# SALR-FORMER: LEARNING RATIO- AND STATE-ADAPTIVE RECONSTRUCTION FOR IMAGE COMPRESSIVE SENSING

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Image compressive sensing (ICS) reconstructs images from highly undersampled measurements. Most ICS networks are optimized for fixed sampling ratios, whereas existing arbitrary-ratio methods mainly make a shared sensing or reconstruction pipeline compatible with multiple ratios. Such compatibility does not specify how reconstruction should change during a finite unfolding process: the sampling ratio changes the measurement-constrained image subspace, while the evolving iterate determines the residual errors that remain to be corrected. We propose SALR-Former, a unified unfolding network for arbitrary-ratio ICS. Its Ratio-Stage Controller (RSC) uses the sampling ratio as a global sensing condition and combines it with the stage index, measurement residual, reconstruction feature, and cross-stage memory to control two complementary mechanisms: a bounded physics-update step size and lightweight FiLM-LoRA modulation of the learned prior. The shared prior integrates local, contextual, and frequency-domain representations, while the predicted controls specialize its behavior without ratio-specific backbones or adapter banks. Experiments on five benchmarks demonstrate strong reconstruction quality across trained and unseen continuous sampling ratios and controlled ablations further show that the sampling ratio determines the dominant reconstruction regime, whereas stage and cross-stage history refine the control trajectory.

## 1 INTRODUCTION

Image compressive sensing (ICS) (Yang et al., 2018; Guo & Gan, 2025) recovers an image from a small number of linear measurements acquired below the Nyquist rate. It reduces acquisition, storage, and transmission costs and has been studied in single-pixel imaging (Duarte et al., 2008), hyperspectral imaging (Liu et al., 2026), and medical imaging (Heckel et al., 2024). The reconstruction is nevertheless severely ill-posed, particularly at low sampling ratios (SRs).

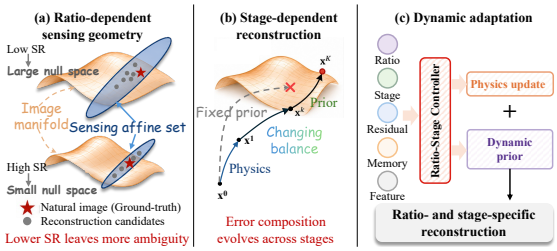



Figure 1: **Motivation for dynamic reconstruction adaptation.** (a) Lower SRs enlarge the sensing null space and increase prior ambiguity. (b) The reconstruction balance changes across stages. (c) RSC generates ratio- and stage-specific controls for the physical update and learned prior.

Following recent ICS systems (Wang & Yuan, 2023; Qu et al., 2026), we consider the separable two-sided sensing model

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x}\mathbf{\Psi}^T, \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^{H\times W}$ is the target image, $\mathbf{y} \in \mathbb{R}^{m\times n}$ is the measurement, and $\mathbf{\Phi} \in \mathbb{R}^{m\times H}$ and $\mathbf{\Psi} \in \mathbb{R}^{n\times W}$ are the vertical and horizontal sensing matrices. The SR is $\tau = mn/(HW)$. Classical methods recover $\mathbf{x}$ by balancing measurement consistency and a handcrafted prior, but their limited representation capacity and iterative computation restrict reconstruction quality and efficiency.

Deep networks have substantially improved ICS reconstruction. CNN-based methods (Shi et al., 2019a) learn direct measurement-to-image mappings, while deep unfolding networks (DUNs)



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.

(Zhang & Ghanem, 2018; Zhang et al., 2020; Song et al., 2021; Guo & Gan, 2024; Gan et al., 2024) map iterative optimization into trainable stages and retain an explicit connection to the sensing model. Transformer-based methods (Ye et al., 2023; Zhang et al., 2023) further improve structural recovery through contextual modeling. Most of these methods are trained for a fixed SR or sensing operator, requiring multiple models when the acquisition budget changes.

Arbitrary-ratio ICS (AR-ICS) aims to reconstruct images under varying acquisition budgets with a single model. Existing AR-ICS methods (You et al., 2021; Wang et al., 2025) make reconstruction compatible with multiple sensing ratios through scalable measurements, ratio-conditioned features, or shared proximal mappings. However, compatibility with a ratio is different from determining the reconstruction policy required at that ratio. In particular, most existing designs retain a largely shared data-consistency update and prior transformation throughout the unfolding trajectory.

Two sources of variation should be distinguished, as summarized in Fig. 1. First, the sampling ratio changes the sensing geometry. After vectorization, the measurement model becomes $\mathbf{y}_v = \mathbf{A}_\tau \mathbf{x}_v$, where $\mathbf{A}_\tau = \mathbf{\Psi}_\tau \otimes \mathbf{\Phi}_\tau$. Reducing $\tau$ reduces the measurement-constrained subspace and enlarges the null space of $\mathbf{A}_\tau$, whose components cannot be determined from the measurements. As conceptually illustrated in Fig. 1(a), low-ratio reconstruction therefore requires a greater contribution from prior-driven inference relative to physics-guided correction. Furthermore, the required balance is not fixed during a finite unfolding process. Early stages primarily suppress large ambiguities and structured artifacts, whereas later stages refine residual errors and image details, as depicted in Fig. 1(b). Therefore, the ratio specifies the global sensing regime but does not fully describe the current reconstruction status. A reconstruction operator shared across all ratios and stages cannot explicitly account for these two sources of variation.

Based on this observation, we propose the Sampling-Aware Low-Rank Modulated Transformer (SALR-Former). As previewed in Fig. 1(c), SALR-Former retains a shared unfolding backbone and uses a Ratio-Stage Controller (RSC) to predict a ratio- and trajectory-conditioned update policy. RSC takes the sampling ratio as the global sensing condition and further incorporates the stage index, measurement residual, reconstruction feature, and cross-stage memory to characterize the current reconstruction status. It jointly controls a bounded physics-guided step size and the FiLM-LoRA variables of a multi-domain learned prior. Consequently, SALR-Former adapts both the measurement update and effective prior operator throughout the unfolding trajectory, without maintaining ratio-specific models or adapter banks.

The main contributions are as follows:

- We identify arbitrary-ratio ICS as a finite-stage reconstruction problem with two coupled sources of variation: the sampling ratio changes the sensing geometry, while the unfolding trajectory changes the residual reconstruction demand.
- We propose SALR-Former, a shared unfolding architecture driven by a Ratio-Stage Controller. RSC predicts a physics-update step size and FiLM-LoRA controls for a multi-domain prior from ratio, stage, residual, feature, and cross-stage memory cues, without ratio-specific backbones or adapter banks.
- Extensive experiments on five benchmarks establish the effectiveness of a single shared model across both trained and unseen sampling ratios, and intervention results further connect the predicted controls to reconstruction behavior, showing the distinct roles of sensing ratio and unfolding trajectory.

## 2 Related Works

### 2.1 ICS Networks

DUNs bridge model-based optimization and data-driven learning by unrolling conventional ICS solvers into trainable iterative networks. Recent methods enhance reconstruction through stronger optimization schemes and learned priors. For example, Guo & Gan (2024) unfolds the Chambolle–Pock algorithm with multi-scale and cross-stage fusion, while Li et al. (2024) combines image- and convolutional-coding-domain priors for cross-stage feature propagation. Despite their strong performance (Shen & Gan, 2025; Liao et al., 2025; Chen et al., 2025; Cui et al., 2026; Wang & Gan, 2026), these methods generally require separate models for different sampling rates. To reduce the

cost of ratio-specific models, Shi et al. (2019b) and Hui et al. (2023) support multiple rates through scalable measurements, hierarchical reconstruction, or ratio-conditioned modulation. Their main objective is to make a shared model compatible with different sensing budgets. In contrast, SALR-Former adapts the reconstruction operator to both the sampling ratio and the evolving unfolding state, thereby adjusting the data-consistency update and learned prior during reconstruction.

### 2.2 Low-Rank Adaptation for Image Restoration

LoRA freezes the original model weights and introduces trainable low-rank updates into linear projections, providing an efficient alternative to full fine-tuning (Hu et al., 2022). Owing to its lightweight and modular design, LoRA has recently been explored in image restoration to adapt pre-trained generative or restoration models to different degradation domains (Fang et al., 2026). For example, LoRA-IR (Ai et al., 2024) introduces degradation-aware low-rank experts for all-in-one restoration, and UIR-LoRA (Zhang et al., 2024) employs multiple LoRA adapters to transfer a shared generative prior across various degradation types. In diffusion-based restoration, LoRA is also used to efficiently adapt large text-to-image priors for real-world image recovery, reducing the cost of task-specific fine-tuning while preserving strong perceptual priors (Zhao, 2024). Existing LoRA-based restoration methods mainly select or learn adapters for discrete tasks and degradation levels. SALR-Former instead reuses shared low-rank factors and predicts low-dimensional coefficients online within an unfolding process, allowing the effective prior to vary continuously with the sensing ratio and reconstruction trajectory.

## 3 Proposed Method

### 3.1 Overall Architecture

The overall architecture of SALR-Former is shown in Fig. 2. Given a sampling ratio $\tau$, the bilateral sensing matrices $\mathbf{H}_\tau$ and $\mathbf{W}_\tau$ are obtained by truncating learnable base matrices, and the measurement is generated by

$$\mathbf{y} = \mathbf{H}_\tau \mathbf{x} \mathbf{W}_\tau^T. \tag{2}$$

The initial reconstruction is obtained by transpose back-projection:

$$\mathbf{x}^0 = \mathbf{H}_\tau^T \mathbf{y} \mathbf{W}_\tau. \tag{3}$$

Starting from $\mathbf{x}^0$, SALR-Former performs $K$ unfolding stages. At the $k$-th stage, the Ratio-Stage Controller (RSC) predicts stage-specific control variables, including the data-consistency step size $\rho_k$, LoRA coefficients $\boldsymbol{\alpha}_k$, and FiLM parameters $\boldsymbol{\gamma}_k, \boldsymbol{\beta}_k$. With the predicted step size, the physics-guided update is formulated as

$$\mathbf{z}^k = \mathbf{x}^k - \rho_k \mathcal{G}\left[\mathbf{H}_\tau^T\left(\mathbf{H}_\tau \mathbf{x}^k \mathbf{W}_\tau^T - \mathbf{y}\right)\mathbf{W}_\tau\right], \tag{4}$$

where $\mathcal{G}(\cdot)$ is a lightweight Gradient Refinement Module (GRM). For a raw sensing gradient $\mathbf{g}^k$, GRM applies a residual depth-wise/point-wise transformation,

$$\mathcal{G}(\mathbf{g}^k) = \mathbf{g}^k + \delta\big(\mathrm{PW}(\mathrm{DW}(\mathbf{g}^k))\big), \tag{5}$$

where $\mathrm{DW}(\cdot)$, $\mathrm{PW}(\cdot)$, and $\delta(\cdot)$ denote a $3 \times 3$ depth-wise convolution, a $1 \times 1$ convolution, and GELU, respectively. GRM learns a local correction to the analytical gradient rather than replacing the sensing model. The intermediate estimate $\mathbf{z}^k$ is then refined by a Rate-adaptive Multi-domain Learning Network (RMLN):

$$\mathbf{x}^{k+1}, \mathbf{f}^{k+1} = \mathcal{R}\left(\mathbf{z}^k; \boldsymbol{\alpha}_k, \boldsymbol{\gamma}_k, \boldsymbol{\beta}_k\right), \tag{6}$$

where $\mathcal{R}(\cdot)$ is a hierarchical U-shaped RMLN, and $\mathbf{f}^{k+1}$ denotes the output feature used by the next stage. Its Prior Aggregation Module (PAM) integrates a Context Modeling Block (CMB), a Low-rank Modulated Transformer (LMT), and a Frequency Operator (FO) to jointly capture local structures, long-range dependencies, and frequency-domain priors.

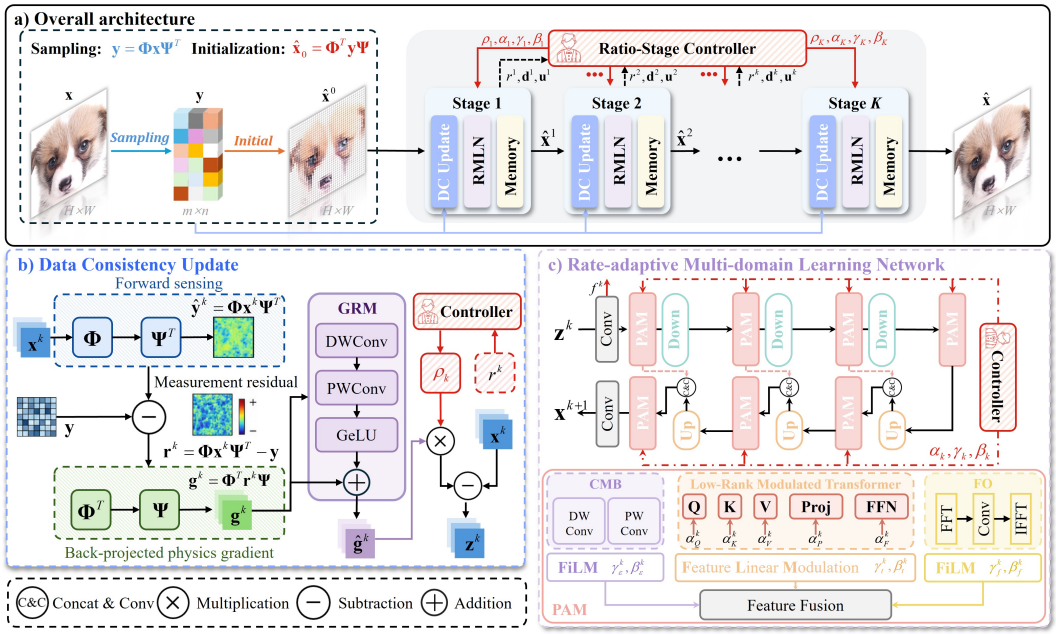



Figure 2: Overall architecture of SALR-Former. Two-sided sampling and transpose back-projection initialize a finite-stage unfolding process. At each stage, RSC conditions the data-consistency step and the shared rate-adaptive multi-domain prior on the sensing ratio and current reconstruction state.

### 3.2 Ratio-Stage Controller

**Why ratio and state should be considered jointly.** The sampling ratio determines the global difficulty of the inverse problem. When $\tau$ decreases, the sensing operator retains fewer independent constraints and admits a larger null space. Errors in this null space are invisible to data consistency and must be resolved by the learned prior. Moreover, the scale of the physical update depends on $\|\mathbf{A}_\tau^T\mathbf{A}_\tau\|_2$, so the update policy should vary with the sensing ratio. A detailed subspace and spectral analysis is provided in Appendix A.1.

The sampling ratio alone, however, does not specify what correction is needed at a particular unfolding stage. Early stages typically suppress large reconstruction errors and artifacts, whereas later stages focus on residual structures and fine details. We therefore use the current measurement residual $r^k$, reconstruction feature $\mathbf{d}^k$, and cross-stage memory $\mathbf{u}^k$ to characterize the evolving reconstruction state. The controller predicts the physical-update parameter and prior-modulation variables from both the global sensing condition and the current state:

$$\rho_k, \boldsymbol{\eta}_k = \Omega(\tau, k, r^k, \mathbf{d}^k, \mathbf{u}^k), \qquad \boldsymbol{\eta}_k = \{\boldsymbol{\alpha}_k, \boldsymbol{\gamma}_k, \boldsymbol{\beta}_k\}. \tag{7}$$

Here, $\rho_k$ controls the physics-guided update, while $\boldsymbol{\eta}_k$ modulates the learned prior.

**Controller instantiation.** As illustrated in Fig. 3, RSC implements $\Omega(\cdot)$ using a shared lightweight MLP. It summarizes measurement inconsistency by

$$r^k = \frac{1}{N}\left\|\mathbf{H}_\tau \mathbf{x}^k \mathbf{W}_\tau^T - \mathbf{y}\right\|_1, \tag{8}$$

where $N$ is the number of measurements. The ratio is encoded by an eight-band Fourier mapping $\mathbf{c}_\tau$, and the stage index is represented by a learned embedding $\mathbf{s}_k$. The current reconstruction and cross-stage history are summarized as

$$\mathbf{d}^k = \mathrm{GAP}(\mathbf{f}^k), \qquad \mathbf{u}^k = \mathrm{GAP}(\mathbf{m}^k), \tag{9}$$

where $\mathbf{f}^k$ and $\mathbf{m}^k$ denote the reconstruction feature and memory state. The controller input is

$$\mathbf{q}^k = [\mathbf{c}_\tau, \mathbf{s}_k, r^k, \mathbf{d}^k, \mathbf{u}^k]. \tag{10}$$

A three-layer MLP with hidden dimensions 96 and 128 maps $\mathbf{q}^k$ to $\rho_k$, five LoRA coefficients $\boldsymbol{\alpha}_k$, and branch-wise FiLM parameters $(\boldsymbol{\gamma}_k, \boldsymbol{\beta}_k)$. The ratio provides the primary global condition because it determines the sensing geometry. The stage embedding and cross-stage memory locate the iterate along the reconstruction trajectory, whereas the residual and feature descriptors provide image-specific refinement cues.

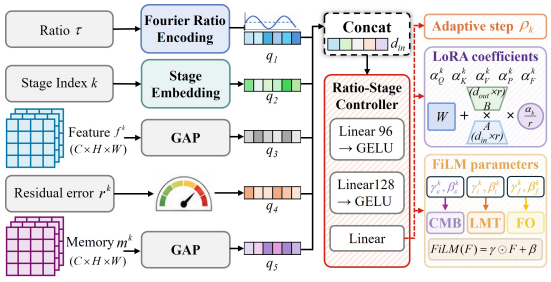



Figure 3: **Ratio-Stage Controller.** A shared MLP predicts controls for the physics-guided update and multi-domain prior from the sensing condition, current reconstruction state, and cross-stage history.

### 3.3 Rate-adaptive Multi-domain Learning Network

After the physics-guided data-consistency update, the intermediate estimate $\mathbf{z}^k$ is refined by the Rate-adaptive Multi-domain Learning Network (RMLN). RMLN acts as a learned image prior that integrates local spatial modeling, window-based global dependency modeling, and frequency-domain enhancement. Conditioned on the modulation variables predicted by RSC, RMLN performs stage- and ratio-adaptive restoration as formulated in equation 6. Its basic PAM consists of three complementary components: CMB, LMT, and FO.

**Context Modeling Block.** CMB is used to capture local structures and fine-grained textures with a lightweight convolutional prior. It adopts a depth-wise separable design, where depth-wise convolutions encode local spatial correlations and point-wise convolutions conduct channel interaction. Given an input feature $\mathbf{F}$, CMB is formulated as

$$\mathbf{F}_c = \mathcal{C}(\mathbf{F}) = \mathrm{PC}\left(\mathrm{DC}\left(\delta\left(\mathrm{PC}\left(\mathrm{DC}(\mathbf{F})\right)\right)\right)\right), \tag{11}$$

where $\mathrm{DC}(\cdot)$ and $\mathrm{PC}(\cdot)$ denote depth-wise and point-wise convolutions, respectively, and $\delta(\cdot)$ is the GELU activation. To adapt the local response to different sampling conditions, FiLM modulation is applied as

$$\widehat{\mathbf{F}}_c = \boldsymbol{\gamma}_\varepsilon^k \odot \mathbf{F}_c + \boldsymbol{\beta}_\varepsilon^k, \tag{12}$$

where $\boldsymbol{\gamma}_\varepsilon^k$ and $\boldsymbol{\beta}_\varepsilon^k$ are generated by RSC. Thus, CMB provides an efficient local prior whose response can be recalibrated according to the current reconstruction state.

**Low-rank Modulated Transformer.** LMT is introduced to model long-range dependencies while preserving the standard Transformer structure. Instead of redesigning the attention block, we inject stage-dependent low-rank updates into the original projection layers, enabling parameter-efficient adaptation across sampling ratios and unfolding stages. Given $\mathbf{F} \in \mathbb{R}^{C\times H\times W}$, we partition it into non-overlapping windows and reshape each window into a token sequence $\mathbf{X} \in \mathbb{R}^{N\times C}$, where $N$ is the number of tokens in each window.

Let $\boldsymbol{\alpha}_k = \{\alpha_q^k, \alpha_k^k, \alpha_v^k, \alpha_o^k, \alpha_{\mathrm{ffn}}^k\}$ denote the LoRA coefficients predicted by RSC for the Transformer branch. For each attention projection, the effective weight is defined as

$$\mathbf{W}_p^k = \mathbf{W}_p + \frac{\alpha_p^k}{r}\mathbf{B}_p\mathbf{A}_p, \quad p \in \{q, k, v, o\}, \tag{13}$$

where $\mathbf{W}_p$ is the base projection, $\mathbf{A}_p \in \mathbb{R}^{r\times C}$ and $\mathbf{B}_p \in \mathbb{R}^{C\times r}$ are learnable low-rank matrices, and $r \ll C$ is the LoRA rank. $\{\mathbf{A}_p, \mathbf{B}_p\}$ are shared across all sampling ratios and unfolding stages. RSC predicts only the corresponding scalar coefficients $\alpha_p^k$. Therefore, the number of adaptation parameters does not grow with the number of supported sampling ratios, unlike maintaining separate backbones or ratio-specific adapter banks. The modulated Query, Key, and Value are computed by

$$\mathbf{Q}^k = \mathbf{X}\mathbf{W}_q^k, \quad \mathbf{K}^k = \mathbf{X}\mathbf{W}_k^k, \quad \mathbf{V}^k = \mathbf{X}\mathbf{W}_v^k. \tag{14}$$

The window attention and output projection are then given by

$$\mathbf{Y}^k = \mathrm{Softmax}\left(\frac{\mathbf{Q}^k(\mathbf{K}^k)^T}{\sqrt{d}}\right)\mathbf{V}^k, \quad \mathbf{O}^k = \mathbf{Y}^k\mathbf{W}_o^k, \tag{15}$$

where $d$ denotes the head dimension. Similarly, the first projection of the feed-forward network is modulated as

$$\mathbf{W}_1^k = \mathbf{W}_1 + \frac{\alpha_{\text{ffn}}^k}{r}\mathbf{B}_1\mathbf{A}_1, \tag{16}$$

and the FFN output is

$$\mathbf{T}^k = \text{FFN}^k(\mathbf{O}^k) = \delta(\mathbf{O}^k\mathbf{W}_1^k)\mathbf{W}_2, \tag{17}$$

where $\mathbf{W}_1$ and $\mathbf{W}_2$ are the expansion and output projections of FFN, respectively. After reversing the window partition, the Transformer response $\mathbf{F}_t^k$ is further recalibrated by FiLM:

$$\widehat{\mathbf{F}}_t^k = \boldsymbol{\gamma}_t^k \odot \mathbf{F}_t^k + \boldsymbol{\beta}_t^k. \tag{18}$$

In this way, LoRA adjusts the effective Transformer parameters, while FiLM modulates the output feature response, allowing LMT to adapt its global modeling behavior with only lightweight additional parameters.

**Frequency Operator.** FO explicitly models spectral information to compensate for high-frequency degradation caused by compressive sampling. Given $\mathbf{F}$, it applies a real-valued FFT and jointly processes the real and imaginary components:

$$\mathbf{S} = \mathcal{F}_r(\mathbf{F}), \qquad [\widetilde{\mathbf{S}}_R, \widetilde{\mathbf{S}}_I] = \mathcal{P}([\text{Re}(\mathbf{S}), \text{Im}(\mathbf{S})]), \tag{19}$$

followed by

$$\mathbf{F}_f = \mathcal{F}_r^{-1}\Big(\widetilde{\mathbf{S}}_R + \text{i}\widetilde{\mathbf{S}}_I; H, W\Big), \tag{20}$$

where $\mathcal{F}_r(\cdot)$ and $\mathcal{F}_r^{-1}(\cdot)$ denote orthonormal `rfft2` and `irfft2`, respectively. $\mathcal{P}(\cdot)$ consists of two $1\times1$ convolutions with a GELU activation and outputs the processed real and imaginary spectra. The recovered frequency feature is further adapted by FiLM:

$$\widehat{\mathbf{F}}_f^k = \boldsymbol{\gamma}_f^k \odot \mathbf{F}_f + \boldsymbol{\beta}_f^k. \tag{21}$$

Thus, FO provides an explicit spectral prior whose response varies with the sampling condition and reconstruction state.

### 3.4 Cross-Stage Memory

RSC should distinguish two iterates that have the same SR and stage index but different reconstruction histories. We therefore maintain a spatial Cross-Stage Memory (CSM) rather than conditioning the controller only on the current feature. The memory is initialized as $\mathbf{m}^0 = \mathbf{0}$. After stage $k$, the feature $\mathbf{f}^{k+1}$ is projected to the memory dimension by $\mathbf{p}^{k+1} = \text{Conv}_{1\times1}(\mathbf{f}^{k+1})$ and incorporated through a convolutional GRU:

$$\begin{aligned}
[\mathbf{z}_m^{k+1}, \mathbf{r}_m^{k+1}] &= \sigma\big(\text{Conv}_{3\times3}([\mathbf{p}^{k+1}, \mathbf{m}^k])\big), \\
\widetilde{\mathbf{m}}^{k+1} &= \tanh\big(\text{Conv}_{3\times3}([\mathbf{p}^{k+1}, \mathbf{r}_m^{k+1} \odot \mathbf{m}^k])\big), \\
\mathbf{m}^{k+1} &= (1 - \mathbf{z}_m^{k+1}) \odot \mathbf{m}^k + \mathbf{z}_m^{k+1} \odot \widetilde{\mathbf{m}}^{k+1}.
\end{aligned} \tag{22}$$

The update and reset gates control how much previous information is retained and how it contributes to the candidate state. RSC pools $\mathbf{m}^{k+1}$ at the next stage through equation 9. CSM thus provides a compact trajectory descriptor while retaining spatial information during its recurrent update.

### 3.5 Loss Function

For each training image $\mathbf{x}$, the network produces the final reconstruction $\mathbf{x}^K$ and intermediate predictions $\{\mathbf{x}^k\}_{k=1}^K$. The final-stage fidelity and physics losses are

$$\mathcal{L}_{\text{rec}} = \|\mathbf{x}^K - \mathbf{x}\|_2^2, \qquad \mathcal{L}_{\text{phy}} = \|\mathcal{A}_\tau(\mathbf{x}^K) - \mathbf{y}\|_2^2. \tag{23}$$

We further supervise the intermediate stages, using linearly increasing normalized stage weights $w_k = k/\sum_{j=1}^K j$. The complete objective is

$$\mathcal{L} = \mathcal{L}_{\text{rec}} + 0.1 \times \mathcal{L}_{\text{phy}} + 0.2 \times \sum_{k=1}^K w_k \|\mathbf{x}^k - \mathbf{x}\|_2^2. \tag{24}$$

Table 1: Quantitative comparison on BSD68 and OST300 across five sampling ratios. Each entry is PSNR/SSIM, and Avg. denotes the arithmetic mean over the five ratios. The best and second-best average results are shown in **red** and **blue**, respectively.

| Method | BSD68 | | | | | | OST300 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.01 | 0.04 | 0.10 | 0.25 | 0.50 | Avg. | 0.01 | 0.04 | 0.10 | 0.25 | 0.50 | Avg. |
| *Fixed-ratio models (independently trained for each sampling ratio)* | | | | | | | | | | | | |
| DGUNet$^+$ (CVPR'22) | 22.55/0.5279 | 25.28/0.5891 | 27.90/0.8119 | 31.80/0.9145 | 36.92/0.9716 | 28.89/0.7630 | 21.59/0.4566 | 24.18/0.6404 | 26.77/0.7844 | 30.62/0.9023 | 35.69/0.9762 | 27.77/0.7520 |
| CSformer (TIP'23) | 22.76/0.5506 | 25.47/0.7071 | 27.77/0.8156 | 31.33/0.9134 | 36.03/0.9691 | 28.67/0.7912 | 21.79/0.4807 | 24.39/0.6636 | 26.72/0.7917 | 30.82/0.9124 | 35.61/0.9697 | 27.87/0.7636 |
| HATNet (CVPR'24) | 23.13/0.5920 | 25.54/0.6887 | 28.93/0.8396 | 32.14/0.9153 | 37.43/0.9714 | 29.43/0.8014 | 21.73/0.5307 | 24.76/0.6553 | 26.94/0.7747 | 30.79/0.9001 | 35.73/0.9639 | 27.99/0.7649 |
| HUNet (CVPR'25) | 22.99/0.5590 | 26.19/0.7102 | 29.19/0.8338 | 33.25/0.9313 | 40.32/0.9749 | **30.39**/0.8018 | 21.74/0.4803 | 23.67/0.6957 | 27.54/0.8251 | 31.01/0.9164 | 36.55/0.9568 | 28.10/0.7749 |
| LGM-Net (TMM'26) | 23.04/0.5561 | 25.70/0.7066 | 28.72/0.8513 | 32.09/0.9185 | 35.34/0.9631 | 28.98/0.7991 | 21.46/0.4927 | 24.37/0.6664 | 26.44/0.8191 | 31.55/0.9180 | 35.08/0.9629 | 27.78/0.7718 |
| RHT-Net (TCSVT'26) | 22.73/0.5328 | 25.62/0.6906 | 28.09/0.8120 | 31.63/0.9116 | 35.42/0.9603 | 28.70/0.7815 | 21.86/0.4619 | 24.40/0.6403 | 26.90/0.7808 | 30.41/0.8986 | 35.02/0.9634 | 27.72/0.7490 |
| ViP-DUN (TMM'26) | 22.87/0.5457 | 25.62/0.7021 | 28.27/0.8197 | 32.14/0.9186 | 37.18/0.9723 | 29.22/0.7917 | 21.86/0.4795 | 24.40/0.6596 | 27.20/0.7968 | 31.05/0.9082 | 36.11/0.9687 | **28.12**/0.7626 |
| SCU-Net (AAAI'26) | 22.65/0.5352 | 25.48/0.6955 | 28.08/0.8145 | 31.86/0.9152 | 36.52/0.9696 | 28.92/0.7860 | 21.73/0.4625 | 24.45/0.6523 | 26.92/0.7869 | 30.60/0.9006 | 35.33/0.9640 | 27.81/0.7533 |
| *Arbitrary-ratio models (one shared model across ratios)* | | | | | | | | | | | | |
| PCNet (TPAMI'24) | 23.14/0.5544 | 26.08/0.7188 | 28.70/0.8380 | 32.74/0.9327 | 38.10/0.9803 | 29.75/**0.8048** | 21.69/0.4742 | 24.30/0.6561 | 26.84/0.7973 | 30.55/0.9113 | 37.39/0.9707 | 28.15/0.7619 |
| ProxUnroll (CVPR'25) | 20.27/0.3793 | 25.59/0.6819 | 28.55/0.8122 | 32.75/0.9167 | 37.70/0.9698 | 28.97/0.7520 | 21.86/0.4651 | 24.39/0.6406 | 26.88/0.7717 | 30.73/0.8943 | 35.72/0.9599 | 27.92/0.7463 |
| SIB-ACS (CVPR'25) | 20.49/0.5454 | 25.84/0.6826 | 29.23/0.8401 | 33.65/0.9312 | 40.37/0.9780 | 29.92/0.7955 | 20.47/0.5307 | 24.24/0.6675 | 27.54/0.8233 | 31.54/0.9213 | 37.68/0.9705 | **28.29**/**0.7827** |
| **SALR-Former (Ours)** | **23.26/0.6065** | **26.34/0.7663** | **29.30/0.8628** | **33.94/0.9406** | **40.47/0.9798** | **30.66/0.8312** | **21.88/0.5396** | **24.46/0.7193** | **27.60/0.8275** | **31.84/0.9220** | **37.73/0.9710** | **28.70/0.7959** |

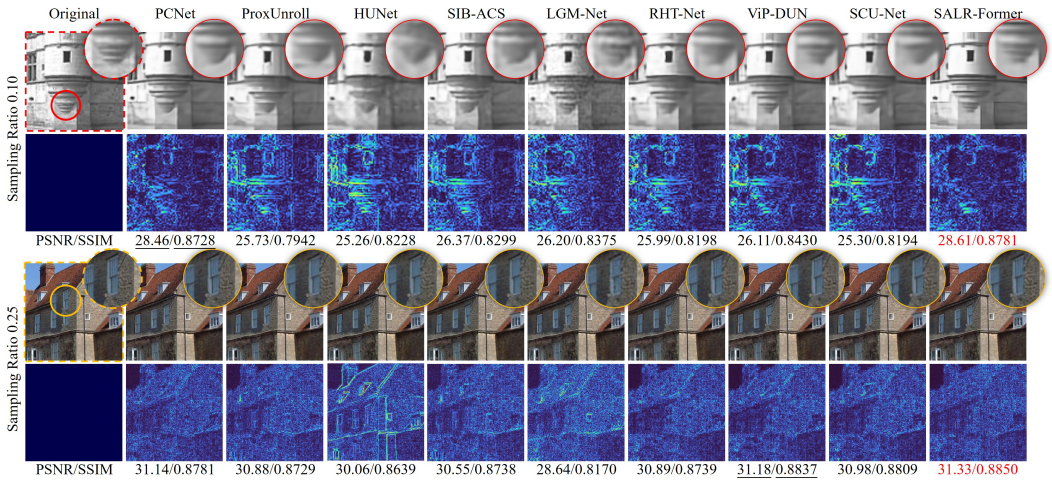



Figure 4: Visual comparisons of our SALR-Former and other competitive methods. From top to bottom, the results are obtained on BSD68 and OST300, respectively.

# 4 Experiments

## 4.1 Implementation Details

Following common ICS settings (Shen et al., 2022; Song et al., 2023; 2026), we use 400 images from BSD500 for training. Images are converted to YCbCr, and sensing and reconstruction are performed on the luminance channel. We repeat the training list 25 times per epoch, yielding 10,000 augmented image instances per epoch, and apply random cropping and horizontal/vertical flipping. Each mini-batch is resized to either $256 \times 256$ or $321 \times 481$. During training, the sampling ratio is randomly selected from $\{0.01, 0.04, 0.10, 0.25, 0.50\}$. Beyond evaluation at these discrete ratios, we further assess and compare the models at continuous unseen sampling ratios during testing. SALR-Former is trained for 300 epochs on NVIDIA RTX 3090 GPUs with a global batch size of 8 using AdamW. The learning rate and weight decay are $2 \times 10^{-4}$ and $10^{-4}$, respectively. The model uses 6 unfolding stages, feature dimension 16, 4 attention heads, window size 8, LoRA rank 4, eight Fourier encoding bands, and memory dimension 24. The shared RSC contains two hidden layers of dimensions 96 and 128.

### 4.2 Comparisons with Other Methods

We conduct performance comparisons on five benchmark datasets, including Set11 (Kulkarni et al., 2016), BSD68 (Martin et al., 2001), General100 (Dong et al., 2016), McM18, and OST300 (Wang et al., 2018), across sampling ratios of 0.01, 0.04, 0.10, 0.25 and 0.50. Eleven representative methods are selected as baselines, including the single-ratio methods DGUNet$^+$ (Mou et al., 2022), CS-former (Ye et al., 2023), HATNet (Qu et al., 2024), HUNet (Shen & Gan, 2025), LGM-Net (Wang et al., 2026), RHT-Net (Song et al., 2026), ViP-DUN (Liao et al., 2026), and SCU-Net (Wang & Gan, 2026); and the arbitrary-ratio methods PCNet (Chen & Zhang, 2024), ProxUnroll (Wang et al., 2025), and SIB-ACS (Tian et al., 2025). For a fair comparison, fixed-ratio baselines are independently trained for each evaluated ratio, whereas arbitrary-ratio methods are trained once. During evaluation, images from Set11 and BSD68 are tested at their original spatial resolutions, while $256 \times 256$ patches are cropped from General100, McM18, and OST300 for testing.

**Comparison protocol.** For a controlled comparison, all methods are trained using the same training data, data augmentation, optimizer, and training budget. Fixed-ratio baselines are trained independently for each evaluated sampling ratio, whereas arbitrary-ratio methods are trained once on the common ratio set and evaluated at every ratio without test-time fine-tuning. Each method is trained for the same prescribed budget, and the final checkpoint is evaluated on the benchmark test sets.

**Comparisons with Arbitrary-ratio Methods.** Table 1 compares SALR-Former with representative arbitrary-ratio ICS methods under the same five sampling ratios. SALR-Former obtains the best average PSNR and SSIM on both BSD68 and OST300. On BSD68, it improves over the best arbitrary-ratio baseline by $0.74$ dB in PSNR and $0.0264$ in SSIM; on OST300, the corresponding gains are $0.41$ dB and $0.0132$. The consistent gains across severe and mild undersampling indicate that a sensing-compatible reconstruction pipeline benefits from explicitly adapting both the physical update and learned prior.

**Comparisons with Fixed-ratio Methods.** Although SALR-Former is trained once over all sampling ratios, it also achieves the best average performance on both datasets. Compared with the best fixed-ratio baseline for each metric, it improves BSD68 by $0.27$ dB in PSNR and $0.0294$ in SSIM, and OST300 by $0.58$ dB and $0.0210$, respectively. These results show that a shared ratio-adaptive model can remain competitive with independently trained ratio-specific networks. As illustrated in Fig. 4, SALR-Former further preserves clearer structural boundaries and facade textures, with fewer structured errors than the displayed single-ratio competitors. *Complexity analysis, detailed results on the remaining datasets, robustness to measurement noise, and reconstruction results at unseen continuous sampling ratios are provided in the Appendix A.2, A.3 and A.4.*

### 4.3 Ablation and Dynamic Adaptation Analysis

**Contribution of adaptive components.** Table 2 disentangles the adaptive components of SALR-Former at sampling ratios 0.20 and 0.40. Removing RSC produces the largest degradation, showing that the controller is central to the complete architecture. Replacing the predicted $\rho_k$ with a fixed step size also causes consistent losses of up to 2.20 dB, while removing CSM weakens the historical information available to RSC. GRM also yields substantial gains by refining the raw sensing gradient before each update, whereas CSM improves all settings by providing the controller with historical stage information. Together, these results show that effective arbitrary-ratio unfolding requires state-dependent gradient correction, update strength, and cross-stage coordination.

**Comparison of modulation strategies.** Table 3 compares different strategies for adapting the reconstruction prior. FiLM-only and LoRA-only retain feature-level and parameter-level modulation, respectively, and both are inferior to their joint use, indicating complementary roles. We also construct a Shared Controller (SC), where one set of prior-modulation variables are shared across all unfolding stages. SC improves over either single mechanism but remains below the complete RSC by 0.34–1.19 dB across the five datasets. This comparison suggests that hybrid modulation benefits from stage- and state-aware control.

**Does the learned control trajectory matter?** To examine whether the controls predicted by RSC are functionally meaningful, we freeze a trained SALR-Former and modify only its control variables during inference. Mean control replaces the stage-wise controls with their within-sample average, Stage shuffle permutes them across stages, Ratio swap applies controls generated at another SR,

Table 2: Component ablation at sampling ratios of 0.20 and 0.40.

| Variant | Set11 | | BSD68 | |
|---|---|---|---|---|
| | 0.20 | 0.40 | 0.20 | 0.40 |
| w/o FiLM | 31.70/0.9468 | 35.04/0.9693 | 30.31/0.9054 | 33.67/0.9555 |
| w/o RSC | 30.79/0.9400 | 33.96/0.9639 | 29.67/0.8964 | 32.91/0.9498 |
| w/o GRM | 31.37/0.9468 | 35.07/0.9678 | 30.40/0.9067 | 33.89/0.9555 |
| w/o LoRA | 32.46/0.9526 | 36.01/0.9731 | 30.73/0.9122 | 34.44/0.9607 |
| w/o CSM | 32.56/0.9536 | 35.92/0.9735 | 30.81/0.9143 | 34.28/0.9611 |
| w/o $\rho$ | 32.41/0.9521 | 35.85/0.9716 | 30.75/0.9124 | 34.38/0.9600 |
| **Complete** | **33.96/0.9579** | **38.05/0.9782** | **31.83/0.9254** | **36.17/0.9688** |

Table 3: Modulation ablation at a sampling ratio of 0.20.

| Dataset | FiLM only | LoRA only | SC | Complete |
|---|---|---|---|---|
| Set11 | 32.39/0.9523 | 32.61/0.9529 | 32.77/0.9553 | **33.96/0.9579** |
| BSD68 | 30.69/0.9121 | 30.81/0.9136 | 31.02/0.9176 | **31.83/0.9254** |
| General100 | 36.12/0.9617 | 36.23/0.9622 | 36.43/0.9641 | **36.99/0.9684** |
| McM18 | 35.70/0.9640 | 35.90/0.9648 | 36.12/0.9669 | **37.05/0.9698** |
| OST300 | 28.76/0.8883 | 28.82/0.8899 | 28.93/0.8933 | **29.27/0.8963** |

Table 4: Post-training RSC-control interventions. Entries are ΔPSNR/SSIM relative to native controls.

| Intervention | General100 | McM18 | OST300 |
|---|---|---|---|
| Native controls | 0/0 | 0/0 | 0/0 |
| Mean control | -1.12/-0.0253 | -0.98/-0.0298 | -0.34/-0.0307 |
| Stage shuffle | -3.25/-0.0499 | -3.02/-0.0625 | -0.93/-0.0506 |
| Ratio swap | -4.68/-0.0930 | -4.68/-0.1078 | -2.00/-0.0907 |
| State masking | -0.10/-0.0009 | -0.03/-0.0007 | -0.02/-0.0007 |

Table 5: Ablation of RSC conditioning variables.

| Variant | $\tau$ | $k$ | $r^k$ | $\mathbf{d}^k$ | $\mathbf{m}^k$ | Set11 | BSD68 |
|---|---|---|---|---|---|---|---|
| Static | – | – | – | – | – | 29.02/0.9184 | 28.06/0.8467 |
| Ratio | ✓ | – | – | – | – | 29.64/0.9269 | 28.36/0.8561 |
| Ratio+Stage | ✓ | ✓ | – | – | – | 29.72/0.9275 | 28.41/0.8574 |
| Ratio+Stage+Memory | ✓ | ✓ | – | – | ✓ | 29.62/0.9263 | 28.42/0.8573 |
| State-aware | ✓ | ✓ | ✓ | ✓ | – | 29.77/0.9276 | 28.45/0.8578 |
| **Full RSC** | ✓ | ✓ | ✓ | ✓ | ✓ | **30.88/0.9325** | **29.30/0.8628** |

and State masking replaces $(r^k, \mathbf{d}^k, \mathbf{u}^k)$ with dataset-level means. As shown in Table 4, Ratio swap causes the largest degradation, indicating that the sensing ratio provides the dominant global condition for the learned controls. Stage shuffle also leads to substantial losses, showing that the control sequence is tied to the unfolding phase rather than being a static ratio-dependent pattern. By contrast, State masking produces relatively small losses under this inference-time intervention, suggesting that the instantaneous residual and feature cues mainly refine a control policy that is already strongly determined by the ratio and stage context.

**Effect of RSC conditioning.** Table 5 studies the inputs to RSC at $\tau = 0.10$, with all variants retrained under the same protocol. The sampling ratio provides the main performance gain over a static controller, while the stage index further accounts for the changing role of each unfolding step. Notably, adding cross-stage memory to the Ratio+Stage controller alone brings little benefit. Memory becomes effective only when paired with the residual and reconstruction-feature descriptors. This result suggests that cross-stage history cannot determine the controls by itself; rather, it complements the current reconstruction state by providing context about the preceding trajectory. The complete RSC therefore uses ratio, stage, current state, and history as complementary cues for stage-wise control. Appendix A.7 further examines the two control targets of RSC and compares them with input-independent stage-wise controls.

## 5 Conclusion

We studied arbitrary-ratio ICS as a finite-stage reconstruction problem in which sensing geometry and reconstruction trajectory impose distinct adaptation requirements. SALR-Former addresses this problem with a shared unfolding backbone and a Ratio-Stage Controller that jointly controls the physics-guided update and FiLM-LoRA-modulated multi-domain prior. Experiments and control interventions show that the sampling ratio determines the global reconstruction regime, while stage and cross-stage history refine the reconstruction policy. The resulting model achieves strong reconstruction quality across trained and unseen continuous sampling ratios without ratio-specific backbones.

## AI USE STATEMENT

We used generative AI tools to assist with manuscript organization and LaTeX formatting. The authors reviewed and verified all AI-assisted text and remain responsible for the technical claims, experimental design, reported results, and final content. No generative AI was used to generate or alter experimental data or results.

### ETHICS STATEMENT

All the authors read and adhere to the ICLR Code of Ethics.

### REPRODUCIBILITY STATEMENT

We ensure the experiments in this paper are reproducible. We provide detailed experimental settings about datasets, training and inference in Section 4.1 and 4.2.

## REFERENCES

Yuang Ai, Huaibo Huang, and Ran He. Lora-ir: Taming low-rank experts for efficient all-in-one image restoration. *arXiv preprint arXiv:2410.15385*, 2024.

Bin Chen and Jian Zhang. Practical compact deep compressed sensing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(3):1610–1626, 2024.

Bin Chen, Zhenyu Zhang, Weiqi Li, Chen Zhao, Jiwen Yu, Shijie Zhao, Jie Chen, and Jian Zhang. Invertible diffusion models for compressed sensing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(5):3992–4006, 2025.

Wenxue Cui, Hualin Li, Yuhang Qin, Yifu Xu, Xiaopeng Fan, and Debin Zhao. Beyond single solution: Multi-hypothesis deep unfolding network for image compressive sensing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5283–5293, 2026.

Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *European conference on computer vision*, pp. 391–407. Springer, 2016.

Marco F Duarte, Mark A Davenport, Dharmpal Takhar, Jason N Laska, Ting Sun, Kevin F Kelly, and Richard G Baraniuk. Single-pixel imaging via compressive sampling. *IEEE signal processing magazine*, 25(2):83–91, 2008.

Chaowei Fang, Bolin Fu, De Cheng, Chengpei Tang, and Guanbin Li. Learning prompt adapters for forgetting-free continual image super-resolution. *IEEE Transactions on Image Processing*, 2026.

Hongping Gan, Zhen Guo, and Feng Liu. Nestd-net: Deep nesta-inspired unfolding network with dual-path deblocking structure for image compressive sensing. *IEEE Transactions on Image Processing*, 33:1923–1937, 2024.

Zhen Guo and Hongping Gan. Cpp-net: Embracing multi-scale feature fusion into deep unfolding cp-ppa network for compressive sensing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 25086–25095, 2024.

Zhen Guo and Hongping Gan. Usb-net: Unfolding split bregman method with multi-phase feature integration for compressive imaging. *IEEE Transactions on Image Processing*, 34:925–938, 2025.

Reinhard Heckel, Mathews Jacob, Akshay Chaudhari, Or Perlman, and Efrat Shimron. Deep learning for accelerated and robust mri reconstruction. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 37(3):335–368, 2024.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022.

Chen Hui, Shengping Zhang, Wenxue Cui, Shaohui Liu, Feng Jiang, and Debin Zhao. Rate-adaptive neural network for image compressive sensing. *IEEE Transactions on Multimedia*, 26:2515–2530, 2023.

Kuldeep Kulkarni, Suhas Lohit, Pavan Turaga, Ronan Kerviche, and Amit Ashok. Reconnet: Non-iterative reconstruction of images from compressively sensed measurements. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 449–458, 2016.

Weiqi Li, Bin Chen, Shuai Liu, Shijie Zhao, Bowen Du, Yongbing Zhang, and Jian Zhang. $D^3c^2$-net: Dual-domain deep convolutional coding network for compressive sensing. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(10):9341–9355, 2024.

Chen Liao, Yan Shen, Dan Li, and Zhongli Wang. Using powerful prior knowledge of diffusion model in deep unfolding networks for image compressive sensing. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 18000–18010, 2025.

Chen Liao, Yan Shen, Zhongli Wang, and Yanbing Li. Adaptive use of convex or non-convex optimization in deep unfolding network for image compressive sensing. *IEEE Transactions on Multimedia*, 28:2850–2864, 2026. doi: 10.1109/TMM.2026.3651069.

Mengzu Liu, Junwei Xu, Weisheng Dong, Le Dong, and Guangming Shi. Learning retinex prior for compressive hyperspectral image reconstruction. *IEEE Transactions on Image Processing*, 35: 1786–1801, 2026.

David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings eighth IEEE international conference on computer vision. ICCV 2001*, volume 2, pp. 416–423. Ieee, 2001.

Chong Mou, Qian Wang, and Jian Zhang. Deep generalized unfolding networks for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17399–17410, 2022.

Gang Qu, Ping Wang, and Xin Yuan. Dual-scale transformer for large-scale single-pixel imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 25327–25337, 2024.

Gang Qu, Ping Wang, Siming Zheng, and Xin Yuan. Breaking measurement barriers: From compressed sensing to deep reconstruction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(11):8631–8639, mar 2026. doi: 10.1609/aaai.v40i11.37815.

Feiyang Shen and Hongping Gan. Hunet: Homotopy unfolding network for image compressive sensing. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 12799–12808, 2025.

Minghe Shen, Hongping Gan, Chao Ning, Yi Hua, and Tao Zhang. Transcs: A transformer-based hybrid architecture for image compressed sensing. *IEEE Transactions on Image Processing*, 31: 6991–7005, 2022. doi: 10.1109/TIP.2022.3217365.

Wuzhen Shi, Feng Jiang, Shaohui Liu, and Debin Zhao. Image compressed sensing using convolutional neural network. *IEEE Transactions on Image Processing*, 29:375–388, 2019a.

Wuzhen Shi, Feng Jiang, Shaohui Liu, and Debin Zhao. Scalable convolutional neural network for image compressed sensing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12290–12299, 2019b.

Heping Song, Jingyao Gong, Hongjie Jia, Xiangjun Shen, Jianping Gou, Hongying Meng, and Le Wang. Representation sampling and hybrid transformer network for image compressed sensing. *IEEE Transactions on Circuits and Systems for Video Technology*, 36(3):3158–3172, 2026. doi: 10.1109/TCSVT.2025.3614371.

Jiechong Song, Bin Chen, and Jian Zhang. Memory-augmented deep unfolding network for compressive sensing. In *Proceedings of the 29th ACM international conference on multimedia*, pp. 4249–4258, 2021.

Jiechong Song, Chong Mou, Shiqi Wang, Siwei Ma, and Jian Zhang. Optimization-inspired cross-attention transformer for compressive sensing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 6174–6184, 2023.

Zhifu Tian, Tao Hu, Chaoyang Niu, Di Wu, and Shu Wang. Sampling innovation-based adaptive compressive sensing. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 2387–2397, 2025.

Chengming Wang, Yan Shen, Chen Liao, Zhongli Wang, Dan Li, and Yanbing Li. Lgm-net: Feature-domain multi-scale network for image compressive sensing based on learned gradient mapping. *IEEE Transactions on Multimedia*, pp. 1–14, 2026. doi: 10.1109/TMM.2026.3680451.

Ping Wang and Xin Yuan. Saunet: Spatial-attention unfolding network for image compressive sensing. In *Proceedings of the 31st ACM International Conference on Multimedia*, pp. 5099–5108, 2023.

Ping Wang, Lishun Wang, Gang Qu, Xiaodong Wang, Yulun Zhang, and Xin Yuan. Proximal algorithm unrolling: Flexible and efficient reconstruction networks for single-pixel imaging. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 411–421, 2025.

Xiaoyang Wang and Hongping Gan. Spectrally adaptive channel-aware unrolling network for compressed sensing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 10190–10198, 2026.

Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 606–615, 2018.

Yan Yang, Jian Sun, Huibin Li, and Zongben Xu. Admm-csnet: A deep learning approach for image compressive sensing. *IEEE transactions on pattern analysis and machine intelligence*, 42 (3):521–538, 2018.

Dongjie Ye, Zhangkai Ni, Hanli Wang, Jian Zhang, Shiqi Wang, and Sam Kwong. Csformer: Bridging convolution and transformer for compressive sensing. *IEEE Transactions on Image Processing*, 32:2827–2842, 2023.

Di You, Jian Zhang, Jingfen Xie, Bin Chen, and Siwei Ma. Coast: Controllable arbitrary-sampling network for compressive sensing. *IEEE Transactions on Image Processing*, 30:6066–6080, 2021.

Cheng Zhang, Dong Gong, Jiumei He, Yu Zhu, Jinqiu Sun, and Yanning Zhang. Uir-lora: Achieving universal image restoration through multiple low-rank adaptation. *arXiv preprint arXiv:2409.20197*, 2024.

Jian Zhang and Bernard Ghanem. Ista-net: Interpretable optimization-inspired deep network for image compressive sensing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1828–1837, 2018.

Kuiyuan Zhang, Zhongyun Hua, Yuanman Li, Yushu Zhang, and Yicong Zhou. Uformer-ics: A u-shaped transformer for image compressive sensing service. *IEEE Transactions on Services Computing*, 17(5):2974–2988, 2023.

Zhonghao Zhang, Yipeng Liu, Jiani Liu, Fei Wen, and Ce Zhu. Amp-net: Denoising-based deep unfolding for compressive image sensing. *IEEE Transactions on Image Processing*, 30:1487–1500, 2020.

Haiyang Zhao. Efficient image restoration through low-rank adaptation and stable diffusion xl. *arXiv preprint arXiv:2408.17060*, 2024.

# A APPENDIX

## A.1 FURTHER ANALYSIS OF RATIO- AND STATE-DEPENDENT RECONSTRUCTION

This section provides a geometric motivation for why arbitrary-ratio ICS should adapt its reconstruction behavior to both the sampling ratio and the current unfolding state.

**Sensing-space decomposition.** The bilateral sensing model can be vectorized as

$$\mathbf{y}_v = \mathbf{A}_\tau \mathbf{x}_v, \qquad \mathbf{A}_\tau = \mathbf{W}_\tau \otimes \mathbf{H}_\tau, \tag{25}$$

where $\mathbf{x}_v = \mathrm{vec}(\mathbf{x})$ and $\otimes$ denotes the Kronecker product. Let $\mathbf{e}^k = \mathbf{x}_v^k - \mathbf{x}_v^\star$ denote the reconstruction error at stage $k$. It can be uniquely decomposed as

$$\mathbf{e}^k = \underbrace{\mathbf{A}_\tau^\dagger \mathbf{A}_\tau \mathbf{e}^k}_{\mathbf{e}^k_{\mathrm{obs}}} + \underbrace{\left(\mathbf{I} - \mathbf{A}_\tau^\dagger \mathbf{A}_\tau\right) \mathbf{e}^k}_{\mathbf{e}^k_{\mathrm{null}}}, \tag{26}$$

where $\mathbf{A}_\tau^\dagger$ is the Moore–Penrose inverse. Both components lie in the image space. The observable component $\mathbf{e}^k_{\mathrm{obs}} \in \mathcal{R}(\mathbf{A}_\tau^T)$ changes the measurements, whereas the null-space component satisfies

$$\mathbf{A}_\tau \mathbf{e}^k_{\mathrm{null}} = \mathbf{0} \tag{27}$$

and is therefore invisible to the sensing system.

For the quadratic data-fidelity term $f_\tau(\mathbf{x}_v) = \frac{1}{2}\|\mathbf{A}_\tau \mathbf{x}_v - \mathbf{y}_v\|_2^2$, the gradient is

$$\nabla f_\tau(\mathbf{x}_v^k) = \mathbf{A}_\tau^T \mathbf{A}_\tau \mathbf{e}^k = \mathbf{A}_\tau^T \mathbf{A}_\tau \mathbf{e}^k_{\mathrm{obs}}. \tag{28}$$

Thus, data consistency can directly correct only the error constrained by the measurements; the null-space component must be inferred from the learned image prior.

Under the nested truncation strategy,

$$\mathrm{rank}(\mathbf{A}_\tau) = \mathrm{rank}(\mathbf{W}_\tau)\,\mathrm{rank}(\mathbf{H}_\tau), \tag{29}$$

$$\dim \mathcal{N}(\mathbf{A}_\tau) = HW - \mathrm{rank}(\mathbf{A}_\tau). \tag{30}$$

Decreasing $\tau$ retains fewer measurements and generally enlarges the null space. Consequently, low-ratio reconstruction places a greater burden on the learned prior, whereas higher ratios allow stronger reliance on the physical measurements.

**Ratio-dependent physical update.** For an unrefined gradient update that produces the intermediate estimate $\mathbf{z}^k$, let $\widetilde{\mathbf{e}}^k = \mathrm{vec}(\mathbf{z}^k) - \mathbf{x}_v^\star$. Expanding the error in the right-singular-vector basis yields

$$\begin{aligned} \widetilde{\mathbf{e}}^k = &\sum_{i=1}^{r_\tau} \left(1 - \rho_k \sigma_{\tau,i}^2\right) a_i^k \mathbf{v}_{\tau,i} \\ &+ \sum_{i=r_\tau+1}^{HW} b_i^k \mathbf{v}_{\tau,i}. \end{aligned} \tag{31}$$

where $\{\mathbf{v}_{\tau,i}\}_{i=1}^{HW}$ denotes a complete orthonormal basis of right singular vectors of $\mathbf{A}_\tau$. For $i \le r_\tau$, $a_i^k = \mathbf{v}_{\tau,i}^T \mathbf{e}^k$ is the coefficient of the observable error component, whereas for $i > r_\tau$, $b_i^k = \mathbf{v}_{\tau,i}^T \mathbf{e}^k$ represents the coefficient in the null space of $\mathbf{A}_\tau$. The observable coefficients are scaled by $1 - \rho_k \sigma_{\tau,i}^2$, whereas the null-space coefficients remain unchanged. Since changing $\tau$ alters both the rank and spectrum of $\mathbf{A}_\tau$, a fixed step size does not necessarily provide an equally effective update across sampling ratios. This analysis motivates ratio-dependent scaling of the physical update, which is further refined by GRM in SALR-Former.

Table A1: Quantitative comparison on Set11, General100, and McM18 across five sampling ratios. Each entry is PSNR/SSIM, and Avg. denotes the arithmetic mean over the five ratios. The best and second-best average results are shown in **red** and **blue**, respectively.

| | | Fixed-ratio models | | | | | | | | Arbitrary-ratio models | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Dataset** | **SR** | DGUNet$^+$ (CVPR'22) | CSformer (TIP'23) | HATNet (CVPR'24) | HUNet (CVPR'25) | LGM-Net (TMM'26) | RHT-Net (TCSVT'26) | ViP-DUN (TMM'26) | SCU-Net (AAAI'26) | PCNet (TPAMI'24) | ProxUnroll (CVPR'25) | SIB-ACS (CVPR'25) | SALR-Former (Ours) |
| Set11 | 0.01 | 20.38/0.5851 | 20.84/0.6347 | 21.35/0.6283 | 21.34/0.6262 | 21.29/0.6485 | 20.82/0.6029 | 21.22/0.6323 | 20.67/0.6034 | 21.30/0.6262 | 21.36/0.6376 | 18.66/0.5844 | 21.40/0.6959 |
| | 0.04 | 24.78/0.7871 | 25.14/0.7871 | 25.22/0.7778 | 26.26/0.8089 | 25.81/0.8159 | 25.56/0.7961 | 25.63/0.8101 | 25.41/0.8038 | 26.58/0.8212 | 26.56/0.8247 | 25.96/0.7933 | 26.69/0.8744 |
| | 0.10 | 28.96/0.8910 | 29.81/0.8943 | 29.47/0.8972 | 30.57/0.9029 | 29.14/0.8762 | 29.01/0.8861 | 29.90/0.9013 | 29.27/0.8947 | 30.45/0.9076 | 30.74/0.9029 | 30.79/0.9111 | 30.88/0.9325 |
| | 0.25 | 33.99/0.9494 | 33.15/0.9471 | 34.24/0.9576 | 35.42/0.9577 | 34.41/0.9516 | 33.58/0.9457 | 34.50/0.9521 | 34.01/0.9493 | 34.86/0.9586 | 35.29/0.9529 | 35.52/0.9567 | 35.77/0.9669 |
| | 0.50 | 39.07/0.9800 | 37.83/0.9780 | 38.61/0.9735 | 41.39/0.9838 | 36.72/0.9706 | 37.86/0.9765 | 39.45/0.9807 | 38.35/0.9769 | 39.79/0.9841 | 40.36/0.9793 | 41.34/0.9830 | 41.43/0.9853 |
| | Avg. | 29.44/0.8385 | 29.35/0.8482 | 29.78/0.8469 | **31.00**/0.8559 | 29.47/0.8526 | 29.37/0.8415 | 30.14/0.8553 | 29.54/0.8456 | 30.60/**0.8595** | 30.86/**0.8595** | 30.45/0.8457 | **31.23/0.8910** |
| General100 | 0.01 | 23.39/0.6022 | 23.94/0.6421 | 24.32/0.6561 | 24.21/0.6319 | 23.96/0.6607 | 24.05/0.6192 | 23.89/0.6301 | 23.86/0.6231 | 24.22/0.6315 | 24.43/0.6397 | 19.86/0.5981 | 24.46/0.6981 |
| | 0.04 | 27.84/0.7888 | 28.25/0.8109 | 28.24/0.7932 | 28.95/0.8135 | 28.94/0.8192 | 28.46/0.7903 | 28.26/0.8076 | 28.47/0.8012 | 29.35/0.8083 | 29.32/0.8147 | 28.69/0.7958 | 29.46/0.8644 |
| | 0.10 | 32.03/0.8962 | 31.93/0.9008 | 31.63/0.8850 | 33.46/0.9116 | 31.81/0.9075 | 32.07/0.8911 | 32.49/0.9089 | 32.49/0.9000 | 33.42/0.9060 | 33.46/0.9053 | 33.50/0.9220 | 33.55/0.9333 |
| | 0.25 | 37.63/0.9625 | 36.59/0.9616 | 37.05/0.9585 | 38.02/0.9674 | 37.83/0.9656 | 36.71/0.9571 | 37.82/0.9675 | 37.80/0.9625 | 38.65/0.9649 | 38.88/0.9643 | 38.87/0.9701 | 39.04/0.9758 |
| | 0.50 | 43.60/0.9884 | 41.87/0.9867 | 42.92/0.9871 | 44.12/0.9893 | 40.50/0.9815 | 41.41/0.9843 | 44.45/0.9893 | 43.13/0.9871 | 44.30/0.9884 | 44.40/0.9879 | 44.76/0.9898 | 44.96/0.9923 |
| | Avg. | 32.90/0.8476 | 32.52/0.8604 | 32.83/0.8560 | 33.75/0.8627 | 32.61/**0.8669** | 32.54/0.8484 | 33.38/0.8607 | 33.15/0.8548 | 33.99/0.8598 | **34.10**/0.8624 | 33.14/0.8552 | **34.29/0.8928** |
| McM18 | 0.01 | 23.31/0.6018 | 23.75/0.6365 | 24.15/0.6510 | 24.20/0.6288 | 24.18/0.6659 | 23.72/0.6113 | 24.21/0.6385 | 23.63/0.6167 | 24.03/0.6263 | 24.22/0.6354 | 20.66/0.6053 | 24.23/0.6906 |
| | 0.04 | 27.48/0.7876 | 27.87/0.8089 | 28.33/0.7908 | 28.23/0.8164 | 28.44/0.8136 | 28.04/0.7894 | 28.76/0.8083 | 28.02/0.8014 | 28.96/0.8102 | 28.90/0.8183 | 28.71/0.7964 | 29.01/0.8663 |
| | 0.10 | 31.73/0.8995 | 31.44/0.9018 | 31.75/0.8851 | 32.42/0.9131 | 32.74/0.9153 | 31.76/0.8933 | 32.92/0.9056 | 32.21/0.9053 | 32.79/0.9110 | 32.71/0.9098 | 33.03/0.9218 | 33.09/0.9373 |
| | 0.25 | 37.26/0.9647 | 36.54/0.9624 | 37.11/0.9581 | 38.29/0.9678 | 37.74/0.9664 | 36.44/0.9595 | 38.27/0.9656 | 37.58/0.9660 | 38.30/0.9675 | 38.44/0.9671 | 38.43/0.9705 | 38.54/0.9779 |
| | 0.50 | 43.35/0.9896 | 42.12/0.9875 | 42.89/0.9867 | 43.68/0.9884 | 41.25/0.9840 | 41.47/0.9855 | 43.90/0.9901 | 43.03/0.9889 | 43.90/0.9893 | 43.95/0.9889 | 44.04/0.9897 | 44.27/0.9930 |
| | Avg. | 32.63/0.8486 | 32.34/0.8594 | 32.85/0.8543 | 33.36/0.8629 | 32.87/**0.8690** | 32.29/0.8478 | 33.61/0.8616 | 32.89/0.8557 | 33.60/0.8609 | **33.64**/0.8639 | 32.97/0.8567 | **33.83/0.8930** |

**Why the reconstruction state is also needed.** The sampling ratio describes the sensing geometry but not the current reconstruction error. Even at the same ratio, early stages mainly remove severe artifacts and recover coarse structures, while later stages operate on refined estimates and focus on residual details. Moreover, images with different contents may reach different reconstruction states at the same stage.

We therefore model the controls as

$$\rho_k, \boldsymbol{\eta}_k = \Omega\left(\tau, k, r^k, \mathbf{d}^k, \mathbf{u}^k\right), \boldsymbol{\eta}_k = \{\boldsymbol{\alpha}_k, \boldsymbol{\gamma}_k, \boldsymbol{\beta}_k\}, \tag{32}$$

where $r^k$ measures the current measurement inconsistency, $\mathbf{d}^k$ summarizes the reconstruction feature, and $\mathbf{u}^k$ records the preceding unfolding history. Notably, $r^k$ reflects only observable inconsistency and cannot directly characterize the null-space error. The feature and memory descriptors therefore provide complementary information about the image content and reconstruction trajectory.

RSC approximates $\Omega(\cdot)$ with a lightweight shared network. The sampling ratio determines what can be constrained by the measurements, while the reconstruction state indicates what remains to be recovered. Their joint use enables SALR-Former to coordinate the physical update and learned prior without introducing independent backbones for different ratios or stages.

### A.2 Additional Quantitative and Visual Comparisons

**Comparisons with Single-ratio Methods.** Table A1 complements the main-paper results with evaluations on Set11, General100, and McM18. SALR-Former attains the best PSNR/SSIM at every evaluated sampling ratio on all three datasets, although it uses a single model for the entire ratio range. On average, it improves over the strongest single-ratio competitors by 0.23 dB/0.0351 on Set11, 0.54 dB/0.0259 on General100, and 0.22 dB/0.0240 on McM18 in PSNR/SSIM, respectively. These results show that dynamic ratio adaptation retains strong reconstruction accuracy across images with diverse structures and textures, without resorting to ratio-specific models.

**Comparisons with Arbitrary-ratio Methods.** SALR-Former also consistently outperforms the arbitrary-ratio baselines with available results. Compared with the strongest competing arbitrary-ratio method, it obtains average gains of 0.37 dB/0.0315 on Set11, 0.19 dB/0.0304 on General100, and 0.19 dB/0.0291 on McM18. The advantage is maintained from severe to mild undersampling, indicating that a unified sensing-compatible model alone is insufficient; the reconstruction operator should further adapt to the sampling condition and the evolving reconstruction state.

**Visual comparisons.** Fig. A1 provides additional visual examples at sampling ratios of 0.01, 0.04, and 0.50. At the extremely low ratio of 0.01, SALR-Former better recovers the radial wing structures and fine boundaries of the butterfly, while several competing methods exhibit stronger blur or structured artifacts. At a ratio of 0.04, it preserves sharper stained-glass boundaries and local textures. At the high-ratio example, SALR-Former recovers finer fur details around the ear and face,

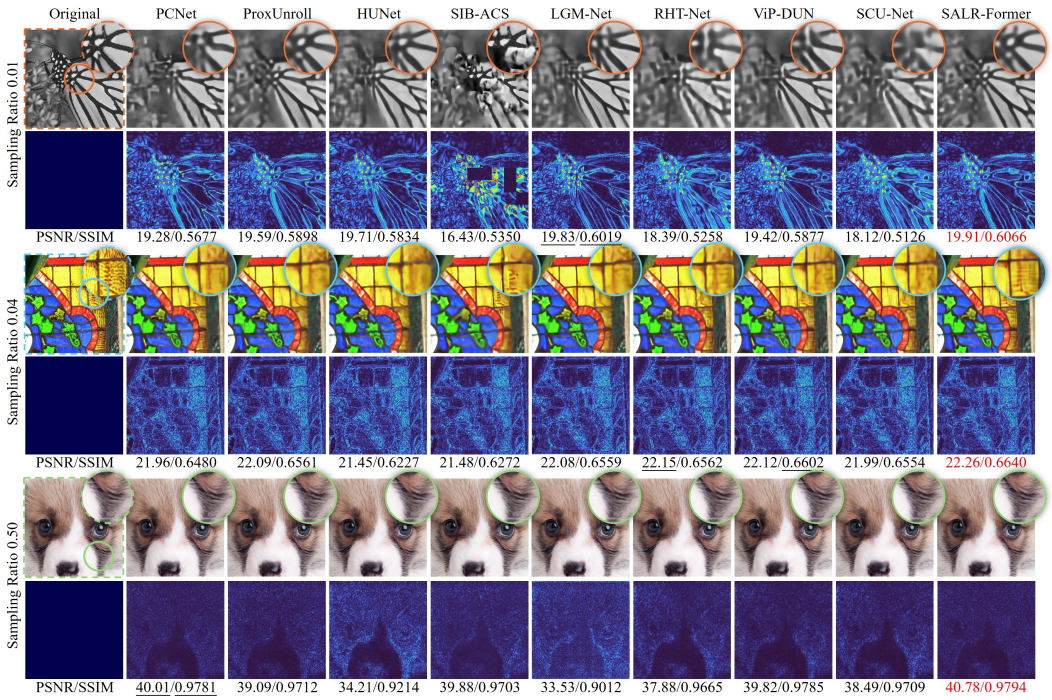



Figure A1: Visual comparisons of our SALR-Former and other competitive methods. From top to bottom, the results are obtained on Set11, McM18 and General100, respectively.

with a less pronounced residual map. These visual observations agree with the quantitative results, showing that the proposed model improves both structural recovery under severe undersampling and detail preservation when more measurements are available.

### A.3 Complexity Analysis

We compare model complexity and inference efficiency at a sampling ratio of 0.01 on BSD68. All models are evaluated in inference mode with a batch size of one on the same NVIDIA RTX 3090. Table A2 shows that SALR-Former achieves the highest reconstruction quality with only 1.00 M parameters and 69.03 GMACs. Compared with the second-lightest SCU-Net, it reduces the parameter count and computation by 17.6% and 67.7%, respectively. Although ViP-DUN and LGM-Net have lower measured latency, they require substantially more parameters and computation while yielding lower reconstruction accuracy. SALR-Former requires 82.59 ms per image, comparable to SCU-Net and SIB-ACS, while providing a more favorable trade-off among reconstruction quality, model size, computational cost, and inference efficiency.

Table A2: Complexity and inference efficiency comparison of SALR-Former with representative methods.

| Method | Params (M) | GMACs | Time (ms) | PSNR (dB) |
|---|---|---|---|---|
| ProxUnroll | 4.56 | 305.96 | 311.67 | 20.27 |
| SIB-ACS | 14.93 | 849.68 | 85.06 | 20.49 |
| LGM-Net | 17.48 | 243.38 | 30.18 | 23.04 |
| ViP-DUN | 14.93 | 240.81 | **21.86** | 22.87 |
| SCU-Net | 1.22 | 213.66 | 82.30 | 22.65 |
| **SALR-Former** | **1.00** | **69.03** | 82.59 | **23.26** |

### A.4 Continuous-ratio Interpolation and Noise Robustness

**Continuous-ratio interpolation.** We further evaluate arbitrary-ratio generalization at continuous sampling ratios that are not used for training. Specifically, for continuous-ratio evaluation, we test at $\{0.02, 0.06, 0.08, \ldots, 0.48\}$, which are excluded from the training ratio set $\{0.01, 0.04, 0.10, 0.25, 0.50\}$. As shown in Fig. A2, all methods improve as more measurements become available; however, SALR-Former maintains a smooth and competitive trajectory across the full ratio range on all five datasets. In particular, its advantage is more apparent at low and in-

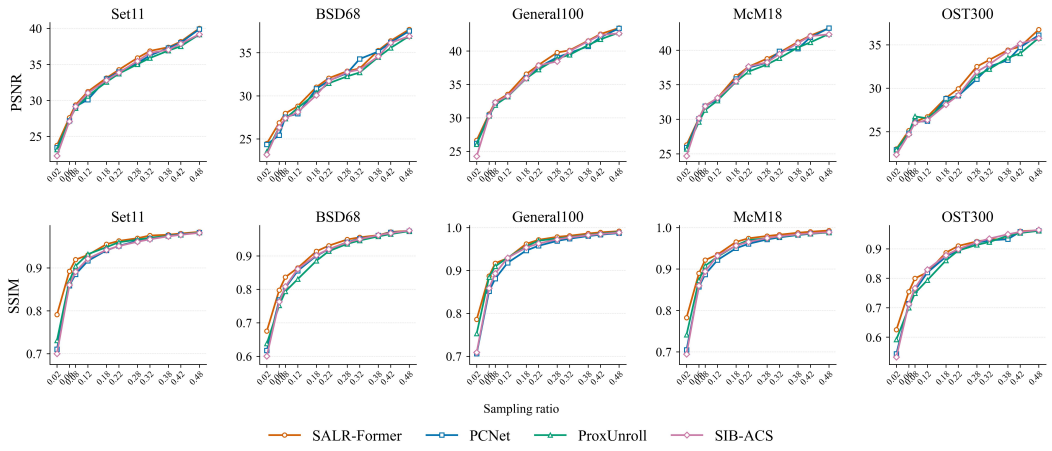



Figure A2: Quantitative comparison with arbitrary-ratio ICS methods at continuous unseen sampling ratios. The top and bottom rows report PSNR and SSIM, respectively, on five benchmark datasets.

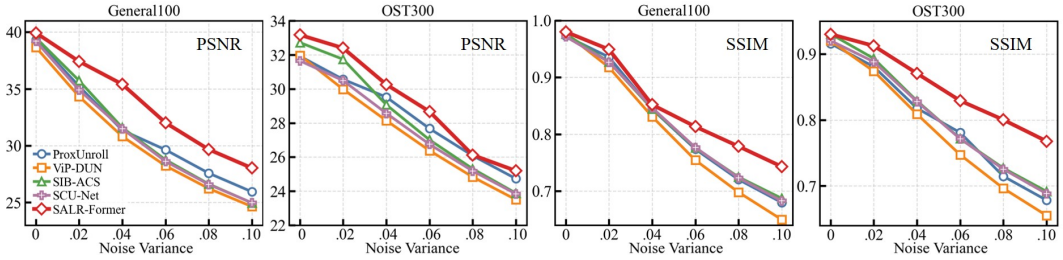



Figure A3: Noise robustness comparison at a sampling ratio of 0.30. PSNR and SSIM are evaluated on General100 and OST300 under increasing measurement-noise variance.

termediate ratios, where the sensing ambiguity is larger and reconstruction requires stronger prior adaptation. These results show that SALR-Former generalizes smoothly to unseen intermediate sampling ratios.

**Noise robustness.** For noise robustness, Gaussian noise is added to the measurements after sensing:

$$\mathbf{y}_\sigma = \mathbf{H}_\tau \mathbf{x} \mathbf{W}_\tau^T + \mathbf{n}, \qquad \mathbf{n} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}), \tag{33}$$

where $\sigma^2 \in \{0, 0.02, 0.04, 0.06, 0.08, 0.10\}$ denotes the measurement-noise variance. All methods are evaluated at $\tau = 0.30$ using the same noisy measurements. Besides, no noise-specific fine-tuning is performed. Fig. A3 compares SALR-Former with representative arbitrary-ratio methods under additive measurement noise at a sampling ratio of 0.30. Although all methods degrade as the noise variance increases, SALR-Former consistently achieves the best PSNR and SSIM on both General100 and OST300. These results show that SALR-Former retains its reconstruction advantage under increasing measurement noise.

### A.5 Role of Multi-Domain Prior Aggregation

Table A3 evaluates different combinations of CMB, LMT, and FO. Among the single-branch variants, LMT generally performs best, highlighting the importance of non-local dependency modeling. Nevertheless, its advantage varies across datasets and remains limited without complementary spatial and spectral cues. Interestingly, simply combining two branches does not always yield monotonic gains, suggesting that the improvement of PAM is not a consequence of increased model capacity alone. Integrating all three branches consistently achieves the best performance. The complete PAM contains only 1.00M parameters and requires 29.42G MACs, providing a favorable trade-off between reconstruction quality and complexity. These results substantiate the complementary roles

Table A3: Ablation of different PAM variants at a sampling ratio of 0.01, where ✓and ✗ denote the presence and absence of each component, respectively.

| Cases | CMB | LMT | FO | General100 | McM18 | OST300 | Params/MACs |
|---|---|---|---|---|---|---|---|
| (a) | ✓ | ✗ | ✗ | 24.22/0.6885 | 23.93/0.6769 | 21.74/0.5306 | 0.89M/19.62G |
| (b) | ✗ | ✓ | ✗ | 24.28/0.6910 | 24.13/0.6843 | 21.78/0.5333 | 0.93M/26.00G |
| (c) | ✗ | ✗ | ✓ | 24.19/0.6871 | 23.91/0.6751 | 21.74/0.5301 | 0.92M/20.29G |
| (d) | ✓ | ✓ | ✗ | 24.21/0.6869 | 23.99/0.6769 | 21.74/0.5296 | 0.95M/27.37G |
| (e) | ✓ | ✗ | ✓ | 24.23/0.6878 | 23.92/0.6754 | 21.74/0.5301 | 0.94M/21.67G |
| (f) | ✗ | ✓ | ✓ | 24.24/0.6889 | 24.06/0.6822 | 21.77/0.5323 | 0.99M/28.04G |
| (g) | ✓ | ✓ | ✓ | **24.46/0.6981** | **24.23/0.6906** | **21.88/0.5396** | 1.00M/29.42G |

Table A4: Ablation study of different LoRA injection positions and rank settings on General100 under different sampling ratios.

| **Cases** | **0.10** | **0.20** | **0.30** | **0.40** | **0.50** |
|---|---|---|---|---|---|
| LoRA-Q | 33.03/0.9297 | 36.18/0.9616 | 39.03/0.9769 | 41.64/0.9810 | 43.47/0.9914 |
| -QK | 33.05/0.9306 | 36.39/0.9627 | 39.20/0.9788 | 41.78/0.9824 | 43.34/0.9913 |
| -QKV | 33.14/0.9315 | 36.34/0.9615 | 39.27/0.9790 | 41.79/0.9821 | 43.53/0.9916 |
| -QKV+P | 32.79/0.9263 | 36.35/0.9631 | 39.00/0.9785 | 41.52/0.9819 | 43.07/0.9910 |
| -QKV+P+F | **33.55/0.9333** | **36.99/0.9684** | **39.94/0.9801** | **41.97/0.9881** | **44.96/0.9923** |
| LoRA-$R_1$ | 32.71/0.9264 | 36.02/0.9610 | 38.77/0.9782 | 39.96/0.9832 | 42.68/0.9908 |
| LoRA-$R_2$ | 33.09/0.9303 | 36.25/0.9594 | 39.23/0.9791 | 40.94/0.9866 | 43.41/0.9915 |
| LoRA-$R_4$ | **33.55/0.9333** | **36.99/0.9684** | **39.94/0.9801** | **41.97/0.9881** | **44.96/0.9923** |
| LoRA-$R_8$ | 33.45/0.9284 | 36.19/0.9637 | 38.52/0.9740 | 40.26/0.9838 | 42.88/0.9911 |

of CMB for local structures, LMT for contextual dependencies, and FO for frequency-detail recovery.

### A.6 Effect of LoRA injection positions and rank

Table A4 studies where to inject dynamic LoRA modulation and how the rank affects reconstruction. Injecting LoRA into Q, K, and V gradually improves the results, whereas adding the output projection alone does not yield further gains. The full "QKV+P+F" configuration performs best across all sampling ratios, indicating that the output and feed-forward projections become beneficial when jointly adapted with the attention projections. For the rank setting, increasing the rank from 1 to 4 consistently improves reconstruction, suggesting that sufficient low-rank capacity is required to represent ratio-dependent adaptations. However, further increasing the rank to 8 does not bring additional gains and even slightly degrades performance. We therefore adopt rank 4 as an empirical balance between modulation capacity and parameter efficiency.

### A.7 Additional Evidence for Dynamic Adaptation

The main paper shows that modifying the predicted controls degrades reconstruction. Here, we further examine whether the gain of RSC can be explained by a fixed stage-wise schedule.

#### A.7.1 Joint Control of Physics Update and Learned Prior

We next examine which outputs of RSC are needed for reconstruction. Table A5 separately enables RSC prediction of the physics-update step size $\rho^k$ and the LoRA–FiLM variables of the learned prior. Predicting either target alone yields limited improvements over the fixed-control baseline. In contrast, jointly predicting both targets consistently gives the best performance across all sampling ratios. This result indicates that adapting the data-consistency update alone is insufficient, and that prior modulation is most effective when it is coordinated with the update strength at each unfolding stage.

#### A.7.2 Static Stage-Wise Controls

We further replace RSC with independently learned controls for each unfolding stage. Specifically, each stage has its own step size, LoRA coefficients, and FiLM parameters, but these parameters

Table A5: Ablation of the RSC control targets on Set11. "Fixed" denotes input-independent learned controls, while "RSC" denotes controls predicted from the RSC inputs.

| Case | $\rho^k$ | Prior modulation | 0.10 | 0.25 | 0.50 |
|---|---|---|---|---|---|
| (a) | Fixed | – | 29.35/0.9229 | 33.70/0.9616 | 38.07/0.9815 |
| (b) | RSC | – | 29.46/0.9238 | 33.87/0.9625 | 38.31/0.9822 |
| (c) | Fixed | RSC | 29.41/0.9245 | 33.83/0.9630 | 38.52/0.9824 |
| (d) | RSC | RSC | **30.88/0.9325** | **35.77/0.9669** | **41.43/0.9853** |

Table A6: Comparison between static stage-wise controls and dynamic RSC at $\tau = 0.25$.

| Control scheme | General100 | McM18 | Set11 | BSD68 | OST300 |
|---|---|---|---|---|---|
| Static stage-wise controls | 37.80/0.9725 | 37.52/0.9751 | 33.94/0.9635 | 32.04/0.9345 | 29.96/0.9141 |
| **Dynamic RSC** | **39.04/0.9758** | **38.54/0.9779** | **35.77/0.9669** | **33.94/0.9406** | **31.84/0.9220** |

remain fixed across input images at the same sampling ratio. Table A6 compares this stage-wise baseline with the complete RSC at $\tau = 0.25$. Although the static variant can learn a different average control pattern for each stage, it consistently underperforms the dynamic RSC on all five datasets. The performance gap shows that stage-dependent parameters alone are insufficient: the reconstruction process also benefits from adapting the controls to the current residual, reconstruction features, and preceding trajectory.

### A.7.3 Stage-wise Reconstruction Trajectories

Fig. A4 empirically characterizes the reconstruction states across sampling ratios and unfolding stages. Lower sampling ratios retain larger image-domain errors throughout the trajectory, consistent with the stronger ambiguity caused by a larger sensing null space. Meanwhile, the measurement residual changes non-monotonically in the early stages and follows markedly different trajectories across ratios. These observations indicate that the relative demand for prior-driven restoration and measurement consistency varies with both the sensing condition and the current iterate.

### A.7.4 Control Visualizations

We further inspect the controls produced by the RSC over sampling ratios and unfolding stages. For each ratio–stage pair, we average the predicted controls over the evaluated images from the five test datasets. Specifically, Fig. A5 reports the mean step size $\rho_k$, the branch-averaged FiLM scale $\gamma_k$, the mean absolute FiLM bias $|\beta_k|$, and the mean absolute LoRA coefficients $|\alpha_k|$. For the vector-valued FiLM variables, the averages are additionally taken over the three prior branches.

As shown in Fig. A5, the predicted controls exhibit clear ratio- and stage-dependent variations. The step size $\rho_k$ changes substantially across sampling ratios, while the FiLM and LoRA variables follow distinct patterns over stages. Moreover, the five LoRA coefficients are not uniformly scaled, indi-

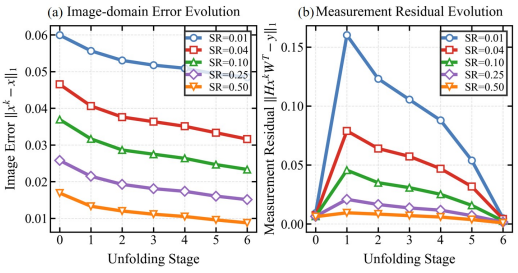



Figure A4: Stage-wise image-domain errors and measurement residuals under different sampling ratios.

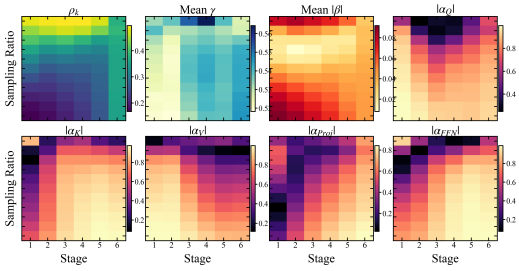



Figure A5: Mean RSC controls across sampling ratios and unfolding stages. $\gamma$ is averaged over the three prior branches, while $|\beta|$ and $|\alpha|$ denote mean absolute values. Each heatmap uses its own color scale.

cating that RSC adapts the attention and feed-forward projections differently. These observations show that the controller learns non-uniform control trajectories for the shared reconstruction backbone. Together with the control-intervention results in the main paper, they support the functional importance of aligning the controls with the sampling condition and unfolding stage.