MAKING EVERY STEP COUNT: SPATIO-TEMPORAL INFORMATION ALLOCATION FOR IMAGING INVERSE PROBLEMS

Anonymous authors
Paper under double-blind review

ABSTRACT

Flow-based generative models have emerged as powerful image priors for training-free inverse problem solving, capturing coherent semantics and fine-grained structure. Despite these strengths, existing flow-based inverse solvers primarily focus on the design of individual updates, largely overlooking spatio-temporal information allocation under a fixed number of function evaluations (NFEs). Temporally, insufficient early exploration can trap the flow trajectory in an incorrect semantic basin, whereas excessive allocation of NFEs to early stages leaves little budget for late-stage refinement. Spatially, data consistency provides direct constraints only within observed regions, whereas the recovery of missing regions relies mainly on the generative prior. To address these two issues, we introduce two complementary and training-free components, i.e., Spectrum-Adaptive Scheduling (SAS) and Measurement-Prioritized Attention (MPA). For temporal allocation, SAS distributes the available NFEs over flow time according to the degradation spectrum and logSNR geometry, thus better balancing semantic exploration and detail refinement. For spatial propagation, MPA exploits data-prior conflicts to guide information toward weakly constrained regions, thereby enhancing semantic and structural fidelity. Extensive experiments on standard image inverse problems, e.g., super-resolution, motion deblurring, and inpainting, demonstrate that the proposed components can be integrated into existing flow-based inverse solvers in a plug-and-play manner without retraining or additional flow-model evaluations, and can also significantly improve the restoration quality of existing solvers.

1 INTRODUCTION

Flow-based generative models Lipman et al. (2023); Liu et al. (2023); Albergo & Vanden-Eijnden (2023), which underpin large-scale text-to-image systems such as Stable Diffusion 3 Esser et al. (2024), have demonstrated remarkable image-generation quality and consequently emerged as powerful priors for ill-posed imaging inverse problems. Accordingly, recent training-free solvers combine a pretrained flow-matching model with a known measurement operator at inference time, avoiding task-specific retraining for super-resolution, deblurring, or inpainting. Within this paradigm, existing methods employ diverse strategies, including posterior guidance, variational optimization, Langevin–proximal refinement, and feasible-set projection Kim et al. (2025); Erbach et al. (2025); Park & Ye (2026); Pourya et al. (2025). Collectively, these approaches have achieved strong restoration performance across a range of inverse problems. Nevertheless, under a finite NFE budget, their designs still center on how each individual update is performed, while much less attention has been paid to their temporal placement along the flow trajectory. The spatial propagation of measurement information is similarly overlooked: data consistency directly constrains only observed regions, leaving missing regions predominantly dependent on the generative prior.

In practice, existing solvers typically adopt a uniform timestep schedule at a fixed NFE budget, treating stages from semantic formation to detail refinement equally. Yet different degradation operators require different balances of these stages. Inpainting provides the clearest example, as illustrated in Figure 1(a). Insufficient early-stage exploration can trap the trajectory in an incorrect semantic basin, causing structures such as sunglass frames to be omitted. Meanwhile, data consistency

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.

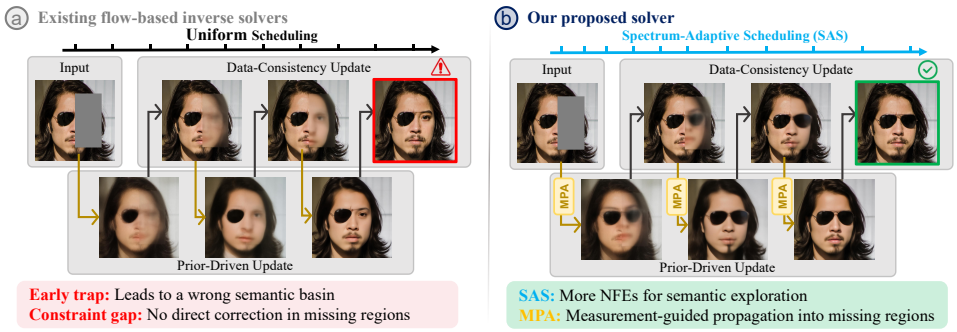


Figure 1: Comparison of existing solvers and our method. (a) Existing solvers adopt uniform scheduling, potentially limiting early exploration, while data consistency constrains only observed regions, leaving missing regions largely dependent on the generative prior. (b) Our method contains two main components, i.e., SAS and MPA. SAS allocates NFEs based on operator demand to balance semantic exploration and detail refinement, while MPA strengthens measurement guidance in weakly constrained regions.

directly constrains only observed pixels and offers no direct correction inside the missing region, which therefore remains predominantly dependent on the generative prior. Other operators also exhibit distinct information deficits. High-factor super-resolution leaves high-frequency details unobserved, whereas motion deblurring preserves coarse semantics but weakly observes many modes. These deficits place different demands on how a finite NFE budget should be allocated, while unobserved or weakly observed details receive little direct correction from data consistency.

These observations suggest that flow inversion under a finite NFE budget is not only an update-design problem, but also a problem of temporal budget allocation and spatial information propagation. *Temporally*, the goal is to allocate the NFE budget based on degradation-specific demand, balancing semantic exploration and detail refinement; *spatially*, it is to strengthen measurement guidance for recovery in weakly constrained regions. To address these challenges, we introduce Spectrum-Adaptive Scheduling (SAS) for NFE budget allocation and Measurement-Prioritized Attention (MPA) for spatial information propagation. Figure 1(b) illustrates the complementary design, which improves restoration quality without retraining or additional flow-model evaluations.

SAS determines how a finite NFE budget should be distributed along flow time according to the degradation spectrum and logSNR geometry. The spectrum distinguishes missing modes that rely more on prior-driven synthesis from attenuated modes that retain partial measurement support. LogSNR characterizes how the demands for semantic exploration and detail refinement change along flow time. Based on these two factors, SAS defines an operator demand and constructs an operator-specific schedule, placing more timesteps in high-demand portions of the trajectory while keeping the total NFE budget unchanged. Meanwhile, MPA directly modulates image-token self-attention within the DiT backbone to strengthen measurement guidance in weakly constrained regions. This design builds on the observation that the prior–data discrepancy highlights structures poorly captured by the current prior. MPA summarizes this discrepancy as a conflict heatmap and transforms it through conflict-guided bias generation into an image-token attention bias, which reshapes the attention weights to improve structural fidelity in weakly constrained regions.

Our contributions are as follows:

- We identify a shared bottleneck in training-free flow inverse solvers and formulate it in terms of temporal allocation of a finite NFE budget and spatial information propagation.
- We introduce SAS and MPA to address the temporal and spatial dimensions, respectively. SAS distributes the NFE budget over flow time according to the degradation spectrum and logSNR geometry, while MPA uses prior–data conflicts to strengthen measurement guidance in weakly constrained regions.
- Both components can be integrated into existing solvers without retraining or additional flow-model evaluations. Experiments on super-resolution, motion deblurring, and inpainting demonstrate consistent improvements in restoration quality and better preservation of instance-specific semantic and structural details.

2 Background and Related Work

2.1 Flow Priors for Inverse Problems

We consider an imaging inverse problem

$$y = \mathcal{A}x + \nu, \tag{1}$$

where $\mathcal{A}$ is the degradation operator and ν is measurement noise. Because $\mathcal{A}$ is generally non-invertible or ill-conditioned, recovery requires a natural-image prior in addition to data fidelity. Diffusion-based methods have established pretrained generative models as broadly applicable priors: RePaint propagates observed-region constraints through resampling, DPS injects likelihood gradients into reverse diffusion, DDRM and DDNM exploit the spectrum or null space of a linear operator, and DAPS decouples noise annealing to make early errors easier to correct Lugmayr et al. (2022); Chung et al. (2023); Kawar et al. (2022); Wang et al. (2023); Zhang et al. (2024).

With recent advances in flow matching, pretrained flow models have also been adopted as generative priors for inverse problems. A common affine probability path is

$$x_t = a_t x_0 + b_t \epsilon, \qquad \epsilon \sim \mathcal{N}(0, I), \tag{2}$$

whose signal-to-noise state is commonly parameterized by the log signal-to-noise ratio (logSNR) Kingma et al. (2021):

$$\ell(t) = \log \frac{a_t^2}{b_t^2}. \tag{3}$$

For the linear path used in this work, $a_t = 1 - t, b_t = t$. A velocity prediction $v_\theta(x_t, t)$ then induces

$$\hat{x}_{0|t} = x_t - t v_\theta(x_t, t), \qquad \hat{\epsilon}_{|t} = x_t + (1 - t) v_\theta(x_t, t). \tag{4}$$

These endpoint estimates provide convenient interfaces for likelihood, projection, or proximal corrections while retaining the pretrained flow trajectory. FlowDPS applies posterior guidance through these endpoint predictions Kim et al. (2025); FLAIR alternates a variational flow regularizer with hard data consistency Erbach et al. (2025); FlowLPS combines Langevin exploration with proximal refinement Park & Ye (2026); and FLOWER alternates flow-consistent denoising with measurement-aware projection Pourya et al. (2025). Despite their different formulations, these solvers all combine prior-driven and data-consistency updates at a sequence of flow timesteps. Under a finite inference budget, however, they largely overlook where these updates should be placed along flow time and how measurement guidance can be strengthened in weakly constrained regions.

2.2 Timestep Scheduling

Sampling schedules determine how a limited evaluation budget is distributed along a generative trajectory. EDM shows that noise parameterization and timestep placement should be designed together, with logSNR providing a natural coordinate for the progression from noise-dominant synthesis to clean-dominant refinement Karras et al. (2022). More recent training-free methods design the grid itself: TORS allocates timesteps according to the geometry of the sampling trajectory Zhou et al. (2026), while Lipschitz-guided schedules use numerical regularity to resolve difficult intervals Chen et al. (2025). These schedules are designed primarily around the generative trajectory or its numerical properties and do not account for the degradation-specific demands of inverse problems. SAS instead combines the degradation spectrum with logSNR geometry to construct an operator-specific schedule that allocates the fixed timestep budget according to restoration demand.

2.3 Attention-Based Spatial Guidance

Attention control provides complementary spatial guidance. Prompt-to-Prompt and Attend-and-Excite manipulate text-conditioned attention, while MasaCtrl and Plug-and-Play Diffusion Features reuse reference attention or intermediate features Hertz et al. (2023); Chefer et al. (2023); Cao et al. (2023); Tumanyan et al. (2023). ControlNet and T2I-Adapter instead learn additional conditioning modules Zhang et al. (2023); Mou et al. (2024). These methods rely on prompts, reference features, or learned conditioning and may require extra optimization, feature injection, or task-specific modules. Existing inverse solvers, meanwhile, use measurements primarily for data-fidelity correction,

and rarely use them to guide image-token attention within the generative backbone. MPA instead converts prior–data conflicts into an image-token attention bias to strengthen measurement guidance in weakly constrained regions, without additional flow-model evaluations. Together, SAS and MPA make more effective use of a finite inference budget through operator-specific temporal allocation and stronger measurement guidance in weakly constrained regions.

3 METHODOLOGY

3.1 OVERVIEW

This section details how SAS and MPA address temporal budget allocation and spatial information propagation, respectively, under a finite NFE budget. Figure 2 illustrates both components: (a) SAS determines *when* inverse updates are executed by distributing timesteps over flow time according to the degradation spectrum and logSNR geometry, while (b) MPA determines *where* measurement guidance should be strengthened by converting prior–data conflicts into an image-token attention bias within the DiT backbone.

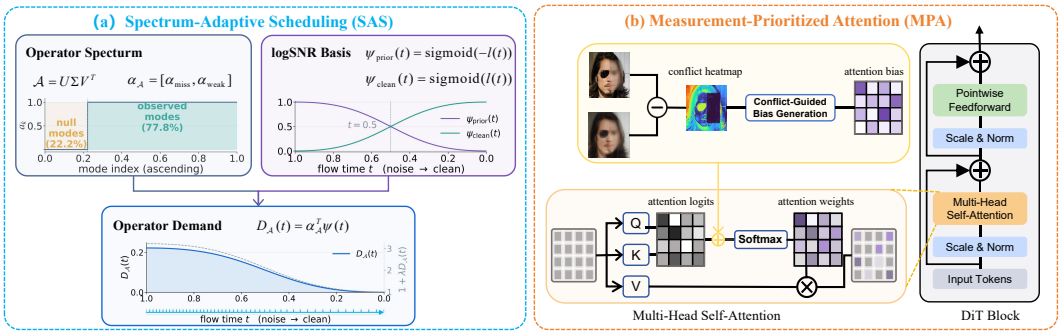


Figure 2: Overview of Spectrum-Adaptive Scheduling (SAS) and Measurement-Prioritized Attention (MPA). (a) SAS combines the operator spectrum and logSNR-derived temporal bases into an operator demand, which is used to distribute the finite NFE budget over flow time. (b) MPA summarizes prior–data conflicts as a conflict heatmap and converts it into an image-token attention bias within the DiT backbone, strengthening measurement guidance in weakly constrained regions.

3.2 SPECTRUM-ADAPTIVE SCHEDULING

SAS constructs an operator-specific schedule by quantifying missing and attenuated spectral modes, mapping their demands over flow time through logSNR, and allocating the fixed timestep budget by equal-mass quantiles.

Operator spectrum. For measurements with $y = \mathcal{A}x_0 + \nu$, the standard quadratic data-consistency objective is

$$\mathcal{L}_{\mathrm{dc}}(x) = \frac{1}{2}\|\mathcal{A}x - y\|_2^2, \tag{5}$$

whose Hessian is

$$\nabla_x^2 \mathcal{L}_{\mathrm{dc}}(x) = \mathcal{A}^\top \mathcal{A}. \tag{6}$$

Hence, for any unit signal direction v, the quadratic form $v^\top \mathcal{A}^\top \mathcal{A} v = \|\mathcal{A}v\|_2^2$ measures how strongly it is constrained by the measurement. Moreover, under isotropic Gaussian noise, $\mathcal{A}^\top \mathcal{A}$ is also proportional to the Fisher information matrix. However, these directions are generally coupled in image coordinates. Following DDRM Kawar et al. (2022), we therefore decouple them via the SVD:

$$\mathcal{A} = U\Sigma V^\top, \qquad \Sigma = \mathrm{diag}(a_1, \ldots, a_d), \tag{7}$$

where U and V provide orthonormal bases for the measurement and image spaces, respectively, and Σ is rectangular diagonal. Let $r = \mathrm{rank}(\mathcal{A})$, with $a_1 \geq \cdots \geq a_r > 0 = a_{r+1} = \cdots = a_d$. Then

$$\mathcal{A}^\top \mathcal{A} = V\,\mathrm{diag}(a_1^2, \ldots, a_d^2)V^\top. \tag{8}$$

Thus, v_k is an independent signal direction and a_k^2 is the measurement strength along that direction. We normalize a_k as $\bar{a}_k = a_k/a_{\max}$, where $a_{\max} = \|\mathcal{A}\|_2$, and obtain

$$G_{\mathcal{A}} := \frac{\mathcal{A}^\top \mathcal{A}}{\|\mathcal{A}\|_2^2} = V \,\mathrm{diag}(\bar{a}_1^2, \ldots, \bar{a}_d^2) V^\top. \tag{9}$$

By construction, each $\bar{a}_k^2 \in [0, 1]$ measures how strongly the k-th direction is observed relative to the strongest direction.

The ordinary rank r counts all nonzero singular directions equally, including those that are severely attenuated. We instead use the stable rank Tropp (2015); Cohen et al. (2016):

$$\mathrm{srank}(\mathcal{A}) := \frac{\|\mathcal{A}\|_F^2}{\|\mathcal{A}\|_2^2} = \mathrm{tr}(G_{\mathcal{A}}) = \sum_{k=1}^{d} \bar{a}_k^2. \tag{10}$$

Stable rank is a scale-invariant effective dimension: a direction contributes according to its relative measurement strength rather than a binary zero/nonzero decision. For every nonzero operator,

$$1 \le \mathrm{srank}(\mathcal{A}) \le r \le d, \tag{11}$$

Relative to full observation, the normalized stable-rank gap separates as

$$\frac{d - \mathrm{srank}(\mathcal{A})}{d} = \underbrace{\frac{d - r}{d}}_{\alpha_{\mathrm{miss}}} + \underbrace{\frac{r - \mathrm{srank}(\mathcal{A})}{d}}_{\alpha_{\mathrm{weak}}}, \tag{12}$$

where α_{miss} is the fraction of unobservable directions, while α_{weak} aggregates attenuation over the observable directions.

LogSNR temporal bases. The spectral coefficients above summarize the operator's missing and attenuated modes. To translate this information into temporal allocation, we first characterize how the role of an update changes along flow time. At the noise-dominant early stage, restoration relies more strongly on the prior for semantic synthesis; toward the clean endpoint, measurement-supported details can be refined more effectively. We describe this transition using the logSNR $\ell(t) = \log(a_t^2/b_t^2)$ Kingma & Gao (2023), which provides a natural coordinate for the progression from the noise-dominant stage to the clean endpoint.

We map logSNR to two complementary temporal bases:

$$\begin{aligned} \psi_{\mathrm{prior}}(t) &= \mathrm{sigmoid}(-\ell(t)) = \frac{b_t^2}{a_t^2 + b_t^2}, \\ \psi_{\mathrm{clean}}(t) &= \mathrm{sigmoid}(\ell(t)) = \frac{a_t^2}{a_t^2 + b_t^2} \end{aligned} \tag{13}$$

For the linear flow path $a_t = 1 - t$ and $b_t = t$, they reduce to

$$\begin{aligned} \psi_{\mathrm{prior}}(t) &= \frac{t^2}{t^2 + (1-t)^2}, \\ \psi_{\mathrm{clean}}(t) &= \frac{(1-t)^2}{t^2 + (1-t)^2}. \end{aligned} \tag{14}$$

Here, the two bases are nonnegative and satisfy $\psi_{\mathrm{prior}}(t) + \psi_{\mathrm{clean}}(t) = 1$: ψ_{prior} emphasizes early prior-driven exploration, whereas ψ_{clean} shifts emphasis toward clean-stage detail refinement.

From operator demand to an adaptive schedule. With the degradation spectrum summarized by the operator-spectrum weights and the restoration demand along flow time characterized by the temporal bases, we define the operator demand as

$$D_{\mathcal{A}}(t) := \alpha_{\mathrm{miss}} \psi_{\mathrm{prior}}(t) + \alpha_{\mathrm{weak}} \psi_{\mathrm{clean}}(t). \tag{15}$$

The pairing reflects the different measurement support of missing and attenuated modes. Missing modes receive no direct measurement correction and hence predominantly rely on prior-driven synthesis, whereas attenuated modes retain partial measurement support and benefit more from clean-stage refinement. The resulting $D_{\mathcal{A}}(t)$ summarizes the operator-specific restoration demand along the flow time and serves as the temporal weighting profile for constructing the adaptive schedule.

Let $\mathcal{T} = [t_{\min}, t_{\max}]$. We convert this demand directly into a normalized allocation density:

$$q_{\mathcal{A},\lambda}(t) = \frac{1 + \lambda D_{\mathcal{A}}(t)}{\int_{t_{\min}}^{t_{\max}} [1 + \lambda D_{\mathcal{A}}(u)]\, du}, \qquad \lambda \geq 0, \tag{16}$$

where λ controls the deviation from uniform allocation, with $\lambda = 0$ recovering the uniform schedule. Given N solver intervals, we take reverse-time equal-mass quantiles of $q_{\mathcal{A},\lambda}$:

$$\begin{aligned} F_{\mathcal{A},\lambda}(t) &:= \int_{t_{\min}}^{t} q_{\mathcal{A},\lambda}(u)\, du, \\ t_i &= F_{\mathcal{A},\lambda}^{-1}\left(1 - \frac{i}{N}\right), \quad i = 0, \ldots, N. \end{aligned} \tag{17}$$

Thus, $t_0 = t_{\max}$ and $t_N = t_{\min}$, while portions of the trajectory with larger operator demand receive denser solver times. Overall, SAS converts the degradation spectrum and logSNR geometry into an operator-specific schedule, allowing the fixed timestep budget to be allocated more effectively between semantic exploration and detail refinement.

3.3 Measurement-Prioritized Attention

MPA strengthens measurement guidance in weakly constrained regions by directly modulating image-token self-attention within the DiT backbone. It first extracts spatial conflicts from the prior–data discrepancy and then converts them into an attention bias that reshapes information flow between image tokens.

Let z_t^{pri} and z_t^{dc} denote the latent estimates produced by the prior and data-consistency updates at step t, respectively. Their token-wise discrepancy measures the correction introduced by data consistency relative to the current prior estimate and we use its magnitude and summarize it as the conflict heatmap

$$c_t = \mathrm{NormPool}\left(\left|z_t^{\mathrm{dc}} - z_t^{\mathrm{pri}}\right|\right) \in [0,1]^n, \tag{18}$$

where $\mathrm{NormPool}$ performs channel pooling, robust normalization, and spatial smoothing.

Suppressing the attention-head index for clarity, let $Q_t, K_t \in \mathbb{R}^{n \times d_h}$ denote the image-token queries and keys, where d_h is the feature dimension of each head. The original attention logits are

$$L_t = \frac{Q_t K_t^\top}{\sqrt{d_h}} \in \mathbb{R}^{n \times n}. \tag{19}$$

The rows and columns of L_t correspond to query and key tokens respectively. These logits capture feature compatibility between image tokens but do not explicitly account for prior–data conflicts.

To match the pairwise attention logits, MPA broadcasts the token-wise map along the query dimension, giving the attention bias

$$B_t = \mathbf{1} c_t^\top \in \mathbb{R}^{n \times n}, \tag{20}$$

where $\mathbf{1} \in \mathbb{R}^n$ broadcasts c_t across all image-token queries, assigning larger key-side biases to locations with stronger measurement responses. MPA then adds B_t to the original image-token attention logits L_t before softmax:

$$\widetilde{P}_t = \mathrm{Softmax}_{\mathrm{row}}\left(L_t + \beta_t B_t\right), \tag{21}$$

where β_t controls the bias strength. The resulting attention weights place greater emphasis on measurement-responsive tokens, thereby strengthening measurement guidance in weakly constrained regions.

Table 1: Quantitative results on FFHQ (1k) and DIV2K (0.8k) at 768×768 resolution. All methods use the same number of function evaluations. "+ Ours" augments the corresponding solver with SAS and MPA.

Method	SR 8×				SR 12×				Motion Deblur				Inpainting			
	PSNR↑	SSIM↑	FID↓	LPIPS↓	PSNR↑	SSIM↑	FID↓	LPIPS↓	PSNR↑	SSIM↑	FID↓	LPIPS↓	PSNR↑	SSIM↑	FID↓	LPIPS↓
FFHQ 1k																
FlowChef	27.33	0.755	41.93	0.340	26.13	0.726	58.99	0.376	27.41	0.756	36.99	0.339	18.96	0.769	65.30	0.407
FlowDPS	27.52	0.702	29.13	0.419	26.62	0.691	31.54	0.453	25.93	0.681	40.58	0.485	19.98	0.761	40.84	0.336
FlowLPS	24.20	0.424	53.05	0.609	23.53	0.405	65.59	0.660	33.18	0.871	31.01	0.247	21.92	0.863	17.31	0.215
FlowLPS + Ours	**26.67**	**0.611**	**36.03**	**0.429**	**26.09**	**0.627**	**39.281**	**0.462**	33.10	**0.872**	**29.47**	0.251	**24.55**	**0.868**	20.86	0.230
FLAIR	29.41	0.796	83.66	0.468	26.68	0.712	58.75	0.381	32.04	0.830	11.26	0.158	23.76	0.849	14.63	0.170
FLAIR + Ours	**29.61**	**0.803**	85.50	**0.449**	**26.81**	**0.712**	**52.65**	**0.346**	**32.10**	0.829	**10.80**	**0.155**	**24.49**	**0.852**	15.28	**0.169**
DIV2K 0.8k																
FlowChef	21.25	0.532	55.81	0.508	20.15	0.490	65.42	0.543	21.38	0.537	55.82	0.506	19.91	0.632	64.82	0.511
FlowDPS	21.93	0.526	43.99	0.482	20.94	0.490	53.91	0.549	20.74	0.493	60.98	0.566	21.39	0.668	41.97	0.333
FlowLPS	20.01	0.382	53.95	0.526	19.40	0.351	68.06	0.585	27.02	0.765	26.96	0.310	23.36	0.840	20.94	0.194
FlowLPS + Ours	19.76	**0.384**	54.45	**0.558**	**19.42**	**0.379**	**65.76**	**0.598**	**27.04**	**0.770**	**25.90**	**0.302**	**24.19**	0.829	**20.86**	0.213
FLAIR	23.44	0.616	58.64	0.585	21.19	0.512	62.10	0.507	26.74	0.743	16.54	0.215	23.83	0.834	17.18	0.164
FLAIR + Ours	**23.51**	**0.622**	59.17	**0.573**	**21.32**	**0.517**	**60.36**	**0.493**	**26.87**	**0.745**	**15.89**	**0.207**	**24.07**	0.833	**16.61**	0.167

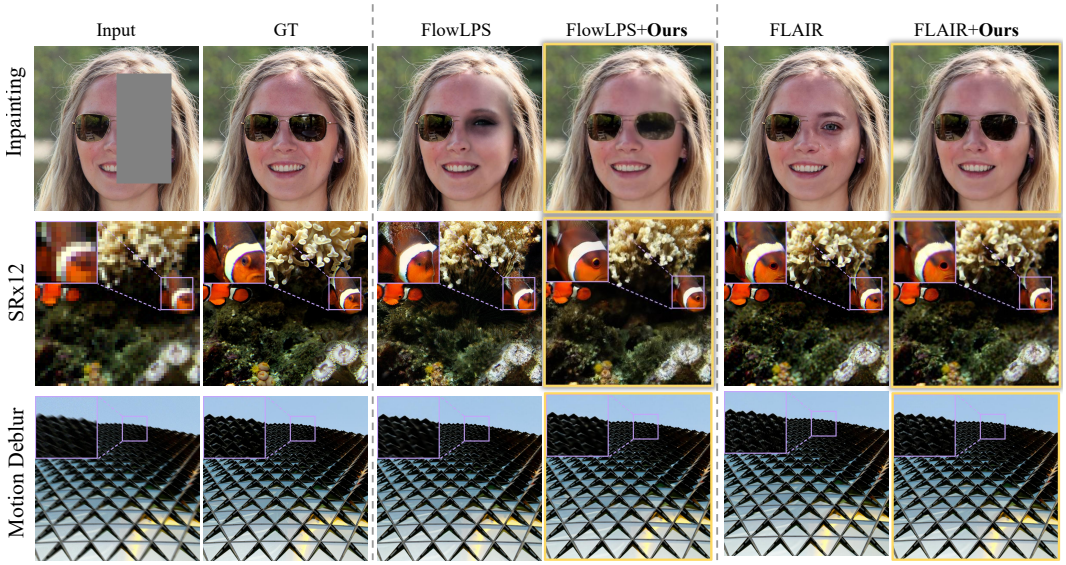


Figure 3: Qualitative comparison across inpainting, $\times 12$ super-resolution, and motion deblurring. Adding our components to FlowLPS or FLAIR improves semantic and structural fidelity while maintaining consistency with the observed content. Best viewed zoomed in.

4 Experiments

4.1 Setup

Datasets and Evaluation Metrics. We conducted comprehensive evaluations across multiple inverse problems using high-resolution images from two datasets: 1k images from FFHQ Karras et al. (2019) and 0.8k images from DIV2K Agustsson & Timofte (2017). All experiments were performed at 768×768 resolution by resizing the original images. Our evaluation employs PSNR and SSIM to quantify pixel-level reconstruction fidelity, complemented by FID and LPIPS for perceptual quality and naturalness.

Baselines. We compare with FlowDPS Kim et al. (2025), FlowChef Patel et al. (2025), FLAIR Erbach et al. (2025), and FlowLPS Park & Ye (2026). The FlowDPS and FlowChef results follow the SD3.0-Medium protocol reported by FlowDPS, whereas FLAIR and FlowLPS use their SD3.5-Medium implementations. Our main controlled comparisons integrate SAS and MPA into FLAIR and FlowLPS, denoted by "+ Ours"; each augmented solver retains the inference settings and number of flow-model evaluations of its corresponding baseline.

Problem Setting. We run and evaluate all methods at a fixed output resolution of 768×768 pixels. For single image super-resolution, we consider scaling factors of $8\times$ and $12\times$. The corresponding low-resolution inputs are generated by bicubic downsampling. Motion blur is simulated with a blur kernel of size 61. For box inpainting, we mask large, continuous rectangles that cover approximately one third of the observation. All synthesized observations are corrupted with additive Gaussian noise, with standard deviation $\sigma_v = 0.3\%$.

Following previous work on flow-based inverse problems Erbach et al. (2025); Park & Ye (2026), we use the fixed prompt "A high quality photo of a face" for FFHQ. For DIV2K, we concatenate "A high quality photo of" with an image-specific description extracted from the observation using DAPE Wu et al. (2024).

4.2 Main Results

Table 1 reports the quantitative results. The gains are most consistent for super-resolution and motion deblurring. Adding our modules to FLAIR improves most metrics for $\times 12$ super-resolution and motion deblurring on both datasets. FlowLPS also obtains clear and consistent improvements on FFHQ super-resolution.

For inpainting, the clearest quantitative gains occur in fidelity: PSNR improves for both augmented solvers. Figure 3 further shows the qualitative advantage. The baselines often fail to recover instance-specific attributes, including sunglasses and facial markings, whereas our method preserves these structures consistently. The baselines use the observed content mainly through data-consistency correction, leaving its structural and semantic cues underused within the generative backbone; their reconstructions can therefore drift toward a generic semantic basin favored by the prior. Under the same NFE budget, SAS provides finer temporal resolution at the noise end for early semantic exploration, while MPA uses measurement-supported cues to guide spatial feature aggregation toward weakly constrained regions. Our method also recovers finer textures in super-resolution and leaves less residual blur in motion deblurring.

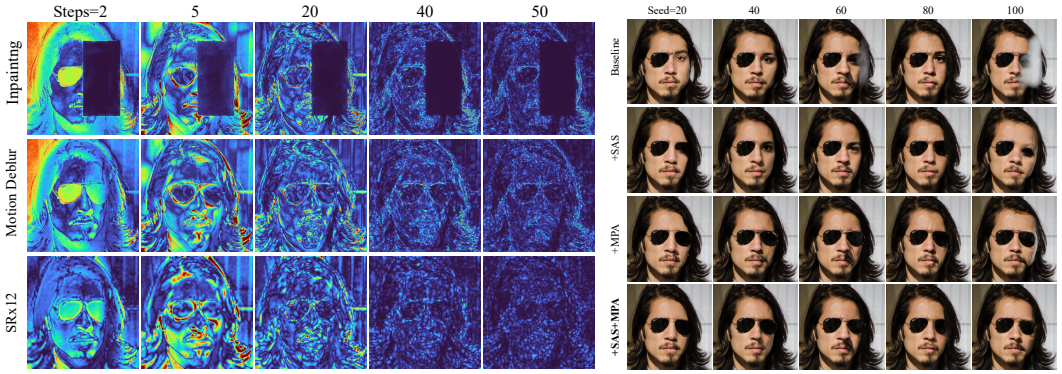
4.3 Ablation Study

Component Ablation. We evaluate the contribution of SAS and MPA in a representative controlled experimental setting based on FlowLPS for FFHQ $\times 12$ super-resolution. All variants use the same 100-image subset and NFE budget. Table 2(a) reports the results when SAS and MPA are enabled individually and jointly. SAS contributes most of the quantitative gain and improves all four metrics over the base schedule. MPA alone produces smaller changes in these aggregate metrics, consistent with its main role in preserving semantic and structural details.

Figure 4(a) visualizes how the prior–data conflicts progressively evolve across solver steps. At early stages, strong responses already delineate major structural contours and highlight high-frequency details that the prior tends to suppress. Toward the clean endpoint, the conflicts diminish and become increasingly concentrated on fine textures. These patterns provide clear spatial evidence for the subsequent attention modulation.

Although SAS provides the largest improvements on aggregate metrics, its semantic recovery remains sensitive to sampling randomness. Allocating more steps to the noise end provides additional opportunities for early semantic exploration, but does not guarantee that the trajectory enters the correct semantic basin. Figure 4(b) examines this limitation on the same challenging box-inpainting case across different random seeds. The base solver consistently converges to a generic face favored by the generative prior, while SAS recovers the complete sunglasses only for some seeds. Changing the seed can still lead to different semantic outcomes, showing that additional early exploration alone does not ensure stable semantic recovery.

MPA complements this temporal allocation by acting as a semantic stabilizer. By strengthening measurement guidance in weakly constrained regions, MPA allows the missing region to make better use of measurement-supported structure from the observed content. As shown in Figure 4(b), MPA consistently recovers the complete sunglasses across seeds. This stronger structural constraint may slightly reduce visual quality in some cases, but it produces markedly more stable semantic recovery, which is not fully reflected by aggregate metrics. Combining SAS and MPA retains this semantic stability while improving visual quality over MPA alone.



(a) Conflict maps (b) Seed stability

Figure 4: Data-consistency conflicts and box-inpainting results across random seeds. (a) Data-consistency conflict heatmaps over the restoration trajectory for inpainting, motion deblurring, and $\times 12$ super-resolution. Brighter values indicate larger corrections between the prior and data-consistent estimates. (b) Box-inpainting results across random seeds. Columns correspond to different seeds, and rows show the Baseline, +SAS, +MPA, and +SAS+MPA settings.

Table 2: Component and schedule ablations for FlowLPS. (a) Ablation study for $\times 12$ super-resolution on FFHQ (100 images). SAS and MPA are individually switched on or off. (b) Schedule ablation for FlowLPS on FFHQ $8\times$ super-resolution using 100 images and 50 NFEs.

(a) Component ablation

SAS	MPA	PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓
×	×	24.34	0.435	0.595	87.11
✓	×	**26.33**	0.574	0.471	**72.94**
×	✓	24.41	0.445	0.585	86.27
✓	✓	26.19	**0.576**	**0.465**	73.38

(b) Schedule ablation

Schedule	PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓
Uniform	26.39	0.573	0.463	73.07
Noise-end dense	26.83	0.616	0.426	68.87
Mid-trajectory dense	23.68	0.378	0.695	95.57
Signal-end dense	24.03	0.396	0.662	90.69
SAS (Ours)	**26.89**	**0.620**	**0.420**	**67.99**

Schedule Ablation. To evaluate whether the NFE distribution produced by SAS reflects operator-specific demand over flow time, we conduct a controlled schedule ablation with FlowLPS on FFHQ $8\times$ super-resolution. The experiment uses a 100-image subset and the same 50-NFE budget for all schedules. All other solver and evaluation settings remain fixed across schedules. We compare SAS with a uniform schedule and three hand-crafted alternatives. Specifically, for the hand-crafted alternatives, we divide the trajectory into three equal segments and assign a larger share of the NFE budget to the noise end, middle, or clean end, respectively.

Table 2(b) shows that SAS performs best on all four metrics, while the noise-end-dense schedule consistently ranks second. This result confirms that allocating more steps to early semantic exploration is beneficial, but also shows that concentrating timesteps in a manually selected segment is insufficient. By distributing the same NFE budget according to operator-specific demand, SAS better balances early semantic exploration and late-stage detail refinement.

5 Conclusion

Existing training-free flow-based inverse solvers largely overlook temporal budget allocation and spatial information propagation under a finite inference budget. To address these limitations, we introduced Spectrum-Adaptive Scheduling (SAS) and Measurement-Prioritized Attention (MPA). SAS combines operator-spectrum weights with logSNR-derived temporal bases to allocate the fixed timestep budget more effectively between semantic exploration and detail refinement. MPA uses prior–data conflicts to bias image-token attention in the DiT backbone, strengthening measurement guidance in weakly constrained regions. Both components integrate into existing solvers without retraining or additional inference steps. Experiments across super-resolution, motion deblurring, and inpainting show consistent improvements in restoration quality and preservation of instance-specific semantic and structural details. These results highlight the value of spatio-temporal information allocation under finite inference budgets.

AI USE STATEMENT

We used generative AI tools, including GPT-6, to polish language, draft selected passages, search literature, format references and LaTeX, and organize the manuscript. The authors reviewed all AI-assisted content and take full responsibility for the final paper.

ETHICS STATEMENT

We use public FFHQ and DIV2K data without new human-subject experiments or personal-data collection. FFHQ raises privacy and representation concerns; restored facial details are not verified attributes of depicted people.

REPRODUCIBILITY STATEMENT

The Methodology, Experiments, and Appendix Sections D and E detail the algorithms, data, solver settings, prompts, seeds, NFE budgets, and evaluation protocols. We will provide code and configurations as supplementary materials for reproduction.

REFERENCES

Eirikur Agustsson and Radu Timofte. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 126–135, 2017.

Michael S. Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *International Conference on Learning Representations*, 2023.

Mingdeng Cao, Xintao Wang, Zhongang Qi, Ying Shan, Xiaohu Qie, and Yinqiang Zheng. Masactrl: Tuning-free mutual self-attention control for consistent image generation and editing. In *IEEE/CVF International Conference on Computer Vision*, 2023.

Hila Chefer, Yuval Alaluf, Yael Vinker, Lior Wolf, and Daniel Cohen-Or. Attend-and-excite: Attention-based semantic guidance for text-to-image diffusion models. In *ACM SIGGRAPH*, 2023.

Yifan Chen, Eric Vanden-Eijnden, and Jiawei Xu. Lipschitz-guided design of interpolation schedules in generative models. *arXiv preprint arXiv:2509.01629*, 2025.

Hyungjin Chung, Jeongsol Kim, Michael Thompson McCann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations*, 2023.

Michael B. Cohen, Jelani Nelson, and David P. Woodruff. Optimal approximate matrix product in terms of stable rank. In *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*, volume 55 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 11:1–11:14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016. doi: 10.4230/LIPIcs.ICALP.2016.11.

Julius Erbach, Dominik Narnhofer, Andreas Dombos, Bernt Schiele, Jan Eric Lenssen, and Konrad Schindler. Solving inverse problems with FLAIR. *arXiv preprint arXiv:2506.02680*, 2025.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, Dustin Podell, Tim Dockhorn, Zion English, Kyle Lacey, Alex Goodwin, Yannik Marek, and Robin Rombach. Scaling rectified flow transformers for high-resolution image synthesis. In *International Conference on Machine Learning*, 2024.

Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Prompt-to-prompt image editing with cross attention control. In *International Conference on Learning Representations*, 2023.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4401–4410, 2019.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Jeongsol Kim, Bryan Sangwoo Kim, and Jong Chul Ye. FlowDPS: Flow-driven posterior sampling for inverse problems. *arXiv preprint arXiv:2503.08136*, 2025.

Diederik Kingma and Ruiqi Gao. Understanding diffusion objectives as the elbo with simple data augmentation. *Advances in Neural Information Processing Systems*, 36:65484–65516, 2023.

Diederik P. Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations*, 2023.

Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11461–11471, 2022.

Chong Mou, Xintao Wang, Liangbin Xie, Yanze Wu, Jian Zhang, Zhongang Qi, and Ying Shan. T2i-adapter: Learning adapters to dig out more controllable ability for text-to-image diffusion models. In *AAAI Conference on Artificial Intelligence*, 2024.

Jonghyun Park and Jong Chul Ye. FlowLPS: Langevin-proximal sampling for flow-based inverse problem solvers. *arXiv preprint arXiv:2512.07150*, 2026.

Maitreya Patel, Song Wen, Dimitris N. Metaxas, and Yezhou Yang. FlowChef: Steering of rectified flow models for controlled generations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 15308–15318, 2025.

Mehrsa Pourya, Bassam El Rawas, and Michael Unser. Flower: A flow-matching solver for inverse problems. *arXiv preprint arXiv:2509.26287*, 2025.

Joel A. Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends in Machine Learning*, 8(1–2):1–230, 2015. doi: 10.1561/2200000048.

Narek Tumanyan, Michal Geyer, Shai Bagon, and Tali Dekel. Plug-and-play diffusion features for text-driven image-to-image translation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *International Conference on Learning Representations*, 2023.

Rongyuan Wu, Tao Yang, Lingchen Sun, Zhengqiang Zhang, Shuai Li, and Lei Zhang. Seesr: Towards semantics-aware real-world image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 25456–25467, 2024.

Bingliang Zhang, Wenda Chu, Julius Berner, Chenlin Meng, Anima Anandkumar, and Yang Song. Improving diffusion inverse problem solving with decoupled noise annealing. *arXiv preprint arXiv:2407.01521*, 2024.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *IEEE/CVF International Conference on Computer Vision*, 2023.

Zhenyu Zhou, Defang Chen, Siwei Lyu, Chun Chen, and Can Wang. Analyzing and improving fast sampling of text-to-image diffusion models. *arXiv preprint arXiv:2603.00763*, 2026.

A Overview

This appendix provides concise analysis and implementation details for SAS and MPA. Sections B and C relate operator spectra and trajectory sensitivity to temporal allocation and analyze conflict-guided attention, respectively. Sections D and E document solver settings and additional experiments.

B Additional Analysis of Spectrum-Adaptive Scheduling

B.1 Operator spectrum and data-consistency deficit

Let $\mathcal{A} \in \mathbb{R}^{m\times d}$ have singular values $a_1 \geq \cdots \geq a_r > 0 = a_{r+1} = \cdots = a_d$ and rank r. With $\bar{a}_k = a_k/a_1$ (and $\bar{a}_k = 0$ for $k > r$), define

$$G_{\mathcal{A}} = \frac{\mathcal{A}^\top \mathcal{A}}{\|\mathcal{A}\|_2^2} = V \operatorname{diag}(\bar{a}_1^2, \ldots, \bar{a}_d^2) V^\top. \tag{22}$$

SAS uses

$$\begin{aligned} \alpha_{\text{miss}} = &\frac{d-r}{d}, \quad \alpha_{\text{weak}} = \frac{r - \operatorname{srank}(\mathcal{A})}{d}, \\ &\operatorname{srank}(\mathcal{A}) = \sum_{k=1}^{r} \bar{a}_k^2. \end{aligned} \tag{23}$$

Proposition 1 (Exact decomposition of measurement deficit)**.** *The SAS spectral weights are nonnegative, invariant to any nonzero scalar rescaling of $\mathcal{A}$, and satisfy*

$$\alpha_{\text{miss}} + \alpha_{\text{weak}} = \frac{1}{d} \operatorname{tr}(I - G_{\mathcal{A}}) = \frac{1}{d} \sum_{k=1}^{d} (1 - \bar{a}_k^2). \tag{24}$$

For isotropic u with $\mathbb{E}[uu^\top] = I/d$,

$$\mathbb{E}\big[u^\top (I - G_{\mathcal{A}}) u\big] = \alpha_{\text{miss}} + \alpha_{\text{weak}}. \tag{25}$$

Proof. Substituting the two definitions yields the sum in Equation 24. The isotropic identity follows from $\mathbb{E}[u^\top M u] = \operatorname{tr}(M)/d$. Normalization by a_1 establishes scale invariance and nonnegativity. □

For $f(x) = \frac{1}{2}\|\mathcal{A}x - \mathcal{A}x^\star\|_2^2$, a gradient step of size $\|\mathcal{A}\|_2^{-2}$ gives

$$e^+ = (I - G_{\mathcal{A}})e, \qquad e_k^+ = (1 - \bar{a}_k^2) e_k \quad \text{in the right-singular basis.} \tag{26}$$

Missing modes receive no correction, whereas attenuated modes contract slowly; α_{miss} and α_{weak} separate these effects.

B.2 LogSNR bases as endpoint sensitivities

Consider the affine flow path

$$x_t = a_t x_0 + b_t \epsilon, \qquad v_t = \dot{a}_t x_0 + \dot{b}_t \epsilon, \tag{27}$$

with $\Delta_t = a_t \dot{b}_t - \dot{a}_t b_t \neq 0$.

Proposition 2 (Velocity-error sensitivity of endpoint estimates)**.** *Perturbing the predicted velocity by δv_t while holding x_t fixed gives endpoint errors*

$$\delta \widehat{x}_{0|t} = -\frac{b_t}{\Delta_t} \delta v_t, \qquad \delta \widehat{\epsilon}_{|t} = \frac{a_t}{\Delta_t} \delta v_t. \tag{28}$$

whose normalized squared sensitivities are

$$\frac{b_t^2}{a_t^2 + b_t^2} = \operatorname{sigmoid}[-\ell(t)], \qquad \frac{a_t^2}{a_t^2 + b_t^2} = \operatorname{sigmoid}[\ell(t)], \tag{29}$$

where $\ell(t) = \log(a_t^2/b_t^2)$.

Proof. Invert Equation 27; normalization cancels the shared factor Δ_t^{-2}. □

For the linear path, define the normalized sensitivities $\psi_{\text{prior}}(t) = t^2/[t^2+(1-t)^2]$ and $\psi_{\text{clean}}(t) = (1-t)^2/[t^2+(1-t)^2]$. Coupling them to the spectral deficits gives

$$D_{\mathcal{A}}(t) = \alpha_{\text{miss}}\psi_{\text{prior}}(t) + \alpha_{\text{weak}}\psi_{\text{clean}}(t). \tag{30}$$

Missing modes therefore shift demand toward the noise endpoint, whereas weak modes shift it toward the clean endpoint.

B.3 Trajectory error and allocation density

Consider $\dot{z}(t) = f(t, z(t))$ and variable-step Euler times $t_{\min} = t_0 < \cdots < t_N = t_{\max}$, with $h_i = t_{i+1} - t_i$.

Proposition 3 (Variable-step trajectory-error bound)**.** *Assume $f(t,\cdot)$ is L-Lipschitz and $\|\ddot{z}(t)\|_2 \le \kappa_i$ on $[t_i, t_{i+1}]$. Starting from the exact initial state, Euler satisfies*

$$\|e_N\|_2 \le \frac{1}{2}\exp[L(t_{\max} - t_{\min})] \sum_{i=0}^{N-1} \kappa_i h_i^2. \tag{31}$$

Proof. Taylor's theorem gives local error at most $\kappa_i h_i^2/2$. Unrolling the Lipschitz recursion and using $1 + Lh_i \le e^{Lh_i}$ proves the claim. □

For a normalized allocation density $q(t) > 0$, locally $h(t) \simeq [Nq(t)]^{-1}$, and Equation 31 approaches

$$\frac{1}{N}\int_{t_{\min}}^{t_{\max}} \frac{\kappa(t)}{q(t)}\, dt. \tag{32}$$

For positive continuous $\kappa(t)$, Cauchy–Schwarz gives the unique normalized density minimizing this surrogate:

$$q^\star(t) = \frac{\sqrt{\kappa(t)}}{\int_{t_{\min}}^{t_{\max}} \sqrt{\kappa(u)}\, du}. \tag{33}$$

SAS uses

$$q_{\mathcal{A},\lambda}(t) = \frac{1 + \lambda D_{\mathcal{A}}(t)}{\int_{t_{\min}}^{t_{\max}} [1 + \lambda D_{\mathcal{A}}(u)]\, du}. \tag{34}$$

Modeling $\sqrt{\kappa_{\mathcal{A}}(t)}$ as $C[1 + \lambda D_{\mathcal{A}}(t)]$ makes SAS optimal for the bound in Proposition 3. Here $D_{\mathcal{A}}$ is an operator-aware proxy rather than the exact local error of every host solver.

B.4 Numerical schedule construction

We evaluate $q_{\mathcal{A},\lambda}$ on an 8192-point grid. For $K = 50$, the cumulative distribution and reverse-time global quantiles are

$$\begin{aligned} F_{\mathcal{A},\lambda}(t) &= \int_{0.18}^{t} q_{\mathcal{A},\lambda}(u)\, du, \\ t_j &= F_{\mathcal{A},\lambda}^{-1}\left(1 - \frac{j}{K+1}\right), \quad j = 1, \ldots, K. \end{aligned} \tag{35}$$

The endpoints are not evaluations, and $\lambda = 0$ recovers the uniform schedule.

C Additional Analysis of Measurement-Prioritized Attention

C.1 Conflict score used by the implementation

At solver step s, let $\delta_s = z_s^{\text{dc}} - z_s^{\text{pri}}$ and $h_s(p) = C^{-1}\sum_c |\delta_{s,c}(p)|$. With threshold τ, scale $v_{\max}$, and exponent γ,

$$\bar{c}_s(p) = m_{\text{known}}(p)\left[\text{clip}\left(\frac{h_s(p) - \tau}{v_{\max} - \tau}, 0, 1\right)\right]^{\gamma}. \tag{36}$$

Average pooling gives the final gate

$$c_s = m_{\text{known}} \odot \text{clip}\left(\text{AvgPool}_k(\bar{c}_s), 0, 1\right). \tag{37}$$

The resulting measurement-responsive gate is area-interpolated to each token grid.

C.2 Gated outer-product attention bias

MPA uses $B_s = g_s c_s^\top$, with $g_s = \mathbf{1}$ for super-resolution and deblurring and $g_s = 1 - m_{\text{known}}$ for inpainting. Appending $\sqrt{\beta_s \sqrt{d_h}} g_s$ and $\sqrt{\beta_s \sqrt{d_h}} c_s$ to the queries and keys gives

$$\frac{Q'_s K'^\top_s}{\sqrt{d_h}} = \frac{Q_s K_s^\top}{\sqrt{d_h}} + \beta_s g_s c_s^\top. \tag{38}$$

without materializing B_s or disabling fused attention.

C.3 MPA as a query-gated exponential tilt

For original attention probabilities $p_{ij} = \text{softmax}_j(L_{ij})$, Equation 38 gives

$$\begin{aligned} \widetilde{p}_{ij}(\beta) &= \frac{p_{ij} \exp(\beta g_i c_j)}{Z_i(\beta)}, \\ Z_i(\beta) &= \sum_k p_{ik} \exp(\beta g_i c_k). \end{aligned} \tag{39}$$

Taking the ratio for keys j and k gives

$$\frac{\widetilde{p}_{ij}(\beta)}{\widetilde{p}_{ik}(\beta)} = \frac{p_{ij}}{p_{ik}} \exp[\beta g_i (c_j - c_k)]. \tag{40}$$

Thus larger c_j increases relative attention when $\beta g_i > 0$, whereas $g_i = 0$ leaves the row unchanged. In inpainting, only missing-region queries prioritize high-conflict observed keys.

Let $\mu_i(\beta) = \mathbb{E}_{\widetilde{p}_i(\beta)}[c]$. Direct differentiation gives

$$\frac{d\mu_i(\beta)}{d\beta} = g_i \, \text{Var}_{\widetilde{p}_i(\beta)}(c) \geq 0. \tag{41}$$

so increasing β monotonically raises the expected attended conflict score for gated queries.

C.4 Conflict correction in FLAIR

Using FLAIR as the host solver, MPA records the native posterior correction

$$\delta_s^{\text{FLAIR}} = z_s^{\text{after data}} - z_s^{\text{after regularizer}}. \tag{42}$$

between its data and regularizer updates. For $f(z) = \frac{1}{2}\|\mathcal{A}z - y\|_2^2$, a gradient correction gives

$$\delta = -\eta \mathcal{A}^\top (\mathcal{A} z^{\text{pri}} - y) \in \text{range}(\mathcal{A}^\top) = \text{null}(\mathcal{A})^\perp. \tag{43}$$

For inpainting this correction is supported only on observed coordinates; Equation 36 selects this measurement-anchored evidence, and Equation 38 directs it through attention.

D Implementation Details

D.1 Common evaluation protocol

All reconstructions are produced at 768×768 resolution. FFHQ uses the fixed prompt "A high quality photo of a face." DIV2K uses an image-specific description extracted from the observation and prefixed by "A high quality photo of," following the main paper. Images, captions, measurement realizations, and seeds are matched across every controlled comparison. Seeds are 3 for super-resolution and 42 for motion deblurring and inpainting. The measurement-noise level and metric definitions are identical to those in the main paper.

The local FLAIR and FlowLPS implementations use Stable Diffusion 3.5-Medium, the TAESD3 lightweight autoencoder, a classifier-free-guidance scale of 2, and an empty negative prompt. Computation uses half-precision model weights and bfloat16 inference states. The FlowDPS and FlowChef protocol below follows the matched Stable Diffusion 3.5-Medium adaptation reported by FLAIR.

The super-resolution operators use bicubic downsampling with factors 8 and 12. Motion deblurring uses a kernel size of 61. FFHQ box inpainting removes rows $128{:}640$ and columns $384{:}640$, corresponding to a missing fraction of $2/9$. DIV2K inpainting uses the same fixed two-rectangle mask for every method. All methods use the same preprocessed targets and observations.

D.2 Host-solver configurations

We integrate SAS and MPA into FLAIR and FlowLPS without changing their within-step update rules.

FLAIR. We use a regularizer SGD learning rate of 1 and 15 data-term steps at every flow time. The data-term learning rate is 12 for $8\times$ and $12\times$ super-resolution and 0.1 for motion deblurring and inpainting. Data-term optimization stops when its loss falls below 10^{-4} times the number of measurements. The regularizer scale is 0.5 with the released time-dependent calibration. The regularizer and data term are optimized sequentially; the conflict correction in Equation 42 is recorded between these two states.

FlowLPS. For $8\times$ and $12\times$ super-resolution, we use 4 Langevin steps and 11 proximal steps, with initial proximal learning rate 0.5 multiplied by 0.85 every five proximal iterations. Motion deblurring uses 6 Langevin steps, 9 proximal steps, and a fixed proximal learning rate of 0.1. Inpainting uses 4 Langevin steps and 11 proximal steps, with initial proximal learning rate 0.1 multiplied by 0.65 every ten proximal iterations. All tasks use a Langevin step size of 10^{-4}, one pCN step, and variance $s_t^2 = t$. The model predicts

$$\widehat{z}_{0|t} = z_t - t v_\theta(z_t, t), \qquad \widehat{z}_{1|t} = z_t + (1-t) v_\theta(z_t, t). \tag{44}$$

The solver applies its Langevin, proximal, and single pCN updates, and records the difference between the proximal and prior endpoint estimates after the proximal update. The next state is

$$z_{t'} = (1-t') z_{0|t}^{\mathrm{prox}} + t' z_{1|t}^{\mathrm{refreshed}}. \tag{45}$$

D.3 SAS parameters

All SAS experiments use 50 solver times, $t_{\min} = 0.18$, $t_{\max} = 1$, $\lambda = 1$, an 8192-point numerical grid, and spectral null threshold 10^{-10}. The numerical construction is detailed in Section B.4. Super-resolution uses

$$\alpha_{\mathrm{miss}} = 1 - \frac{1}{s^2}, \qquad \alpha_{\mathrm{weak}} = 0 \tag{46}$$

for scale s. Inpainting reads the missing fraction directly from the mask. The super-resolution expression is the ideal rank-deficit approximation: it treats the $1/s^2$ retained samples as observed modes and does not assign a separate weak-mode term to the bicubic prefilter. This avoids making the schedule depend on padding and boundary conventions while retaining the dominant null-space fraction. For motion deblurring, the singular values are the magnitudes of the two-dimensional discrete Fourier transform of the normalized length-61 kernel at image resolution. This is the circulant spectral approximation used only to construct the schedule; it does not alter the forward operator used for evaluation. Each singular value is divided by its maximum before applying Equation 23.

D.4 MPA parameters

Table 3 gives all task-dependent settings. Shared settings are $\gamma = 0.7$, no exponential moving average, gate updates after completed solver steps 2 through 35, all transformer layers, and application only to the conditional half of classifier-free guidance. A gate computed after one solver step is consumed by the following transformer evaluation. The same settings are used for both FLAIR and FlowLPS.

Table 3: Task-dependent MPA parameters. "Pool" is the side length of the average-pooling kernel in latent space.

Task	β	τ	$v_{\max}$	Pool	Query gate
SR 8×/12×	2.0	0.15	1.0	17	all
Motion deblur	0.5	0.02	0.5	17	all
Inpainting	4.0	0.15	1.0	9	missing

Table 4: Normalized spectral statistics used by SAS at 768×768 resolution. Super-resolution uses the ideal rank model, motion deblurring uses the circulant spectrum of the length-61 kernel, and inpainting uses the fixed FFHQ box mask.

Operator	Normalized squared spectrum	r/d	$\mathrm{srank}(\mathcal{A})/d$	α_{miss}	α_{weak}
SR 8×	$1\ (d/64),\ 0\ (63d/64)$	0.01563	0.01563	0.98438	0
SR 12×	$1\ (d/144),\ 0\ (143d/144)$	0.00694	0.00694	0.99306	0
Motion deblur	$\lvert\mathcal{F}k_{61}\rvert^2/\lVert\mathcal{F}k_{61}\rVert_\infty^2$	1.00000	0.01639	0	0.98361
Inpainting (FFHQ)	$1\ (7d/9),\ 0\ (2d/9)$	0.77778	0.77778	0.22222	0

D.5 Baseline configurations

For both baselines, we use Stable Diffusion 3.5-Medium and the same task definitions, observations, captions, and seeds as the host solvers.

FlowDPS. We use the standard FlowDPS implementation adapted by FLAIR, with a classifier-free-guidance scale of 2 and 50 NFEs. The update step size is 15 for inpainting and 10 for the other tasks.

FlowChef. We use the FLAIR implementation with a classifier-free-guidance scale of 2 and update step size 1. FlowChef uses 50 NFEs for super-resolution and motion deblurring and 200 NFEs for inpainting.

E Additional Experiments

E.1 Operator spectra and schedules

Table 4 reports the normalized spectra and the resulting SAS coefficients. These quantities depend only on the measurement operator and require no flow-model evaluation.

Super-resolution. For scale s, SAS uses the ideal rank surrogate

$$\mathcal{A}_{\mathrm{SR},s} = \left(I_{d/s^2} \otimes h_s^\top\right) P_s, \qquad h_s = s^{-2}\mathbf{1}_{s^2}, \tag{47}$$

where P_s groups nonoverlapping $s \times s$ patches. Since $h_s^\top$ has one singular value $1/s$, the normalized spectrum contains d/s^2 ones and $d(1 - 1/s^2)$ zeros. Hence $(\alpha_{\mathrm{miss}}, \alpha_{\mathrm{weak}}) = (1 - 1/s^2, 0)$.

Motion deblurring. Under the circular approximation, the length-61 line-kernel surrogate is Fourier diagonal:

$$\mathcal{A}_{\mathrm{blur}} = \mathcal{F}^* \,\mathrm{diag}(\widehat{k}_{61})\mathcal{F}, \qquad \bar{a}_\omega^2 = \frac{|\widehat{k}_{61}(\omega)|^2}{\|\widehat{k}_{61}\|_\infty^2}. \tag{48}$$

All sampled magnitudes are nonzero, so $r = d$; Parseval's identity gives $\mathrm{srank}(\mathcal{A})/d = 1/61$. Therefore $(\alpha_{\mathrm{miss}}, \alpha_{\mathrm{weak}}) = (0, 60/61)$.

Inpainting. If P_m moves the r observed pixels first, the exact decomposition is

$$\mathcal{A}_{\mathrm{inp}} = [I_r \quad 0]\, P_m. \tag{49}$$

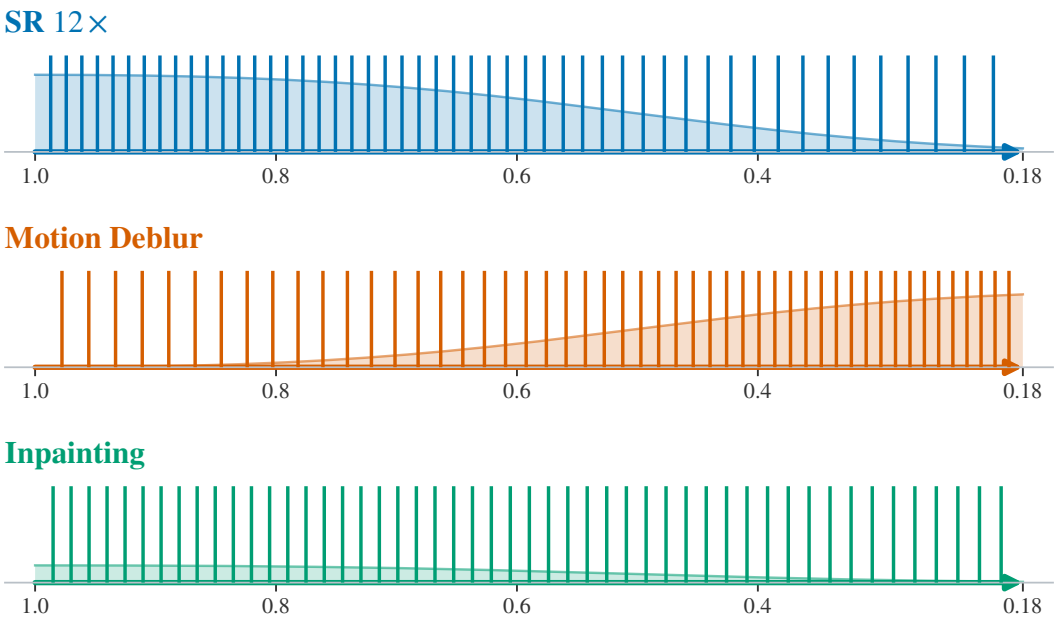


Figure 5: Global SAS schedules for SR 12×, motion deblurring, and inpainting. Each vertical mark denotes one of the 50 solver times; colors distinguish the degradation operators, and the translucent profile shows the normalized local density of the displayed times. Flow time proceeds from noise to clean.

The spectrum thus has r ones and $d-r$ zeros. The fixed FFHQ box removes $512 \times 256/768^2 = 2/9$ of the pixels, giving $(\alpha_{\text{miss}}, \alpha_{\text{weak}}) = (2/9, 0)$.

Accordingly, Figure 5 allocates SR and inpainting by missing modes, with inpainting closer to uniform because its coefficient is smaller, while motion deblurring is driven by weak modes.

E.2 Validating the data-consistency deficit

This diagnostic tests the first link in SAS: operator spectrum → spectral coefficients. It verifies that $\alpha_{\text{miss}} + \alpha_{\text{weak}}$ measures the average error retained by one normalized data-consistency update, giving the coefficients an operational meaning. It does not evaluate schedule quality or reconstruction quality; those are tested by the schedule ablation. For an isotropic error u, a larger retained residual means that data consistency leaves more work to the flow prior. The measured coefficients subsequently weight the two temporal bases in $D_{\mathcal{A}}(t)$. We define

$$R_{\text{lin}}(u) = \frac{u^\top (I - G_{\mathcal{A}})u}{\|u\|_2^2}, \qquad R_{\text{sq}}(u) = \frac{\|(I - G_{\mathcal{A}})u\|_2^2}{\|u\|_2^2}. \tag{50}$$

The expectations are $d^{-1}\,\text{tr}(I - G_{\mathcal{A}})$ and $d^{-1}\,\text{tr}[(I - G_{\mathcal{A}})^2]$, respectively. We draw 100 Gaussian errors at 768×768 resolution and evaluate the update in the right-singular basis, without a flow model or decoder.

E.3 MPA spatial localization and bias strength

We use one 100-image FFHQ box-inpainting experiment to distinguish spatially meaningful conflict guidance from a generic increase in attention bias, and to examine sensitivity to the bias strength β. All variants use FLAIR, Global SAS with $\lambda = 1$, 50 NFEs, $\tau = 0.15$, $v_{\max} = 1$, $\gamma = 0.7$, a 9×9 pooling kernel, missing-region queries, completed solver steps 2–35, and all transformer layers.

Table 5: Theory and numerical measurement of the residual after one normalized data-consistency update. Values are means and standard deviations over 100 isotropic errors. SR uses the ideal rank model, motion deblurring uses the circulant spectrum, and inpainting uses the FFHQ box mask.

Operator	R_{lin}		R_{sq}	
	Theory	Measured	Theory	Measured
SR $8\times$	0.98438	0.98439 ± 0.00023	0.98438	0.98439 ± 0.00023
Motion deblur	0.98361	0.98361 ± 0.00019	0.97814	0.97815 ± 0.00023
Inpainting	0.22222	0.22224 ± 0.00081	0.22222	0.22224 ± 0.00081

At the default $\beta = 4$, we compare:

1. **No MPA**: the attention bias is disabled.
2. **Correct**: use the conflict heatmap in Equation 37.
3. **Shuffled**: deterministically permute the heatmap values within the known region for each image and active step, preserving its histogram and total mass but destroying spatial correspondence.
4. **Uniform**: replace all known-region values by their spatial mean,

$$c_s^{\mathrm{uni}} = m_{\mathrm{known}} \frac{\sum_j c_{s,j}}{\sum_j m_{\mathrm{known},j}}, \tag{51}$$

preserving support and total heatmap mass while removing localization.

The spatial controls test whether MPA benefits from where the conflict occurs, rather than merely from adding a positive key-side bias. We separately evaluate bias-strength sensitivity with the correct heatmap for $\beta \in \{0, 2, 3, 4, 6\}$. The $\beta = 0$ endpoint is equivalent to disabling MPA; all positive values use the same heatmap, threshold, normalization, pooling, query gate, active steps, and layers. Figure 6 reports all four restoration metrics as functions of β, keeping the strength analysis separate from the spatial controls.

Table 6: MPA spatial controls for FLAIR on 100 FFHQ box-inpainting images. Correct, shuffled, and uniform heatmaps all use $\beta = 4$ and preserve the remaining MPA settings. Bold denotes the best result in each column.

Heatmap	β	PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓
None	0	23.52	0.8500	0.1692	32.44
Correct	4	**23.89**	**0.8529**	0.1655	32.15
Shuffled	4	23.88	0.8528	**0.1654**	32.41
Uniform	4	23.83	0.8526	0.1660	**32.08**

The correct heatmap gives the highest PSNR and SSIM, while the shuffled and uniform controls are close and give the best LPIPS and FID, respectively. Thus spatially matched conflict guidance is most evident in the fidelity metrics under this 100-image protocol. In the strength sweep, $\beta = 4$ achieves the best PSNR, LPIPS, and FID; the small further SSIM gain at $\beta = 6$ is accompanied by worse values for the other three metrics.

E.4 Runtime, memory, and NFE accounting

Table 8 separates computational overhead from reconstruction quality. We benchmark FLAIR with batch size one on SR $8\times$, SR $12\times$, motion deblurring, and box inpainting using an NVIDIA RTX 4090. For each setting, runtime is averaged over the same 10 FFHQ images. Model loading, image I/O, and metric computation are excluded; CUDA is synchronized before and after each reconstruction, and three warm-up trials are discarded. Peak allocated memory is reset before every measured image, and we report the maximum of the 10 per-image peaks. All variants use the same 50 flow-model evaluations.

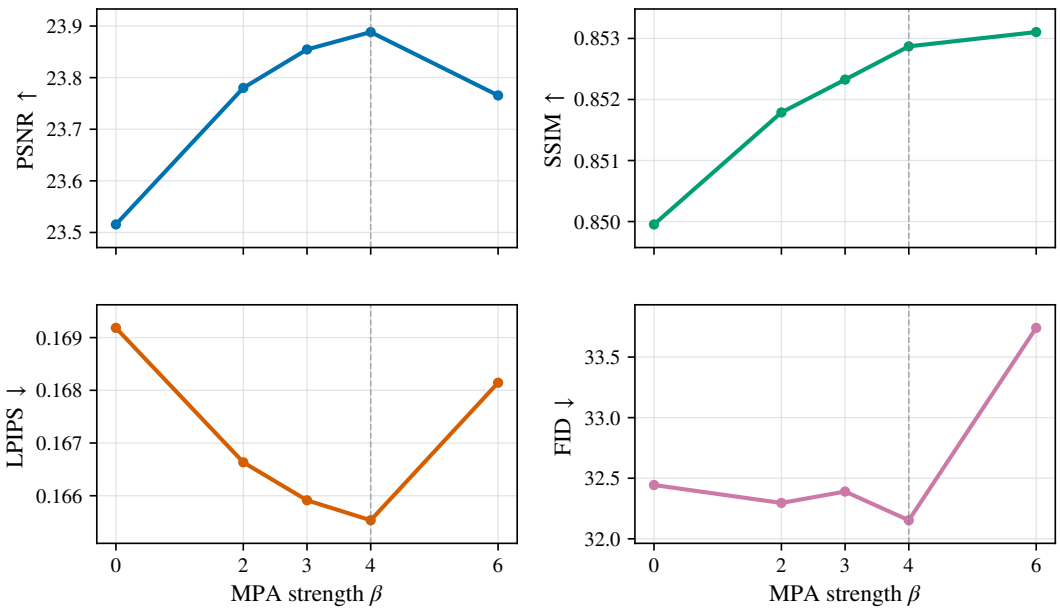


Figure 6: MPA bias-strength sensitivity with the correct heatmap on 100 FFHQ box-inpainting images using FLAIR and Global SAS with $\lambda = 1$. The four panels report PSNR, SSIM, LPIPS, and FID over $\beta \in \{0, 2, 3, 4, 6\}$; $\beta = 0$ is the no-MPA endpoint.



Figure 7: Layout for the MPA spatial-control comparison. The final figure should show the key-side heatmap, reconstruction, and enlarged missing-region crop for the same input and seed. This placeholder must be replaced with experimental outputs before submission.

Across all four tasks, SAS changes runtime by only -0.02 to $+0.14$ seconds per image and leaves peak memory unchanged at the reported precision. The measurable overhead comes from MPA: Full adds 3.44–3.60 seconds per image and approximately 0.54 GiB of peak memory, while preserving the 50-NFE budget. The small signed differences for SAS are within the run-to-run variation of this timing protocol.

E.5 Additional Inpainting Results

Figure 8 provides further qualitative evidence for the inpainting behavior discussed in the main paper. Under matched observations, seeds, and the same 50-NFE budget, the proposed components more consistently preserve instance-specific semantic and structural cues across the masked boundary, rather than allowing the completion to drift toward a generic solution favored by the prior. These examples reinforce that the improvement extends beyond pixel fidelity to semantic and structural consistency within the missing region.



Figure 8: Additional qualitative results on FFHQ box inpainting. Under the same observations, seeds, and NFE budget, the proposed components more consistently preserve instance-specific semantic and structural cues across the masked boundary, reinforcing the inpainting observations in the main paper.

E.6 Sensitivity to the SAS strength λ

We examine the density-strength parameter using FLAIR on 100 FFHQ $8\times$ super-resolution images. All variants use the same images, measurements, seeds, 50-NFE budget, and numerical schedule construction, with MPA disabled. We vary only $\lambda \in \{0, 0.5, 1, 2, 4\}$ in $1 + \lambda D_{\mathcal{A}}(t)$ and hold all host-solver settings fixed. Because these are test images, we report sensitivity rather than use them for hyperparameter selection.

Increasing λ steadily improves PSNR, SSIM, and LPIPS over the tested range, while FID varies by only 0.51 without a monotonic trend. These results show that stronger operator-aware allocation has a consistent but gradual effect on the per-image metrics rather than introducing an abrupt operating point.

Table 7: Sensitivity to the SAS strength on 100 FFHQ $8\times$ super-resolution images using FLAIR, 50 NFEs, and no MPA. Bold denotes the best result in each column.

λ	PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓
0	29.67	0.8067	0.4481	**116.77**
0.5	29.74	0.8094	0.4407	116.79
1	29.80	0.8116	0.4348	117.28
2	29.85	0.8134	0.4278	117.04
4	**29.93**	**0.8163**	**0.4190**	117.14

E.7 Statistical reporting for 100-image experiments

For each comparison, we form 100 paired per-image differences using the same images, measurements, and seeds. We resample these 100 pairs with replacement 10,000 times (seed 2027) and report the observed mean difference with the 2.5th and 97.5th percentiles as a paired 95% bootstrap confidence interval. For PSNR and SSIM, the difference is Ours minus baseline; for LPIPS, it is baseline minus Ours, so positive values always indicate improvement. An interval excluding zero is treated as significant at the 5% level. FID is reported only as a set-level metric because 100 samples do not provide a reliable per-image significance test.

F Computational Overhead

SAS is computed once per operator and does not evaluate the flow model. Its spectral calculations are analytic for super-resolution and inpainting and use one FFT of the blur kernel for motion deblurring. The one-dimensional schedule construction is performed on an 8192-point grid.

MPA reuses the prior and data-consistent estimates already produced by the host solver. Constructing the conflict gate costs $O(nd_z)$ for n image tokens and latent width d_z. Equation 38 adds one feature to each query and key, changing a head's dot-product cost from $O(n^2 d_h)$ to $O[n^2(d_h+1)]$. It does not materialize the outer product, does not add a transformer call, and introduces no additional flow-model evaluations. Both components preserve the fixed NFE budget.

Table 8: Runtime and peak allocated GPU memory for FLAIR on 10 FFHQ images per setting. Times are seconds per image; each parenthesized change is measured relative to the corresponding Base row. Full denotes SAS+MPA, and all variants use 50 NFEs.

Task	Variant	Time (ΔTime)	GPU memory
SR $8\times$	Base	23.79	15.89
SR $8\times$	+SAS	23.78 (0.00)	15.89
SR $8\times$	+MPA	27.23 (+3.44)	16.43
SR $8\times$	Full	27.24 (+3.45)	16.43
SR $12\times$	Base	23.81	15.89
SR $12\times$	+SAS	23.78 (−0.02)	15.89
SR $12\times$	+MPA	27.20 (+3.39)	16.43
SR $12\times$	Full	27.25 (+3.44)	16.43
Motion deblur	Base	32.27	15.88
Motion deblur	+SAS	32.37 (+0.10)	15.88
Motion deblur	+MPA	35.89 (+3.63)	16.42
Motion deblur	Full	35.87 (+3.60)	16.42
Inpainting	Base	22.70	15.89
Inpainting	+SAS	22.84 (+0.14)	15.89
Inpainting	+MPA	26.08 (+3.38)	16.43
Inpainting	Full	26.14 (+3.44)	16.43

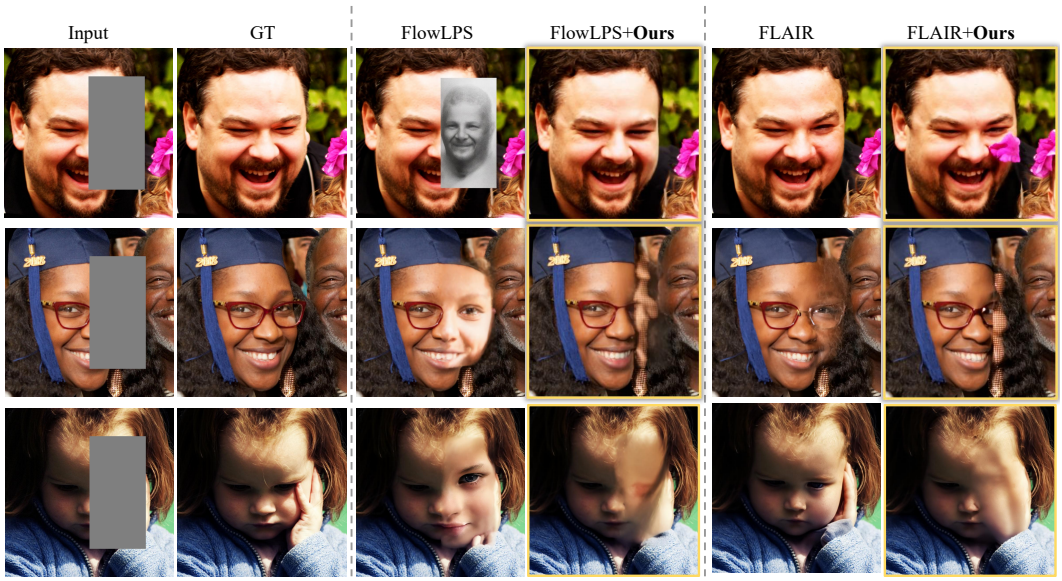


Figure 9: Qualitative failure case on FFHQ box inpainting. Columns show the target, masked observation, host baseline, and full method for the same image and seed. When several similar contours overlap near the masked boundary, the full method can produce a plausible but incorrect contour association inside the missing region.

G Limitations of the Analysis

The trajectory result in Proposition 3 is a first-order discretization bound and does not prove that the proposed demand equals the exact local error of a nonlinear latent inverse solver. The link to SAS is conditional on an explicit operator-aware difficulty model. Likewise, the MPA analysis characterizes the attention redistribution but does not by itself guarantee improvement in the final decoded image. Their role is to establish the direction, selectivity, and stability of the intervention; the controlled experiments evaluate whether those properties improve restoration in practice.

H Inpainting Failure Case

The remaining failures tend to occur near complex mask boundaries where multiple visually similar contours overlap or intersect. Because the missing region removes the connections between observed contour fragments, several continuations may remain plausible, making it difficult to associate each fragment with its correct counterpart. The full method may consequently merge or misconnect these contours even when the overall completion remains semantically plausible. Figure 9 shows a representative example of this boundary ambiguity.