DIRECTIONAL POSITIVE-DEFINITE METRIC CONSENSUS FOR TIME-MODULATED MULTI-FRAME HYPERSPECTRAL RECONSTRUCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Time-modulated hyperspectral imaging records multiple spatially aligned measurements of the same scene under different spectral responses for reconstruction. However, correlations among these responses lead to nonuniform sensitivity across spectral directions, while existing reconstruction algorithms generally coordinate learned priors using shared scalar penalties. Here, we propose the Time-Modulated Calibration-Conditioned Spatial–Spectral Reconstruction Network (TCSS), an ADMM-inspired unfolding network built on Directional Positive-Definite Metric Consensus. Using calibration-derived directions in spectral space, TCSS constructs branch-specific positive-definite metrics that jointly control prior strength and spectral–spatial allocation across directions. Spectral-Shape-Guided Dual-Branch Reconstruction (SDBR) and Cross-Stage Reconstruction Increment Propagation (CSIP) further improve spatial–spectral prior extraction and information propagation across reconstruction stages. Extensive experiments show that TCSS achieves superior reconstruction performance with a compact architecture and generalizes consistently across datasets and the tested calibrated response sets. The code is released at `https://anonymous.4open.science/r/test24/`.

1 INTRODUCTION

Hyperspectral images (HSIs) capture both spatial information and spectral features, enabling accurate material discrimination (Hong et al., 2026). They can be applied to various fields, including precision agriculture (Lu et al., 2020), environmental monitoring (Rajabi et al., 2024), chemical analysis (Wright et al., 2019), and medical imaging (Fei, 2019).

Traditional hyperspectral data are acquired by scanning along the spatial or spectral dimension, resulting in slow acquisition (Yuan et al., 2021). In recent years, advances in compressive sensing (CS) and optoelectronic technology (Lin et al., 2025) have promoted the development of computational hyperspectral imaging systems (Cao et al., 2016; Bacca et al., 2023), including coded aperture snapshot spectral imaging (CASSI) (Gehm et al., 2007; Wagadarikar et al., 2008), spatially modulated hyperspectral imaging (Lin et al., 2014), and time-modulated hyperspectral imaging (Zhou et al., 2024; Zhang et al., 2025). CASSI uses a coded aperture and spectral dispersion to encode the scene (Wagadarikar et al., 2009; Kittle et al., 2010). Spatially modulated systems typically assign different broadband spectral responses to different detector pixels, whereas time-modulated systems use reconfigurable optical elements to switch a spatially shared broadband spectral response between exposures. Since these systems acquire encoded measurements rather than the full data cube, reconstruction algorithms are required to recover the HSI.

A variety of deep learning algorithms have been developed for hyperspectral reconstruction, mainly following two routes: end-to-end (E2E) networks (Cai et al., 2022a; Yao et al., 2024) and deep unfolding (DU) networks (Cai et al., 2022b; Li et al., 2023; Zhang et al., 2024). E2E methods directly learn a mapping from encoded measurements to HSIs, while DU methods unroll iterative reconstruction algorithms into multiple trainable stages that alternate between imaging-model-based updates and learned prior refinement. These methods have significantly improved hyperspectral reconstruction quality.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Despite this progress, the similarity of spectral responses across different exposures can affect reconstruction performance in time-modulated HSI, yet this issue is generally not explicitly considered by existing reconstruction methods. More specifically, two challenges arise, as shown in Fig. 1. **(1) Nonuniform Measurement Sensitivity.** Highly correlated spectral response curves make some spectral changes much less distinguishable in the measurements than others, corresponding to low-sensitivity directions in spectral space (Fig. 1(a,b)). Measurement noise can therefore cause much larger reconstruction errors along these directions, even when the number of measurements exceeds the number of spectral bands. **(2) Direction-Uniform Prior Coordination.** When multiple priors are coordinated using scalar penalties, both the overall prior strength and spectral–spatial allocation remain shared across directions (Fig. 1(c)). This limits direction-specific prior coordination under nonuniform measurement sensitivity.

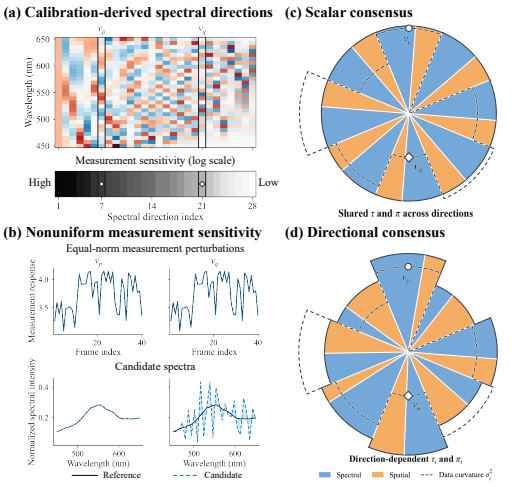


Figure 1: **Motivation for Directional Positive-Definite Metric Consensus.** (a) Calibration-derived directions in spectral space with nonuniform measurement sensitivity. (b) Under equal measurement perturbations, lower-sensitivity directions correspond to larger spectral deviations. (c) Scalar consensus uses a shared total penalty τ and spectral–spatial allocation ratio π across directions. (d) Directional consensus allows τ_i and π_i to vary across directions.

To address these challenges, we propose the Time-Modulated Calibration-Conditioned Spatial–Spectral Reconstruction Network (**TCSS**). TCSS adopts a consensus-splitting structure inspired by the Alternating Direction Method of Multipliers (ADMM) (Boyd et al., 2011), combining measurements and spatial–spectral priors through a shared image variable. Its core mechanism, **Directional Positive-Definite Metric Consensus**, replaces scalar penalties with direction-dependent positive-definite metrics, allowing both the overall prior strength and prior allocation to vary across spectral directions, as shown in Fig. 1(d). Specifically, TCSS generates direction-wise metric weights from calibration features. The resulting metrics are parameterized in the right-singular-vector basis of the sensing matrix, enabling a closed-form image update evaluated direction-wise.

Directional consensus depends on the quality of the branch outputs it coordinates. To improve these outputs, we further introduce two supporting mechanisms. First, Spectral-Shape-Guided Dual-Branch Reconstruction (**SDBR**) captures dependencies across the full spectrum in the spectral branch and uses spectral-shape similarity to guide neighborhood aggregation in the spatial branch. Second, Cross-Stage Reconstruction Increment Propagation (**CSIP**) propagates the reconstruction increment from the preceding shared-image update to both branches at the next stage.

In summary, our contributions are as follows:

- We propose TCSS for time-modulated hyperspectral reconstruction, centered on Directional Positive-Definite Metric Consensus. Its branch-specific positive-definite metrics jointly control prior strength and spectral–spatial allocation across calibration-derived spectral directions and admit a closed-form image update.
- We introduce SDBR and CSIP to effectively extract spatial–spectral prior information and enhance information propagation across reconstruction stages.
- We systematically evaluate TCSS through comprehensive experiments. TCSS achieves superior reconstruction quality over the evaluated methods with a compact architecture and maintains consistent performance across datasets and the tested calibrated response sets.

2 Related Work

Hyperspectral Imaging Systems. Traditional hyperspectral imaging systems acquire hyperspectral measurements through row- or column-wise scanning (Yuan et al., 2021). With the development of compressive sensing (CS) and optoelectronic technologies, coded hyperspectral imaging has advanced

rapidly (Xu et al., 2024; Cao et al., 2011). As shown in Figure 2, coded HSI systems compressively encode spatial–spectral information into measurements for computational HSI reconstruction, greatly improving system efficiency and the level of integration. Among them, CASSI combines a coded aperture with dispersive optical elements to form mixed coded measurements through spatial coding and wavelength-dependent shifts (Gehm et al., 2007; Wagadarikar et al., 2008; 2009; Kittle et al., 2010). Broadband filter-array systems use filters with different spectral transmittances to encode spectral information spatially (Lin et al., 2014), as shown in Figure 2(b). Recent hardware developments further extend spectral encoding through programmable nanophotonic pixels, metasurface imagers, and integrated hyperspectral microsystems (Guo et al., 2024; He et al., 2024; Ding et al., 2025; Bian et al., 2026). In recent years, researchers have developed various sensors with reconfigurable spectral responses (Mu et al., 2026; Yu et al., 2025), laying the foundation for the time-modulated hyperspectral imaging system shown in Figure 2(c). By sequentially applying K calibrated spectral response states to the same scene, this scheme acquires spatially aligned multi-frame measurements for joint hyperspectral reconstruction (Li et al., 2026). However, broadband spectral responses reported in recent devices can be highly correlated (Mu et al., 2026; Wu et al., 2026), causing some spectral changes to produce only small differences in the measurements, corresponding to low-sensitivity spectral directions. Compared with the spatial–spectral mixing problem in CASSI, this nonuniform sensitivity poses a different challenge for reconstruction algorithms.

Hyperspectral Reconstruction Algorithms. Early model-based methods address computational spectral reconstruction through rank minimization, approximate message passing, and low-rank tensor recovery (Liu et al., 2019; Tan et al., 2016; Zhang et al., 2019). Learning-based methods can be broadly divided into end-to-end (E2E) networks and deep unfolding (DU) networks. E2E methods directly learn a mapping from encoded measurements to hyperspectral images and improve reconstruction through spatial–spectral modeling, attention mechanisms, and multiscale feature extraction. Representative methods include TSA-Net (Meng et al., 2020), MST++ (Cai et al., 2022a), SPECAT (Yao et al., 2024), and S^2-Transformer (Wang et al., 2025). Other learning-based methods explore dual-domain representation, self-supervised reconstruction, and multiscale transformer modeling (Hu et al., 2022; Meng et al., 2021; Luo et al., 2024). DU methods unfold traditional iterative reconstruction algorithms into multiple trainable stages, alternating between imaging-model-based updates and learned prior refinement at each stage (Ma et al., 2019; Wang et al., 2020; Meng et al., 2023). For example, DAUHST (Cai et al., 2022b) and PADUT (Li et al., 2023) incorporate imaging degradation information into the unfolding process, allowing the network to adapt reconstruction to different imaging conditions, while DPU-SCI (Zhang et al., 2024) further introduces multiple learned priors to improve iterative reconstruction. EDUNet refines gradient and proximal steps (Qin et al., 2024), while DLADUN updates degradation information across unfolding stages (Liu et al., 2026). Other unfolding models explore Gaussian-scale-mixture priors (Huang et al., 2021), mixed spatial–spectral priors (Dong et al., 2023), Mamba-inspired joint unfolding (Qin et al., 2025), and low-rank unfolding (Huang et al., 2026). These methods model spatial–spectral structures and imaging degradation in different ways, but generally do not use operator-derived directional sensitivity to adapt prior strength and spectral–spatial allocation across spectral directions. When the spectral responses used for encoding are highly correlated, weakly observed spectral variations can significantly affect reconstruction performance.

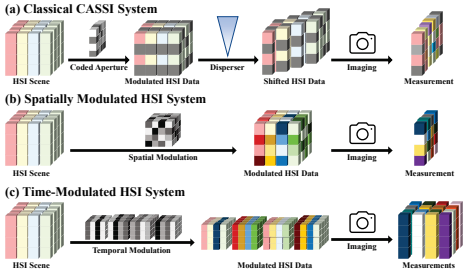


Figure 2: Representative hyperspectral encoding schemes: (a) classical CASSI, (b) spatial modulation, and (c) time modulation.

ADMM-based unrolling has also been used for spectral image fusion from compressive measurements, as in LADMM-Net (Ramirez et al., 2021). Beyond hyperspectral reconstruction, related inverse-problem formulations have explored nonuniform prior control in iterative reconstruction. Consensus equilibrium provides a general framework for combining multiple reconstruction agents (Buzzard et al., 2018), while preconditioned PnP-ADMM adapts denoising strength to spatially varying error statistics through a diagonal preconditioner (Le Pendu & Guillemot, 2023). Within this broader line

of work, TCSS derives branch-specific positive-definite metrics from the calibrated sensing operator to adapt both prior strength and spectral–spatial allocation across spectral directions.

3 METHOD

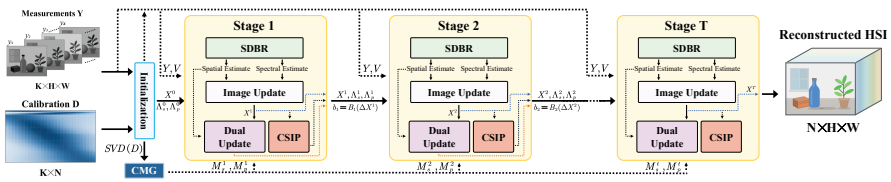


Figure 3: TCSS architecture. A one-time SVD of the response matrix D provides calibration features to the calibration-conditioned metric generator (CMG), which produces stage-specific directional metrics. Each reconstruction stage contains SDBR, image and dual updates, and CSIP.

3.1 TIME-MODULATED IMAGING MODEL AND MEASUREMENT SENSITIVITY

We consider time-modulated hyperspectral imaging, whose acquisition process provides multiple direct spectral measurements at each spatial location. Appendix A.1 compares single-frame spatial modulation with aligned multi-frame acquisition, while Appendix A.3 examines operator observability for these configurations and CASSI.

Let $X \in \mathbb{R}^{N\times P}$ denote an $H \times W$ hyperspectral image with N bands and $P = HW$. For K calibrated response states, $Y \in \mathbb{R}^{K\times P}$ and $D \in \mathbb{R}^{K\times N}$ denote the measurements and response matrix, respectively. The k-th row of D contains the spectral response of state k. The measurements follow

$$Y = DX + E, \tag{1}$$

where $E \in \mathbb{R}^{K\times P}$ denotes measurement noise.

To characterize the measurement response to spectral variations, we decompose the Gram matrix as

$$D^\top D = V\,\mathrm{diag}(\sigma_1^2, \ldots, \sigma_N^2)V^\top, \tag{2}$$

where $V = [v_1, \ldots, v_N]$ is a complete orthonormal basis of right singular vectors of D, with corresponding singular values $\sigma_1 \geq \cdots \geq \sigma_N \geq 0$. Each v_i represents a spectral variation direction across the N bands. An arbitrary variation can be expanded as $\delta x = \sum_{i=1}^{N} c_i v_i$, where c_i is its coefficient along v_i. The resulting measurement change satisfies

$$\|D\delta x\|_2^2 = \delta x^\top D^\top D\,\delta x = \sum_{i=1}^{N} \sigma_i^2 c_i^2. \tag{3}$$

Setting $\delta x = v_i$ in Eq. (3) gives $\|Dv_i\|_2^2 = \sigma_i^2$. Here, σ_i^2 characterizes the measurement sensitivity along v_i.

3.2 DIRECTIONAL POSITIVE-DEFINITE METRIC CONSENSUS

From Scalar to Directional Metrics. At stage t, the spectral and spatial branches produce estimates $Z_s^{t+1}, Z_p^{t+1} \in \mathbb{R}^{N\times P}$. For branch $a \in \{s, p\}$, let $M_a^t \in \mathbb{R}^{N\times N}$ denote the positive-definite metric matrix that weights the difference between the shared image X and Z_a^{t+1}. The quadratic penalty is $\frac{1}{2}\|X - Z_a^{t+1}\|_{M_a^t}^2$, where $\|A\|_M^2 = \mathrm{tr}(A^\top MA)$ for $A \in \mathbb{R}^{N\times P}$.

A scalar metric $M_a^t = \rho_a^t I$, with $\rho_a^t > 0$ and identity matrix $I \in \mathbb{R}^{N\times N}$, applies the same penalty strength to all spectral directions. We instead define

$$M_a^t = V\,\mathrm{diag}(m_{a,1}^t, \ldots, m_{a,N}^t)V^\top, \qquad m_{a,i}^t > 0, \quad a \in \{s, p\}, \tag{4}$$

where $m_{a,i}^t$ controls the penalty of branch a along v_i. We parameterize these metrics in the right-singular-vector basis of D.

Calibration-Driven Metric Weights. The calibration-conditioned metric generator (CMG) uses a feature $\phi_i(D) \in \mathbb{R}^2$ containing the log-scale ratio σ_i^2/σ_1^2 and the overall scale of D; Appendix B.1 gives its exact definition. We denote the MLP at stage t by MLP_t and share it across directions. CMG predicts the metric eigenvalues from singular-value features, while the operator-specific spectral

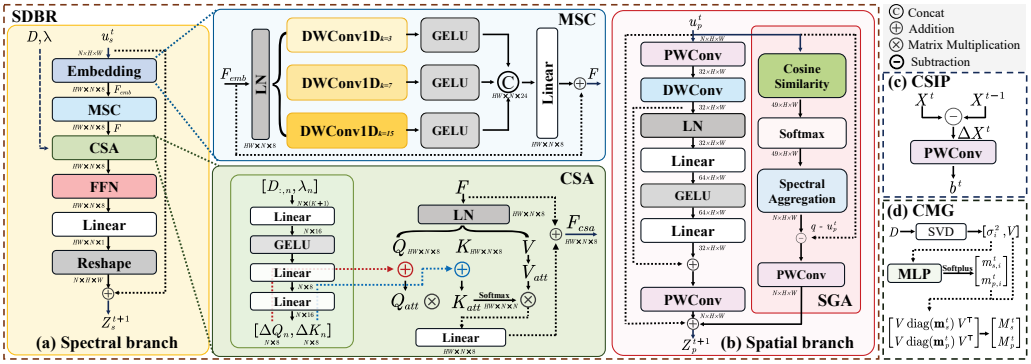


Figure 4: Detailed components of TCSS. (a) Spectral branch with multi-scale spectral convolution and calibration-conditioned spectral attention. (b) Spatial branch with spectral-shape-guided aggregation. (c) CSIP. (d) Calibration-conditioned metric generator.

orientation enters through the basis V in Eq. 4. To keep the two metric matrices positive definite, we apply Softplus and a positive lower bound to its outputs:

$$\begin{bmatrix} m_{s,i}^t \\ m_{p,i}^t \end{bmatrix} = \mathrm{softplus}(\mathrm{MLP}_t(\phi_i(D))) + \epsilon \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \epsilon = 10^{-4}. \tag{5}$$

Image and Dual Updates. Let $\Lambda_a^t \in \mathbb{R}^{N\times P}$ denote the unscaled Lagrange multiplier for the constraint $X = Z_a$. Given the two branch estimates, we obtain the following image update using ADMM (Boyd et al., 2011):

$$\begin{aligned} X^{t+1} = (D^\top D + M_s^t + M_p^t)^{-1} \\ \cdot (D^\top Y + M_s^t Z_s^{t+1} + M_p^t Z_p^{t+1} - \Lambda_s^t - \Lambda_p^t). \end{aligned} \tag{6}$$

The coefficient matrix is positive definite, so the solution is unique. Appendix B.2 gives the subproblem and derivation. The multipliers are then updated as

$$\Lambda_a^{t+1} = \Lambda_a^t + M_a^t(X^{t+1} - Z_a^{t+1}), \qquad a \in \{s, p\}. \tag{7}$$

Total Penalty and Spectral–Spatial Allocation. We express the two weights using the total consensus penalty τ_i^t and the spectral–spatial allocation π_i^t, the fraction assigned to the spectral branch:

$$\tau_i^t = m_{s,i}^t + m_{p,i}^t, \qquad \pi_i^t = \frac{m_{s,i}^t}{\tau_i^t}. \tag{8}$$

The branch weights are $m_{s,i}^t = \tau_i^t \pi_i^t$ and $m_{p,i}^t = \tau_i^t(1-\pi_i^t)$. Because $D^\top D$, M_s^t, and M_p^t share the basis V, the image update can be computed separately along each direction. Let $\widetilde{x}_i$, $\widetilde{z}_{a,i}$, and $\widetilde{\lambda}_{a,i}$ denote the i-th rows of $V^\top X$, $V^\top Z_a$, and $V^\top \Lambda_a$, respectively, and let q_i be the i-th row of $V^\top D^\top Y$. Equation (6) becomes

$$\widetilde{x}_i^{t+1} = \frac{q_i + \tau_i^t[\pi_i^t \widetilde{z}_{s,i}^{t+1} + (1-\pi_i^t)\widetilde{z}_{p,i}^{t+1}] - \widetilde{\lambda}_{s,i}^t - \widetilde{\lambda}_{p,i}^t}{\sigma_i^2 + \tau_i^t}. \tag{9}$$

Both the total consensus penalty and spectral–spatial allocation can vary across directions and stages (Fig. 1(d)).

3.3 Spectral-Shape-Guided Dual-Branch Reconstruction and Cross-Stage Reconstruction Increment Propagation

At stage t, branch a receives

$$u_a^t = X^t + (M_a^t)^{-1}\Lambda_a^t + b^t, \qquad a \in \{s, p\}, \tag{10}$$

where $(M_a^t)^{-1}\Lambda_a^t$ is the dual correction and b^t is the shared input generated by CSIP from the preceding reconstruction increment. Figure 4 shows the two reconstruction branches and CSIP.

Spectral Branch. Multi-Scale Spectral Convolution (MSC) applies parallel one-dimensional depth-wise convolutions with kernel sizes of 3, 7, and 15 to the embedded band features, followed by concatenation and linear projection. Calibration-Conditioned Spectral Attention (CSA) then computes attention across all N bands at each spatial location. Let $F \in \mathbb{R}^{N\times d}$ be its input features, where d is the feature dimension. A condition encoder maps each band's response column $D_{:,n}$ and wavelength encoding to query and key offsets, stacked as ΔQ and ΔK. The queries, keys, and values are

$$Q_{\mathrm{att}} = FW_Q + \Delta Q, \qquad K_{\mathrm{att}} = FW_K + \Delta K, \qquad V_{\mathrm{att}} = FW_V, \tag{11}$$

where W_Q, W_K, W_V are learnable linear projections. Standard scaled dot-product attention (Vaswani et al., 2017) aggregates the spectral features using these calibration-adjusted queries and keys. A feed-forward block and output projection produce a spectral residual, which is added to u_s^t to obtain Z_s^{t+1}. Appendix B.3 gives the attention and residual equations.

Spatial Branch. A local residual path extracts spatial features in parallel with Spectral-Shape-Guided Aggregation (SGA). For pixel r, let $u_{p,r}^t \in \mathbb{R}^N$ denote its input spectrum, $\bar{u}_{p,r}^t = u_{p,r}^t / \max(\|u_{p,r}^t\|_2, 10^{-4})$ its normalized spectrum, and $\mathcal{N}(r)$ its 7×7 neighborhood. SGA computes the similarity s_{rj}^t and weight w_{rj}^t for each neighbor j as

$$s_{rj}^t = \langle \bar{u}_{p,r}^t, \bar{u}_{p,j}^t \rangle, \qquad w_{rj}^t = \frac{\exp(\beta^t s_{rj}^t)}{\sum_{\ell\in\mathcal{N}(r)} \exp(\beta^t s_{r\ell}^t)}, \quad j \in \mathcal{N}(r), \tag{12}$$

where $\beta^t = \mathrm{softplus}(\theta_\beta^t) + 10^{-3}$ is a learned stage-specific similarity scale. These weights are applied to the original, unnormalized spectra:

$$q_r^t = \sum_{j\in\mathcal{N}(r)} w_{rj}^t u_{p,j}^t. \tag{13}$$

Let q^t collect the aggregated spectra over all pixels. A zero-initialized 1×1 convolution projects $q^t - u_p^t$ to a correction, which is added to the local residual path's estimate to produce Z_p^{t+1}. Appendix B.3 provides the detailed spatial branch structure.

Cross-Stage Reconstruction Increment Propagation. CSIP passes the preceding reconstruction increment ΔX^t to both branches through

$$\Delta X^t = X^t - X^{t-1}, \qquad b^t = B^t(\Delta X^t), \qquad t \geq 1, \tag{14}$$

where B^t is a stage-specific, zero-initialized 1×1 convolution, and $b^0 = 0$.

3.4 Reconstruction and Training

TCSS uses T reconstruction stages (Fig. 3). We decompose D once and reuse its right-singular-vector basis and singular values across stages. The initialization is

$$X^0 = \frac{D^\top Y}{\|D\|_F^2}, \qquad \Lambda_s^0 = \Lambda_p^0 = 0. \tag{15}$$

The final output X^T is reshaped into an $H\times W\times N$ hyperspectral image. Algorithm 1 in Appendix B.4 summarizes the reconstruction procedure.

Training Loss. Only the final output is supervised, using

$$\mathcal{L} = \mathcal{L}_{\mathrm{MSE}} + \eta\mathcal{L}_{\mathrm{shape}}, \tag{16}$$

where $\mathcal{L}_{\mathrm{MSE}}$ is the mean squared error against the ground truth, $\mathcal{L}_{\mathrm{shape}}$ is the mean cosine distance over pixels whose ground-truth spectral norm is at least 10^{-3}, and η weights the shape term. Appendix B.5 gives the complete loss definition.

4 Experiments

4.1 Experimental Setup

Datasets and Measurement Settings. Following the CAVE-to-KAIST evaluation protocol used in prior HSI reconstruction studies (Li et al., 2023; Yao et al., 2024), we use all 32 CAVE scenes (Yasuma et al., 2010) for training and 10 KAIST scenes (Choi et al., 2017) for testing. KAIST evaluation uses the central 256×256 crop of each test scene. We further use Harvard (Chakrabarti & Zickler, 2011) and the ICVL dataset (Arad & Ben-Shahar, 2016) for cross-dataset evaluation without fine-tuning. All experiments use $N = 28$ spectral bands spanning 450–650 nm. Unless otherwise specified, we use $K = 40$ calibrated spectral responses and simulate measurements using Eq. 1 with Gaussian noise of standard deviation 0.01. Results under different numbers of observations are provided in the Appendix.

Competing Methods. We compare TCSS with nonnegative Tikhonov, LMMSE/Wiener, TSA-Net (Meng et al., 2020), MST++ (Cai et al., 2022a), SPECAT (Yao et al., 2024), S^2-Transformer (Wang et al., 2025), SRNet (Bian et al., 2024), PG-SVRT (Cai et al., 2026), DAUHST (Cai et al., 2022b), PADUT (Li et al., 2023), and DPU-SCI (Zhang et al., 2024). All learning-based baselines are adapted to the time-modulated forward model and retrained using the same data split and measurement simulation protocol. For the adapted unfolding baselines, the quadratic data updates use the full calibrated response matrix D; response summaries are used only as conditioning features. Appendix A.4 details each adaptation.

Implementation and Evaluation. All learning-based models are implemented in PyTorch and trained on one NVIDIA A100 GPU for 300 epochs using AdamW (Loshchilov & Hutter, 2019) with a constant learning rate of 10^{-4}. Reconstruction quality is evaluated using mean peak signal-to-noise ratio (MPSNR), structural similarity index (SSIM) (Wang et al., 2004), and spectral angle mapper (SAM) (Kruse et al., 1993). For each scene, MPSNR averages PSNR over all spectral bands. Unless otherwise specified, all ablations use the same data split and measurement settings as the main experiments. Complete training and implementation settings are provided in the Appendix.

4.2 Reconstruction Performance and Efficiency

Table 1: Quantitative comparison averaged over the ten KAIST test scenes ($K = 40$, $N = 28$). Best and second-best results are bold and underlined; "–" marks unavailable entries.

Method	MPSNR (dB)↑	SAM (°)↓	SSIM↑	Params (M)	GFLOPs
Non-negative Tikhonov	26.588	38.494	0.610328	–	–
LMMSE/Wiener	44.343	10.906	0.983773	–	–
TSA-Net	38.044	7.417	0.977796	41.79	188.44
MST++	41.571	7.056	0.987761	1.37	37.58
SPECAT	39.887	14.140	0.965607	0.48	26.42
S^2-Transformer	40.432	14.765	0.959717	4.15	465.62
SRNet	39.494	7.524	0.983701	1.91	10.22
PG-SVRT	38.707	6.359	0.981029	2.33	57.55
DAUHST-5stg	45.050	7.156	0.990014	3.44	69.14
PADUT-5stg	42.955	13.070	0.971050	2.24	59.38
DPU-SCI-5stg	46.093	7.074	0.989399	1.66	59.19
TCSS-3stg	<u>47.340</u>	3.650	0.994983	0.03	23.54
TCSS-5stg	47.305	<u>3.186</u>	<u>0.995219</u>	0.06	39.21
TCSS-9stg	**47.553**	**2.619**	**0.995657**	0.10	70.54

Quantitative Results and Efficiency. Table 1 compares reconstruction accuracy and computational cost. TCSS achieves the best MPSNR, SAM, and SSIM among the evaluated methods. TCSS-9stg reaches 47.553 dB MPSNR and 2.619° SAM, while TCSS-3stg already reaches 47.340 dB and 3.650° SAM with only 0.03 M parameters and 23.54 GFLOPs. The three stage configurations provide different accuracy–computation trade-offs while remaining compact. Across three training seeds, TCSS-5stg consistently outperforms DPU-SCI-5stg under the unified training protocol (Appendix A.6).



Figure 5: Qualitative reconstruction comparison on selected 256×256 KAIST crops. The reported MPSNR and SAM are computed on the displayed regions.

Qualitative Results. Figure 5 shows that TCSS preserves finer spatial structures and reduces spectral reconstruction errors compared with the evaluated methods. The reconstructed spectra also follow the reference spectral variations more closely.

Overall Component Ablation. Table 2 progressively constructs TCSS by introducing the Generated-V metric, SGA, and CSIP into the dual-branch baseline. The Generated-V metric, which instantiates the proposed Directional Positive-Definite Metric Consensus, produces the largest reduction in SAM, while SGA and CSIP provide complementary improvements. These results establish the metric-consensus mechanism as the central reconstruction component of TCSS, with SGA and CSIP further refining spatial–spectral prior modeling and cross-stage information propagation.

Table 2: Progressive component ablation of TCSS ($K = 40$, $N = 28$, $T = 5$). Each row cumulatively adds the indicated component to the preceding configuration.

Variant	SAM (°)↓	MPSNR↑	SSIM↑
Scalar	5.938	46.317	0.99280
+ Generated-V Metric	3.460	46.714	0.99500
+ SGA	3.313	47.280	0.99511
+ CSIP (TCSS)	**3.186**	**47.305**	**0.99522**

4.3 Metric Parameterization and Directional Consensus Analysis

To clarify the role of the proposed metric design, we disentangle three factors: metric parameterization, direction-dependent weighting, and calibration-feature correspondence. We then examine how the directional formulation controls total prior strength and spectral–spatial allocation.

Metric Parameterization and Directional Conditioning. We compare free and generated metric parameterizations under scalar and directional formulations, together with a shuffled-feature variant that disrupts the correspondence between calibration features and spectral directions. Table 3(a) shows that the results separate primarily by metric parameterization: the generated variants achieve lower mean SAM than the free-parameter variants, while scalar, directional, and shuffled-feature generation remain comparable on the reference response. Across unseen response sets, the generated variants retain stronger transfer than the free scalar model, while their relative ordering varies across training seeds. These results identify metric generation as the main performance distinction in this ablation family.

Directional Prior Strength and Allocation. We next examine the two degrees of freedom within the generated directional formulation. Directional-total varies the total penalty τ_i across directions, Directional-allocation varies the spectral–spatial allocation π_i, and Full varies both. Jointly adapting penalty strength and branch allocation gives the strongest performance among the directional configurations, indicating complementary roles of the two controls.

Table 3: Controlled analysis of metric design. (a) Metric parameterization and directional conditioning. (b) Directional control of total penalty strength and spectral–spatial allocation.

(a) Metric parameterization and directional conditioning (mean ± sample SD over three training seeds)

Variant	Metric design	SAM (°)↓	MPSNR↑	SSIM↑
Scalar	Free scalar	5.962 ± 0.524	46.317 ± 0.188	0.992795 ± 0.000249
Free-Native	Free diag., wavelength basis	6.214 ± 0.498	46.254 ± 0.165	0.992317 ± 0.000156
Free-V	Free diag., V basis	6.116 ± 0.546	46.395 ± 0.147	0.992736 ± 0.000224
Generated-Scalar	Generated scalar	3.691 ± 0.371	46.528 ± 0.127	0.994734 ± 0.000172
Generated-V	Generated diag., V basis	3.676 ± 0.383	46.517 ± 0.196	0.994826 ± 0.000266
Shuffled-V	Generated diag., shuffled features	3.570 ± 0.371	46.518 ± 0.187	0.994864 ± 0.000187

(b) Directional prior strength and allocation

Variant	Dir. total τ	Dir. alloc. π	SAM (°)↓	MPSNR↑	SSIM↑
Dir.-total	✓	–	4.095	46.421	0.99423
Dir.-alloc.	–	✓	3.898	46.254	0.99438
Full	✓	✓	**3.460**	**46.714**	**0.99500**

Directional Consensus Behavior. Figure 6 illustrates how the full directional formulation coordinates the two reconstruction branches. The learned metrics produce different branch mixtures and measurement–prior balances across spectral directions, showing how directional consensus is realized in the unfolded update.

4.4 Generalization and Robustness

Cross-Dataset Generalization. We evaluate CAVE-trained models without fine-tuning on Harvard and ICVL after wavelength-based resampling to the shared 28-band grid, using the main response and noise settings. TCSS achieves the lowest SAM and highest MPSNR on both datasets, with the best SSIM on ICVL and second-best on Harvard (Table 4).

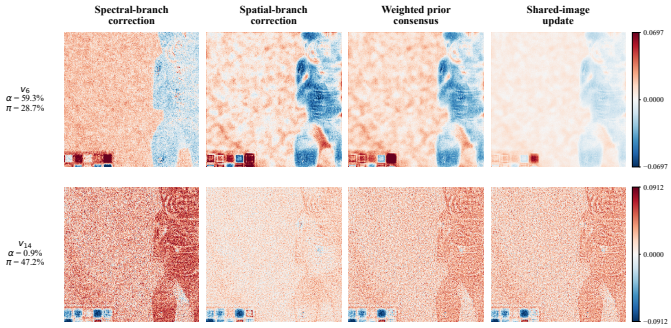


Figure 6: Directional consensus at Stage 3 for v_6 and v_{14}, where $\alpha_i^t = \sigma_i^2/(\sigma_i^2 + \tau_i^t)$ denotes the relative measurement weight in the image update.

Table 4: Cross-dataset reconstruction on Harvard and ICVL without fine-tuning.

(a) Harvard

Method	SAM (°)↓	MPSNR↑	SSIM↑
TSA-Net	5.164	33.115	0.885736
MST++	4.373	35.323	0.902960
SPECAT	5.788	32.996	0.852900
S^2-Transformer	4.580	35.199	0.895686
SRNet	5.814	32.200	0.892542
PADUT-5stg	3.721	37.507	0.931760
DPU-SCI-5stg	3.269	38.147	**0.937956**
TCSS-5stg	**2.625**	**39.301**	0.937115

(b) ICVL

Method	SAM (°)↓	MPSNR↑	SSIM↑
TSA-Net	2.210	37.470	0.980979
MST++	1.573	43.687	0.992802
SPECAT	1.941	39.576	0.968321
S^2-Transformer	1.641	40.902	0.976981
SRNet	3.652	33.127	0.979356
PADUT-5stg	1.340	42.852	0.981704
DPU-SCI-5stg	1.078	45.726	0.995262
TCSS-5stg	**0.841**	**47.399**	**0.995456**

Cross-Response Transfer and Noise Robustness. We test ten unseen $K = 40$ response sets from the same library of 64 measured response curves, using each supplied response for both measurement simulation and frozen reconstruction. TCSS-5stg averages 3.684° SAM, versus 3.186° on the reference response, and has the lowest unseen-response SAM (Fig. 7(a)). Across the tested Gaussian noise levels, TCSS also leads in SAM, MPSNR, and SSIM (Fig. 7(b)–(d)). Complete transfer metrics and additional robustness results are in the Appendix.

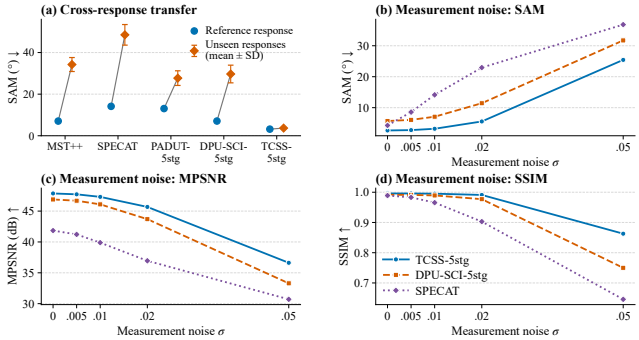


Figure 7: Cross-response transfer and robustness to measurement noise. (a) SAM on the reference response and ten unseen response sets. (b)–(d) SAM, MPSNR, and SSIM under Gaussian measurement noise.

4.5 Extension to Spatial Modulation

Directional metric consensus can also be applied to single-frame spatial modulation. To realize this extension, we construct directional consensus metrics from calibrated responses within each local neighborhood. The resulting extension achieves the lowest SAM and highest MPSNR among the compared methods; Appendix A.2 provides the full comparison.

5 Conclusion

In this work, we presented TCSS, a compact unfolding framework for time-modulated hyperspectral reconstruction. At its core, Directional Positive-Definite Metric Consensus uses branch-specific positive-definite metrics to jointly control prior strength and spectral–spatial allocation across calibration-derived spectral directions. Together with SDBR and CSIP, TCSS achieves superior reconstruction quality over the evaluated methods and maintains strong performance across datasets and the tested calibrated response sets. These results support Directional Positive-Definite Metric Consensus as an effective strategy for hyperspectral reconstruction under correlated spectral responses.

AI USE STATEMENT

Large language models assisted with literature organization, research ideation, mathematical reasoning, implementation support, experimental planning and analysis, and manuscript drafting and editing. The authors formulated the method, selected and checked the experimental protocol, verified sources and mathematical statements, and made the final decisions about the scientific content. Language models were not treated as authors or independent sources of evidence.

ETHICS STATEMENT

This work uses publicly available hyperspectral image datasets and simulated measurements generated from calibrated spectral response curves. It does not involve human participants, animal subjects, medical intervention, or personally identifiable information. We follow the access and usage conditions associated with the source datasets.

REPRODUCIBILITY STATEMENT

The main text and appendices specify the data split, wavelength grid, response construction, noise model, observation configurations, training objective, optimizer, update budget, evaluation metrics, and hardware used for the experiments.

REFERENCES

Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural RGB images. In *Computer Vision – ECCV 2016*, volume 9911, pp. 19–34, 2016. doi: 10.1007/978-3-319-46478-7_2.

Jorge Bacca, Emmanuel Martinez, and Henry Arguello. Computational spectral imaging: A contemporary overview. *Journal of the Optical Society of America A*, 40(4):C115–C125, 2023. doi: 10.1364/JOSAA.482406.

Liheng Bian, Zhen Wang, Yuzhe Zhang, Lianjie Li, Yinuo Zhang, Chen Yang, Wen Fang, Jiajun Zhao, Chunli Zhu, Qinghao Meng, Xuan Peng, and Jun Zhang. A broadband hyperspectral image sensor with high spatio-temporal resolution. *Nature*, 635(8037):73–81, 2024. doi: 10.1038/s41586-024-08109-1.

Liheng Bian, Qinghao Meng, Lianjie Li, Zhen Wang, Yibo Feng, Xuan Peng, Jiajun Zhao, Jingyi Wang, Zhu Yang, and Jun Zhang. Hypervision: An on-chip hyperspectral microsystem for online video-rate computational imaging. *Science*, 393(6814):888–894, 2026. doi: 10.1126/science.aef8268.

Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011. doi: 10.1561/2200000016.

Gregery T. Buzzard, Stanley H. Chan, Suhas Sreehari, and Charles A. Bouman. Plug-and-play unplugged: Optimization-free reconstruction using consensus equilibrium. *SIAM Journal on Imaging Sciences*, 11(3):2001–2020, 2018. doi: 10.1137/17M1122451.

Lijing Cai, Zhan Shi, Chenglong Huang, Jinyao Wu, Qiping Li, Zikang Huo, Linsen Chen, Chongde Zi, and Xun Cao. Exploring spatiotemporal feature propagation for video-level compressive spectral reconstruction: Dataset, model and benchmark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12522–12532, June 2026.

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

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*, volume 35, pp. 37749–37761, 2022b. doi: 10.52202/068431-2736.

Xun Cao, Hao Du, Xin Tong, Qionghai Dai, and Stephen Lin. A prism-mask system for multispectral video acquisition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12): 2423–2435, 2011.

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. doi: 10.1109/MSP.2016.2582378.

Ayan Chakrabarti and Todd Zickler. Statistics of real-world hyperspectral images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 193–200, 2011.

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. doi: 10.1145/3130800.3130810.

Kaiyang Ding, Qian Zhou, Mengyuan Chen, Kuizhi Shao, Xiaohao Wang, Xiaojun Liang, Kai Ni, and Benfeng Bai. Deep learning empowered parallelized metasurface computed tomography snapshot spectral imaging. *Advanced Materials*, 37(27):2419383, 2025. doi: 10.1002/adma.202419383.

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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22262–22271, June 2023.

Baowei Fei. Hyperspectral imaging in medical applications. In *Data Handling in Science and Technology*, volume 32, pp. 523–565. Elsevier, 2019. doi: 10.1016/B978-0-444-63977-6.00021-3.

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. *Optics Express*, 15(21):14013–14027, 2007. doi: 10.1364/oe.15.014013.

Tingbiao Guo, Zhi Zhang, Zijian Lin, Jiahan Tian, Yi Jin, Julian Evans, Yinghe Xu, and Sailing He. Durable and programmable ultrafast nanophotonic matrix of spectral pixels. *Nature Nanotechnology*, 19(11):1635–1643, 2024. doi: 10.1038/s41565-024-01756-5.

Haoyang He, Yuzhe Zhang, Yujie Shao, Yan Zhang, Guangzhou Geng, Junjie Li, Xin Li, Yongtian Wang, Liheng Bian, Jun Zhang, and Lingling Huang. Meta-attention network based spectral reconstruction with snapshot near-infrared metasurface. *Advanced Materials*, 36(23):2313357, 2024. doi: 10.1002/adma.202313357.

Danfeng Hong, Chenyu Li, Naoto Yokoya, Bing Zhang, Xiuping Jia, Antonio Plaza, Paolo Gamba, Jon Atli Benediktsson, and Jocelyn Chanussot. Hyperspectral imaging. *Nature Reviews Methods Primers*, 6(1):19, 2026. doi: 10.1038/s43586-026-00470-x.

Xiaowan Hu, Yuanhao Cai, Jing Lin, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. HDNet: High-resolution dual-domain learning for spectral compressive imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17542–17551, June 2022.

He Huang, Yujun Guo, and Wei He. LRDUN: A low-rank deep unfolding network for efficient spectral compressive imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10556–10566, June 2026.

Tao Huang, Weisheng Dong, Xin Yuan, Jinjian Wu, and Guangming Shi. Deep gaussian scale mixture prior for spectral compressive imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16216–16225, June 2021.

David Kittle, Kerkil Choi, Ashwin Wagadarikar, and David J. Brady. Multiframe image estimation for coded aperture snapshot spectral imagers. *Applied Optics*, 49(36):6824–6833, 2010. doi: 10.1364/ao.49.006824.

Fred A. Kruse, A. B. Lefkoff, J. W. Boardman, K. B. Heidebrecht, A. T. Shapiro, P. J. Barloon, and Alexander F. H. Goetz. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment*, 44(2–3):145–163, 1993. doi: 10.1016/0034-4257(93)90013-N.

Mikael Le Pendu and Christine Guillemot. Preconditioned plug-and-play ADMM with locally adjustable denoiser for image restoration. *SIAM Journal on Imaging Sciences*, 16(1):393–422, 2023. doi: 10.1137/22M1504809.

Fangning Li, Heyan Meng, Jiajie Zhang, Xianye Li, Zhangjun Wang, Guoliang Shentu, Baoqing Sun, and Yupeng Wang. Hyperspectral image reconstruction algorithms for broadband spectral modulation: A comprehensive review. *Advanced Imaging*, 3(1):A00003, 2026. doi: 10.3788/AI.2025.20005.

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

Xing Lin, Yebin Liu, Jiamin Wu, and Qionghai Dai. Spatial-spectral encoded compressive hyperspectral imaging. *ACM Transactions on Graphics*, 33(6):1–11, 2014. doi: 10.1145/2661229.2661262.

Zijian Lin, Tingbiao Guo, Zhi Zhang, Yuan Zhang, Haochen Chu, Yijia Zeng, Xiao Chen, Nan Wang, Ruili Zhang, and Sailing He. A spatiotemporal tunable filter array chip for video-rate hyperspectral imaging. *Nano Letters*, 25(9):3455–3463, 2025. doi: 10.1021/acs.nanolett.4c05603.

Lei Liu, Ping Xu, Xin Yuan, Chen Xu, and Haifeng Zheng. Degradation learning adaptive deep unfolding network for spectral compressive imaging. *Pattern Recognition*, 175:113058, 2026. doi: 10.1016/j.patcog.2026.113058.

Yang Liu, Xin Yuan, Jinli Suo, David J. Brady, and Qionghai Dai. Rank minimization for snapshot compressive imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(12):2990–3006, 2019. doi: 10.1109/tpami.2018.2873587.

Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A ConvNet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11976–11986, 2022.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL `https://openreview.net/forum?id=Bkg6RiCqY7`.

Bing Lu, Phuong D. Dao, Jiangui Liu, Yuhong He, and Jiali Shang. Recent advances of hyperspectral imaging technology and applications in agriculture. *Remote Sensing*, 12(16):2659, 2020. doi: 10.3390/rs12162659.

Fulin Luo, Xi Chen, Xiuwen Gong, Weiwen Wu, and Tan Guo. Dual-window multiscale transformer for hyperspectral snapshot compressive imaging. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 3972–3980, 2024. doi: 10.1609/aaai.v38i4.28190.

Jiawei Ma, Xiao-Yang Liu, Zheng Shou, and Xin Yuan. Deep tensor ADMM-net for snapshot compressive imaging. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

Ziyi Meng, Jiawei Ma, and Xin Yuan. End-to-end low cost compressive spectral imaging with spatial-spectral self-attention. In *Computer Vision – ECCV 2020*, 2020.

Ziyi Meng, Zhenming Yu, Kun Xu, and Xin Yuan. Self-supervised neural networks for spectral snapshot compressive imaging. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2622–2631, October 2021.

Ziyi Meng, Xin Yuan, and Shirin Jalali. Deep unfolding for snapshot compressive imaging. *International Journal of Computer Vision*, 131(11):2933–2958, 2023. doi: 10.1007/s11263-023-01844-4.

Ge Mu, Cheng Bi, Jintao Zou, Yanfei Liu, Qun Hao, and Xin Tang. Hyperspectral quantum-dot image sensors via in-pixel reconfigurable band-alignment. *Nature Photonics*, 20(5):523–531, 2026. doi: 10.1038/s41566-026-01860-z.

Mengjie Qin, Yuchao Feng, Zongliang Wu, Yulun Zhang, and Xin Yuan. Detail matters: Mamba-inspired joint unfolding network for snapshot spectral compressive imaging. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 6594–6602, 2025. doi: 10.1609/aaai.v39i6.32707.

Xinran Qin, Yuhui Quan, and Hui Ji. Enhanced deep unrolling networks for snapshot compressive hyperspectral imaging. *Neural Networks*, 174:106250, 2024. doi: 10.1016/j.neunet.2024.106250.

Roozbeh Rajabi, Amin Zehtabian, Keshav D. Singh, Alireza Tabatabaeenejad, Pedram Ghamisi, and Saeid Homayouni. Editorial: Hyperspectral imaging in environmental monitoring and analysis. *Frontiers in Environmental Science*, 11:1353447, 2024. doi: 10.3389/fenvs.2023.1353447.

Juan Marcos Ramirez, Jose Ignacio Martinez-Torre, and Henry Arguello. LADMM-Net: An unrolled deep network for spectral image fusion from compressive data. *Signal Processing*, 189:108239, 2021. doi: 10.1016/j.sigpro.2021.108239.

Jin Tan, Yanting Ma, Hoover Rueda, Dror Baron, and Gonzalo R. Arce. Compressive hyperspectral imaging via approximate message passing. *IEEE Journal of Selected Topics in Signal Processing*, 10(2):389–401, 2016. doi: 10.1109/JSTSP.2015.2500190.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

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. doi: 10.1364/ao.47.000b44.

Ashwin A. Wagadarikar, Nikos P. Pitsianis, Xiaobai Sun, and David J. Brady. Video rate spectral imaging using a coded aperture snapshot spectral imager. *Optics Express*, 17(8):6368–6388, 2009. doi: 10.1364/oe.17.006368.

Jiamian Wang, Kunpeng Li, Yulun Zhang, Xin Yuan, and Zhiqiang Tao. s^2-transformer for mask-aware hyperspectral image reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(6):4299–4316, 2025. doi: 10.1109/TPAMI.2025.3543842.

Lizhi Wang, Chen Sun, Maoqing Zhang, Ying Fu, and Hua Huang. DNU: Deep non-local unrolling for computational spectral imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

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. doi: 10.1109/tip.2003.819861.

Stephanie L. Wright, Joseph M. Levermore, and Frank J. Kelly. Raman spectral imaging for the detection of inhalable microplastics in ambient particulate matter samples. *Environmental Science & Technology*, 53(15):8947–8956, 2019. doi: 10.1021/acs.est.8b06663.

Jianghong Wu, Bangjie Shao, Yuting Ye, Qingyan Deng, Jialing Jian, Xixi Jiang, Yafei Wu, Lan Li, and Yang Chai. Broadband and high-resolution snapshot spectroscopy with high-index transition metal dichalcogenides. *Nature Communications*, 17:1955, 2026. doi: 10.1038/s41467-026-68685-w.

Yibo Xu, Liyang Lu, Vishwanath Saragadam, and Kevin F. Kelly. A compressive hyperspectral video imaging system using a single-pixel detector. *Nature Communications*, 15(1):1456, 2024. doi: 10.1038/s41467-024-45856-1.

Zhiyang Yao, Shuyang Liu, Xiaoyun Yuan, and Lu Fang. SPECAT: Spatial-spectral cumulative-attention transformer for high-resolution hyperspectral image reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25368–25377, June 2024.

Fumihito Yasuma, Tomoo Mitsunaga, Daisuke Iso, and Shree K. Nayar. Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum. *IEEE Transactions on Image Processing*, 19(9):2241–2253, 2010. doi: 10.1109/TIP.2010.2046811.

Huabin Yu, Muhammad Hunain Memon, Mingjia Yao, Zhixiang Gao, Yuanmin Luo, Yang Kang, Qianqian Zhan, Wei Chen, Yinpeng Chen, Sheng Liu, Zongyin Yang, Tawfique Hasan, and Haiding Sun. A miniaturized cascaded-diode-array spectral imager. *Nature Photonics*, 19:1322–1329, 2025. doi: 10.1038/s41566-025-01754-6.

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. doi: 10.1109/msp.2020.3023869.

Chong Zhang, Wenjing Liu, Juntao Li, Siqi Li, Lizhi Wang, Hua Huang, Yuanjin Zheng, Yongtian Wang, Jinli Suo, and Weitao Song. Tunable optimally-coded snapshot hyperspectral imaging for scene adaptation. *Laser & Photonics Reviews*, 19(11):2401921, 2025. doi: 10.1002/lpor.202401921.

Jiancheng Zhang, Haijin Zeng, Jiezhang Cao, Yongyong Chen, Dengxiu Yu, and Yin-Ping Zhao. Dual prior unfolding for snapshot compressive imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 25742–25752, 2024. doi: 10.1109/CVPR52733.2024.02432.

Shipeng Zhang, Lizhi Wang, Ying Fu, Xiaoming Zhong, and Hua Huang. Computational hyperspectral imaging based on dimension-discriminative low-rank tensor recovery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

Zhou Zhou, Yiheng Zhang, Yingxin Xie, Tian Huang, Zile Li, Peng Chen, Yan-Qing Lu, Shaohua Yu, Shuang Zhang, and Guoxing Zheng. Electrically tunable planar liquid-crystal singlets for simultaneous spectrometry and imaging. *Light: Science & Applications*, 13(1):242, 2024. doi: 10.1038/s41377-024-01608-w.

APPENDIX / SUPPLEMENTARY MATERIAL

Summary. In this supplementary material, we provide additional analyses, experimental results, and implementation details for TCSS. Section A.1 compares spatial sampling with aligned multi-frame acquisition, while Sec. A.2 extends the proposed directional metric consensus to spatial modulation. Section A.3 analyzes the observability of single-frame, multi-frame, and CASSI imaging operators. Section A.4 describes the baseline adaptations used for fair comparison, and Sec. A.5 reports scene-wise reconstruction results. Section A.6 provides supplementary ablation studies, and Sec. A.7 analyzes the learned directional metrics and numerical stability. Section A.8 presents cross-response transfer results and cross-dataset visual comparisons. Section A.9 studies different observation configurations and transfer under low-observation settings, while Sec. A.10 reports supplementary robustness evaluations. Section A.11 gives the training and implementation details. Part B provides the implementation and mathematical derivations of the proposed method.

A.1 Spatial sampling versus aligned multi-frame acquisition
A.2 Extension to spatial modulation
A.3 Observability of single-frame, multi-frame, and CASSI operators
A.4 Baseline adaptations
A.5 Scene-wise reconstruction results
A.6 Supplementary Ablation Studies
A.7 Directional metrics and numerical stability
A.8 Cross-response transfer and cross-dataset visual comparisons

A.9 Observation configurations and low-observation transfer
A.10 Supplementary Robustness Evaluations
A.11 Training and Implementation Details
B Implementation and derivation details

A.1 Spatial Sampling versus Aligned Multi-Frame Acquisition

Figure A.1 compares the two acquisition configurations illustrated in Figure 2(b,c) of the main paper. The single-frame configuration applies the measured filter responses according to their spatial arrangement, yielding one spectrally coded value at each pixel. The multi-frame configuration selects 40 calibrated response states and applies each response across the full image, yielding 40 spatially aligned values per pixel.

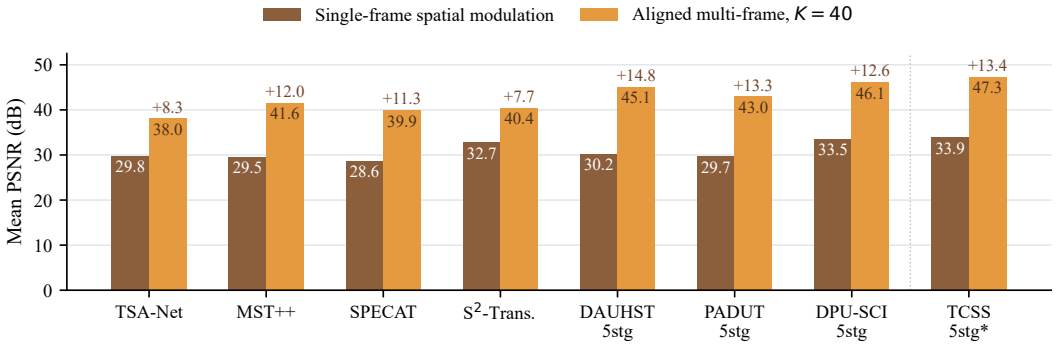


Figure A.1: MPSNR under single-frame spatial modulation and aligned 40-frame acquisition. Paired bars show independently trained variants of each method. The asterisk marks TCSS, whose single-frame variant uses window-conditioned directional metrics.

All eight methods attain higher MPSNR with the aligned multi-frame configuration. This empirical comparison illustrates the reconstruction value of obtaining multiple complete, spatially aligned response images for the same scene, motivating the time-modulated setting studied in the main text. For TCSS-5stg, the single-frame configuration yields 33.890 dB and 7.896° SAM, whereas the multi-frame configuration yields 47.305 dB and 3.186° SAM.

A.2 Extension to Spatial Modulation

We instantiate directional metric consensus for single-frame spatial modulation, whose per-pixel forward operator has rank one. Overlapping neighborhoods of calibrated filter responses provide local spectral directions for positive-definite consensus metrics. The window metrics are combined at each pixel, while the image update retains the exact rank-one data step for its own measurement. On the ten held-out KAIST scenes, this TCSS adaptation achieves 7.896° SAM, 33.890 dB MPSNR, and 0.953 SSIM. Table A.1 reports the complete single-frame comparison corresponding to the left bars of Figure A.1. TCSS obtains the lowest SAM and highest MPSNR, while DPU-SCI obtains the highest SSIM. These results show that directional metric consensus can also be instantiated for spatially varying single-frame encoding.

A.3 Observability of Single-Frame, Multi-Frame, and CASSI Operators

This section compares the observability of single-frame spatial modulation, aligned multi-frame acquisition, and CASSI. The three configurations exhibit different forms of ambiguity: insufficient per-pixel spectral measurements in single-frame spatial modulation, direction-dependent spectral sensitivity in aligned multi-frame acquisition, and spatial–spectral mixing in CASSI.

Table A.1: Single-frame spatial-modulation reconstruction on the ten held-out KAIST scenes. Each metric is averaged across scenes.

Method	SAM (°)↓	MPSNR (dB)↑	SSIM↑
TSA-Net	16.222	29.775	0.900
MST++	16.743	29.528	0.912
SPECAT	15.338	28.617	0.879
S^2-Transformer	11.803	32.730	0.942
DAUHST-5stg	11.981	30.225	0.932
PADUT-5stg	21.605	29.692	0.880
DPU-SCI-5stg	8.198	33.488	**0.957**
TCSS-5stg	**7.896**	**33.890**	0.953

Table A.2: Spectral rank and conditioning under the evaluated sensing configurations and assumptions.

Response set and assumption	Rank	Condition number
Single-frame spatial modulation, one pixel	1	∞
Single-frame, 7×7 shared spectrum	28	1.07×10^7
Single-frame, 15×15 shared spectrum	28	6.45×10^5
Aligned 40-frame acquisition, one pixel	28	1.13×10^7
CASSI, spatially constant spectrum	28	9.17

Single-frame spatial modulation. At each pixel, one measurement provides a single inner product $y_p = m_p^\top x_p$ for the 28-band spectrum. The local operator has rank one and leaves 27 spectral directions unobserved. Combining calibrated responses within a 7×7 or 15×15 neighborhood provides full spectral rank under the assumption that neighboring pixels share the same spectrum. Across 256 sampled windows, the median condition numbers are 1.07×10^7 and 6.45×10^5, respectively. Figure A.2(a) shows the median normalized singular spectrum for the 7×7 windows.

Aligned multi-frame acquisition. With $K = 40$ aligned measurements, the same $D \in \mathbb{R}^{40\times 28}$ acts independently at every spatial location, providing full spectral rank at each pixel. The calibrated responses remain highly correlated: the spectral condition number is 1.13×10^7 despite only a 1.11-fold variation in column-wise RMS gain. Thus the weak directions arise mainly from response similarity rather than isolated low-response bands. Figure A.2(a) shows the normalized singular spectrum of D alongside the local-window spectrum.

CASSI. The coded aperture and two-pixel wavelength-dependent shifts jointly mix spatial and spectral information. A 28-dimensional spectral restriction for spatially constant inputs has condition number 9.17, as listed in Table A.2. For a 256×256 crop, the full operator A maps $28 \cdot 256^2 = 1{,}835{,}008$ unknowns to $256(256 + 54) = 79{,}360$ detector measurements. Its exact null space occupies 95.68% of the ambient spatial–spectral domain. We further evaluate this ambiguity on ten hyperspectral test scenes by measuring the energy outside the noiseless minimum-norm projection:

$$\eta_j = \frac{\|x_j - A^\dagger A x_j\|_2^2}{\|x_j\|_2^2}.$$

The fractions range from 48.79% to 72.25%, with a mean of 59.70% (Figure A.2(b)).

Single-frame spatial modulation is locally rank deficient and gains spectral constraints through spatial sharing. Aligned multi-frame acquisition provides full per-pixel spectral rank but retains weak directions from correlated responses. CASSI instead has a high-dimensional spatial–spectral null space under compressive mixing.

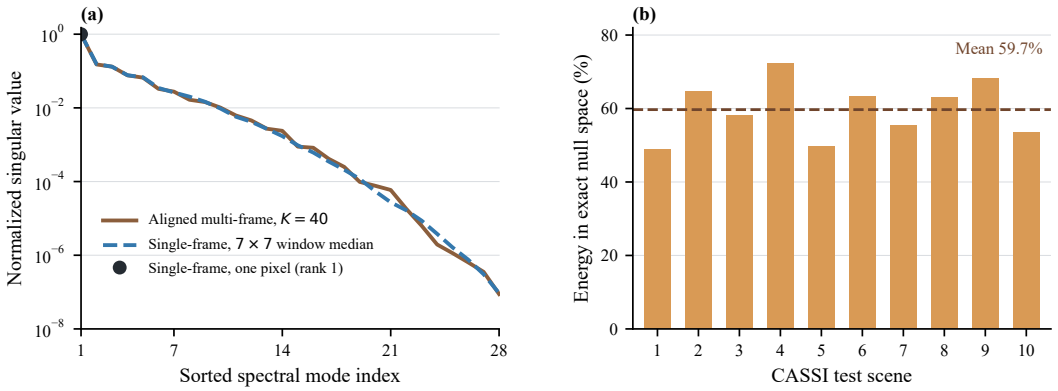


Figure A.2: Observability diagnostics for the evaluated sensing configurations. (a) Normalized singular spectra for a single-pixel spatial measurement, the median over 256 single-frame 7×7 shared-spectrum windows, and aligned $K = 40$ multi-frame acquisition. Spectra are sorted independently. The single-pixel operator has 27 exact zero singular values, omitted from the logarithmic axis. (b) Fraction of scene energy in the exact null space of the full CASSI operator for ten test scenes; the dashed line denotes the mean.

A.4 Baseline Adaptations

All learning-based baselines are adapted to the time-modulated forward model $Y = DX + E$. The adaptations retain each method's reconstruction structure and replace its sensing interface with the shared measurements and calibration.

Nonnegative Tikhonov. For each pixel, we solve

$$\hat{x} = \arg\min_{x \geq 0} \{ \|y - Dx\|_2^2 + 0.01 \|Lx\|_2^2 \},$$

where L is the second-order divided-difference operator on the fixed wavelength grid. The regularization weight 0.01 is fixed before evaluation, and the constrained problem is solved by projected gradient iterations.

LMMSE/Wiener. The spectral mean μ and covariance C are estimated from the CAVE training set and fixed during evaluation. Given a response matrix D, reconstruction is computed as

$$\hat{x} = \mu + CD^\top \left(DCD^\top + \sigma_n^2 I \right)^{-1} (y - D\mu),$$

where $\sigma_n = 0.01$ matches the measurement-noise setting used in the main experiments.

TSA-Net. The attention blocks are retained. A learned $K \to 28$ projection adapts the measurement stack, replacing the native CASSI shifts and masks.

MST++. The native MST body and residual path are retained. A learned measurement projection and an ordered encoding of D condition the initial feature path.

SPECAT. The spectral and spatial attention, FFN, encoder/decoder, and output residual are retained. A learned measurement projection and ordered D encoding modulate the attention values, with a per-sample spatial guide.

S^2-Transformer. The spatial–spectral attention body receives a learned $K \to 28$ projection and a fixed scalar mean(D) proxy.

SRNet. The measurement and mask embeddings, spectral-attention U-Net, skip connections, and output mapping are retained. The ordered K-frame stack replaces the single-channel input, while an ordered D encoding supplies the calibration condition and the output is mapped to 28 bands.

PG-SVRT. The CDPA windows, bridged-token propagation, MDFFN, and U-Net body are retained. The sequence axis is aligned with the ordered rows of D, and the K fused proposals are averaged into one 28-band estimate.

DAUHST. The HST denoiser and unfolding structure are retained. Each stage performs the quadratic data update

$$z + D^\top (DD^\top + \rho I)^{-1}(y - Dz),$$

using the full calibrated response matrix D, with a positive scalar ρ estimated by the model. Response-dependent calibration features are used for conditioning and do not replace the full operator in the data step.

PADUT. The native MPR, SAM, fusion blocks, and pixel-adaptive gates are retained. Each stage first evaluates the same full-D quadratic data update and then blends the correction with the current estimate through the learned gate. The normalized column-wise response norm is used as a conditioning feature rather than as a substitute for the data operator.

DPU-SCI. The prior, dual, and feedback modules are retained. Each stage evaluates the full-D quadratic data update before the learned feedback fusion and dual update. Response energy and signed column-sum features condition the learned modules but do not replace D in the data-consistency step.

Table A.3: Baseline adaptation details under the unified time-modulated forward model.

Method	Input construction	Explicit D	Physics / data consistency	Updated when D changes
LMMSE/Wiener	Y, D, μ, C	Yes	Closed-form linear MMSE estimate	Wiener gain through D
TSA-Net	Learned Y projection	No	None	None; adapted measurement only
MST++	Projected Y plus D encoding	Yes	None	Ordered calibration encoding
SPECAT	Projected Y with D-gated attention	Yes	None	Attention calibration buffers
S^2-Transformer	Projected Y with mean(D) proxy	No	None	Mean(D) mask scalar
SRNet	Ordered Y with calibration embedding	Yes	None	Calibration embedding buffer
PG-SVRT	Ordered Y/D proposals	Yes	None	Response rows and observation scaling
DAUHST	Y, D; column-sum conditioning	Yes	Exact quadratic prox for learned scalar ρ	Response, sensitivity, adjoint scale, ρ floor
PADUT	Y, D; column-norm conditioning	Yes	Full-D prox followed by learned gate interpolation	Response, adjoint scale, sensitivity proxy
DPU-SCI	Y, D; energy/sum conditioning	Yes	Full-D prox followed by feedback and dual updates	Response, adjoint scale, energy/sum features, ρ floor
TCSS	Y, D, λ with SPD solve	Yes	Closed-form image update	Response, wavelengths, V, σ^2, features, M_s^t, and M_p^t

Response-Row Permutation Check. We apply five permutations to the rows of both the measurement and response matrix, $(Y, D) \mapsto (PY, PD)$, which leaves the sensing problem unchanged. Table A.4 reports the mean change in SAM relative to the unpermuted input over the ten KAIST scenes. TCSS, DPU-SCI, and PADUT change little, whereas MST++, SPECAT, and S^2-Transformer are sensitive to the input row order under their current adapters.

Table A.4: SAM change under five equivalent response-row permutations, averaged over permutations and ten KAIST scenes.

Method	Mean ΔSAM ($^\circ$)
MST++	$+40.595$
SPECAT	$+46.619$
S^2-Transformer	$+35.836$
PADUT-5stg	$+0.023$
DPU-SCI-5stg	≈ 0
TCSS-5stg	$+0.008$

A.5 Scene-Wise Reconstruction Results

The entries below expand the main comparison into its ten individual scenes and their equal-weight average.

Table A.5: Scene-wise KAIST reconstruction results on central 256×256 test crops under the shared measurement and evaluation settings. Each cell lists MPSNR, SAM, and SSIM from top to bottom; S1–S10 denote KAIST scenes 06, 07, 12, 14, 16, 17, 23, 24, 25, and 29.

Method	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Avg
	27.729	25.136	27.755	24.373	30.127	24.646	27.136	26.089	26.612	26.274	26.588
	43.472	29.736	43.799	29.871	51.446	29.186	38.616	39.193	40.273	39.350	38.494
Non-negative Tikhonov	0.551537	0.574925	0.725157	0.484489	0.738745	0.714328	0.535122	0.653861	0.626641	0.498472	0.610328
	46.201	44.873	44.865	40.707	45.107	44.134	44.956	43.757	42.563	46.267	44.343
	8.982	6.771	15.959	4.992	20.540	5.530	11.014	12.327	13.176	9.774	10.906
LMMSE/Wiener	0.983140	0.986718	0.983747	0.981072	0.981029	0.993713	0.980716	0.984402	0.980132	0.983061	0.983773
	41.292	36.706	38.740	34.932	39.522	35.634	40.325	37.825	36.433	39.033	38.044
	11.130	5.937	8.968	5.860	9.578	5.895	6.410	6.275	7.772	6.346	7.417
TSA-Net	0.975732	0.980368	0.984837	0.965727	0.983562	0.975130	0.978400	0.982357	0.969185	0.982666	0.977796
	45.597	41.405	41.055	37.724	43.185	40.082	43.373	40.692	38.926	43.671	41.571
	5.640	4.751	9.443	4.207	12.232	4.994	7.133	7.526	8.615	6.022	7.056
MST++	0.992692	0.990607	0.990067	0.978914	0.988764	0.987114	0.986781	0.989286	0.981365	0.992016	0.987761
	42.404	39.773	40.405	36.525	40.844	38.431	41.428	39.744	38.417	40.902	39.887
	12.170	9.425	20.023	6.947	25.286	8.136	14.020	15.834	16.627	12.929	14.140
SPECAT	0.967621	0.972019	0.965851	0.957514	0.960968	0.978113	0.961325	0.967413	0.958194	0.967056	0.965607
	42.722	40.056	40.451	37.059	41.649	39.608	41.083	40.122	39.563	42.012	40.432
	12.729	9.540	20.653	7.036	26.610	8.054	15.063	16.777	17.575	13.609	14.765
S^2-Transformer	0.961381	0.969560	0.951322	0.963135	0.943664	0.980239	0.953907	0.960729	0.950584	0.962654	0.959717
	42.595	37.581	40.810	35.558	41.115	37.740	41.536	39.862	38.015	40.130	39.494
	7.932	5.514	9.264	4.781	11.128	6.479	7.161	7.633	9.180	6.164	7.524
SRNet	0.988840	0.986298	0.986190	0.974406	0.987181	0.975030	0.985093	0.987012	0.976203	0.990754	0.983701
	42.222	38.936	39.649	33.825	40.005	35.472	41.114	39.074	36.853	39.916	38.707
	4.861	4.497	7.556	4.840	9.844	7.398	6.722	6.399	6.818	4.656	6.359
PG-SVRT	0.992429	0.985981	0.987138	0.959688	0.986114	0.967424	0.983860	0.985706	0.971277	0.990669	0.981029
	48.107	44.574	45.145	40.752	46.540	44.393	45.464	44.207	42.804	48.518	45.050
	5.820	4.841	10.153	3.818	13.117	4.080	7.407	7.773	8.588	5.965	7.156
DAUHST-5stg	0.992164	0.991864	0.990162	0.985986	0.987847	0.995291	0.987873	0.990292	0.986084	0.992577	0.990014
	44.624	42.757	43.715	39.434	44.423	42.357	43.209	42.659	41.299	45.069	42.955
	11.420	8.655	18.701	6.026	21.591	6.728	14.666	13.712	17.653	11.546	13.070
PADUT-5stg	0.971046	0.977341	0.969494	0.975853	0.967057	0.988655	0.964350	0.975515	0.948011	0.973181	0.971050
	48.570	46.352	46.349	41.161	47.650	44.745	46.697	45.642	44.002	49.763	46.093
	5.017	4.983	10.322	4.065	11.361	4.634	7.049	8.781	8.925	5.605	7.074
DPU-SCI-5stg	0.993375	0.991486	0.988464	0.986122	0.989812	0.994168	0.988241	0.985294	0.984023	0.993007	0.989399
	50.089	47.620	47.738	41.913	50.011	45.857	47.745	46.628	45.324	50.469	47.340
	3.548	2.378	4.512	2.876	5.748	2.543	3.844	3.863	4.308	2.875	3.650
TCSS-3stg	0.996742	0.995944	0.997422	0.987933	0.997269	0.997634	0.993106	0.995070	0.991601	0.997104	0.994983
	50.099	47.512	47.419	42.135	50.522	46.369	47.540	46.618	45.108	49.734	47.305
	3.616	2.287	3.747	2.765	4.494	2.278	3.173	3.194	3.672	2.631	3.186
TCSS-5stg	0.996598	0.996093	0.997712	0.988056	0.997796	0.997864	0.993563	0.995332	0.991844	0.997330	0.995219
	50.351	48.133	47.697	42.098	50.666	46.389	47.791	46.615	45.615	50.177	47.553
	2.705	1.821	2.856	2.605	3.495	2.099	2.877	2.761	2.862	2.112	2.619
TCSS-9stg	0.997674	0.996509	0.998234	0.988240	0.998371	0.998056	0.993854	0.995516	0.992371	0.997749	0.995657

A.6 SUPPLEMENTARY ABLATION STUDIES

Training Objective Ablation. We examine the effect of spectral-shape supervision using the same five-stage TCSS architecture and training protocol. Adding the spectral-shape term improves SAM while preserving MPSNR and SSIM, showing that the auxiliary objective primarily improves spectral fidelity.

Table A.6: Loss ablation with the final shared-CSIP architecture ($K = 40$, $N = 28$, $T = 5$).

Training objective	SAM (°)↓	MPSNR (dB)↑	SSIM↑
MSE	3.337	47.246	0.99516
MSE + spectral-shape	**3.186**	**47.305**	**0.99522**

Dual-Branch Ablation. We evaluate the spectral and spatial reconstruction branches individually and jointly. Combining the two branches gives the strongest performance across all three metrics, showing that they provide complementary spatial–spectral prior information.

Table A.7: Spectral and spatial branch ablation ($K = 40$, $N = 28$, $T = 5$).

Variant	Spec.	Spat.	SAM (°)↓	MPSNR↑	SSIM↑
Spectral-only	✓	–	4.807	46.234	0.99289
Spatial-only	–	✓	6.855	45.692	0.99115
Dual-branch	✓	✓	**3.460**	**46.714**	**0.99500**

Metric Variant Definitions. Scalar and Free-V directly optimize positive metric weights, initialized with $m_s = m_p = 0.025$. Generated-Scalar uses the stage-wise metric generator to produce direction-shared branch weights, whereas Generated-V produces direction-dependent weights from the calibration features in Eq. (B1). Shuffled-V uses the same generated directional form but randomly permutes the relative singular-value feature across directions during training while retaining the global-scale feature. Generated-Scalar feeds only the global calibration-scale feature to the generator, while Shuffled-V uses one fixed, checkpointed permutation shared across stages at evaluation.

Metric Transfer and Parameterization Diagnostics. Table A.8 reports frozen transfer of the basic dual-branch variants to ten unseen $K = 40$ response sets. The generated variants have lower mean SAM than the free Scalar baseline; their ordering differs from that on the reference response in Table 3(a) of the main paper.

Table A.8: Frozen transfer to ten unseen $K = 40$ response sets. Each metric first averages the response-set means within each training seed, then averages across three training seeds. The two SAM SD columns separately report training-seed and response-set variation.

Variant	Train seeds	Mean SAM (°)↓	Mean MPSNR (dB)↑	Mean SSIM↑	Train-seed SD (SAM)	Response-set SD (SAM)
Scalar	3	12.205	43.866	0.969426	3.472	2.532
Generated-Scalar	3	5.367	44.426	0.991877	1.602	0.339
Generated-V	3	5.807	43.972	0.990690	1.225	0.533
Shuffled-V	3	4.803	44.489	0.992855	0.792	0.537

We also instantiate Free-V using the metric weights produced by Generated-V while keeping the remaining network parameters identical. Across the ten KAIST scenes, the maximum relative reconstruction difference is 1.26×10^{-6}, confirming numerical equivalence between the two parameterizations when the metric weights are matched.

Training-Seed Stability. We repeat TCSS-5stg and DPU-SCI-5stg over three independent training seeds under the unified training protocol. TCSS maintains lower SAM together with higher MPSNR and SSIM across the repeated training runs.

Table A.9: Training-seed stability under the unified training protocol. Results are mean ± sample SD over three independent training seeds.

Method	SAM (°)↓	MPSNR (dB)↑	SSIM↑
DPU-SCI-5stg	7.067 ± 0.202	45.934 ± 0.199	0.989804 ± 0.000611
TCSS-5stg	$\mathbf{3.046 \pm 0.148}$	$\mathbf{47.416 \pm 0.112}$	$\mathbf{0.995318 \pm 0.000089}$

A.7 Directional Metrics and Numerical Stability

Learned Metric Behavior. Table A.10 summarizes metric behavior for the basic dual-branch ablation configurations used in Sec. 4.3. Figures A.3–A.4 and Table A.11 instead analyze the full TCSS-5stg model with SGA and CSIP. The free Scalar model remains close to its small initial metric scale across stages, whereas Generated-Scalar learns a pronounced stage-dependent schedule. Generated-V additionally exhibits clear variation across spectral directions, particularly in later stages. In contrast, Shuffled-V produces nearly direction-uniform weights within each stage, despite retaining the same directional metric form. These behaviors complement the performance ablations in Sec. 4.3 by showing how the different parameterizations translate into effective metric schedules.

Table A.10: Effective total metric penalty across reconstruction stages. Generated-Scalar is reported as mean ± sample SD over three training seeds. Generated-V and Shuffled-V report the median and 5–95% range across the 28 spectral directions for representative metric-ablation checkpoints. Scalar is direction-shared. All metric variants are initialized with $m_s = m_p = 0.025$, corresponding to $\tau = 0.05$.

Stage	Scalar τ^t	Generated-Scalar τ^t	Generated-V τ_i^t	Shuffled-V τ_i^t
1	0.0679	0.185 ± 0.061	0.206 [0.100, 0.239]	0.109 [0.108, 0.110]
2	0.0651	0.623 ± 0.627	0.133 [0.122, 0.161]	0.143 [0.143, 0.143]
3	0.0622	0.419 ± 0.132	0.724 [0.502, 1.081]	0.866 [0.863, 0.873]
4	0.0619	0.861 ± 0.510	0.666 [0.459, 0.793]	0.987 [0.975, 0.990]
5	0.0749	2.822 ± 1.694	8.577 [3.707, 16.432]	3.445 [3.407, 3.460]

Figure A.3 visualizes the directional metrics of the full TCSS-5stg model. Both the total penalty τ_i^t and spectral–spatial allocation π_i^t vary across directions and stages, resulting in direction-dependent measurement–prior weighting through α_i^t.

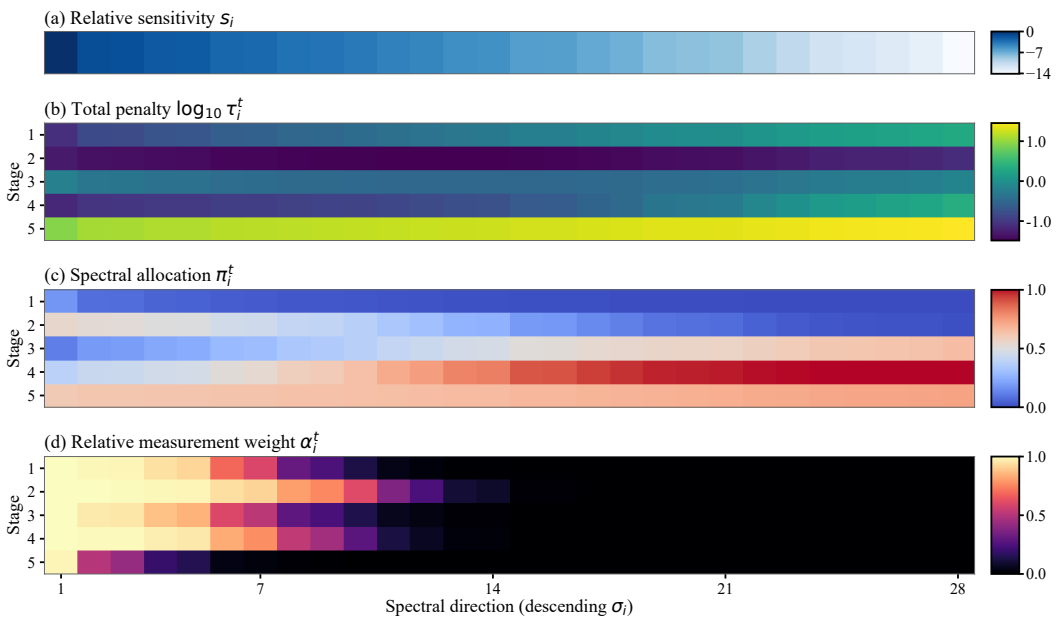


Figure A.3: Directional sensitivity and consensus weights for the reference response D_0. (a) Relative sensitivity $s_i = \log_{10}[(\sigma_i^2 + \epsilon)/(\sigma_1^2 + \epsilon)]$, with $\epsilon = 10^{-16}$; (b) $\log_{10} \tau_i^t$; (c) spectral allocation π_i^t; (d) relative measurement weight $\alpha_i^t = \sigma_i^2/(\sigma_i^2 + \tau_i^t)$ in the image update. Directions are ordered by descending σ_i; the π and α color scales are fixed to $[0, 1]$.

Direction-Wise Reconstruction Behavior. We next compare reconstruction errors in the weakest and strongest 20% of sensing directions. As shown in Figure A.4(a), errors are larger along weakly observed directions, with the largest separation occurring at Stage 2. The gap decreases in later stages as reconstruction proceeds. This provides a reconstruction-level counterpart to the nonuniform sensitivity of the sensing operator.



Figure A.4: Reconstruction and inference diagnostics across 11 calibrated responses. (a) Reconstruction errors in the weakest and strongest 20% sensing directions. (b) Normalized reconstruction increment, spectral and spatial consensus residuals, and dual correction across stages. Scene averages are computed within each response before response-level medians and 5–95% ranges are formed.

Inference Dynamics and Numerical Stability. Figure A.4(b) tracks the reconstruction increment, spectral and spatial consensus residuals, and dual correction across stages. The correction magnitudes vary nonmonotonically across stages and become smaller at the final stage. The image-update coefficient matrix is

$$H^t = D^\top D + M_s^t + M_p^t = V \, \mathrm{diag}\big(\sigma_i^2 + \tau_i^t\big) \, V^\top,$$

whose eigenvalues $\sigma_i^2 + \tau_i^t$ are positive because $\tau_i^t > 0$. Table A.11 confirms this over all 11 calibrated $K = 40$ response matrices: no non-finite values occur, while conditioning varies considerably across stages. Stage 2 has the largest condition numbers, whereas Stage 5 has a lower condition number.

Table A.11: Stage-wise conditioning of the image update over 11 calibrated $K = 40$ response matrices. The τ column reports the 5–95% range with the median in parentheses; the π column reports the fraction of directions with $\pi_i^t < 0.05$ or $\pi_i^t > 0.95$.

Stage	$\lambda_{\min}(H^t)$	$\lambda_{\max}(H^t)$	$\kappa_2(H^t)$	NaN/Inf	π extreme	τ range (median)
1	0.401–0.408	481–516	1.20×10^3–1.26×10^3	0	82.1%	0.148–1.72 (0.564)
2	0.0344–0.0347	481–516	1.40×10^4–1.50×10^4	0	22.1%	0.0327–0.0681 (0.0404)
3	0.309–0.315	482–517	1.56×10^3–1.64×10^3	0	0.0%	0.310–0.629 (0.376)
4	0.161–0.163	481–516	2.98×10^3–3.18×10^3	0	39.3%	0.0929–1.76 (0.207)
5	14.1–14.3	489–524	34.7–36.6	0	0.0%	11.4–26.8 (17.5)

A.8 Cross-Response Transfer and Cross-Dataset Visual Comparisons

The ten response sets are predefined subsets of the same 64 measured curves. For each D'_r, both the forward process and model input use that matrix: $Y_r = D'_r X + E$ and $D_{\text{input}} = D'_r$. The reconstruction weights remain frozen.

Each set-level metric first averages the ten KAIST scenes. The reported mean and sample standard deviation summarize the ten response-set means. The response-row permutation check in Appendix A.4 shows that the current MST++, SPECAT, and S^2-Transformer adapters are sensitive to equivalent input-row reorderings. Their cross-response changes in this section therefore combine response transfer with ordered-input sensitivity; the comparisons among TCSS, DPU-SCI, and PADUT are less affected by this interface behavior.

Table A.12: Frozen cross-response transfer relative to the reference response D_{ref}. Reference metrics follow Table 1 of the main paper; Δ denotes new-response minus reference.

Method	Metric	Reference D_{ref}	New responses, mean $\pm$ SD	Δ
MST++	SAM (°) ↓	7.056	34.241 $\pm$ 3.357	+27.185
	MPSNR (dB) ↑	41.571	19.489 $\pm$ 1.256	−22.082
	SSIM ↑	0.987761	0.741783 $\pm$ 0.057332	−0.245978
SPECAT	SAM (°) ↓	14.140	48.512 $\pm$ 4.950	+34.372
	MPSNR (dB) ↑	39.887	16.900 $\pm$ 1.353	−22.987
	SSIM ↑	0.965607	0.597821 $\pm$ 0.030622	−0.367786
PADUT-5stg	SAM (°) ↓	13.070	27.705 $\pm$ 3.523	+14.635
	MPSNR (dB) ↑	42.955	36.867 $\pm$ 1.307	−6.088
	SSIM ↑	0.971050	0.797078 $\pm$ 0.041668	−0.173972
DPU-SCI-5stg	SAM (°) ↓	7.074	29.682 $\pm$ 4.252	+22.607
	MPSNR (dB) ↑	46.093	37.849 $\pm$ 2.257	−8.244
	SSIM ↑	0.989399	0.792381 $\pm$ 0.068150	−0.197018
TCSS-5stg	**SAM (°) ↓**	**3.186**	**3.684 $\pm$ 0.259**	**+0.498**
	MPSNR (dB) ↑	**47.305**	**45.959 $\pm$ 0.739**	−1.346
	SSIM ↑	**0.995219**	**0.994454 $\pm$ 0.000460**	−0.000765

A.8.1 Cross-Dataset Visual Comparisons

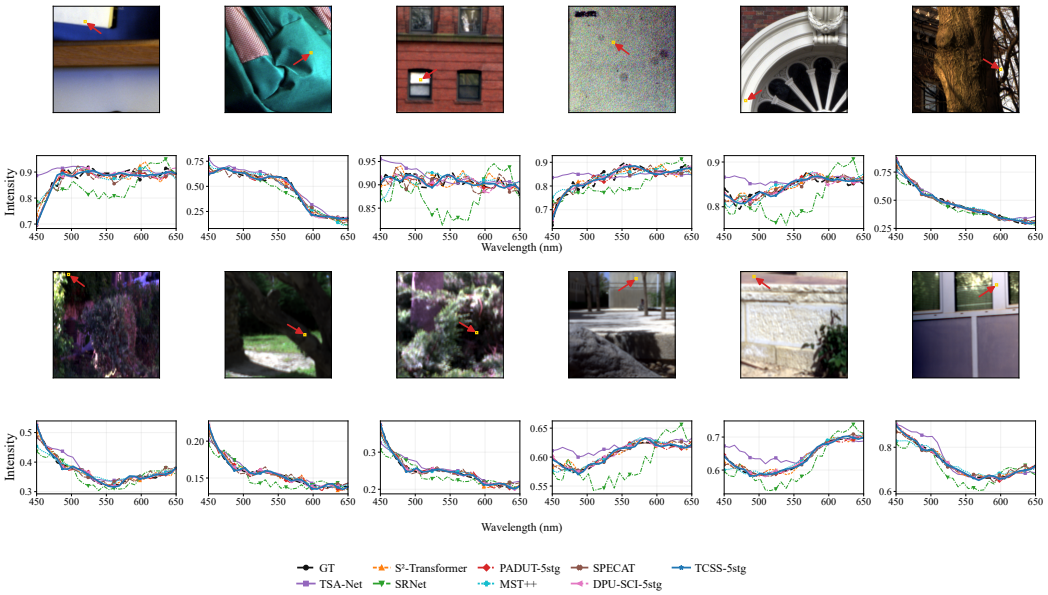


Figure A.5: Additional cross-dataset spectral comparisons on Harvard and ICVL. Spectra are averaged over the marked 5×5 regions.

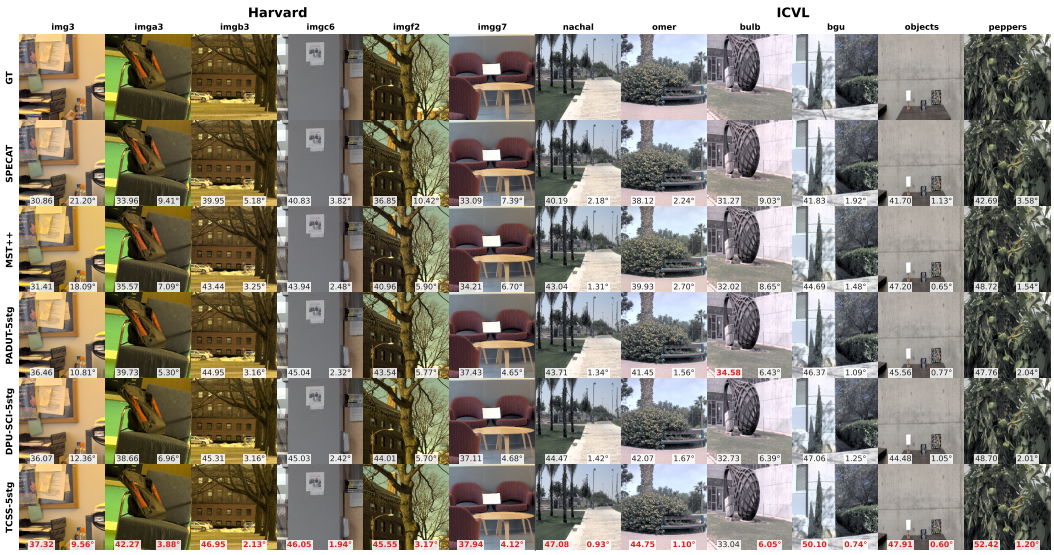


Figure A.6: Qualitative reconstruction of six Harvard and six ICVL scenes.

A.9 Observation Configurations and Low-Observation Transfer

Low-Observation Cross-Response Transfer. A TCSS-5stg model trained at $K = 16$ is evaluated on ten predefined $K = 16$ response sets. For each response matrix D'_r, the same matrix generates the measurements and conditions the reconstruction model.

Table A.13: TCSS transfer at $K = 16$: mean $\pm$ sample SD across ten response sets.

Method	SAM (°)↓	MPSNR↑	SSIM↑
TCSS-5stg	7.865 $\pm$ 1.039	42.598 $\pm$ 0.692	0.984858 $\pm$ 0.002997

The tested $K = 16$ setting exhibits a larger cross-response transfer gap than the $K = 40$ setting.

Effect of the Number of Observations. We evaluate six predefined observation configurations with $K \in \{8, 16, 24, 32, 40, 48\}$ for $N = 28$ using the same five reconstruction methods. As shown in Table A.14, TCSS achieves the best SAM, MPSNR, and SSIM across all tested configurations. The gains beyond $K = 40$ are modest for the reference setting used in the main experiments.

Table A.14: Reconstruction performance under different observation numbers K with $N = 28$.

Method	Metric	K=8	K=16	K=24	K=32	K=40	K=48
MST++	SAM (°)↓	10.366	7.079	7.128	7.618	7.056	7.260
	MPSNR↑	37.842	41.311	40.884	41.618	41.571	41.916
	SSIM↑	0.97210	0.98666	0.98662	0.98594	0.98776	0.98694
SPECAT	SAM (°)↓	15.547	14.311	14.868	14.063	14.140	14.494
	MPSNR↑	35.560	38.475	38.693	39.526	39.887	38.960
	SSIM↑	0.94526	0.95779	0.95789	0.96341	0.96561	0.96156
PADUT-5stg	SAM (°)↓	13.629	14.372	14.836	14.516	13.070	13.683
	MPSNR↑	39.757	41.979	41.968	42.572	42.955	42.837
	SSIM↑	0.96577	0.96043	0.95896	0.96118	0.97105	0.96664
DPU-SCI Full-5stg	SAM (°)↓	8.278	7.485	8.246	7.289	7.074	6.684
	MPSNR↑	41.678	44.946	44.908	45.679	46.093	46.348
	SSIM↑	0.98444	0.98771	0.98566	0.98867	0.98940	0.99063
TCSS-5stg	SAM (°)↓	**4.167**	**3.432**	**3.069**	**3.065**	**3.186**	**3.251**
	MPSNR↑	**43.720**	**46.510**	**46.492**	**46.741**	**47.305**	**47.833**
	SSIM↑	**0.99223**	**0.99450**	**0.99475**	**0.99503**	**0.99522**	**0.99544**

A.10 Supplementary Robustness Evaluations

A.10.1 Spatial Alignment Robustness

To evaluate sensitivity to imperfect spatial alignment, we perturb the $K = 40$ measurement frames with independent global subpixel translations. Frame 0 remains unshifted; each non-reference frame receives a displacement bounded by $r \in \{0.25, 0.5, 1.0\}$ pixels. TCSS achieves the lowest SAM at every tested misregistration magnitude, while its SAM rises from $3.189°$ at 0 px to $21.931°$ at 1.0 px. Metrics use the central 248×248 crop rather than the central 256×256 test crops used for Table 1; the fixed-noise 0-px case also undergoes zero-displacement resampling.

Table A.15: SAM (°) under shared frame-wise subpixel translations.

Method	0 px	0.25 px	0.5 px	1.0 px
SPECAT	14.172°	19.078°	23.358°	29.298°
PADUT-5stg	13.119°	16.373°	19.446°	24.175°
DPU-SCI-5stg	7.013°	18.028°	26.780°	36.935°
TCSS-5stg	3.189°	9.393°	14.471°	21.931°

A.10.2 Calibration Mismatch Robustness

We evaluate $K = 40$ checkpoints when the response used to generate the measurements differs from the calibration supplied to reconstruction. We compare TCSS-5stg, DPU-SCI-5stg, PADUT-5stg, and SPECAT. Correct calibration uses $D_{\text{input}} = D_{\text{true}}$.

For the smooth spectral-shape experiment, the true response is fixed at D_0 and the same measurements $Y = D_0X + E$ are used at every level. Correct calibration supplies $D_{\text{input}} = D_0$; mismatch changes only the supplied response on the 28 spectral samples:

$$D_{\text{input}} = D_0 \odot (1 + \epsilon_s G), \qquad \epsilon_s \in \{0, 1, 3, 5, 10\}\%,$$

where G is the archived smooth shape template. In the response-gain experiment, row-wise gains perturb the true response, and measurements are regenerated as $Y = D_{\text{true}}X + E$. Correct calibration supplies $D_{\text{input}} = D_{\text{true}}$, while mismatch supplies the nominal $D_{\text{input}} = D_0$. We use $\sigma_g \in \{1\%, 2\%, 5\%\}$. Metrics use the central 248×248 crop rather than the central 256×256 test crops used for Table 1, with fixed noise residuals.

Table A.16: Calibration mismatch robustness measured by SAM (°). (a) Smooth spectral-shape mismatch: the 0% column gives each method's correct-calibration reference once; other columns use the perturbed response as input. (b) Response-gain mismatch: each cell lists correct/mismatched calibration for the indicated gain level.

(a) Smooth spectral-shape mismatch

Method	0%	1%	3%	5%	10%
SPECAT	14.142	14.142	14.142	14.142	14.142
PADUT-5stg	13.064	12.916	13.778	16.206	23.132
DPU-SCI-5stg	7.087	7.686	10.274	12.619	19.371
TCSS-5stg	3.185	3.319	4.024	4.884	7.308

(b) Response-gain mismatch: correct/mismatched

Method	$\sigma_g = 1\%$	$\sigma_g = 2\%$	$\sigma_g = 5\%$
SPECAT	15.901 / 15.901	20.834 / 20.835	30.234 / 30.234
PADUT-5stg	13.200 / 14.112	13.049 / 16.944	13.749 / 24.689
DPU-SCI-5stg	6.990 / 21.296	7.219 / 30.377	7.345 / 45.558
TCSS-5stg	3.189 / 5.867	3.193 / 11.138	3.190 / 23.649

Across the tested mismatched-calibration settings, TCSS has the lowest SAM among the compared methods, although its absolute SAM increases as the mismatch grows. Row-wise response-gain mismatch produces a larger degradation than the smooth spectral-shape perturbations tested here. SPECAT changes little between correct and mismatched calibration within each gain condition, consistent with the current adapter's limited dependence on the explicit response input.

A.11 Training and Implementation Details

Table A.17: Training and implementation details for the controlled protocol.

Item	Setting
Training data	CAVE, 32 training scenes; fixed KAIST ten-scene test split
Spectral grid	28 bands, 450–650 nm; $K = 40$ calibrated responses
Measurement noise	Additive Gaussian, standard deviation 0.01
Patch / evaluation	64×64 training patches; central 256×256 test crops
Augmentation	Horizontal flip, vertical flip, and horizontal–vertical flip
Optimizer	AdamW; weight decay 0
Learning-rate schedule	Unified recipe: constant 10^{-4} throughout training
Training duration	300 epochs; 19,200 optimizer updates
Effective batch size	32 for all TCSS stage variants
Stage-specific batch setup	TCSS-3stg/5stg: physical batch 32; TCSS-9stg: physical batch 16 with gradient accumulation 2
Objective	Final-output MSE plus cosine spectral-shape loss, $\eta = 6.5 \times 10^{-5}$

All reported models use the final checkpoint after 19,200 optimizer updates; no separate validation split or metric-based checkpoint selection is used in this protocol.

SAM Evaluation. For each pixel, SAM is the arccosine in degrees of the predicted and reference spectral cosine, with a 10^{-8} denominator guard and cosine clipped to $[-1, 1]$. Pixel angles are averaged within each scene, then equally across the ten KAIST scenes. SAM is computed over all image pixels. To verify that black or near-black pixels do not dominate the angular error, we additionally evaluate SAM after excluding pixels whose ground-truth spectral norm is below 10^{-3}. No pixels in the ten KAIST test scenes are excluded under this threshold, so the reported SAM is unaffected by this masking criterion.

Learning-Rate Recipe Sensitivity. We further evaluate DAUHST-5stg and DPU-SCI-5stg using their recommended learning-rate recipes while keeping the data split, forward model, adaptation, and training budget unchanged. For both methods, the recommended recipe uses an initial learning rate of 4×10^{-4} with epoch-wise cosine annealing to 10^{-6} over 300 epochs. Both baselines improve under these recipes.

Table A.18: Learning-rate recipe sensitivity of the adapted unfolding baselines. DPU-SCI results are mean ± sample SD over three training seeds; DAUHST entries report one run per recipe.

Method	LR recipe	SAM (°)↓	MPSNR (dB)↑	SSIM↑
DAUHST-5stg	Unified	7.156	45.050	0.990014
	Recommended	**5.097**	**45.479**	**0.993755**
DPU-SCI-5stg	Unified	7.067 ± 0.202	45.934 ± 0.199	0.989804 ± 0.000611
	Recommended	$\mathbf{6.106 \pm 0.418}$	$\mathbf{46.528 \pm 0.132}$	$\mathbf{0.991990 \pm 0.000836}$

Inference Cost. We measure end-to-end inference on an NVIDIA A100 GPU for 256×256 inputs with batch size 1. After 8 warm-up runs, latency is reported as the median over 30 CUDA-synchronized runs. Peak allocated GPU memory is recorded during the same inference procedure.

Table A.19: Inference cost of TCSS stage variants.

Stages	Latency (ms)	Peak memory (GiB)
3	106.2	1.166
5	177.5	1.180
9	317.8	1.208

Peak inference memory varies little with stage depth, and the V-basis projection is not the dominant memory contributor under this setting.

Appendix B. Implementation and Derivation Details

B.1 Calibration Features and Metric Generation

For the ordered singular values defined in Sec. 3.1, the two-dimensional calibration feature is

$$\phi_i(D) = \begin{bmatrix} \log_{10}\left(\max\left\{\frac{\sigma_i^2}{\sigma_1^2}, 10^{-16}\right\}\right) \\ \ln(1+\sigma_1^2) \end{bmatrix} \in \mathbb{R}^2. \tag{B1}$$

The first component records the base-10 logarithm of the squared singular value relative to the strongest direction. The lower bound prevents taking the logarithm of zero. The second component records the overall response scale and is shared across directions; it uses the natural logarithm, matching `log1p` in the implementation.

Each reconstruction stage uses its own MLP with layer widths $2 \to 32 \to 2$, shared across all spectral directions within that stage. Its two outputs are transformed by Softplus and the positive lower bound in Eq. (5) to generate $m_{s,i}^t$ and $m_{p,i}^t$.

B.2 Image Update and Dual Correction

With the branch estimates and metrics fixed, the image update solves

$$\begin{aligned} X^{t+1} = \arg\min_X \frac{1}{2}\|Y - DX\|_F^2 \\ + \sum_{a\in\{s,p\}} \left[\langle \Lambda_a^t, X - Z_a^{t+1}\rangle_F + \frac{1}{2}\|X - Z_a^{t+1}\|_{M_a^t}^2\right], \end{aligned} \tag{B2}$$

where $\langle A, B\rangle_F = \mathrm{tr}(A^\top B)$ is the Frobenius inner product. The matrices Λ_s^t and Λ_p^t are the unscaled Lagrange multipliers for the spectral and spatial constraints.

Setting the gradient with respect to X to zero gives

$$D^\top(DX - Y) + \sum_{a\in\{s,p\}} \left[\Lambda_a^t + M_a^t(X - Z_a^{t+1})\right] = 0. \tag{B3}$$

Rearranging yields

$$(D^\top D + M_s^t + M_p^t)X^{t+1} = D^\top Y + M_s^t Z_s^{t+1} + M_p^t Z_p^{t+1} - \Lambda_s^t - \Lambda_p^t. \tag{B4}$$

Since $D^\top D$ is positive semidefinite and both metrics are positive definite, the coefficient matrix is positive definite. Solving Eq. (B4) gives the unique update in Eq. (6).

The common basis V also gives

$$D^\top D + M_s^t + M_p^t = V \,\mathrm{diag}(\sigma_1^2 + \tau_1^t, \ldots, \sigma_N^2 + \tau_N^t) V^\top. \tag{B5}$$

Multiplying Eq. (B4) by $V^\top$ and substituting $m_{s,i}^t = \tau_i^t \pi_i^t$ and $m_{p,i}^t = \tau_i^t(1-\pi_i^t)$ gives the direction-wise update in Eq. (9). Its denominators are positive because $\sigma_i^2 \geq 0$ and $\tau_i^t > 0$.

Dual Correction in Branch Inputs. For a branch estimate Z_a, completing the square in its linear and quadratic terms gives

$$\begin{aligned} &\langle \Lambda_a^t, X - Z_a \rangle_F + \frac{1}{2}\|X - Z_a\|_{M_a^t}^2 \\ &\quad = \frac{1}{2}\left\|Z_a - X - (M_a^t)^{-1}\Lambda_a^t\right\|_{M_a^t}^2 - \frac{1}{2}\left\langle \Lambda_a^t, (M_a^t)^{-1}\Lambda_a^t \right\rangle_F. \end{aligned} \tag{B6}$$

The last term is independent of Z_a. The image supplied to branch a is therefore corrected by $(M_a^t)^{-1}\Lambda_a^t$, before adding the shared CSIP input as in Eq. (10).

B.3 Spectral and Spatial Branch Details

Spectral Branch. At each spatial location, every band is embedded into a d-dimensional feature vector, with $d = 8$. MSC, CSA, and the subsequent feed-forward network use residual connections. Let $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_N]^\top$ denote the wavelength vector, where λ_n is the wavelength of band n. The condition encoder takes $D_{:,n}$ and the wavelength encoding of λ_n to generate the query and key offsets in Eq. (11).

Given Q_{att}, K_{att}, and V_{att}, the attention weights and output features are

$$A = \mathrm{softmax}\left(\frac{Q_{\mathrm{att}} K_{\mathrm{att}}^\top}{\sqrt{d_{\mathrm{att}}}}\right), \qquad F_{\mathrm{attn}} = \mathrm{Proj}(A V_{\mathrm{att}}), \tag{B7}$$

where d_{att} is the query/key feature dimension, $A \in \mathbb{R}^{N\times N}$ contains the inter-band attention weights, and Proj is a learnable output projection. Softmax is applied over the key bands. The values contain the current spectral features; calibration and wavelength information enter only the queries and keys.

Let R_s^t denote the spectral residual mapping formed by the band embedding, MSC, CSA, feed-forward network, and final output projection. The spectral estimate is

$$Z_s^{t+1} = u_s^t + R_s^t(u_s^t; D, \boldsymbol{\lambda}). \tag{B8}$$

Spatial Branch. Following the depthwise-convolution design of ConvNeXt (Liu et al., 2022), the base path embeds its input into 32-channel features, applies a 7×7 depthwise convolution and a $32 \to 64 \to 32$ channel transformation, and projects the result back to the spectral bands. Adding the input gives the base estimate, denoted by $N_{p,\mathrm{base}}^t(u_p^t)$.

Let A^t denote the zero-initialized 1×1 convolution that processes the aggregation residual. With q^t formed by Eq. (13), the complete spatial estimate is

$$Z_p^{t+1} = N_{p,\mathrm{base}}^t(u_p^t) + A^t(q^t - u_p^t). \tag{B9}$$

B.4 Complete Reconstruction Procedure

Algorithm 1: TCSS Reconstruction Procedure

Input: measurements Y, calibrated response matrix D, wavelength vector $\boldsymbol{\lambda}$, and number of stages T.

Output: reconstructed hyperspectral image $X^T \in \mathbb{R}^{N\times P}$.

1. Decompose D to obtain V and the singular values; compute $\phi_i(D)$ using Eq. (B1) and $q_i = (V^\top D^\top Y)_{i,:}$.
2. Initialize $X^0 = D^\top Y / \|D\|_F^2$ and $\Lambda_s^0 = \Lambda_p^0 = 0$.
3. For $t = 0, \ldots, T-1$:
   (a) Generate $m_{s,i}^t, m_{p,i}^t$ using Eq. (5), then construct M_s^t, M_p^t using Eq. (4).
   (b) Set $b^t = 0$ when $t = 0$; otherwise compute $b^t = B^t(X^t - X^{t-1})$.
   (c) Form u_s^t, u_p^t using Eq. (10).
   (d) Compute the spectral and spatial estimates using Eqs. (B8)–(B9).
   (e) Compute τ_i^t, π_i^t using Eq. (8), then evaluate each row $\widetilde{x}_i^{t+1}$ using Eq. (9).
   (f) Stack these rows into $\widetilde{X}^{t+1}$ and set $X^{t+1} = V\widetilde{X}^{t+1}$.
   (g) Update $\Lambda_s^{t+1}, \Lambda_p^{t+1}$ using Eq. (7).
4. Return X^T, reshaped to $H \times W \times N$ for image output.

B.5 Training Loss

Only the final reconstruction X^T is supervised. For one image, the mean squared error is

$$\mathcal{L}_{\mathrm{MSE}} = \frac{1}{NP}\|X^T - X_{\mathrm{gt}}\|_F^2, \tag{B10}$$

where $X_{\mathrm{gt}} \in \mathbb{R}^{N\times P}$ denotes the ground-truth image. During minibatch training, this term is averaged over the images.

For the shape term, let $\hat{\mathbf{x}}_r, \mathbf{x}_r \in \mathbb{R}^N$ denote the reconstructed and ground-truth spectra, with r indexing all pixels in the current minibatch. We define the valid-pixel set as

$$\Omega = \{r : \|\mathbf{x}_r\|_2 \geq 10^{-3}\}. \tag{B11}$$

For each $r \in \Omega$, the clipped cosine similarity is

$$c_r = \mathrm{clip}\left(\frac{\hat{\mathbf{x}}_r^\top \mathbf{x}_r}{\max(\|\hat{\mathbf{x}}_r\|_2, 10^{-3})\max(\|\mathbf{x}_r\|_2, 10^{-3})}, -1, 1\right), \tag{B12}$$

where $\mathrm{clip}(z, -1, 1)$ restricts z to $[-1, 1]$. The spectral-shape loss is

$$\mathcal{L}_{\mathrm{shape}} = \frac{\sum_{r\in\Omega}(1 - c_r)}{\max(|\Omega|, 1)}. \tag{B13}$$

The norm lower bounds prevent division by zero. Pixels with a ground-truth spectral norm below 10^{-3} do not contribute to the average. If the minibatch contains no valid pixels, the shape loss is zero. The two terms are combined using Eq. (16).