decompose the optimization problem into two sub-problems: the data fidelity term, as shown in Eq. (7); and the prior term, as shown in Eq. (8).

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi x}\|_2^2 + \mu^{k+1}\left\|\mathbf{z}^k - \mathbf{x}\right\|_2^2, \quad (7)$$

$$\mathbf{z}^{k+1} = \arg\min_{\mathbf{z}} \frac{1}{2}\mu^{k+1}\left\|\mathbf{z} - \mathbf{x}^{k+1}\right\|_2^2 + \eta^{k+1} J(\mathbf{z}), \quad (8)$$

the superscript k indicates the stage index number, and $\mathbf{z}$ is an auxiliary variable introduced to facilitate the optimization. The parameter μ^{k+1} is a relaxation parameter that governs the alignment between the current estimate $\mathbf{x}$ and the auxiliary variable $\mathbf{z}$. The parameter η^{k+1} is a stage-specific parameter for the regularization coefficient η.

The data fidelity term can be directly calculated in closed form, as demonstrated in Eq. (9).

$$\mathbf{x}^{k+1} = \left(\mathbf{\Phi}^{\mathsf{T}}\mathbf{\Phi} + \mu\mathbf{I}\right)^{-1}\left(\mathbf{\Phi}^{\mathsf{T}}\mathbf{y} + \mu^{k+1}\mathbf{z}^k\right), \quad (9)$$

in the context of HSI reconstruction in CASSI, the matrix $\mathbf{\Phi}^{\mathsf{T}}\mathbf{\Phi}$ is strictly diagonal. However, with the addition of the PAN camera branch in DCCHI, the matrix $\mathbf{\Phi}^{\mathsf{T}}\mathbf{\Phi}$ becomes only diagonally dominant, and thus a closed-form solution cannot be explicitly calculated.

Given the non-diagonal form of $\mathbf{\Phi}^{\mathsf{T}}\mathbf{\Phi}$, we employ the conjugate gradient (CG) method [34] for the efficient resolution of the data fidelity term. The CG method is well suited for this problem due to its effectiveness in solving large-scale, sparse linear systems.

To fully explore the guidance of the PAN image $\mathbf{y}^{\mathbf{p}}$ on prior term, we design a PAN-guided Gaussian denoiser to estimate $\mathbf{z}^{k+1}$. This is formulated as Eq. (10):

$$\mathbf{z}^{k+1} = \text{Denoiser}_{k+1}\left(\mathbf{x}^{k+1}, \mathbf{GFE}(\mathbf{y}^{\mathbf{p}}), \sigma^{k+1}\right), \quad (10)$$

where $\mathbf{GFE}(\mathbf{y}^{\mathbf{p}})$ serves as structural guidance for the denoising, which incorporates multi-scale spatial details from the high-resolution $\mathbf{y}^{\mathbf{p}}$ data into the denoiser, aiming to refine the fidelity of the reconstructed HSI by adding contextual information. The Gaussian noise level of the denoiser, denoted as σ^{k+1}, is equal to $\sqrt{\eta^{k+1}/\mu^{k+1}}$. The μ^{k+1} and σ^{k+1} are the elements of vectors $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, which are inferred by the InitialNet of PGDU.

The feature extraction from $\mathbf{y}^{\mathbf{p}}$ is delineated as follows: within the unrolling framework, the feature representation of $\mathbf{y}^{\mathbf{p}}$ is extracted singularly and iteratively shared across multiple stages. A stack of convolutional blocks is utilized to harvest feature maps at varying resolutions, assembling a guided feature pyramid $[\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3]$. The guided feature pyramid is integral to the subsequent denoising stage, aiding in the preservation of intricate spatial details during the denoising.
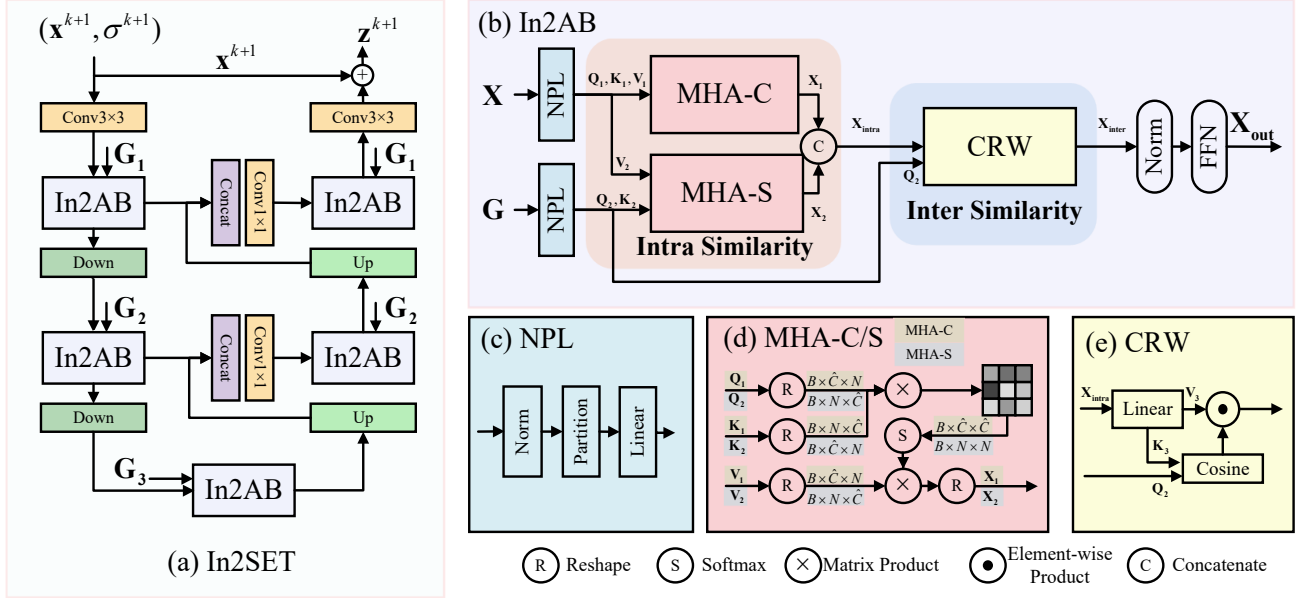
Figure 4. Diagram of In2SET architecture. (a) The U-shaped In2SET structure. (b) In2AB, consisting of two normalization layers, an intra-similarity attention module, an inter-similarity attention module, and an FFN layer. (c) The components of NPL. (d) The multi-head self-attention in channel (MHA-C) and multi-head cross-attention in spatial (MHA-S). (e) The cosine similarity reweighting (CRW) mechanism.

4. In2SET

4.1. Overall Architecture

In this section, we introduce the denoiser In2SET within the PGDU framework. As shown in Figure 4 (a), the proposed In2SET adopts a U-shaped architecture [36], involving the intra-inter attention block (In2AB) as the foundational units. This architecture employs multi-scale feature extraction coupled with skip connections between the encoder and decoder, enhancing the efficiency in information processing and feature extraction. To be specific, the denoiser initializes feature map $\mathbf{X}$ from the spectral image $\mathbf{x}^{k+1}$ and the noise level σ^{k+1}, passing through a convolutional layer with a kernel size of 3×3. The input fed into each In2AB consists of guided feature $\mathbf{G_i}$, feature maps extracted from the PAN image, and HSI feature $\mathbf{X}$. The denoised image $\mathbf{z}^{k+1}$ is the sum of $\mathbf{x}^{k+1}$ and the output of the last In2AB, calculated by a convolutional layer with a kernel size of 3×3. The In2AB structure is shown in Figure 4 (b). Our In2AB module includes two normalization layers, one intra-similarity attention module, one inter-similarity attention module, and one feed-forward fully connected (FFN) layer. In the following, we introduce the detail of the intra-similarity and inter-similarity.

4.2. Intra-Similarity Attention

Our intra-similarity attention module is grounded on two key principles: firstly, for exploring intra-spatial similarity,

the comprehensive and continuous spatial data provided by the PAN image are invaluable. This allows the spatial similarity in the PAN image to serve as an effective proxy for approximating the spatial similarity of the HSI. Secondly, in addressing intra-spectral similarity, analyzing the reconstructed HSI becomes significantly more effective, given the PAN image's deficiency in spectral color information.

As shown in Figure 4 (b), our intra-similarity attention module includes two branches: multi-head self-attention in channel (MHA-C) and multi-head cross-attention in spatial (MHA-S), representing the exploration of spectral and spatial self-similarity, respectively. The MHA-C is based on the work in MST [7], which has been demonstrated to effectively explore spectral self-similarity.

The input features $\mathbf{X}$ and $\mathbf{G}$ have dimensions $\mathbb{R}^{H \times W \times C}$ and $\mathbb{R}^{H \times W \times \hat{C}}$, respectively, where $\hat{C} = \frac{C}{2}$. After normalization and partitioning, the dimensions become $\mathbb{R}^{B \times N \times C}$ and $\mathbb{R}^{B \times N \times \hat{C}}$. In the first and last layers of HSAB within In2SET, $B = \frac{HW}{M^2}$ and $N = M^2$. For the remaining layers, dimensions are recalibrated to $B = M^2$ and $N = \frac{HW}{M^2}$. Following linear projection as shown in Eq. (11) and Eq. (12).

$$\mathbf{Q_1}, \mathbf{K_1}, \mathbf{V_1}, \mathbf{V_2} = L_{Q_1, K_1, V_1, V_2}(\mathbf{X}), \qquad (11)$$

$$\mathbf{Q_2}, \mathbf{K_2} = L_{Q_2, K_2}(\mathbf{G}), \qquad (12)$$

where $\mathbf{X}$ is projected to $\mathbf{Q_1}, \mathbf{K_1}, \mathbf{V_1} \in \mathbb{R}^{B \times \hat{C} \times N}$ and

Method	GFLOPs	Scene										Average
		01	02	03	04	05	06	07	08	09	10	
PFusion-RGB [21]	-	40.09	38.84	38.70	46.65	32.07	37.12	39.74	36.75	34.52	35.53	38.00
		0.979	0.968	0.966	0.936	0.980	0.980	0.964	0.965	0.931	0.979	0.965
PIDS-RGB [13]	-	42.09	40.08	41.50	48.55	40.05	39.00	36.63	37.02	38.82	38.64	40.24
		0.983	0.949	0.968	0.989	0.982	0.974	0.940	0.948	0.953	0.980	0.967
TV-DC [39]	-	35.81	33.22	31.07	40.11	33.32	34.62	31.09	32.31	29.36	33.84	33.47
		0.947	0.885	0.879	0.947	0.944	0.943	0.885	0.916	0.862	0.953	0.910
PIDS-DC [13]	-	39.82	37.07	37.72	46.78	37.45	37.74	32.90	31.66	34.35	38.58	37.41
		0.977	0.921	0.950	0.978	0.973	0.963	0.896	0.915	0.902	0.972	0.945
BiSRNet-DC [10]	1.33	35.02	34.13	31.50	35.88	33.70	35.58	32.31	32.73	31.37	34.48	33.67
		0.945	0.914	0.883	0.895	0.935	0.925	0.900	0.903	0.899	0.936	0.914
CST-DC [6]	25.40	37.44	38.91	36.79	42.27	36.57	38.91	36.87	35.91	35.87	37.93	37.75
		0.975	0.978	0.969	0.983	0.982	0.984	0.967	0.980	0.973	0.990	0.978
HDNet-DC [22]	144.31	38.06	39.79	38.21	42.79	37.22	39.26	37.41	36.51	36.64	37.52	38.34
		0.976	0.982	0.973	0.983	0.983	0.986	0.968	0.982	0.976	0.988	0.980
MST-DC [7]	25.77	38.38	40.45	37.63	42.88	37.72	39.66	37.56	37.40	37.86	38.10	38.76
		0.976	0.976	0.966	0.975	0.978	0.976	0.965	0.971	0.971	0.981	0.974
MST++-DC [8]	17.69	38.38	40.47	37.70	43.88	37.75	39.42	37.48	37.38	38.82	39.04	39.03
		0.977	0.980	0.968	0.984	0.983	0.985	0.964	0.982	0.980	0.989	0.980
DAUHST-DC-2stg [9]	16.79	40.78	43.13	41.73	47.09	39.84	40.90	39.75	38.98	41.29	40.04	40.22
		0.983	0.987	0.980	0.990	0.987	0.986	0.976	0.981	0.983	0.988	0.983
DAUHST-DC-3stg [9]	24.70	40.22	43.52	41.74	47.07	38.81	40.16	39.86	38.21	40.63	39.32	40.95
		0.983	0.989	0.981	0.993	0.985	0.987	0.978	0.981	0.983	0.990	0.985
DAUHST-DC-5stg [9]	40.51	40.74	44.00	41.58	46.84	39.66	40.89	40.21	38.72	39.98	40.10	41.27
		0.984	0.989	0.981	0.991	0.986	0.987	0.979	0.983	0.982	0.989	0.916
DAUHST-DC-9stg [9]	72.11	41.59	45.19	43.47	48.92	40.27	41.17	40.73	40.11	43.50	41.33	42.62
		0.985	0.991	0.984	0.993	0.988	0.988	0.979	0.986	0.988	0.990	0.987
In2SET-2stg (Ours)	14.35	40.33	42.30	40.34	47.24	39.42	40.61	39.46	38.42	40.37	39.96	40.84
		0.983	0.985	0.977	0.991	0.986	0.986	0.975	0.979	0.981	0.988	0.983
In2SET-3stg (Ours)	20.79	40.78	43.13	41.73	47.09	39.84	40.90	39.75	38.98	41.29	40.04	41.35
		0.983	0.987	0.980	0.990	0.987	0.986	0.976	0.981	0.983	0.988	0.984
In2SET-5stg (Ours)	33.66	41.13	44.43	42.74	47.29	40.33	40.95	40.49	39.15	42.07	39.44	41.80
		0.985	0.990	0.983	0.993	0.988	0.987	0.979	0.982	0.985	0.987	0.985
In2SET-9stg (Ours)	59.40	**42.56**	**46.42**	**44.55**	**50.63**	**42.01**	**42.49**	**41.59**	**40.53**	**43.83**	**42.33**	**43.69**
		0.989	**0.994**	**0.986**	**0.996**	**0.992**	**0.991**	**0.983**	**0.989**	**0.990**	**0.994**	**0.990**

Table 1. Comparison of In2SET with SOTA DCCHI methods across 10 simulated scenes, including FLOPS, PSNR (upper entry), and SSIM (lower entry). "Method-RGB" indicates the use of RGB observations, while "Method-DC" denotes grayscale observations in PAN images. It is noted that "Method-DC" is modified from other HSI reconstruction method for the DCCHI reconstruction task.

$\mathbf{V_2} \in \mathbb{R}^{B \times N \times \hat{C}}$. $\mathbf{G}$ is projected to $\mathbf{Q_2}, \mathbf{K_2} \in \mathbb{R}^{B \times N \times \hat{C}}$.

The computations of MHA-C and MHA-S can be represented by Eq. (13) and Eq. (14),

$$\mathbf{X_1} = \mathrm{Softmax} \left(\frac{\mathbf{Q_1} \mathbf{K_1}^T}{\sqrt{d_{h1}}} + \mathbf{P_1} \right) \mathbf{V_1}, \qquad (13)$$

$$\mathbf{X_2} = \mathrm{Softmax} \left(\frac{\mathbf{Q_2} \mathbf{K_2}^T}{\sqrt{d_{h2}}} + \mathbf{P_2} \right) \mathbf{V_2}, \qquad (14)$$

where $\sqrt{d_{h1}}$ and $\sqrt{d_{h2}}$ are scalars as defined in [9], $\mathbf{P_1} \in \mathbb{R}^{B \times \hat{C} \times \hat{C}}$, and $\mathbf{P_2} \in \mathbb{R}^{B \times N \times N}$ denote learnable positional encodings.

The output of intra-attention $\mathbf{X_{intra}}$ can be obtained by Eq. (15).

$$\mathbf{X_{intra}} = \mathrm{Concat} \left(\mathbf{X_1}, \mathbf{X_2} \right). \qquad (15)$$

4.3. Inter-Similarity Attention

We establish our inter-similarity attention module based on one principle: the shallow semantic features of the PAN image and the corresponding HSI area are similar, including

consistency in texture, shape features, and low-level image attributes (brightness and contrast). The credibility of the reconstructed area is directly proportional to this consistency. Inter-similarity attention explores cross-modal similarity using cosine similarity reweighting (CRW).

After exploiting intra-similarity, the output of the intra-similarity attention module $\mathbf{X_{intra}}$ is projected to obtain $\mathbf{K_3}$ and $\mathbf{V_3}$,

$$\mathbf{K_3}, \mathbf{V_3} = L_{K_3, V_3}(\mathbf{X_{intra}}), \qquad (16)$$

where $\mathbf{K_3} \in \mathbb{R}^{B \times N \times \hat{C}}$ and $\mathbf{V_3} \in \mathbb{R}^{B \times N \times C}$.

Then, the computations of CRW can be represented as follows:

$$\mathbf{X_{inter}} = \frac{\mathbf{Q_2} \cdot \mathbf{K_3}}{\|\mathbf{Q_2}\| \|\mathbf{K_3}\|} \odot \mathbf{V_3}, \qquad (17)$$

where the cosine similarity computed along the last dimension between $\mathbf{Q_2}$ and $\mathbf{K_3}$ quantifies the feature alignment in corresponding areas. A higher similarity score reflects a stronger feature correspondence, which, in turn, informs the enhanced weighting of $\mathbf{V_3}$, improving the fidelity of the region being reconstructed.

5. Experiments

In this section, we begin by introducing detailing the network training process. Subsequently, we evaluate the performance of In2SET on both simulation and real-world DC-CHI datasets. Then, we validate the effectiveness of the intra-similarity block using statistical simulation data. Finally, we perform an ablation study to demonstrate the effectiveness of our proposed method.

Simulation Dataset. Two distinct datasets are used: CAVE [50] and KAIST [14]. The CAVE dataset is made up of 32 hyperspectral images with 512×512 spatial resolution. The KAIST dataset contains 30 hyperspectral images, each with a larger spatial dimension of 2704×3376. The CAVE dataset is used as the training data, while a subset of 10 scenes crop from the KAIST dataset is used for the evaluation phase, in accordance with the protocols established in references [7, 9, 22, 23, 31].

Implementation. Our method is implemented with Pytorch and trained with Adam [24] optimizer for 300 epochs. During training, the learning rate is 4×10^{-4} using the cosine annealing scheduler, and the loss function for network training is L_1 loss. We randomly extract patches from 3D HSI cubes to serve as training samples. The dimensions of these patches are $256 \times 256 \times 28$ for the simulation experiment and $350 \times 260 \times 26$ for the real-world experiment. In simulation imaging model, we configure a dispersion shift step d of 2, directing the dispersion to the right. In real-word imaging model, the dispersion shift step d is set to 1, with the dispersion oriented upwards.

5.1. Simulation Data Results

The reconstruction quality of hyperspectral images are assessed using peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [48] metrics.

The Table 1 clearly shows that In2SET excels in all test scenarios, particularly in In2SET-9stg, achieving the best performance across all scenes with an average PSNR of 43.69 dB and a SSIM of 0.990. This significantly surpasses model optimization methods like TV-DC and PIDS-DC. Compared to deep learning-based methods, such as various stages of DAUHST-DC [9], In2SET demonstrates remarkable superiority. In2SET-3stg achieves performance comparable to that of DAUHST-DC-5stg [9], but with only 51.3% of the cost, amounting to 20.78 GFLOPs compared to 40.50 GFLOPs.

To facilitate a direct visual quality assessment, we analyzed the distinction between Ground Truth (GT) and the reconstructions produced by various open-source methods. Figure 5 presents reconstruction results for selected spectral bands. These results are generated from two model-based approaches, TV-DC [39] and PIDS-DC [13], five end-to-end network algorithms, BiSRNet-DC [10], CST-DC [6], HDNet-DC [22], MST-DC [7], MST++-DC [8], as well as

	RMSE	Correlation	PSNR(dB)
CAVE	0.034	0.999	36.28
KAIST	0.025	0.996	35.74
ICVL	0.076	0.997	31.10

Table 2. Comparison of correlation maps between HSI and PAN image, including average Root Mean Squared Error (RMSE), average correlation, and PSNR.

Baseline	CRW	MHA-C	MHA-S	PSNR(dB)	SSIM	GFLOPs
✓				37.23	0.944	9.87
✓	✓			40.05	0.980	11.19
✓	✓	✓		40.46	0.982	14.06
✓	✓	✓	✓	40.84	0.983	14.35

Table 3. Break-down ablation study on individual components of the proposed method.

Methods	CG-1	CG-2	CG-5	CG-10
PSNR(dB)	40.42	40.66	40.84	41.05
SSIM	0.982	0.983	0.983	0.984
FPS	26.45	22.79	19.16	13.99

Table 4. Ablation study on iterative performance of conjugate gradient descent for data item.

from two deep unfolding methods, namely DAUHST-DC-9stg [9] and our In2SET-9stg. The comparison underlines the exceptional ability of our In2SET approach to maintain spatial and spectral fidelity.

We suggest that the intra similarity in PAN image can effectively approximates that in HSI. To support this, we experiment with three public datasets: CAVE [50], KAIST [14] and ICVL [1]. We compute spatial correlation maps for each scene, comparing the correlation map of the HSI with that of the corresponding PAN image. The results of this comparison are displayed in Table 2, indicating a high degree of similarity between the intra-spatial correlations in PAN images and HSI. More visual results are presented in the supplementary materials.

5.2. Real Data Results

In this research, we used a real-world DCCHI measurement *Ninja*, taken from publicly available data as detailed in reference [42]. Figure 6 illustrates the reconstruction results for four spectral bands in this scene, using various DCCHI reconstruction algorithms. The comparison highlights the superior image restoration quality of our model over other methods, validating its effectiveness and reliability in real-world applications.

5.3. Ablation Study

To verify the effectiveness of our proposed method, we conducted ablation studies for the In2SET method. All evaluations are conducted on simulated datasets.

Break-down Ablation. The break-down study, as delineated in Table 3, offers a detailed examination of how each component within our proposed method influences reconstruction performance. Starting from a baseline derived by removing all ablated components from In2SET-2stg, the
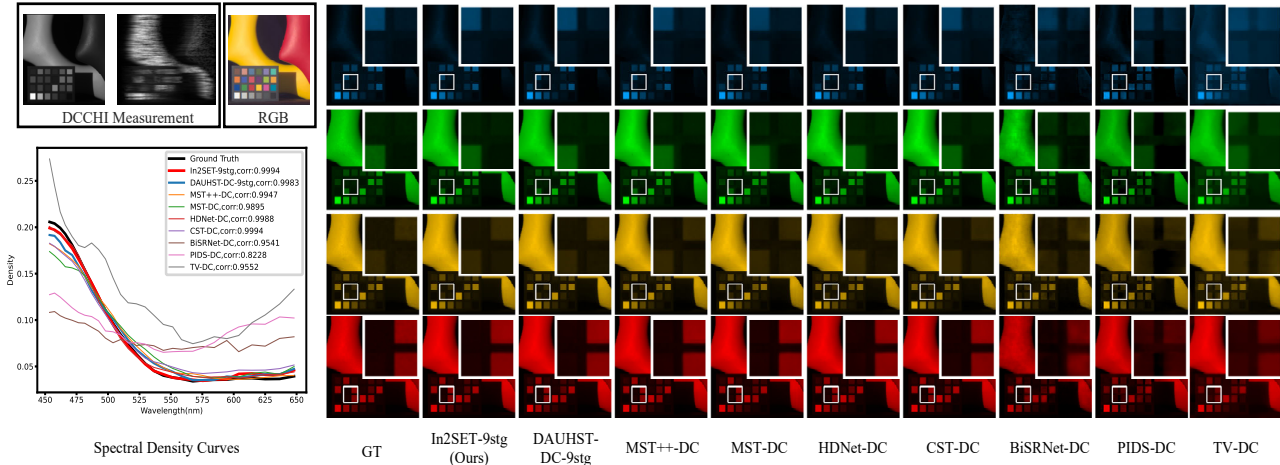
Figure 5. Comparative reconstruction results of different reconstruction methods for Scene 3 from the KAIST dataset at spectral bands 476.5nm, 536.5nm, 584.5nm, and 625.0nm. The spectral density curves are plotted from the blue region in the colorchecker.
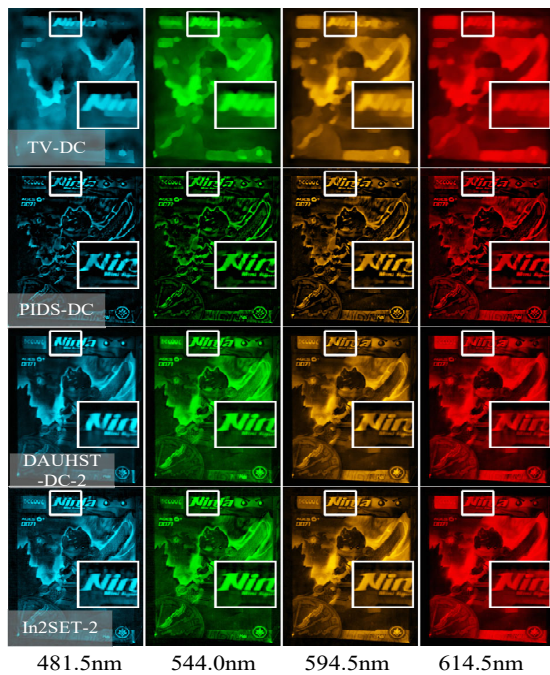

Figure 6. Visualization of reconstruction performance across different spectral bands on real data. The spectral bands selected for comparison are 481.5nm, 544.0nm, 594.5nm, and 614.5nm.

model achieves a PSNR of 37.23 dB. The addition of the CRW component leads to a significant increase in PSNR by +2.82 dB, albeit at an increased computational cost of +1.32 GFLOPs. The subsequent integration of MHA-C further elevates the PSNR by +0.41 dB. The final inclusion of MHA-S brings an additional PSNR gain of +0.38 dB. These step-by-step enhancements highlight the critical role each component plays in enhancing the overall capability of the model for HSI reconstruction, demonstrating a progressive improvement in performance with each added component.

Ablation Study of CG Iterations. Table 4 presents a comparative analysis of PSNR, SSIM, and Frames Per Second (FPS) metrics across varying numbers of CG iterations for HSI reconstruction. The CG-1, which conducts a single iteration, essentially operates as a vanilla gradient descent. These experiments are conducted on a system equipped with a TITAN Xp GPU (12GB) and an Intel (R) Xeon (R) Platinum 8358P CPU @ 2.60GHz.

There is a clear trend of improvement in PSNR as the number of CG iterations increases from 1 to 10, with values escalating from 40.42 dB to 41.05 dB. However, an increase in iterations from CG-1 to CG-10 leads to a decrease in FPS from 26.45 dB to 13.99 dB, indicating a trade-off between reconstruction quality and computational speed. A notable observation is that CG-5 offers an optimal balance between time efficiency and performance, making it the preferred choice for our In2SET network.

6. Conclusion

In this paper, we propose the In2SET for DCCHI reconstruction. Our method maximizes PAN image utility for HSI reconstruction by: (1) approximating HSI intra-similarity using PAN image, addressing the unreliability of direct computation from intermediate results. (2) Leveraging inter-similarity between HSI and PAN images to accurately reconstruct regions, providing cues for uncertain areas. Integrating In2SET into a PGDU framework allowed us to substantially enhance the spatial-spectral fidelity and detail of reconstructed images. Experiments on real and simulated datasets show that our method consistently outperforms the state-of-the-art while maintaining lower computational complexity.

References

[1] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *European Conference on Computer Vision*, pages 19–34, 2016. 7

[2] Gonzalo R Arce, David J Brady, Lawrence Carin, Henry Arguello, and David S Kittle. Compressive coded aperture spectral imaging: An introduction. *IEEE Signal Processing Magazine*, 31(1):105–115, 2013. 1

[3] Henry Arguello, Hoover Rueda, Yuehao Wu, Dennis W Prather, and Gonzalo R Arce. Higher-order computational model for coded aperture spectral imaging. *Applied Optics*, 52(10):D12–D21, 2013. 2

[4] José M Bioucas-Dias and Mário AT Figueiredo. A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Transactions on Image Processing*, 16 (12):2992–3004, 2007. 2

[5] José M Bioucas-Dias, Antonio Plaza, Gustavo Camps-Valls, Paul Scheunders, Nasser Nasrabadi, and Jocelyn Chanussot. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and Remote Sensing Magazine*, 1 (2):6–36, 2013. 1

[6] Yuanhao Cai, Jing Lin, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Coarse-to-fine sparse transformer for hyperspectral image reconstruction. In *European Conference on Computer Vision*, pages 686–704, 2022. 3, 6, 7

[7] Yuanhao Cai, Jing Lin, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 17502–17511, 2022. 3, 5, 6, 7

[8] Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, Radu Timofte, and Luc Van Gool. Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition Workshop*, pages 745–755, 2022. 6, 7

[9] Yuanhao Cai, Jing Lin, Haoqian Wang, Xin Yuan, Henghui Ding, Yulun Zhang, Radu Timofte, and Luc Van Gool. Degradation-aware unfolding half-shuffle transformer for spectral compressive imaging. In *Advances in Neural Information Processing Systems*, pages 37749–37761, 2022. 3, 6, 7

[10] Yuanhao Cai, Yuxin Zheng, Jing Lin, Xin Yuan, Yulun Zhang, and Haoqian Wang. Binarized spectral compressive imaging. In *Advances in Neural Information Processing Systems*, pages 38335–38346, 2023. 6, 7

[11] Mihaela A Calin, Sorin V Parasca, Dan Savastru, and Dragos Manea. Hyperspectral imaging in the medical field: Present and future. *Applied Spectroscopy Reviews*, 49(6):435–447, 2014. 1

[12] Xun Cao, Tao Yue, Xing Lin, Stephen Lin, Xin Yuan, Qionghai Dai, Lawrence Carin, and David J Brady. Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world. *IEEE Signal Processing Magazine*, 33 (5):95–108, 2016. 1

[13] Yurong Chen, Yaonan Wang, and Hui Zhang. Prior image guided snapshot compressive spectral imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 2, 3, 6, 7

[14] Inchang Choi, Daniel S Jeon, Giljoo Nam, Diego Gutierrez, and Min H Kim. High-quality hyperspectral reconstruction using a spectral prior. *ACM Transactions on Graphics*, 36 (6):1–13, 2017. 7

[15] Weisheng Dong, Huan Wang, Fangfang Wu, Guangming Shi, and Xin Li. Deep spatial–spectral representation learning for hyperspectral image denoising. *IEEE Transactions on Computational Imaging*, 5(4):635–648, 2019. 1

[16] Yubo Dong, Dahua Gao, Tian Qiu, Yuyan Li, Minxi Yang, and Guangming Shi. Residual degradation learning unfolding framework with mixing priors across spectral and spatial for compressive spectral imaging. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 22262–22271, 2023. 3

[17] Mário AT Figueiredo, Robert D Nowak, and Stephen J Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of Selected Topics in Signal Processing*, 1(4): 586–597, 2007. 2

[18] Ying Fu, Yinqiang Zheng, Imari Sato, and Yoichi Sato. Exploiting spectral-spatial correlation for coded hyperspectral image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3727–3736, 2016. 2

[19] Michael E Gehm, Renu John, David J Brady, Rebecca M Willett, and Timothy J Schulz. Single-shot compressive spectral imaging with a dual-disperser architecture. *Optics Express*, 15(21):14013–14027, 2007. 1

[20] Wei He, Quanming Yao, Chao Li, Naoto Yokoya, and Qibin Zhao. Non-local meets global: An integrated paradigm for hyperspectral denoising. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 6868–6877, 2019. 2

[21] Wei He, Naoto Yokoya, and Xin Yuan. Fast hyperspectral image recovery of dual-camera compressive hyperspectral imaging via non-iterative subspace-based fusion. *IEEE Transactions on Image Processing*, 30:7170–7183, 2021. 2, 3, 6

[22] Xiaowan Hu, Yuanhao Cai, Jing Lin, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Hdnet: High-resolution dual-domain learning for spectral compressive imaging. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 17542–17551, 2022. 6, 7

[23] Tao Huang, Weisheng Dong, Xin Yuan, Jinjian Wu, and Guangming Shi. Deep gaussian scale mixture prior for spectral compressive imaging. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 16216–16225, 2021. 2, 7

[24] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint*, 2014. 7

[25] Edmund Y Lam. Computational photography with plenoptic camera and light field capture: Tutorial. *Journal of the Optical Society of America A*, 32(11):2021–2032, 2015. 1

[26] Jun Li, José M Bioucas-Dias, and Antonio Plaza. Hyperspectral image segmentation using a new bayesian approach with active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 49(10):3947–3960, 2011. 1

[27] Miaoyu Li, Ying Fu, Ji Liu, and Yulun Zhang. Pixel adaptive deep unfolding transformer for hyperspectral image reconstruction. In *IEEE International Conference on Computer Vision*, pages 12959–12968, 2023. 3

[28] Robert P Lin, Brian R Dennis, Gordon J Hurford, DM Smith, Alex Zehnder, PR Harvey, David W Curtis, Dave Pankow, Paul Turin, M Bester, et al. The reuven ramaty high-energy solar spectroscopic imager (rhessi). In *The Reuven Ramaty High-Energy Solar Spectroscopic Imager (RHESSI): Mission Description and Early Results*, pages 3–32, 2003. 1

[29] Xing Lin, Yebin Liu, Jiamin Wu, and Qionghai Dai. Spatial-spectral encoded compressive hyperspectral imaging. *ACM Transactions on Graphics*, 33(6):1–11, 2014. 1

[30] Yang Liu, Xin Yuan, Jinli Suo, David Brady, and Qionghai Dai. Rank minimization for snapshot compressive imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(12):2990–3006, 2018. 2

[31] Ziyi Meng, Jiawei Ma, and Xin Yuan. End-to-end low cost compressive spectral imaging with spatial-spectral self-attention. In *European Conference on Computer Vision*, pages 187–204, 2020. 2, 7

[32] Ziyi Meng, Zhenming Yu, Kun Xu, and Xin Yuan. Self-supervised neural networks for spectral snapshot compressive imaging. In *IEEE International Conference on Computer Vision*, pages 2622–2631, 2021. 3

[33] Xin Miao, Xin Yuan, Yunchen Pu, and Vassilis Athitsos. Lambda-net: Reconstruct hyperspectral images from a snapshot measurement. In *IEEE International Conference on Computer Vision*, pages 4059–4069, 2019. 2

[34] John L Nazareth. Conjugate gradient method. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(3):348–353, 2009. 4

[35] Zhihong Pan, Glenn Healey, Manish Prasad, and Bruce Tromberg. Face recognition in hyperspectral images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1552–1560, 2003. 1

[36] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pages 234–241, 2015. 5

[37] Michal Shimoni, Rob Haelterman, and Christiaan Perneel. Hypersectral imaging for military and security applications: Combining myriad processing and sensing techniques. *IEEE Geoscience and Remote Sensing Magazine*, 7(2):101–117, 2019. 1

[38] Ashwin Wagadarikar, Renu John, Rebecca Willett, and David Brady. Single disperser design for coded aperture snapshot spectral imaging. *Applied Optics*, 47(10):B44–B51, 2008. 1

[39] Lizhi Wang, Zhiwei Xiong, Dahua Gao, Guangming Shi, and Feng Wu. Dual-camera design for coded aperture snapshot spectral imaging. *Applied Optics*, 54(4):848–858, 2015. 1, 2, 6, 7

[40] Lizhi Wang, Zhiwei Xiong, Dahua Gao, Guangming Shi, Wenjun Zeng, and Feng Wu. High-speed hyperspectral video acquisition with a dual-camera architecture. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4942–4950, 2015. 1

[41] Lizhi Wang, Zhiwei Xiong, Guangming Shi, Wenjun Zeng, and Feng Wu. Simultaneous depth and spectral imaging with a cross-modal stereo system. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(3):812–817, 2016. 2

[42] Lizhi Wang, Zhiwei Xiong, Guangming Shi, Feng Wu, and Wenjun Zeng. Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(10):2104–2111, 2017. 2, 3, 7

[43] Lizhi Wang, Zhiwei Xiong, Hua Huang, Guangming Shi, Feng Wu, and Wenjun Zeng. High-speed hyperspectral video acquisition by combining nyquist and compressive sampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(4):857–870, 2018. 1

[44] Lizhi Wang, Tao Zhang, Ying Fu, and Hua Huang. Hyperreconnet: Joint coded aperture optimization and image reconstruction for compressive hyperspectral imaging. *IEEE Transactions on Image Processing*, 28(5):2257–2270, 2018. 2

[45] Lizhi Wang, Chen Sun, Ying Fu, Min H Kim, and Hua Huang. Hyperspectral image reconstruction using a deep spatial-spectral prior. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 8032–8041, 2019. 2, 3

[46] Lizhi Wang, Chen Sun, Maoqing Zhang, Ying Fu, and Hua Huang. Dnu: deep non-local unrolling for computational spectral imaging. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1661–1671, 2020. 2

[47] Lizhi Wang, Shipeng Zhang, and Hua Huang. Adaptive dimension-discriminative low-rank tensor recovery for computational hyperspectral imaging. *International Journal of Computer Vision*, 129(10):2907–2926, 2021. 2

[48] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 7

[49] Qingsong Xu, Xin Yuan, Chaojun Ouyang, and Yue Zeng. Attention-based pyramid network for segmentation and classification of high-resolution and hyperspectral remote sensing images. *Remote Sensing*, 12(21):3501, 2020. 1

[50] Fumihito Yasuma, Tomoo Mitsunaga, Daisuke Iso, and Shree K Nayar. Generalized assorted pixel camera: Post-capture control of resolution, dynamic range, and spectrum. *IEEE Transactions on Image Processing*, 19(9):2241–2253, 2010. 7

[51] Xin Yuan. Generalized alternating projection based total variation minimization for compressive sensing. In *IEEE International Conference on Image Processing*, pages 2539–2543, 2016. 2

[52] Xin Yuan, Tsung-Han Tsai, Ruoyu Zhu, Patrick Llull, David Brady, and Lawrence Carin. Compressive hyperspectral imaging with side information. *IEEE Journal of Selected Topics in Signal Processing*, 9(6):964–976, 2015. 2

[53] Xin Yuan, David J Brady, and Aggelos K Katsaggelos. Snapshot compressive imaging: Theory, algorithms, and applications. *IEEE Signal Processing Magazine*, 38(2):65–88, 2021. 1

[54] Maoqing Zhang, Lizhi Wang, Lei Zhang, and Hua Huang. Compressive hyperspectral imaging with non-zero mean noise. *Optics Express*, 27(13):17449–17462, 2019. 2

[55] Shipeng Zhang, Hua Huang, and Ying Fu. Fast parallel implementation of dual-camera compressive hyperspectral imaging system. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(11):3404–3414, 2018. 2

[56] Shipeng Zhang, Lizhi Wang, Ying Fu, Xiaoming Zhong, and Hua Huang. Computational hyperspectral imaging based on dimension-discriminative low-rank tensor recovery. In *IEEE International Conference on Computer Vision*, pages 10183–10192, 2019. 2

[57] Weihang Zhang, Jinli Suo, Kaiming Dong, Lianglong Li, Xin Yuan, Chengquan Pei, and Qionghai Dai. Handheld snapshot multi-spectral camera at tens-of-megapixel resolution. *Nature Communications*, 14(1):5043, 2023. 1