BEYOND FEATURE MODULATION: PHYSICS-CONDITIONED STATE EVOLUTION FOR HYPERSPECTRAL COMPRESSIVE IMAGING

Anonymous authors
Paper under double-blind review

ABSTRACT

Physical acquisition information provides valuable guidance for reconstructing hyperspectral images from compressed measurements. Existing methods have incorporated acquisition information through mask-guided feature modulation or adaptive optimization, yet such strategies mainly affect current representations rather than explicitly governing state evolution that carries accumulated information into subsequent inference. Motivated by this, we propose **Physics-Conditioned State Evolution** (Φ2State), a framework that uses state space models to extend physical guidance from feature modulation to state evolution. Specifically, we design a deep unfolding network that couples explicit physical modeling and implicit prior inference through a shared effective coding operator. For explicit physical modeling, a bounded correction of the nominal mask yields an effective operator that guides data-consistency updates. For implicit representation, the same corrected mask conditions selective state space models within the proximal network to regulate information injection, retention, and readout throughout the entire state evolution process. The evolving states enable acquisition-dependent information to be progressively accumulated, updated, and reused during proximal inference. Extensive experiments on hyperspectral compressive reconstruction with varying spectral dimensionalities demonstrate the effectiveness of the proposed framework. Ablation studies and state evolution analyses further validate the proposed guidance mechanism and characterize its role in shaping inference dynamics.

1 INTRODUCTION

Hyperspectral imaging captures spatial structures together with material-specific spectral signatures across contiguous wavelengths, enabling fine-grained material identification and scene understanding. However, acquiring high-dimensional spectral cubes with sufficient spatial and spectral resolution remains challenging. Coded aperture snapshot spectral imaging (CASSI) addresses this limitation by encoding spectral information with a coded aperture and recording it as a two-dimensional measurement (Wagadarikar et al., 2008). This compressed measurement process leads to a severely ill-posed inverse problem, where spatial and spectral components become entangled in the recorded observation. Recovering the spectral cube consequently requires jointly leveraging the physical acquisition model and hyperspectral priors (Liu et al., 2019; Cai et al., 2022a).

Deep unfolding offers a principled framework for integrating physical acquisition models with learned priors by transforming iterative reconstruction algorithms into finite-depth trainable networks. Each stage typically combines a data-consistency update derived from the sensing operator with a learned proximal mapping that restores spatial–spectral structures from intermediate estimates (Huang et al., 2021; Cai et al., 2022b; Dong et al., 2023). Although these two components address complementary aspects of the inverse problem, existing unfolding architectures mainly incorporate acquisition information through representation modulation or adaptive optimization variables, leaving the evolution of intermediate inference states insufficiently coupled with the sensing process. Consequently, the learned proximal mapping lacks an explicit mechanism to preserve, update, and reuse acquisition-dependent evidence during reconstruction.

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.

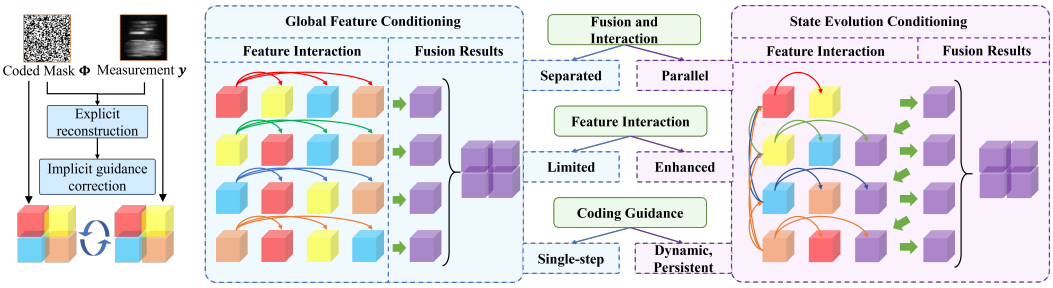

Figure 1: Motivation of Φ2State. Previous methods mainly provide static physical guidance through feature modulation or parameter adaptation, without explicitly modeling acquisition-conditioned state evolution. Our framework incorporates the shared coding condition into inference dynamics to enable progressive accumulation and reuse of reconstruction evidence.

In this context, existing methods have made substantial progress toward acquisition-aware reconstruction. MST incorporates mask guidance into spectral attention (Cai et al., 2022a); DAUHST estimates degradation-adaptive iteration parameters (Cai et al., 2022b); RDLUF learns corrections to the sensing operator (Dong et al., 2023); and DERNN jointly estimates sensing errors and denoising conditions (Dong et al., 2024b). Together, these advances integrate physical information into both explicit optimization and learned restoration. However, the acquisition condition is mainly introduced through feature modulation or parameter adaptation, while its role in explicitly governing the evolution dynamics of intermediate inference states has not been fully explored. As illustrated in Figure 1, existing strategies influence the current reconstruction but lack an explicit mechanism to preserve and reuse acquisition-dependent information for subsequent inference states.

Therefore, we introduce physics-conditioned unfolding that extends acquisition awareness from feature conditioning to state evolution. The key idea is to couple a shared effective coding operator with both the explicit data-consistency update and the implicit state transition within the proximal inference process. A bounded correction of the nominal mask constructs the effective operator for data consistency, while the same acquisition-aware descriptor conditions selective state transitions to control information injection, retention, and readout. This coupling enables the sensing process to directly govern the accumulation and propagation of reconstruction evidence during inference, moving beyond previous acquisition-aware unfolding methods that mainly influence feature representations or optimization parameters. The main contributions can be summarized as follows:

- We propose a physics-conditioned unfolding framework that couples physical acquisition and learned inference, extending acquisition-aware reconstruction from feature conditioning to state evolution. This design enables acquisition-dependent reconstruction evidence to be progressively accumulated and reused during inference.
- We develop a shared physical guidance mechanism in which a bounded mask correction improves data consistency, while the corresponding acquisition descriptor regulates selective state transitions in the proximal network.
- We demonstrate the effectiveness of the proposed framework on hyperspectral compressive reconstruction across diverse spectral dimensionalities, ranging from 31 to 144 spectral bands. Extensive ablations and state evolution analyses validate the proposed design and reveal how physical guidance shapes inference dynamics.

2 Related Work

Hyperspectral compressive reconstruction. Hyperspectral compressive reconstruction has evolved from handcrafted priors toward learned spatial–spectral modeling and physics-aware iterative reconstruction. Early model-based methods exploit structural priors such as nonlocal self-similarity and low-rankness (Liu et al., 2019), while plug-and-play approaches incorporate learned denoisers into measurement-consistency iterations (Yuan et al., 2020). Subsequent methods increasingly integrate learnable priors with the sensing process. DGSMP learns Gaussian scale mixture parameters within iterative reconstruction (Huang et al., 2021), MST introduces mask-guided spec-

tral modeling (Cai et al., 2022a), and unfolding methods such as DAUHST and PADUT adapt iterative updates according to acquisition conditions (Cai et al., 2022b; Li et al., 2023). More recent studies further improve acquisition modeling and reconstruction priors: RDLUF learns residual corrections to the sensing operator (Dong et al., 2023), DERNN jointly estimates sensing errors and noise conditions (Dong et al., 2024b), LADE-DUN incorporates a latent diffusion prior (Wu et al., 2024), and LRDUN combines low-rank decomposition with physics-based unfolding (Huang et al., 2026). SGDE further models geometric mask misalignment and jointly estimates degradation and reconstruction (He et al., 2026). These advances demonstrate the importance of coupling learned reconstruction with acquisition information. Our framework follows this direction but uses a shared effective coding operator to jointly govern explicit data consistency and the internal dynamics of the learned proximal prior.

State space models for reconstruction. State space models provide efficient long-range modeling through recurrent state propagation (Gu et al., 2022), while Mamba introduces input-dependent selective mechanisms for controlling state evolution (Gu & Dao, 2023). Recent hyperspectral reconstruction methods incorporate such models into unfolding or prior networks, including DHM, VmambaSCI, and MiJUN (Dong et al., 2024a; Zhang et al., 2024; Qin et al., 2025). Sp3ctralMamba further introduces physical and structural priors into state-space reconstruction (Meng et al., 2025), while CAMM conditions hyperspectral state updates on auxiliary RGB and semantic information (Zhang et al., 2025). Related developments such as MambaCS also combine Mamba blocks with multiscale unfolded reconstruction (Yang & Gan, 2026). Different from these approaches, our method directly derives state-conditioning signals from the effective coding operator used in data consistency and injects them into the selective step size and state input/readout mappings. In this way, acquisition guidance is extended from feature conditioning to state evolution, enabling sensing-dependent information to be continuously propagated and reused during proximal inference.

3 METHOD

3.1 OVERVIEW OF PHYSICS-GUIDED UNFOLDING FRAMEWORK

Hyperspectral compressive imaging reconstructs a high-dimensional spectral image from measurements generated by a coded physical acquisition system. Let $\mathbf{Z}_c = \mathcal{S}_c(\mathrm{Pad}(\mathbf{X}_c))$ denote the spatially shifted representation of the c-th spectral band. The measurement process is formulated as

$$\mathbf{y} = \mathcal{A}_{\mathbf{\Phi}}(\mathbf{X}) + \mathbf{n} = \sum_{c=1}^{C} \mathbf{\Phi}_c \odot Z_c + \mathbf{n}, \tag{1}$$

where $\mathcal{S}_c(\cdot)$ denotes the band-dependent spatial shifting operation induced by the dispersive optics, $\mathbf{\Phi}_c$ represents the coding pattern corresponding to the c-th spectral band, and $\odot$ denotes element-wise multiplication. The sensing process therefore mixes all shifted spectral bands into a single measurement according to their coding responses. The reconstruction problem can be formulated as

$$\min_{\mathbf{X}} \frac{1}{2}\|\mathcal{A}_{\mathbf{\Phi}}(\mathbf{X}) - \mathbf{y}\|_2^2 + \lambda \mathcal{R}(\mathbf{X}), \tag{2}$$

where the first term enforces measurement consistency and the regularization term characterizes the unknown spatial-spectral prior. Following the half-quadratic splitting (HQS) framework, the optimization is decomposed into two alternating subproblems:

$$\begin{aligned} \mathbf{Z}^k &= \arg\min_{\mathbf{Z}} \frac{1}{2}\|\mathcal{A}_{\mathbf{\Phi}}(\mathbf{Z}) - \mathbf{y}\|_2^2 + \frac{\mu_k}{2}\|\mathbf{Z} - \mathbf{X}^{k-1}\|_2^2, \\ \mathbf{X}^k &= \arg\min_{\mathbf{X}} \lambda \mathcal{R}(\mathbf{X}) + \frac{\mu_k}{2}\|\mathbf{X} - \mathbf{Z}^k\|_2^2. \end{aligned} \tag{3}$$

The first subproblem performs explicit reconstruction under the physical sensing model, while the second corresponds to implicit prior inference for spatial-spectral recovery.

We unfold the above optimization into a physics-guided network, where the effective coding condition is shared by the explicit data-consistency update and the implicit proximal mapping. As illustrated in Figure 2, this shared physical condition connects the explicit reconstruction and implicit

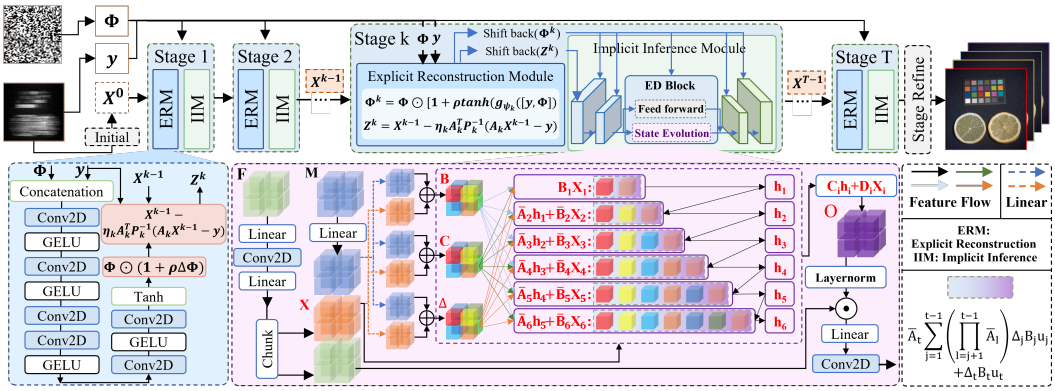

Figure 2: Overview of the proposed Φ2State framework for hyperspectral compressive reconstruction. The reconstruction process is unfolded into multiple stages, each consisting of an explicit reconstruction module (ERM) and an implicit inference module (IIM). The effective coding condition is shared between physical data consistency and state evolution, enabling acquisition-dependent reconstruction evidence to be progressively accumulated, propagated, and reused during iterative inference.

inference processes across unfolding stages. Each stage alternates between an explicit reconstruction step and an implicit proximal inference step, which can be formulated as

$$\begin{aligned} \mathbf{Z}^k &= \mathcal{D}_k(\mathbf{X}^{k-1}, \mathbf{y}, \mathbf{\Phi}^k), \\ \mathbf{X}^k &= \mathbf{Z}^k + \mathcal{P}_{\theta_k}(\mathbf{Z}^k, \mathbf{\Phi}^k). \end{aligned} \tag{4}$$

where $\mathcal{D}_k(\cdot)$ denotes the unfolded explicit update derived from the physical sensing model, $\mathbf{\Phi}^k$ is the effective coding mask, and $\mathcal{P}_{\theta_k}(\cdot)$ represents the implicit proximal mapping.

3.2 Shared Effective Coding Operator for Explicit Reconstruction

The explicit reconstruction module in equation 4 performs data-consistency updates according to the physical sensing model. Recent unfolding methods have explored acquisition-aware reconstruction by learning adaptive sensing operators (Dong et al., 2023). Following this direction, we estimate an effective coding mask $\mathbf{\Phi}^k$ at each unfolding stage. The stage-specific estimation allows different unfolding iterations to adapt to the evolving reconstruction state. Unlike previous operator correction methods, the estimated coding condition is shared by both explicit reconstruction and implicit state evolution. Specifically, the effective coding mask is estimated as

$$\mathbf{\Phi}^k = \mathbf{\Phi} \odot [1 + \rho \tanh\left(g_{\psi_k}([\mathbf{y}, \mathbf{\Phi}])\right)], \tag{5}$$

where $g_{\psi_k}(\cdot)$ denotes a lightweight convolutional estimator and ρ controls the correction range. The corrected coding mask defines a stage-specific sensing operator $\mathcal{A}_{\mathbf{\Phi}^k}$. For convenience, we denote the corresponding linear measurement matrix as $\mathbf{A}_k$, which characterizes the physical mapping from the hyperspectral cube to the compressed measurement under the effective coding condition $\mathbf{\Phi}^k$. Different from the nominal operator induced by the original mask, $\mathbf{A}_k$ represents the acquisition condition adaptively estimated at the k-th unfolding stage. The explicit update is then formulated as

$$\begin{aligned} \mathbf{Z}^k &= \mathcal{D}_k(\mathbf{X}^{k-1}, \mathbf{y}, \mathbf{\Phi}^k) \\ &= \mathbf{X}^{k-1} - \eta_k \mathbf{A}_k^T \mathbf{P}_k^{-1}(\mathbf{A}_k \mathbf{X}^{k-1} - \mathbf{y}), \end{aligned} \tag{6}$$

where $\mathbf{A}_k^T$ denotes the adjoint operation of the effective sensing matrix for back-projecting measurement residuals, $\mathbf{P}_k$ normalizes the coding energy of the effective operator, and η_k denotes the learnable update step. The estimated operator therefore provides a unified physical representation, which is subsequently converted into descriptors for state-conditioned prior inference.

3.3 Physics-Conditioned State Evolution for Implicit Prior Inference

Although the explicit module enforces measurement consistency, the implicit proximal mapping should further adapt its inference to the acquisition condition. Therefore, we employ a physics-

Table 1: Summary of hyperspectral datasets used in experiments.

Dataset	Scene	Resolution	Bands	Training	Testing
CAVE	Natural scene	N/A	31	21 (512×512)	10 (128×128)
UH2018	Urban scene	1.0 m	48	15 (256×256)	12 (128×128)
PC	Urban scene	1.3 m	100	7 (256×256)	4 (128×128)
UH2013	Complex urban scene	1.5 m	144	50 (256×256)	7 (128×128)

conditioned U-shaped proximal network to refine $\mathbf{Z}^k$ under the guidance of $\mathbf{\Phi}^k$. Within each encoder-decoder block of $\mathcal{P}_{\theta_k}$, image features $\mathbf{F}$ and the corresponding physical descriptors $\mathbf{M}$ are processed by a physics-conditioned state evolution branch $\mathcal{S}_\theta(\cdot)$ and a nonlinear feed-forward branch $\mathcal{F}_\theta(\cdot)$:

$$\mathbf{F}^{out} = \mathcal{S}_\theta(\mathbf{F}, \mathbf{M}) + \mathcal{F}_\theta(\mathbf{F}) \tag{7}$$

where $\mathbf{M} = \Gamma(\mathbf{\Phi}^k)$ denotes the physical descriptor extracted from the $\mathbf{\Phi}^k$. Different from feature-level modulation, the state evolution branch introduces the physical descriptor into the generation of selective state parameters. Given the input feature state $\mathbf{u}_t$ and the physical embedding $\mathbf{m}_t$, the selective parameters are obtained by

$$\begin{aligned}
\Delta_t &= \Delta_t(\mathbf{u}_t, \mathbf{m}_t) = W_\Delta(f_\Delta(\mathbf{u}_t) + \alpha_\Delta g_\Delta(\mathbf{m}_t)), \\
\mathbf{B}_t &= \mathbf{B}_t(\mathbf{u}_t, \mathbf{m}_t) = f_B(\mathbf{u}_t) + \alpha_B g_B(\mathbf{m}_t), \\
\mathbf{C}_t &= \mathbf{C}_t(\mathbf{u}_t, \mathbf{m}_t) = f_C(\mathbf{u}_t) + \alpha_C g_C(\mathbf{m}_t),
\end{aligned} \tag{8}$$

where $f(\cdot)$ denotes feature-driven parameter generation and $g(\cdot)$ represents the acquisition-dependent modulation induced by the physical descriptor. The learnable coefficients α_Δ, α_B, and α_C control the influence of the physical condition on state transition, state writing, and state readout, respectively. The resulting selective state evolution is formulated as

$$\begin{aligned}
\mathbf{o}_t &= \mathbf{C}_t(\mathbf{u}_t, \mathbf{m}_t)(\bar{\mathbf{A}}_t \mathbf{h}_{t-1} + \Delta_t(\mathbf{u}_t, \mathbf{m}_t)\mathbf{B}_t(\mathbf{u}_t, \mathbf{m}_t)\mathbf{u}_t) \\
&= \mathbf{C}_t(\bar{\mathbf{A}}_t(\bar{\mathbf{A}}_{t-1}\mathbf{h}_{t-2} + \Delta_{t-1}\mathbf{B}_{t-1}\mathbf{u}_{t-1}) + \Delta_t \mathbf{B}_t \mathbf{u}_t) \\
&= ... \\
&= \mathbf{C}_t(\bar{\mathbf{A}}_t \sum_{j=1}^{t-1} \left(\prod_{\ell=j+1}^{t-1} \bar{\mathbf{A}}_\ell \right) \Delta_j \mathbf{B}_j \mathbf{u}_j + \Delta_t \mathbf{B}_t \mathbf{u}_t)
\end{aligned} \tag{9}$$

where $\mathbf{h}_t$ denotes the hidden reconstruction state that aggregates the current acquisition-conditioned information and previously accumulated reconstruction evidence, and $\bar{\mathbf{A}}_t = \exp(\Delta_t \mathbf{A})$ is the discretized state transition matrix. Unlike static feature modulation, the recursive state transition enables acquisition-dependent evidence to be progressively propagated across inference steps and influence subsequent reconstruction.

4 Experiments

4.1 Experiment Settings

Datasets. We evaluate the proposed Φ2State model on four hyperspectral datasets, including CAVE, UH2018, Pavia Center (PC), and UH2013. These datasets are selected to cover diverse spectral dimensionalities and scene characteristics, with spectral bands ranging from 31 to 144. As illustrated in Figure 3, the CAVE dataset contains natural scenes with rich spatial structures and fine spectral variations. Meanwhile, the remote sensing datasets shown in Figure 4 include complex urban and terrestrial scenes with different spatial resolutions and spectral dimensions, providing challenging reconstruction scenarios with varying spectral complexities and spatial structures. This diversity enables a comprehensive evaluation of the proposed framework under different spectral complexities and physical reconstruction conditions. The detailed information, including scene characteristics, spatial resolutions, spectral bands, and training/testing settings, is summarized in Table 1.

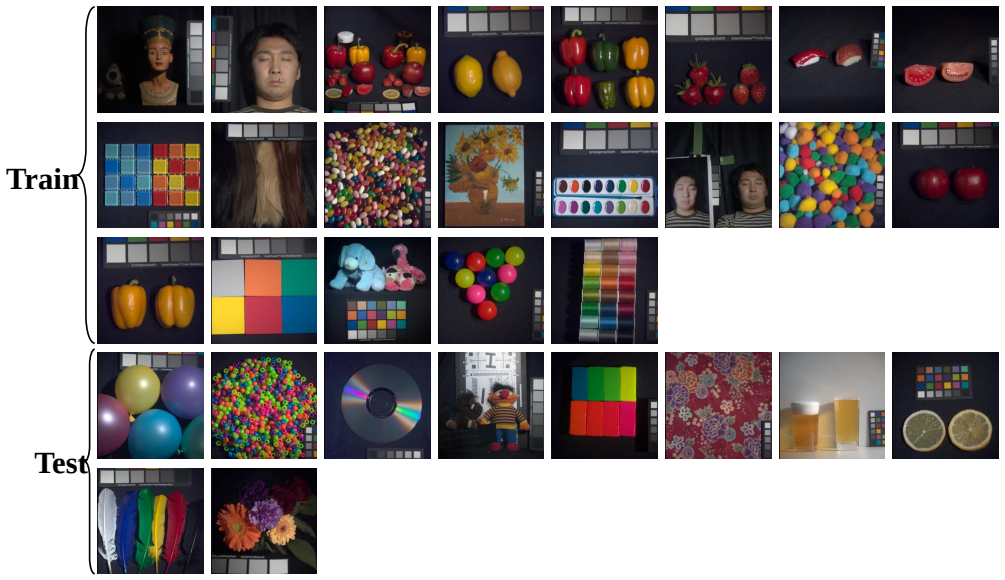

Figure 3: Visualization of the CAVE dataset. The dataset contains natural scenes with 31 spectral bands, providing a representative benchmark for evaluating hyperspectral reconstruction on low-dimensional spectral data.

Comparison Methods. We compare Φ2State with eight state-of-the-art hyperspectral compressive reconstruction methods. MST (Cai et al., 2022a) and DSMT (Luo et al., 2025) employ attention-based architectures to capture long-range spatial-spectral dependencies. RDLUF (Dong et al., 2023), PADUT (Li et al., 2023), RCUMP (Zhao et al., 2024), and DERNN (Dong et al., 2024b) are representative deep unfolding methods that improve reconstruction by incorporating acquisition information, adaptive optimization, or degradation modeling. MiJUN (Qin et al., 2025) and SCM-DUN (Feng et al., 2026) further exploit state-space modeling and efficient unfolding strategies for hyperspectral reconstruction. All compared methods are implemented following their original experimental settings for fair comparison.

Implementation Details. All experiments are conducted using PyTorch on a Linux server equipped with a single NVIDIA H100 GPU. Reconstruction quality is evaluated using three commonly used metrics, including peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). Computational efficiency is measured in terms of floating-point operations (GFLOPs) and the number of trainable parameters (Params). The proposed model is optimized using the Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The initial learning rate is set to 4×10^{-4} and gradually reduced using a cosine annealing schedule with warmup. The proposed model is trained for 300 epochs with a batch size of 2. The Charbonnier loss is adopted as the reconstruction objective.

4.2 Comparison with State-of-the-Art Methods

Table 2 reports quantitative comparisons on different hyperspectral compressive reconstruction benchmarks. Our method achieves consistently competitive performance across datasets with different spectral dimensions, especially on high-dimensional PC and UH2013 datasets. Figure 5 provides qualitative comparisons on the CAVE and PC datasets. The spectral profiles show that our method better preserves spectral variations and achieves closer agreement with the ground truth. The reconstructed band images and MSE maps further demonstrate improved recovery of spatial details, especially for object boundaries and fine structures. Although SCMDUN achieves competitive spectral reconstruction on CAVE, its performance degrades on the higher-dimensional PC dataset. In contrast, our method maintains stable reconstruction quality, indicating stronger capability in modeling complex spectral-spatial correlations.These results demonstrate that explicitly incorporating acquisition conditions into state evolution enables the model to better preserve spectral-spatial correlations

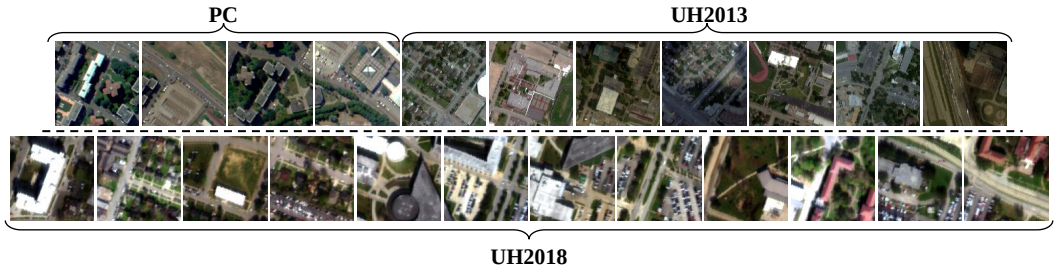

Figure 4: Visualization of hyperspectral test datasets used in this work. The examples cover diverse scene characteristics and spectral dimensionalities, including complex urban environments from PC, UH2013, and UH2018.

Table 2: Quantitative comparison on the UH2018, PC, and UH2013 datasets. The best and second-best results are highlighted in **bold** and underlined, respectively.

UH2018 (48 bands)

Method	Metric	1	2	3	4	5	6	7	8	9	10	11	12	Avg
MST-l	PSNR	37.92	38.34	38.31	38.80	35.42	38.63	37.77	37.77	36.17	36.63	36.50	38.17	37.54
	SSIM	0.9383	0.9396	0.9367	0.9370	0.9210	0.9367	0.9355	0.9276	0.9325	0.9289	0.9249	0.9346	0.9328
DSMT	PSNR	38.69	39.36	39.04	39.41	36.35	39.66	38.54	38.83	36.84	37.43	37.47	39.30	38.41
	SSIM	0.9479	0.9512	0.9463	0.9463	0.9353	0.9487	0.9458	0.9422	0.9424	0.9395	0.9388	0.9477	0.9443
RDLUF-9stg	PSNR	37.37	37.86	37.92	38.35	35.01	38.22	37.40	37.53	35.65	36.21	35.81	37.73	37.09
	SSIM	0.9311	0.9357	0.9307	0.9316	0.9187	0.9297	0.9302	0.9233	0.9264	0.9205	0.9147	0.9267	0.9266
PADUT-9stg	PSNR	40.84	41.67	40.34	41.17	**38.46**	41.15	40.25	40.43	38.78	**39.19**	**39.38**	40.66	40.19
	SSIM	**0.9661**	0.9689	0.9598	0.9631	**0.9597**	0.9624	0.9611	0.9590	0.9614	0.9567	**0.9604**	0.9621	**0.9617**
RCUMP	PSNR	39.46	39.96	39.34	40.07	37.63	39.88	39.31	39.29	38.03	38.12	37.99	39.59	39.06
	SSIM	0.9568	0.9581	0.9522	0.9550	0.9512	0.9527	0.9544	0.9484	0.9550	0.9489	0.9482	0.9533	0.9529
DERNN-9stg	PSNR	40.19	40.87	39.69	40.40	37.84	40.34	39.90	39.62	38.30	38.48	38.58	40.05	39.52
	SSIM	0.9598	0.9631	0.9531	0.9564	0.9527	0.9554	0.9577	0.9503	0.9562	0.9504	0.9527	0.9558	0.9553
MiJUN-9stg	PSNR	40.65	41.33	40.37	40.97	38.32	41.23	39.89	**40.49**	38.94	39.14	39.31	**40.85**	40.13
	SSIM	0.9640	0.9669	0.9597	0.9622	0.9579	0.9626	0.9590	**0.9595**	0.9616	**0.9568**	0.9586	**0.9625**	0.9609
SCMDUN-9stg	PSNR	40.62	41.41	40.02	40.92	38.26	40.98	40.26	39.99	38.63	38.68	39.03	40.47	39.94
	SSIM	0.9634	0.9672	0.9562	0.9612	0.9567	0.9608	0.9597	0.9549	0.9606	0.9524	0.9567	0.9596	0.9591
Φ2State	PSNR	**40.91**	**41.91**	**40.54**	**41.40**	38.30	**41.51**	**40.60**	40.44	**39.36**	39.15	39.32	40.80	**40.35**
	SSIM	0.9652	**0.9695**	**0.9604**	**0.9644**	0.9552	**0.9638**	**0.9619**	0.9576	**0.9621**	0.9557	0.9579	0.9621	0.9613

		PC (100 bands)					UH2013 (144 bands)							
Method	Metric	1	2	3	4	Avg	1	2	3	4	5	6	7	Avg
MST-l	PSNR	29.15	30.37	30.74	29.38	29.91	35.02	36.67	36.40	39.97	33.12	35.27	38.66	36.44
	SSIM	0.8623	0.8153	0.8635	0.8400	0.8453	0.8648	0.9048	0.8956	0.9334	0.8712	0.8770	0.9294	0.8966
DSMT	PSNR	28.68	30.08	30.44	29.41	29.65	36.76	38.70	38.17	41.65	34.40	37.16	40.78	38.23
	SSIM	0.8477	0.8134	0.8505	0.8398	0.8379	0.9089	0.9347	0.9285	0.9553	0.9091	0.9163	0.9535	0.9295
RDLUF-9stg	PSNR	28.96	29.90	29.20	28.90	29.24	36.67	38.32	37.83	41.51	34.75	36.82	40.41	38.04
	SSIM	0.8529	0.7968	0.8158	0.8178	0.8208	0.9008	0.9247	0.9209	0.9528	0.9009	0.9084	0.9491	0.9225
PADUT-9stg	PSNR	29.67	30.73	31.07	29.76	30.31	37.01	38.63	38.31	42.00	35.33	36.94	40.55	38.39
	SSIM	0.8732	0.8240	0.8713	0.8532	0.8554	0.9064	0.9289	0.9272	0.9566	0.9120	0.9113	0.9521	0.9278
RCUMP	PSNR	29.75	30.92	30.97	30.31	30.49	37.10	39.11	38.24	41.63	35.32	37.32	40.73	38.49
	SSIM	0.8690	0.8294	0.8617	0.8626	0.8557	0.9147	0.9407	0.9322	0.9542	0.9176	0.9205	0.9545	0.9335
DERNN-9stg	PSNR	30.78	31.24	**31.70**	30.08	30.95	37.64	39.36	38.86	42.10	35.75	37.78	41.61	39.01
	SSIM	**0.8958**	0.8362	**0.8861**	0.8607	0.8697	0.9170	0.9396	0.9349	0.9598	0.9207	0.9232	0.9591	0.9363
MiJUN-9stg	PSNR	30.44	31.10	31.51	29.70	30.68	**37.81**	39.20	39.15	42.28	36.15	37.99	41.70	39.18
	SSIM	0.8929	0.8413	0.8796	0.8508	0.8662	**0.9226**	0.9397	0.9399	0.9609	0.9265	0.9276	0.9598	0.9396
SCMDUN-9stg	PSNR	30.49	**31.35**	31.47	30.44	30.94	37.51	39.36	38.75	41.97	36.06	37.81	41.32	38.97
	SSIM	0.8898	**0.8483**	0.8811	**0.8729**	**0.8730**	0.9177	0.9398	0.9330	0.9579	0.9255	0.9250	0.9567	0.9365
Φ2State	PSNR	**30.87**	31.20	31.65	**30.65**	**31.09**	37.78	**39.64**	**39.32**	**42.80**	**36.19**	**38.11**	**42.17**	**39.43**
	SSIM	0.8886	0.8421	0.8835	0.8709	0.8713	0.9213	**0.9431**	**0.9409**	**0.9655**	**0.9272**	**0.9281**	**0.9634**	**0.9414**

during iterative reconstruction, especially for hyperspectral data with higher spectral dimensionalities.

4.3 Ablation Study

Effect of Shared Effective Coding Condition. Table 3 investigates the contribution of the proposed shared effective coding condition. The Fixed Φ variant follows the conventional unfolding paradigm, where the nominal coding mask is only used for enforcing measurement consistency. Introducing the learnable effective coding mask improves the PSNR from 39.81 dB to 40.07 dB, demonstrating that adaptively estimating the acquisition condition provides a more accurate physical representation for reconstruction. When the estimated coding condition is further introduced into state evolution, the State Evolution Φ variant achieves an additional improvement of 0.18 dB. This

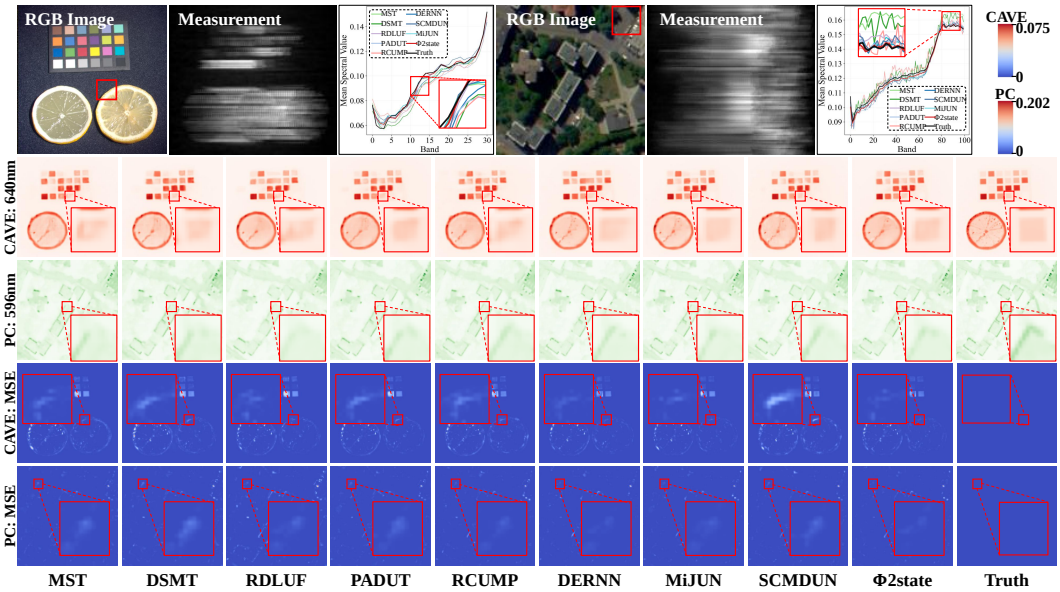

Figure 5: Qualitative comparison of hyperspectral compressive reconstruction methods on CAVE and PC datasets. The top panels show RGB images, coded measurements, and spectral profiles from the regions marked by red boxes. The lower panels present representative spectral band reconstructions and the pixel-wise MSE maps calculated over the entire reconstructed images.

Table 3: Ablation study of shared effective coding condition on the UH2018 dataset.

Variant	Φ_{eff}	State Condition	Stage Refine	PSNR	SSIM	ΔPSNR (dB)
Fixed Φ	No	No	No	39.81	0.9584	–
Learnable Φ	Yes	No	No	40.07	0.9601	0.26
State Evolution Φ	Yes	Yes	No	40.25	0.9609	0.18
Φ2State	Yes	Yes	Yes	40.35	0.9613	0.10

result indicates that physical information is not only beneficial for correcting explicit data consistency, but can also provide meaningful guidance for the evolution of implicit reconstruction states. Finally, the complete Φ2State framework achieves the highest performance of 40.35 dB PSNR and 0.9613 SSIM by jointly coupling effective coding estimation, state conditioning, and stage-wise refinement. These results verify the necessity of sharing the physical condition between explicit reconstruction and implicit inference.

Effect of State-level Physical Guidance. Table 4 compares different strategies for incorporating physical information into the inference process. Feature concatenation and cross-attention introduce acquisition information at the representation level, which can enhance feature interaction but does not explicitly regulate the evolution of hidden reconstruction states. The Mamba State variant further improves the modeling of long-range dependencies through state space modeling; however, the state transition is still mainly driven by image features without explicit physical constraints. In contrast, the Physics-only State introduces acquisition information into state evolution, leading to improved reconstruction performance. By jointly modeling the reconstruction state and acquisition condition, the proposed Physics-conditioned State achieves the best performance, demonstrating that physical guidance is more effective when embedded into the inference dynamics rather than only modifying intermediate representations.

Effect of Physics-guided State Transition Modulation. Table 5 further analyzes how the physical condition affects different components of selective state evolution.Specifically, B, C, and Δt correspond to state writing, state reading, and state transition dynamics, respectively. Modulating

Table 4: Ablation study of physics-conditioned state evolution strategies on the UH2018 dataset.

State Evolution Strategy	PSNR (dB)	SSIM
Feature Concatenation	40.15	0.9604
Cross Attention	40.20	0.9613
Mamba State	40.18	0.9605
Physics-only State	40.23	0.9609
Physics-conditioned State	40.35	0.9613

Table 5: Ablation study of physical modulation on state transition components on the UH2018 dataset.

Variant	B	C	Δt	PSNR (dB)	SSIM
Only B	Yes	No	No	40.21	0.9604
Only C	No	Yes	No	40.26	0.9612
Only Δt	No	No	Yes	40.20	0.9603
$B + C$	Yes	Yes	No	40.23	0.9604
$B + \Delta t$	Yes	No	Yes	40.17	0.9601
$C + \Delta t$	No	Yes	Yes	40.26	0.9611
Full	Yes	Yes	Yes	40.35	0.9613

each individual component already improves reconstruction quality, indicating that acquisition information can effectively regulate different aspects of the inference process. Among single-parameter variants, introducing physical guidance into C achieves relatively better performance, suggesting that acquisition-dependent information is particularly beneficial for controlling state readout. Combining multiple state parameters further improves reconstruction quality, and the full model with joint modulation of B, C, and Δt achieves the best performance. This confirms that the effectiveness of the proposed framework originates from jointly controlling information injection, transformation, and readout during state evolution.

5 Conclusion

In this work, we proposed Φ2State, a physics-conditioned unfolding framework for hyperspectral compressive reconstruction that extends acquisition-aware modeling from feature conditioning to state evolution. By constructing a shared effective coding operator, the proposed framework enables the acquisition condition to jointly regulate explicit data consistency and implicit prior inference. Specifically, the corrected coding information is converted into physical descriptors to condition selective state transitions, allowing acquisition-dependent reconstruction evidence to be progressively accumulated, preserved, and reused throughout inference. Extensive experiments on hyperspectral datasets with diverse spectral dimensionalities demonstrate the effectiveness of the proposed framework compared with existing reconstruction methods. Comprehensive ablation studies and state evolution analyses further validate the proposed guidance mechanism and demonstrate how physical acquisition information influences iterative inference dynamics. The proposed framework provides a new perspective for integrating physical acquisition models with learned priors in inverse problems. Future work will explore more adaptive physical representations and lightweight state evolution mechanisms for large-scale hyperspectral reconstruction and other physics-driven imaging tasks.

AI use statement

In this work, generative AI tools were used for language refinement, proofreading, and improving the clarity and organization of the manuscript. They were also used to assist manuscript presentation, including suggesting possible figure layouts and visualization designs.

Generative AI tools were not used to generate experimental results, perform data analysis, develop the core methodology, derive theoretical results, or make scientific claims. All algorithmic designs,

mathematical formulations, implementations, experiments, and conclusions were conducted and verified by the authors.

All AI-assisted suggestions were carefully reviewed and revised by the authors. AI-assisted writing suggestions were checked against the original research content and relevant literature to ensure accuracy and consistency. Any AI-assisted code suggestions, when used, were manually inspected, modified, and tested by the authors before integration. We take full responsibility for the final content of this work, including all text, claims, experimental results, and supplementary research outputs.

ETHICS STATEMENT

This work does not involve human subjects, private data collection, or personally identifiable information. All datasets used in our experiments are publicly available benchmark datasets for hyperspectral image reconstruction, and we follow their original usage protocols. The proposed method focuses on improving computational reconstruction quality and does not require additional collection or processing of privacy-sensitive information.

We conduct this research following principles of research integrity, including accurate reporting of experimental results, transparent evaluation, and responsible use of publicly available resources.

REPRODUCIBILITY STATEMENT

We provide detailed descriptions of the proposed model, optimization formulation, training settings, and evaluation protocols in the main paper and supplementary material. All experiments are conducted on publicly available hyperspectral benchmarks with documented preprocessing and evaluation procedures. Upon acceptance, we will release the source code and implementation details to facilitate reproduction and further research.

REFERENCES

Yuanhao Cai, Jing Lin, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 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.

Jiahua Dong, Hui Yin, Hongliu Li, Wenbo Li, Yulun Zhang, Salman Khan, and Fahad Shahbaz Khan. Dual hyperspectral Mamba for efficient spectral compressive imaging. *arXiv preprint arXiv:2406.00449*, 2024a. URL https://arxiv.org/abs/2406.00449.

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*, pp. 22262–22271, 2023.

Yubo Dong, Dahua Gao, Yuyan Li, Guangming Shi, and Danhua Liu. Degradation estimation recurrent neural network with local and non-local priors for compressive spectral imaging. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 2024b.

Yuchao Feng, Mengjie Qin, Zongliang Wu, Yuxiang Yang, Junhua Gao, and Xin Yuan. 2d-slice and 3d-cube mamba network for snapshot spectral compressive imaging. *IEEE Transactions on Circuits and Systems for Video Technology*, 36(4):4049–4063, 2026.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.

Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. In *International Conference on Learning Representations*, 2022.

Yuqiao He, Xiaoyan Liu, Jianxu Mao, Yaonan Wang, Hui Zhang, Lizhu Liu, Yurong Chen, and Wenbin He. SGDE: Self-supervised geometry degradation estimation framework for coded aperture compressive spectral imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 34084–34094, June 2026.

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*, 2021.

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*, 2023.

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.

Fulin Luo, Xi Chen, Tan Guo, Xiuwen Gong, Lefei Zhang, and Ce Zhu. Dsmt: Dual-stage multiscale transformer for hyperspectral snapshot compressive imaging. *IEEE Transactions on Image Processing*, 34:2473–2486, 2025.

Ge Meng, Jingyan Tu, Jingjia Huang, Yunlong Lin, Yingying Wang, Xiaotong Tu, Yue Huang, and Xinghao Ding. Sp3ctralMamba: Physics-driven joint state space model for hyperspectral image reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 6108–6116, 2025.

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.

Ashwin Wagadarikar, Renu John, Rebecca Willett, and David Brady. Single disperser design for coded aperture snapshot spectral imaging. *Appl. Opt.*, 47(10):B44–B51, Apr 2008.

Zongliang Wu, Ruiying Lu, Ying Fu, and Xin Yuan. Latent diffusion prior enhanced deep unfolding for snapshot spectral compressive imaging. In *European Conference on Computer Vision*, pp. 164–181, 2024.

Le Yang and Hongping Gan. Multi-scale gradient-guided unrolling architecture with adaptive Mamba for compressive sensing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 41794–41803, June 2026.

Xin Yuan, Yang Liu, Jinli Suo, and Qionghai Dai. Plug-and-play algorithms for large-scale snapshot compressive imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1447–1457, 2020.

Mingjin Zhang, Longyi Li, Wenxuan Shi, Jie Guo, Yunsong Li, and Xinbo Gao. VmambaSCI: Dynamic deep unfolding network with Mamba for compressive spectral imaging. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 2024.

Mingjin Zhang, Longyi Li, Fei Gao, Qiming Zhang, and Jie Guo. Multimodal prior learning with double constraint alignment for snapshot spectral compressive imaging. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 2359–2367, 2025.

Yin-Ping Zhao, Jiancheng Zhang, Yongyong Chen, Zhen Wang, and Xuelong Li. Rcump: Residual completion unrolling with mixed priors for snapshot compressive imaging. *IEEE Transactions on Image Processing*, 33:2347–2360, 2024.

Table 6: Effect of the number of unfolding stages on the UH2018 dataset.

Number of Stages	3	5	7	9
PSNR (dB)	39.20	39.96	40.08	40.35

Table 7: Quantitative comparison on the CAVE dataset (31 bands). The best and second-best results are highlighted in **bold** and underline, respectively.

CAVE (31 bands)												
Method	Metric	1	2	3	4	5	6	7	8	9	10	Avg
MST	PSNR	31.79	19.72	25.30	26.91	33.14	27.51	31.65	32.52	26.55	30.46	28.56
	SSIM	0.9387	0.4906	0.8829	0.8679	0.9119	0.5908	0.9267	0.9196	0.8415	0.8789	0.8250
DSMT	PSNR	31.71	20.05	25.04	27.95	33.16	27.51	32.04	32.57	27.59	30.47	28.81
	SSIM	0.9357	0.4677	0.8709	0.8762	0.9127	0.5850	0.9263	0.9195	0.8580	0.8733	0.8225
RDLUF	PSNR	33.84	19.13	24.97	27.50	36.46	27.81	33.57	32.86	28.45	31.41	29.60
	SSIM	0.9581	0.4434	0.8810	0.8869	0.9446	0.6184	0.9467	0.9226	0.8942	0.8938	0.8390
PADUT	PSNR	34.00	20.23	25.05	28.25	35.46	28.09	33.62	33.42	28.61	31.42	29.81
	SSIM	0.9571	0.5093	0.8833	0.8931	0.9379	0.6304	0.9468	0.9310	0.8835	0.8954	0.8468
RCUMP	PSNR	33.11	19.79	25.63	28.00	34.89	27.34	32.65	32.77	28.70	31.15	29.40
	SSIM	0.9415	0.4545	0.8578	0.8654	0.9094	0.5802	0.9348	0.9162	0.8657	0.8684	0.8194
DERNN	PSNR	34.86	19.89	**26.56**	28.94	**38.13**	**28.33**	34.28	33.89	29.48	**32.54**	30.69
	SSIM	0.9657	0.5103	**0.9053**	0.9095	0.9570	0.6528	0.9543	0.9355	**0.9184**	**0.9187**	0.8628
MiJUN	PSNR	35.16	**20.45**	25.22	29.26	37.67	28.18	**34.87**	34.54	**29.65**	31.78	30.68
	SSIM	0.9702	**0.5241**	0.8981	0.9111	0.9584	**0.6742**	**0.9567**	0.9431	0.9152	0.9077	**0.8659**
SCM-DUN	PSNR	33.53	19.53	24.64	28.21	35.99	27.74	33.23	32.76	28.51	31.12	29.53
	SSIM	0.9535	0.5035	0.8725	0.8918	0.9354	0.6270	0.9426	0.9251	0.8839	0.8865	0.8422
Φ2State	PSNR	**35.36**	20.42	26.44	**29.31**	38.09	28.32	34.55	**34.59**	28.97	31.53	**30.76**
	SSIM	**0.9715**	0.5240	0.8995	**0.9132**	0.9600	0.6566	0.9530	**0.9438**	0.9036	0.9037	0.8629

A Derivation of the Explicit Reconstruction Update

Effect of the Number of Unfolding Stages. We further investigate the influence of the number of unfolding stages on reconstruction performance. As shown in Table 6, increasing the number of stages consistently improves the reconstruction quality. The PSNR increases from 39.20 dB with 3 stages to 40.35 dB with 9 stages, indicating that additional iterative refinement enables the model to progressively recover spatial-spectral information from compressed measurements. The performance improvement gradually saturates with more stages, suggesting that the proposed framework can achieve a favorable trade-off between reconstruction accuracy and computational complexity.

B Additional Results on the CAVE Dataset

To further evaluate the reconstruction performance on natural hyperspectral scenes, we provide additional quantitative comparisons on the CAVE dataset with 31 spectral bands in Table 7. Compared with existing methods, Φ2State achieves the highest average PSNR of 30.76 dB, demonstrating superior spectral reconstruction accuracy across diverse natural scenes. In particular, the proposed method obtains the best PSNR performance on multiple test scenes and consistently outperforms most existing unfolding-based approaches.

For structural reconstruction quality, Φ2State achieves competitive SSIM results and obtains the best performance on several individual scenes. The results indicate that introducing acquisition conditions into state evolution can effectively preserve spatial–spectral structures during iterative reconstruction. Although the performance differences become smaller on the CAVE dataset compared with complex urban scenes, the proposed method still maintains stable reconstruction capability, verifying the generalization of the proposed physics-conditioned state evolution mechanism across different scene types and spectral dimensionalities.