LEARNING TO TRUST KRYLOV REFINEMENT IN SECOND-ORDER DIFFUSION INVERSE PROBLEMS

Anonymous authors
Paper under double-blind review

ABSTRACT

Second-order diffusion inverse solvers exploit local covariance information to improve posterior conditioning beyond mean-only or isotropic approximations. However, this covariance is often obtained from derivatives of a pretrained denoiser whose Jacobian is not explicitly constrained as a covariance operator. The resulting second-order information can therefore be useful yet imperfect. We find that this creates a safety–quality tradeoff in the inner Krylov solver: shallow conjugate gradient (CG) can leave useful refinement unexploited, whereas deeper CG often improves reconstruction but can occasionally amplify covariance errors into catastrophic image-space corrections. We introduce KRC (Krylov Reliability Controller), a lightweight controller that determines which Krylov proposals should be trusted while preserving useful deeper exploration. A 7.8K-parameter two-head MLP uses only causal scalar solver/proposal statistics to predict the value of the current proposal and the recoverable value of deeper iterations. Additionally, a separate trusted correction allows Krylov exploration to continue without injecting unreliable proposals into the diffusion trajectory. Experiments on random-mask single-pixel imaging (SPI), with zero-shot transfer to cake-cutting Hadamard SPI, super-resolution, and deblurring, demonstrate improved reconstruction quality and reliability over fixed-depth and residual-based Krylov control without operator-specific retraining.

1 INTRODUCTION

Diffusion and score-based generative models provide strong image priors that can be reused for inverse problems without task-specific retraining (Sohl-Dickstein et al., 2015; Song & Ermon, 2019; Ho et al., 2020; Song et al., 2020b; Nichol & Dhariwal, 2021; Song et al., 2020a; Dhariwal & Nichol, 2021). Existing methods incorporate known measurements through projection, null-space constraints, likelihood gradients, pseudoinverse guidance, plug-and-play updates, or variational objectives (Kawar et al., 2022; Wang et al., 2022b; Chung et al., 2022; Song et al., 2023; Zhu et al., 2023; Mardani et al., 2024). A common ingredient is the denoised estimate $\mu_t \approx \mathbb{E}[x_0 \mid x_t]$, which summarizes where the clean image is expected to lie at the current diffusion state.

The conditional mean alone, however, does not describe uncertainty around that estimate. Recent work therefore incorporates second-order information, typically through an approximation of the local conditional covariance $\mathrm{Cov}(x_0 \mid x_t)$ (Boys et al., 2023; Rozet et al., 2024; Peng et al., 2024; Rout et al., 2024; Hamidi & Yang, 2024; Rissanen et al., 2025; Yismaw et al., 2025). Such information describes how uncertainty varies across image-space directions and can substantially improve measurement-conditioned reconstruction. In practice, however, second-order structure must be estimated or approximated from pretrained diffusion models and is therefore not perfectly reliable. This creates a tension between exploiting richer local geometry and limiting the effect of errors in that geometry.

This tension becomes especially important when a covariance-aware update is implemented with an inner Krylov solver. A shallow solve is conservative but may underuse useful second-order information: in our population study, 80.85% of diffusion-timestep trajectories attain their best observed proposal beyond CG3. Conversely, deeper conjugate-gradient (CG) refinement often improves reconstruction but can occasionally produce severe image-space degradation. Thus, Krylov

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.

depth is not merely a numerical-accuracy parameter; it also determines how strongly an inverse solver relies on the estimated second-order structure.

We study this problem in random-mask single-pixel imaging (SPI), where spatial patterns produce scalar bucket measurements (Duarte et al., 2008; Shapiro, 2008; Bromberg et al., 2009; Gibson et al., 2020; Lu et al., 2020). Its strongly underdetermined sensing makes unreliable posterior corrections particularly visible. Starting from a covariance-aware diffusion solver, we record inner-CG trajectories across images, sampling ratios, diffusion timesteps, and solver depths to obtain a population-level view of Krylov refinement.

This analysis reveals three findings with direct consequences for inverse reconstruction. First, numerical convergence of the inner system does not guarantee a reliable image-space posterior correction; consequently, residual-based stopping alone can accept harmful updates. Second, CG exhibits an early correction-establishment regime followed by later posterior refinement; therefore, uniformly shallow solves can terminate before useful second-order information has been fully exploited. Third, an unreliable current proposal does not necessarily invalidate deeper iterations, because some trajectories recover with continued Krylov exploration; immediately stopping at the first poor proposal can therefore discard recoverable reconstruction quality.

Together, these observations motivate KRC (Figure 1), a two-head Reliability Controller based only on causal scalar solver/proposal statistics. Specifically, it predicts both the immediate value of the current proposal and the value that may still be recovered through deeper refinement. Because these predictions require neither ground truth nor future CG states, the controller can make reliability decisions online. Crucially, KRC separates Krylov exploration from posterior acceptance: the internal recursion may continue through an unreliable proposal, while only the deepest trusted image-space correction is injected into the diffusion trajectory.

Our contributions are:

- We characterize the reliability of depth-dependent Krylov refinement in covariance-aware diffusion inverse problems and show that useful deep refinement can coexist with rare but severe image-space failures.
- We formulate Krylov trust through immediate proposal value and future recoverability, predicted by a lightweight 7.8K-parameter controller using only test-time observable scalar diagnostics.
- We introduce KRC, which separates Krylov exploration from posterior acceptance and transfers without retraining from random-mask SPI to cake-cutting Hadamard SPI, super-resolution, and deblurring.

2 Related Work

Diffusion inverse problems and second-order uncertainty. Training-free diffusion inverse solvers differ in how measurement information and denoising uncertainty are approximated. Projection and null-space methods directly impose linear constraints (Kawar et al., 2022; Wang et al., 2022b), whereas DPS uses likelihood gradients around a denoised estimate (Chung et al., 2022). Gaussian posterior approximations subsequently make uncertainty explicit (Song et al., 2023; Peng et al., 2024). Second-order methods retain different structures: TMPD uses higher-order Tweedie moments and scalable covariance approximations (Boys et al., 2023); STSL retains a Hessian-trace surrogate (Rout et al., 2024); Peng et al. construct structured posterior covariance (Peng et al., 2024); CA-DPS uses finite-difference covariance (Hamidi & Yang, 2024); Free Hunch estimates covariance without explicit denoiser Jacobians (Rissanen et al., 2025); and CoDPS introduces covariance-corrected Gaussian likelihoods while avoiding gradients through the diffusion model (Yismaw et al., 2025).

Krylov refinement and adaptive guidance. MMPS is closest to our computational setting: its formulation accesses full directional covariance implicitly through denoiser Jacobian products and uses an iterative linear solve (Rozet et al., 2024). This exposes a practical gap between ideal covariance structure and learned derivatives, motivating conservative iterative refinement in prior work. Our work instead asks how proposal reliability evolves when an approximate second-order operator is queried progressively by Krylov iterations. Krylov solvers also appear in DDS for data-consistency refinement (Chung et al., 2024), but without this learned-covariance reliability question. ZAPS adapts timestep likelihood guidance with an efficient Hessian approximation (Alçalar & Akçakaya, 2024),

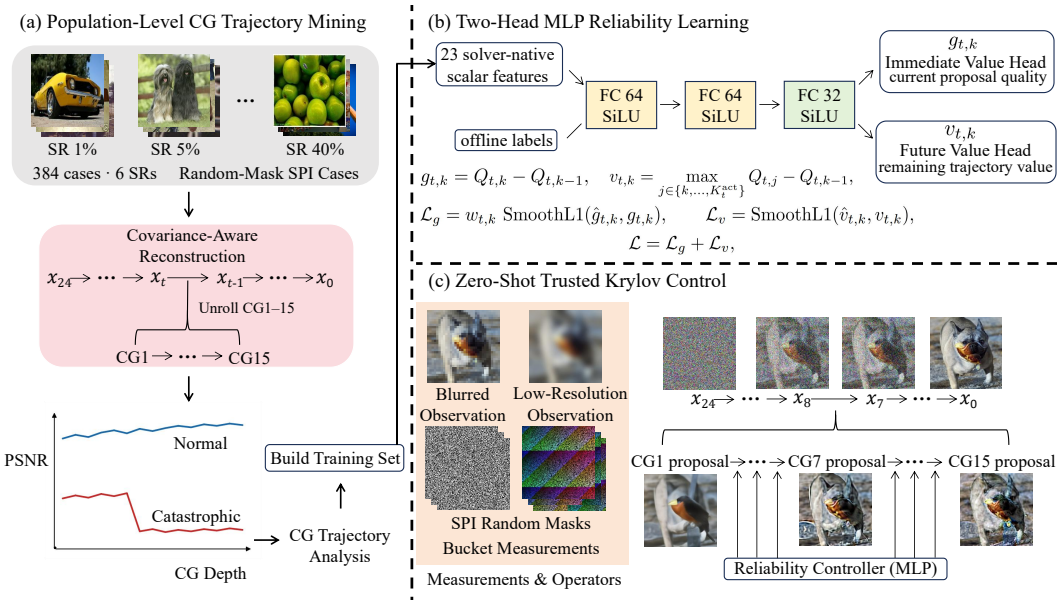


Figure 1: **Overview of KRC.** (a) CG trajectories form the training set. (b) A Reliability Controller predicts current value and future recoverability. (c) The controller transfers zero-shot across operators.

while ZADS automatically tunes data-fidelity weights across diffusion inverse solvers (Alçalar et al., 2025); both act at the outer guidance level rather than controlling inner Krylov proposal acceptance. Classical iterative regularization and semi-convergence likewise show that solver depth can regularize inverse problems (Paige & Saunders, 1982; Hanke, 2017; Hansen, 1998).

Single-pixel imaging. SPI reconstructs images from spatial patterns and scalar bucket measurements and is closely related to compressive sensing and computational ghost imaging (Donoho, 2006; Candès et al., 2006; Duarte et al., 2008; Shapiro, 2008; Bromberg et al., 2009; Gibson et al., 2020). Representative reconstruction approaches include correlation, Fourier regularization, plug-and-play denoising, learned constraints, and unrolled optimization (Ferri et al., 2010; Pastuszczak et al., 2021; Zhang et al., 2021; Wang et al., 2022a; 2025; Zhang et al., 2025). We use random-mask SPI as the primary underdetermined testbed and structured Hadamard SPI to evaluate operator transfer.

3 Method

3.1 Second-order diffusion conditioning

We consider a linear inverse problem

$$y = Ax_0 + \eta, \tag{1}$$

where x_0 is the unknown clean image, A is known, and η denotes measurement error. For SPI, each row of A is an illumination pattern and each entry of y is the associated bucket measurement. For deblurring and super-resolution, A denotes the corresponding blur or blur/downsampling operator.

For a variance-preserving diffusion process,

$$x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon, \tag{2}$$

the frozen denoiser ϵ_θ gives the Tweedie estimate

$$\mu_t(x_t) = \frac{x_t - \sqrt{1-\bar{\alpha}_t}\, \epsilon_\theta(x_t, t)}{\sqrt{\bar{\alpha}_t}}. \tag{3}$$

Second-order Tweedie identities relate the local conditional uncertainty of x_0 to derivatives of the denoising estimate (Boys et al., 2023; Rozet et al., 2024). Writing $J_t = \partial \mu_t / \partial x_t$, we use the

symmetrized covariance action

$$\Sigma_t^{\mathrm{sym}} v = \frac{1-\bar{\alpha}_t}{2\sqrt{\bar{\alpha}_t}} \left(J_t v + J_t^\top v \right), \tag{4}$$

throughout. This action is evaluated implicitly using JVPs and VJPs without forming a dense covariance matrix. Symmetrization defines the common covariance-aware base solver used throughout our fixed-depth and adaptive-control experiments; it enforces self-adjointness of the action but does not by itself guarantee positive semidefiniteness.

Using the local approximation $p(x_0 \mid x_t) \approx \mathcal{N}(\mu_t, \Sigma_t^{\mathrm{sym}})$, the measurement-conditioned mean is

$$\mu_t^{\mathrm{post}} = \mu_t + \Sigma_t^{\mathrm{sym}} A^\top \left(A \Sigma_t^{\mathrm{sym}} A^\top + \tau I \right)^{-1} (y - A\mu_t), \tag{5}$$

where τI is numerical damping. Since the covariance is implicit, we solve the measurement-space system with CG:

$$K_t u_t = r_t, \qquad K_t = A \Sigma_t^{\mathrm{sym}} A^\top + \tau I, \qquad r_t = y - A\mu_t. \tag{6}$$

The depth-k CG iterate $u_{t,k}$ produces the image-space proposal

$$\Delta_{t,k} = \Sigma_t^{\mathrm{sym}} A^\top u_{t,k}, \qquad \mu_{t,k}^{\mathrm{post}} = \mu_t + \Delta_{t,k}. \tag{7}$$

This distinction is central to our method. CG directly reduces $\|K_t u_{t,k} - r_t\|_2$ in measurement space, but reverse diffusion receives the transported image-space correction $\Delta_{t,k}$. Thus, a numerically better linear solve need not be a better posterior update: an imperfect learned covariance can amplify or misdirect the transported correction. We therefore do not treat CG convergence as a sufficient acceptance criterion. Instead, we learn when a proposal is likely useful and when additional Krylov exploration remains worthwhile. Damping and hard curvature safeguards are specified in Appendix A.

3.2 Population-level proposal landscape

We first collect a population-level trajectory dataset rather than designing a controller from isolated failures. The frozen ADM prior is run on random-mask SPI, recording every valid proposal up to CG15 at every DDIM timestep. For offline analysis only, let $Q_{t,k}$ denote the PSNR of the candidate produced by accepting proposal k at timestep t. The immediate proposal value is

$$g_{t,k} = Q_{t,k} - Q_{t,k-1}. \tag{8}$$

A proposal is labelled catastrophic when $g_{t,k} < -3$ dB. These PSNR labels make it possible to study which observable solver statistics precede damaging updates, but they are never available to the online controller.

To characterize proposal geometry without ground truth, we use accumulated and incremental correction-growth ratios,

$$R_{\mathrm{acc}}(t,k) = \frac{\|\Delta_{t,k}\|_2}{\|\Delta_{t,k-1}\|_2 + \epsilon}, \qquad R_{\mathrm{inc}}(t,k) = \frac{\|\Delta_{t,k} - \Delta_{t,k-1}\|_2}{\|\Delta_{t,k-1}\|_2 + \epsilon}. \tag{9}$$

The first few CG iterations form an *establishment* regime: relatively large correction growth can be normal because the measurement constraint is first being imposed. Later iterations enter a tighter *refinement* regime (Figure 2), where comparable growth is more indicative of an unstable update. Consequently, a single depth-independent guard is inappropriate, while uniformly shallow CG discards useful refinement.

We use two hand-crafted controls to test this observation. MIN3HARD trusts the first three proposals and then applies a global correction-growth guard. PERCGHARD instead uses a depth-specific normal envelope. Both depend mainly on the current proposal and cannot distinguish a damaging proposal from one whose trajectory may recover at a deeper CG iterate. This distinction motivates separate prediction of immediate damage and remaining recoverable value.

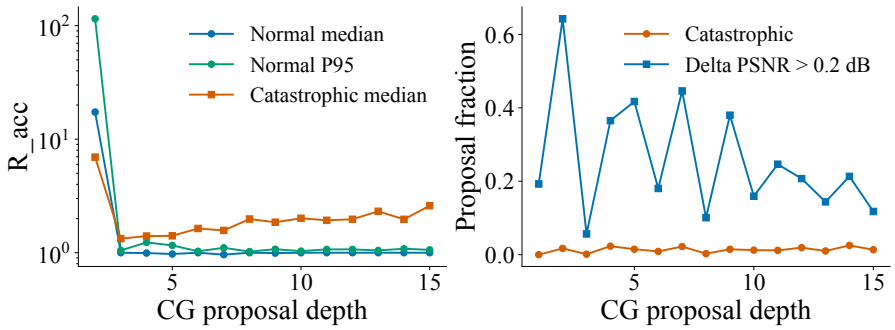


Figure 2: **Correction establishment and posterior refinement.** Correction growth and proposal outcomes across CG depth.

3.3 Reliability Controller

The Reliability Controller is trained from the recorded trajectories. One training example is one CG proposal at one diffusion timestep. Its input contains 23 causal scalar features covering CG phase, diffusion phase, sampling ratio, correction geometry, recursive CG state, measurement consistency, and candidate stability. These features are deliberately scalar and causal: they can be computed during inference without storing an image ensemble or accessing future CG proposals. The complete schema is given in Appendix C.2. At inference, the controller receives no raw image, ADM feature tensor, correction map, ground-truth PSNR, or future CG state.

The controller predicts the immediate value $g_{t,k}$ in Eq. equation 8 and the future recoverable value

$$v_{t,k} = \max_{j \in \{k,\dots,K_t^{\mathrm{act}}\}} Q_{t,j} - Q_{t,k-1}, \tag{10}$$

where $K_t^{\mathrm{act}} \leq 15$ is the deepest valid recorded proposal. Intuitively, $g_{t,k}$ asks whether accepting the current proposal helps or harms, whereas $v_{t,k}$ asks whether useful value may still remain deeper in the same trajectory. The two targets are complementary: a proposal can be damaging now but followed by recovery, or safe now while leaving little value in additional CG iterations.

A shared two-head MLP predicts $(\hat{g}_{t,k}, \hat{v}_{t,k})$ from the causal features. It has 7,842 trainable parameters and is trained by weighted Smooth-L_1 regression, with greater weight on damaging proposals. The image-disjoint split, architecture, loss weights, normalization, and optimization details are given in Appendix C. The controller is trained only on random-mask SPI and is used unchanged in all transfer experiments.

3.4 Trusted-correction control

At test time, the two predicted outcomes are converted into a current-damage score, a no-recovery score, and a joint risk:

$$d_{t,k} = \sigma(-\hat{g}_{t,k} - 1), \qquad h_{t,k} = \sigma(-\hat{v}_{t,k}), \qquad \rho_{t,k} = d_{t,k} h_{t,k}. \tag{11}$$

A high $d_{t,k}$ means that the current proposal is predicted to be harmful. A high $h_{t,k}$ means that further Krylov iterations are not predicted to recover useful value. Their product becomes high only when the proposal is both unsafe now and unlikely to become useful through continued exploration.

The first three CG proposals establish a coarse correction and are trusted unconditionally. From CG4 onward, a low-risk proposal replaces the stored trusted correction $\Delta^{\mathrm{trusted}}$. A risky proposal is *shadowed*: the raw CG recursion continues, but the diffusion state is not updated with that proposal. This preserves a trusted image-space fallback without prematurely discarding a trajectory that may recover. Trust can be restored by a later individually safe proposal whose predicted cumulative gain compensates for the shadowed interval. Exploration terminates only after persistent high risk, natural convergence, a hard numerical safeguard, or the maximum CG depth. Exact calibrated thresholds are reported in Appendix C.

Thus, Krylov exploration and posterior acceptance are explicitly separated. The diffusion step always receives the deepest trusted correction, rather than necessarily the final raw CG iterate. This

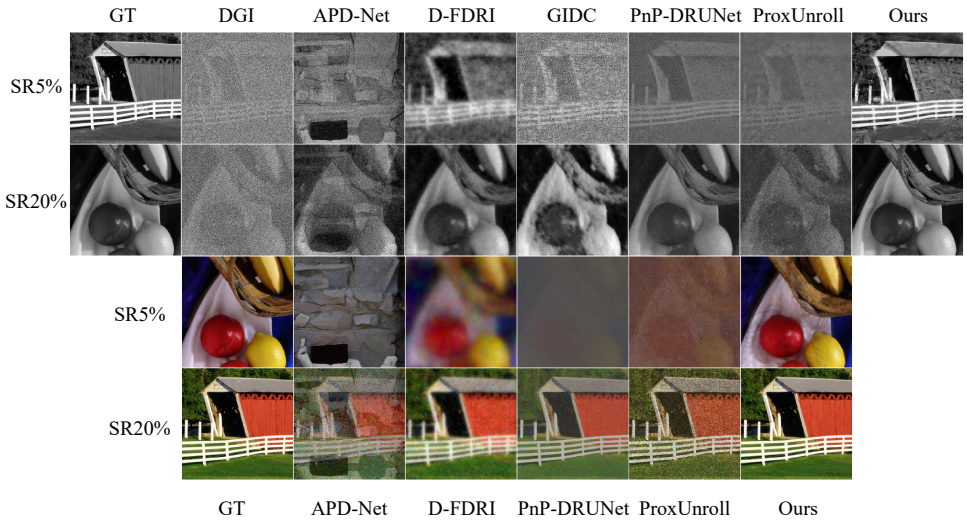


Figure 3: **Random-mask SPI on McMaster.** Grayscale and color reconstructions at the indicated sampling ratios using identical measurements.

permits useful deep refinement while limiting unstable transported updates caused by imperfect local covariance estimation.

4 Experiments

4.1 Experimental setup

All diffusion-based methods use the same frozen OpenAI guided-diffusion 256×256 unconditional ImageNet ADM prior (Deng et al., 2009; Dhariwal & Nichol, 2021). KRC and all matched fixed-depth/control variants use deterministic DDIM25 ($\eta = 0$) (Song et al., 2020a); DPS and ΠGDM retain their DDPM1000 configurations and serve as reconstruction-quality rather than compute-matched references. Main experiments use six NVIDIA RTX 4090 GPUs.

The ImageNet landscape uses random-mask SPI at SRs 1/3/5/10/20/40%, while McMaster (Zhang et al., 2011) benchmarks use 1/5/10/20/40/60%. We also test cake-cutting Hadamard SPI at 5/20/60%, Gaussian deblurring, and $\times 4/\times 8$ super-resolution, reporting PSNR, SSIM (Wang et al., 2004), and LPIPS (Zhang et al., 2018).

SPI baselines are DGI (Ferri et al., 2010), APD-Net (Zhang et al., 2025), D-FDRI (Pastuszczak et al., 2021), GIDC (Wang et al., 2022a), PnP-DRUNet (Zhang et al., 2021), ProxUnroll (Wang et al., 2025), and the matched Fixed-CG2 control; transfer baselines are DPS, ΠGDM, and TMPD. All SPI methods share the same sensing matrix and bucket measurements. ProxUnroll covers all 216 random-mask cases, with no missing or misaligned outputs. Further details are in Appendix A.

4.2 Population-Level Failure Landscape

We analyze 384 ImageNet random-mask SPI cases across six SRs, 9,600 DDIM timesteps, and 135,566 valid recorded CG proposals; 1,749 incur a one-step drop below -3 dB. The full landscape is in Appendix B.1. The trajectories show a clear establishment–refinement transition: correction growth is broad at CG2 but contracts near one by CG4, so a single depth-independent threshold is inappropriate. Yet deeper search remains useful: 80.85% of trajectories peak beyond CG3, 59.06% at CG8+, and 44.16% at CG12+.

Table 1: Held-out regression of immediate proposal value and future-best value. The split is disjoint by base image.

Target	MAE	RMSE	Pearson	Spearman
Immediate $g_{t,k}$	.212	.404	.904	.688
Future-best $v_{t,k}$	.601	.999	.828	.815

Table 2: **Random-mask SPI on McMaster.** All methods use identical Bernoulli sensing matrices and corresponding bucket measurements at each SR. Representative columns show PSNR (dB); averages use all six SRs. Lower LPIPS is better; best/second-best are **bold**/underlined.

Method	SR1	SR5	SR20	SR60	Avg. PSNR ↑	Avg. SSIM ↑	Avg. LPIPS ↓
Grayscale							
DGI	10.929	11.101	12.135	13.102	11.891	0.135	1.119
APD-Net	12.531	13.305	16.042	26.342	17.171	0.297	0.733
D-FDRI	17.666	20.505	23.746	28.805	23.167	0.576	0.551
GIDC	10.704	11.788	17.036	20.109	15.522	0.340	0.689
PnP-DRUNet	14.209	15.515	22.238	**38.235**	23.476	0.560	0.517
ProxUnroll	14.140	14.606	16.044	20.223	16.348	0.281	0.702
FIXED-CG2	17.004	21.456	25.096	26.519	23.213	0.637	0.343
Ours	**17.772**	**23.169**	**29.915**	36.604	**28.004**	**0.719**	**0.230**
Color							
APD-Net	11.263	11.543	12.453	14.965	12.623	0.144	0.793
D-FDRI	15.276	17.453	19.967	22.555	19.231	0.429	0.624
PnP-DRUNet	12.487	12.864	14.400	19.139	14.809	0.270	0.772
ProxUnroll	13.428	13.883	15.224	19.157	15.520	0.269	0.788
FIXED-CG2	15.868	18.850	22.510	24.093	20.958	0.544	0.375
Ours	**15.937**	**20.931**	**26.531**	**31.847**	**24.809**	**0.639**	**0.262**

Standard numerical guards are not enough: non-positive curvature never precedes the first catastrophic recorded proposal, whereas image-space correction statistics become strongly discriminative after establishment. This motivates controlling the transported correction rather than the measurement-space recurrence alone.

As summarized in Table 1, on the image-disjoint split, the immediate head reaches 0.212 dB MAE/0.904 Pearson and the future head 0.601 dB MAE/0.815 Spearman; $-\hat{g}$ yields 0.9966/0.8261 AUROC/AUPRC for catastrophe ranking. Thus causal scalar diagnostics are sufficient to learn proposal reliability and remaining trajectory value without image features.

4.3 SINGLE-PIXEL IMAGING

We evaluate all methods on 18 McMaster images; Figure 3 shows representative reconstructions, with complete six-SR results in Appendix D.

Table 2 shows that KRC reaches 28.004 dB average PSNR in grayscale and 24.809 dB in color. Relative to the matched FIXED-CG10 control in Table 5, this is a gain of 1.70 dB and 1.96 dB, respectively, while reducing final-reconstruction collapses from 5/108 and 6/108 to 0/108. Relative to the strongest external SPI baseline, the grayscale average gain is 4.53 dB.

Structured-pattern transfer. Replacing Bernoulli masks with cake-cutting Hadamard patterns (Zhang et al., 2025), we reuse the controller, normalization, and calibration without retraining. Table 3 shows that Ours attains the highest mean PSNR/SSIM at the three tested SRs; at SR=60%, the PSNR margin to PnP-DRUNet is only 0.06 dB, so we interpret this result primarily as evidence of transfer rather than a large performance separation. Qualitative results are in Appendix D.1.

Table 3: **Cake-cutting Hadamard SPI on McMaster (grayscale).** Each cell reports PSNR/SSIM. Average metrics are computed over SR=5%, 20%, and 60%. Best and second-best results are shown in **bold** and underline, respectively.

Method	SR5	SR20	SR60	Avg.
DGI	18.13/.525	20.38/.748	25.70/.925	21.40/.733
APD-Net	23.13/.594	28.29/.847	36.66/.971	29.36/.804
D-FDRI	23.81/.650	27.89/.851	34.08/.963	28.59/.821
GIDC	19.62/.609	22.00/.826	23.57/.915	21.73/.783
PnP-DRUNet	24.91/.714	30.30/.903	38.61/.980	31.27/.866
ProxUnroll	23.82/.660	28.26/.876	37.55/.984	29.88/.840
Ours	**25.91/.737**	**30.74/.911**	**38.67/.989**	**31.77/.879**

4.4 ZERO-SHOT TRANSFER TO SUPER-RESOLUTION AND DEBLURRING

We test zero-shot transfer on McMaster for two Gaussian deblurring settings and $\times 4/\times 8$ super-resolution, reusing the SPI-trained controller without recalibration. We compare with DPS (Chung et al., 2022), ΠGDM (Song et al., 2023), and TMPD (Boys et al., 2023). The original TMPD

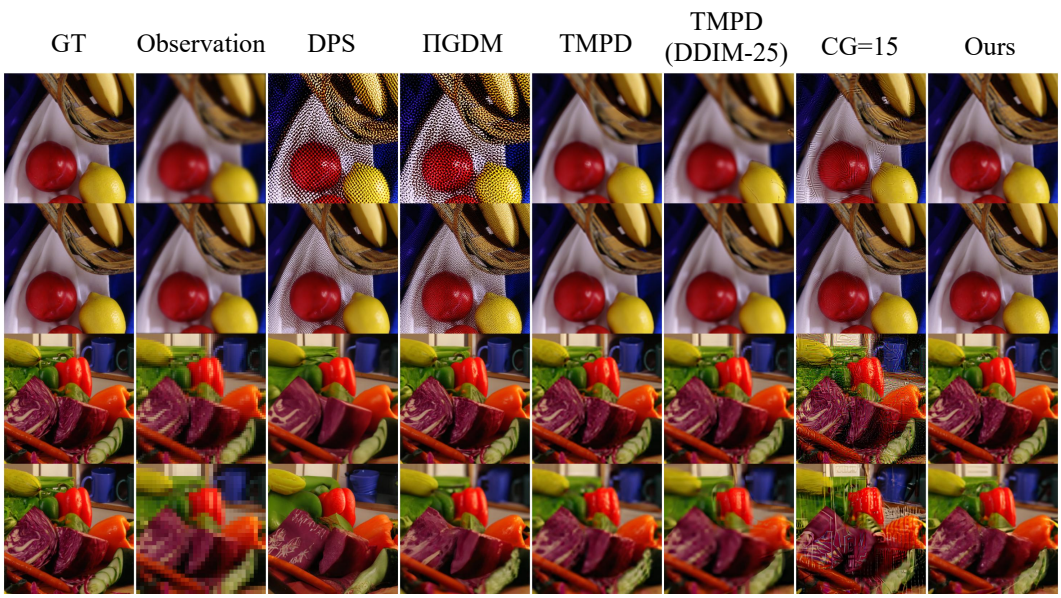


Figure 4: **Zero-shot transfer on McMaster.** Rows: heavy deblurring (61, $\sigma = 3.0$), moderate deblurring (31, $\sigma = 1.5$), $\times 4$ SR, and $\times 8$ SR. FIXED-CG15 (25-step) shows visible texture artifacts.

Table 4: **Zero-shot transfer on McMaster.** Quantitative results are reported for two Gaussian deblurring settings and $\times 4$/$\times 8$ super-resolution.

	Heavy Deblur		Moderate Deblur		$\times 4$ SR		$\times 8$ SR	
Method	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
DPS	20.364	0.352	22.726	0.524	23.264	0.650	20.351	0.490
ΠGDM	21.369	0.589	24.823	0.703	24.556	0.741	20.743	0.563
TMPD	22.605	0.713	27.679	0.875	24.619	0.743	21.358	0.602
TMPD (25-step)	21.265	0.680	26.327	0.823	23.894	0.738	21.217	0.585
FIXED-CG15 (25-step)	19.293	0.531	22.309	0.719	19.753	0.512	17.051	0.379
Ours	**26.085**	**0.788**	**31.164**	**0.919**	**25.281**	**0.772**	**21.489**	**0.616**

baseline retains its original sampling configuration, while TMPD (25-step), FIXED-CG15 (25-step), and Ours use the matched 25-step schedule. The FIXED-CG15 control isolates unrestricted deep Krylov refinement under the same schedule used by our controller.

These operators induce Krylov dynamics unlike random-mask SPI; successful transfer therefore tests whether the learned reliability cues transfer across substantially different linear operators.

FIXED-CG15 may sharpen structure but introduces high-frequency artifacts (Figure 4), whereas trusted refinement avoids accepting unstable proposals. Ours attains the highest mean PSNR and SSIM in all four zero-shot settings (Table 4); the smallest margins are interpreted cautiously given the 18-image evaluation set.

4.5 ABLATION OF KRYLOV-DEPTH CONTROL

We compare FIXED-CG2, FIXED-CG10, and FIXED-CG12, a classical relative-residual stopping rule (RESIDUALSTOP), MIN3HARD, PERCGHARD, and our Reliability Controller under the same prior, operator, covariance, and DDIM schedule. RESIDUALSTOP begins testing after CG3 and terminates once the cached relative CG residual falls below a validation-frozen threshold; its exact rule and calibration are given in Appendix A.1.

To expose stability beyond average reconstruction quality, we additionally measure the lower tail relative to the conservative FIXED-CG2 control. For each matched image–SR pair i and method m, let

$$\Delta_i^{(m)} = \mathrm{PSNR}_i^{(m)} - \mathrm{PSNR}_i^{(\text{Fixed-CG2})}. \tag{12}$$

Table 5: **CG-control ablation on random-mask SPI.** Representative columns report PSNR; average metrics use all six SRs. Safety statistics are computed against matched FIXED-CG2 reconstructions: $\Delta = \mathrm{PSNR}_m - \mathrm{PSNR}_{\text{Fixed-CG2}}$, P_5 is its fifth percentile, and **Coll.** counts cases with $\Delta < -3$ dB. Lower LPIPS/time and higher P_5/Worst are better.

Method	SR1	SR5	SR20	SR60	Avg. P ↑	Avg. S ↑	Avg. L ↓	Time ↓	$P_5\Delta$ ↑	Worst Δ ↑	Coll. ↓
Grayscale											
FIXED-CG2	17.004	21.456	25.096	26.519	23.213	0.637	0.343	**168.1**	0.00	0.00	0/108
FIXED-CG10	17.165	21.591	27.904	34.761	26.304	0.710	0.242	314.2	-0.86	-4.86	5/108
FIXED-CG12	17.015	20.538	25.092	31.050	24.370	0.656	0.290	474.9	-4.89	-12.89	21/108
ResidualStop	**18.130**	22.769	26.848	31.565	25.688	0.655	0.296	170.7	-5.71	-14.385	9/108
Min3Hard	17.081	22.030	28.811	34.144	26.694	0.715	0.240	364.4	-0.57	-3.13	1/108
PerCGHard	17.283	22.076	28.883	34.983	26.684	0.713	0.239	368.5	-0.71	-3.18	1/108
Ours	17.772	**23.169**	**29.915**	**36.604**	**28.004**	**0.719**	**0.230**	306.7	-0.44	-1.33	0/108
Color											
FIXED-CG2	15.868	18.850	22.510	24.093	20.958	0.544	0.375	393.3	0.00	0.00	0/108
FIXED-CG10	15.219	19.023	24.540	29.271	22.853	0.614	0.284	794.4	-4.39	-6.10	6/108
FIXED-CG12	14.927	18.349	23.041	23.037	20.671	0.546	0.343	1025.0	-6.73	-13.75	17/108
ResidualStop	**16.176**	20.824	24.593	23.305	21.442	0.516	0.371	**307.7**	-1.47	-19.75	11/108
Min3Hard	15.035	19.912	26.019	28.854	23.542	0.634	0.272	409.0	-0.57	-2.41	0/108
PerCGHard	15.138	19.930	25.947	28.580	23.513	**0.640**	**0.262**	410.9	-0.53	-2.49	0/108
Ours	15.937	**20.931**	**26.531**	**31.847**	**24.809**	0.639	0.262	534.7	-0.30	-1.03	0/108

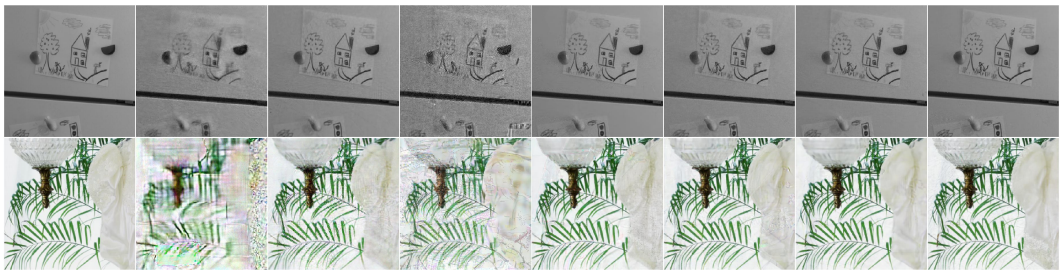


Figure 5: **CG-control ablation at SR=40%.** Grayscale (top) and color (bottom) for fixed and adaptive CG control.

Table 5 reports its fifth percentile, worst case, and the number of final-reconstruction collapses $\Delta_i^{(m)} < -3$ dB together with the conventional reconstruction metrics. Fixed-depth performance is strongly non-monotonic: useful deep refinement can coexist with a pronounced failure tail. KRC retains the benefit of deeper refinement while avoiding severe final-reconstruction collapse, yielding the strongest overall quality–reliability tradeoff among the evaluated controls.

5 Discussion and Conclusion

Our results show that Krylov depth in covariance-aware diffusion inference is more than a numerical-accuracy parameter. Although deeper refinement can better exploit useful second-order information, a smaller inner-system residual does not necessarily yield a more reliable image-space posterior correction. CG optimizes the measurement-space solve, whereas reverse diffusion is driven by the resulting covariance-transported correction. This motivates controlling which Krylov proposals are trusted rather than relying on fixed depth or residual-based stopping alone.

We therefore introduce KRC, a lightweight Reliability Controller that separates internal Krylov exploration from posterior acceptance using only causal scalar diagnostics. This design preserves useful deeper refinement while preventing unreliable proposals from directly perturbing the diffusion trajectory. Trained only on random-mask SPI, the same controller transfers without retraining to cake-cutting Hadamard SPI, super-resolution, and deblurring. These results indicate that reliability-aware Krylov control provides a general strategy for exploiting approximate second-order diffusion information beyond conventional fixed-depth refinement.

AI USE STATEMENT

Generative AI tools were used only for language editing and literature retrieval. All scientific claims, citations, implementations, and experimental results were reviewed and verified by the authors.

ETHICS STATEMENT

This work uses public natural-image datasets and involves no human subjects or sensitive personal data. Unstable generative reconstructions can be misleading if interpreted as measured evidence; improving measurement-conditioned reliability is therefore part of the motivation for this work.

REPRODUCIBILITY STATEMENT

The supplementary release specifies the diffusion checkpoint and schedule, sensing operators, covariance/CG settings, 23-feature schema, image-disjoint split, trained controller, feature normalization, frozen calibration, and evaluation entry point; the appendix records training details and population-data provenance.

REFERENCES

Yaşar Utku Alçalar and Mehmet Akçakaya. Zero-shot adaptation for approximate posterior sampling of diffusion models in inverse problems. In *European Conference on Computer Vision*, pp. 444–460. Springer, 2024.

Yaşar Utku Alçalar, Junno Yun, and Mehmet Akçakaya. Automated tuning for diffusion inverse problem solvers without generative prior retraining. In *2025 IEEE 10th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 241–245. IEEE, 2025.

Benjamin Boys, Mark Girolami, Jakiw Pidstrigach, Sebastian Reich, Alan Mosca, and O Deniz Akyildiz. Tweedie moment projected diffusions for inverse problems. *arXiv preprint arXiv:2310.06721*, 2023.

Yaron Bromberg, Ori Katz, and Yaron Silberberg. Ghost imaging with a single detector. *Physical Review A—Atomic, Molecular, and Optical Physics*, 79(5):053840, 2009.

Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006.

Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022.

Hyungjin Chung, Suhyeon Lee, and Jong Chul Ye. Decomposed diffusion sampler for accelerating large-scale inverse problems. In *International conference on learning representations*, volume 2024, pp. 38922–38949, 2024.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

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.

Fabio Ferri, D Magatti, LA Lugiato, and A Gatti. Differential ghost imaging. *Physical review letters*, 104(25):253603, 2010.

Graham M Gibson, Steven D Johnson, and Miles J Padgett. Single-pixel imaging 12 years on: a review. *Optics express*, 28(19):28190–28208, 2020.

Shayan Mohajer Hamidi and En-Hui Yang. Enhancing diffusion models for inverse problems with covariance-aware posterior sampling. *arXiv preprint arXiv:2412.20045*, 2024.

Martin Hanke. *Conjugate gradient type methods for ill-posed problems*. Routledge, 2017.

Per Christian Hansen. *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*. SIAM, 1998.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *Advances in neural information processing systems*, 35:23593–23606, 2022.

Tian'ao Lu, Ziheng Qiu, Zibang Zhang, and Jingang Zhong. Comprehensive comparison of single-pixel imaging methods. *Optics and Lasers in Engineering*, 134:106301, 2020.

Morteza Mardani, Jiaming Song, Jan Kautz, and Arash Vahdat. A variational perspective on solving inverse problems with diffusion models. In *International Conference on Learning Representations*, volume 2024, pp. 28027–28053, 2024.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pp. 8162–8171. PmLR, 2021.

Christopher C Paige and Michael A Saunders. Lsqr: An algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software (TOMS)*, 8(1):43–71, 1982.

Anna Pastuszczak, Rafał Stojek, Piotr Wróbel, and Rafał Kotyński. Differential real-time single-pixel imaging with fourier domain regularization: applications to vis-ir imaging and polarization imaging. *Optics Express*, 29(17):26685–26700, 2021.

Xinyu Peng, Ziyang Zheng, Wenrui Dai, Nuoqian Xiao, Chenglin Li, Junni Zou, and Hongkai Xiong. Improving diffusion models for inverse problems using optimal posterior covariance. *arXiv preprint arXiv:2402.02149*, 2024.

Severi Rissanen, Markus Heinonen, and Arno Solin. Free hunch: Denoiser covariance estimation for diffusion models without extra costs. In *International Conference on Learning Representations*, volume 2025, pp. 90079–90112, 2025.

Litu Rout, Yujia Chen, Abhishek Kumar, Constantine Caramanis, Sanjay Shakkottai, and Wen-Sheng Chu. Beyond first-order tweedie: Solving inverse problems using latent diffusion. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9472–9481. IEEE, 2024.

François Rozet, Gérôme Andry, François Lanusse, and Gilles Louppe. Learning diffusion priors from observations by expectation maximization. *Advances in Neural Information Processing Systems*, 37:87647–87682, 2024.

Jeffrey H Shapiro. Computational ghost imaging. *Physical Review A—Atomic, Molecular, and Optical Physics*, 78(6):061802, 2008.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pp. 2256–2265. pmlr, 2015.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020a.

Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International conference on learning representations*, 2023.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020b.

F Wang, C Wang, M Chen, W Gong, Y Zhang, S Han, and G Situ. Far-field super-resolution ghost imaging with a deep neural network constraint, light: Sci. In *Appl*, volume 11, pp. 1–11, 2022a.

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 *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 411–421. IEEE, 2025.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. *arXiv preprint arXiv:2212.00490*, 2022b.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Nebiyou Yismaw, Ulugbek S Kamilov, and M Salman Asif. Gaussian is all you need: A unified framework for solving inverse problems via diffusion posterior sampling. *IEEE Transactions on Computational Imaging*, 2025.

Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. Plug-and-play image restoration with deep denoiser prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6360–6376, 2021.

Lei Zhang, Xiaolin Wu, Antoni Buades, and Xin Li. Color demosaicking by local directional interpolation and nonlocal adaptive thresholding. *Journal of Electronic imaging*, 20(2):023016–023016, 2011.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *2018 IEEE/CVF conference on computer vision and pattern recognition*, pp. 586–595. IEEE, 2018.

Yifei Zhang, Yingxin Li, Zonghao Liu, Fei Wang, Guohai Situ, Mu Ku Chen, Haoqiang Wang, and Zihan Geng. Physics and data-driven alternative optimization enabled ultra-low-sampling single-pixel imaging. *Advanced Photonics Nexus*, 4(3):036005–036005, 2025.

Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1219–1229. IEEE, 2023.

A Additional Experimental Details

All main experiments were run on six NVIDIA RTX 4090 GPUs. KRC keeps the ADM prior frozen. The released random-mask SPI protocol uses 256×256 center crops, Bernoulli $0/1$ masks with probability 0.5, DDIM25 with $\eta = 0$, symmetric covariance, $K_{\max} = 15$, CG tolerance 10^{-5}, and zero added measurement noise. The shift τ is numerical damping, not observation-noise variance: at each timestep we set $\tau = 0.01 s_t$, where s_t is the median of two positive Rademacher Rayleigh probes of $A\Sigma_t^{\mathrm{sym}} A^\top$. A failed probe or non-positive CG curvature is a hard numerical safeguard; the latter truncates to the deepest trusted correction.

KRC adds no controller-only denoiser evaluation. For a CG search direction p_k, the covariance action $w_k = \Sigma_t^{\mathrm{sym}} A^\top p_k$ is already computed when forming the matrix-vector product, and the image-space proposal is updated as $\Delta_k = \Delta_{k-1} + \alpha_k w_k$. The same cached quantities provide the transported-proposal features.

A.1 Residual-based stopping baseline

We additionally implement a classical measurement-space stopping baseline, ResidualStop, using exactly the same covariance-aware conditioning operator, symmetrized covariance, damping, DDIM schedule, and numerical safeguards as the fixed-depth runs. Let

$$K_t = A\Sigma_t^{\mathrm{sym}} A^\top + \tau I \tag{13}$$

and let

$$\xi_{t,k} = r_t - K_t u_{t,k} \tag{14}$$

denote the algebraic residual of the inner CG solve after depth k. The stopping statistic is the relative residual

$$e_{t,k} = \frac{\|\xi_{t,k}\|_2}{\|\xi_{t,0}\|_2 + 10^{-12}}. \tag{15}$$

The returned depth is

$$k_t^\star = \min\left\{k \in \{3, \dots, 15\} : e_{t,k} \le \epsilon_{\mathrm{res}}\right\}, \tag{16}$$

with CG15 used when the threshold is not reached, unless an existing hard numerical safeguard terminates the solve earlier. The first three CG iterations are therefore retained as a common correction-establishment phase.

Importantly, $\xi_{t,k}$ is the residual vector already maintained by the standard CG recurrence. ResidualStop therefore requires no additional application of K_t, no additional covariance action, and no additional denoiser JVP/VJP. It uses neither ground-truth image quality nor KRC predictions.

The stopping threshold is selected before McMaster evaluation using only the frozen image-disjoint ImageNet validation trajectories. We evaluate the predefined grid

$$\epsilon_{\mathrm{res}} \in \left\{10^{-1}, 5{\times}10^{-2}, 10^{-2}, 5{\times}10^{-3}, 10^{-3}, 5{\times}10^{-4}, 10^{-4}\right\}, \tag{17}$$

with $k_{\min} = 3$ and $K_{\max} = 15$. The released validation procedure freezes

$$\boxed{\epsilon_{\mathrm{res}} = 0.05} \tag{18}$$

before any McMaster reconstruction is evaluated. No McMaster PSNR, SSIM, or LPIPS is used to tune this threshold. The same frozen stopping rule is then used for every grayscale/color image and sampling ratio.

A.2 SPI comparison protocol

For grayscale SPI we compare DGI, APD-Net, D-FDRI, GIDC, PnP-DRUNet, and the matched Fixed-CG2 control; color experiments use APD-Net, D-FDRI, PnP-DRUNet, ProxUnroll, and Fixed-CG2. The CG-control ablation additionally compares Fixed-CG2, Fixed-CG10, and Fixed-CG12, ResidualStop, Min3Hard, PerCGHard, and KRC.

For every SPI comparison reported in this paper, all methods are evaluated using exactly the same sensing matrix and the corresponding bucket measurements at each image and sampling ratio. No

method is paired with a method-specific or privileged sensing pattern. Although some baseline methods were originally developed or validated under particular sensing designs, the implementations included in our tables operate on the shared sensing operator used for that benchmark.

For APD-Net, we use the same frozen diffusion backbone as our method while retaining its original conditioning formulation. We also tested direct adaptations of additional diffusion baselines to random-mask SPI. Only implementations that passed basic reconstruction sanity checks under the same shared Bernoulli measurements are included in the SPI quantitative comparison; this exclusion concerns methods not reported in the SPI tables and does not change the shared-measurement protocol among the reported methods.

B Additional Population-Level Landscape Analysis

B.1 Full timestep–depth failure landscape

Figure 6 provides the complete population view across diffusion timestep and CG proposal depth.

The main paper uses the population study to motivate state-aware Krylov control. The final landscape contains all 384 planned valid ImageNet random-mask SPI cases: 64 images at SR=1%, 3%, 5%, 10%, 20%, and 40%, giving 9,600 DDIM timesteps and 135,566 valid recorded CG proposals. The count is below the nominal $9{,}600 \times 15$ maximum because natural convergence and hard numerical safeguards can truncate an inner trajectory before CG15. Artifact-integrity checks require complete finite timestep records; unless stated otherwise, all analyses below use this final population.

B.2 Best observed depth and remaining timestep-level headroom

Overall, 80.85% of timestep trajectories achieve their best observed candidate beyond CG3, 59.06% at CG8 or deeper, and 44.16% at CG12 or deeper. This confirms that a uniformly shallow inner solve discards substantial useful refinement.

We additionally compute oracle headroom relative to several fixed CG depths within each stored timestep trajectory. This quantity measures how much value remains available deeper in the same local Krylov trajectory. It is not a final-reconstruction oracle because changing the accepted proposal at one timestep also changes all subsequent diffusion states.

B.3 Dependence on sampling ratio

The failure landscape changes with sensing ratio, but not monotonically. Figure 7 shows that catastrophic incidence generally increases toward larger SRs while the mean best observed depth peaks at SR=10%. At the same time, R_{acc} remains strongly discriminative in the CG4+ refinement regime across all six SRs. These trends motivate using SR as contextual input to the Reliability Controller rather than assigning a fixed depth to each SR.

B.4 Failure signals and curvature timing

Figure 8 compares simple scalar warning signals when all valid recorded CG depths are pooled. In this pooled setting, clip-ratio jump is stronger than the correction-growth ratios because normal CG1–CG2 proposals belong to the broad establishment regime. Once the analysis is restricted to CG4+, the ordering changes: R_{acc} reaches AUROC/AUPRC 0.958/0.623 and R_{inc} reaches 0.969/0.551. Thus correction growth is most useful after the correction has been established rather than as a depth-agnostic threshold from the first CG iteration onward.

Non-positive curvature remains an important hard numerical safeguard, but it is not a timely image-space failure predictor in this dataset. Among trajectories that contain a catastrophic recorded proposal, the first curvature warning appears before the catastrophe in 0% of cases, only afterward in 59.48%, and never in 40.52%. This confirms that curvature is useful as a hard numerical safeguard rather than as an early predictor of posterior failure.

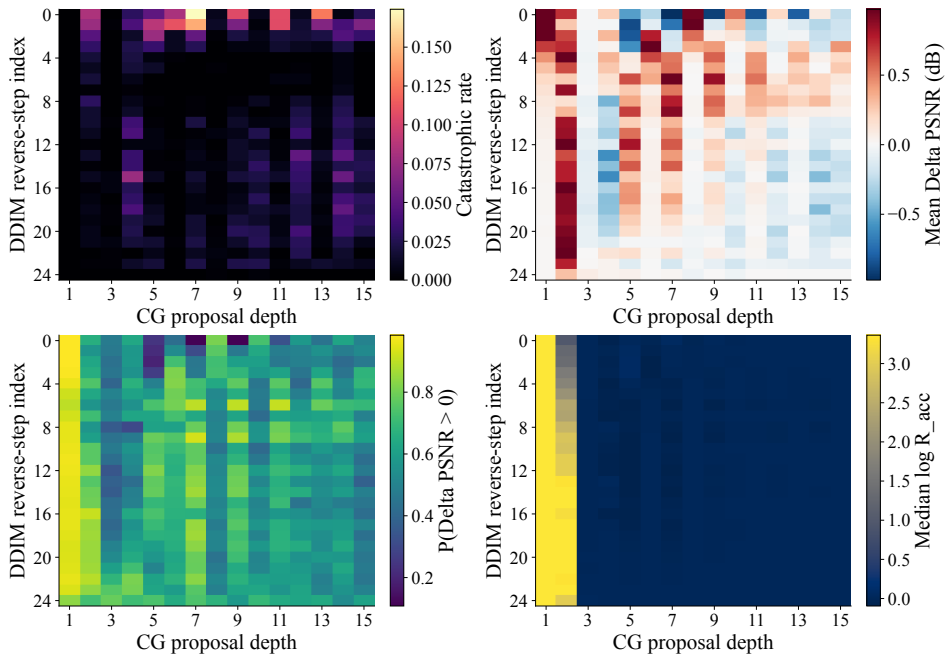


Figure 6: **Population-level CG failure landscape.** Statistics over 384 ImageNet random-mask SPI cases across diffusion timestep and CG depth.

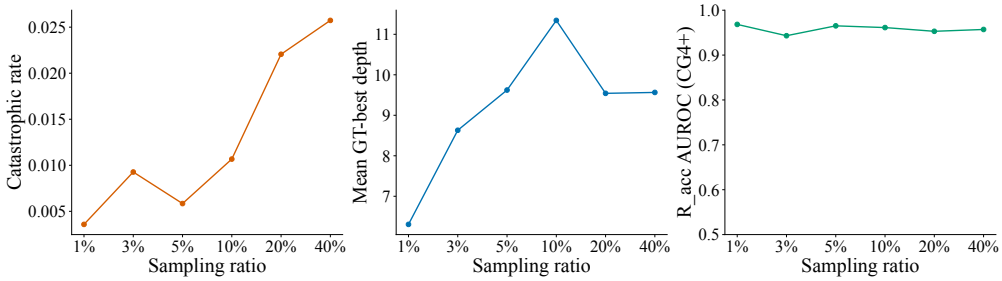


Figure 7: **Sampling-ratio dependence of the final population landscape.** Left: catastrophic proposal rate. Middle: mean best observed CG depth among recorded proposals up to depth 15. Right: AUROC of R_{acc} for catastrophe ranking restricted to the CG4+ refinement regime. The dependence is non-monotonic, so SR is informative context but not a sufficient control rule.

B.5 Recovery and offline shrinkage evidence

A catastrophic proposal does not always make the remaining Krylov trajectory useless. In the final population, 216 of 1,749 catastrophes (12.35%) later recover beyond the pre-drop candidate by more than 0.2 dB; 1,510 have non-positive future gain and 23 are only marginally recoverable. An additional offline soft-shrink oracle evaluates $\Delta_k^{\text{soft}} = \Delta_{k-1} + \lambda(\Delta_k - \Delta_{k-1})$. It selects an intermediate $0 < \lambda < 1$ for 65.69% of catastrophes, with median $\lambda = 0.10$ and interquartile range $[0, 0.20]$. These are GT-dependent offline diagnostics, not deployable controller decisions; they only motivate preserving a trusted fallback while allowing limited shadow exploration.

C Additional Reliability-Controller Diagnostics

The Reliability Controller was frozen from an earlier acquisition snapshot, before the final 384-case sweep had completed. Cases were admitted only if they were already marked complete and passed pre-specified artifact checks (25 DDIM timestep records, contiguous valid CG proposals, and finite stored values); incomplete cases were excluded by acquisition status rather than reconstruction quality. The resulting snapshot contains 364 cases from 63 base images, 9,100 diffusion timesteps,

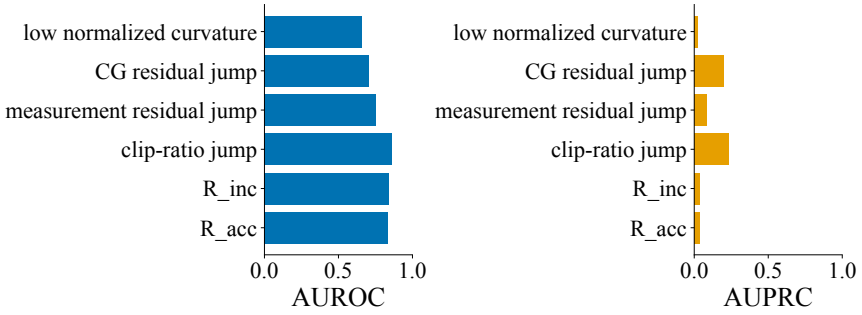


Figure 8: **Single-signal catastrophe ranking pooled over all valid recorded CG depths.** Left: AUROC. Right: AUPRC. The pooled comparison includes the broad CG1–CG2 establishment regime, in which correction-growth ratios are naturally less selective; their stronger CG4+ performance is reported in the text.

Table 6: Complete grayscale random-mask SPI results. Each cell reports PSNR / SSIM / LPIPS.

Method	SR1	SR5	SR10	SR20	SR40	SR60
DGI	10.929/.058/1.185	11.101/.089/1.168	11.642/.111/1.155	12.135/.143/1.122	12.440/.188/1.063	13.102/.219/1.020
APD-Net	12.531/.167/.700	13.305/.136/.874	14.180/.149/.951	16.042/.215/.925	20.625/.427/.635	26.342/.692/.311
D-FDRI	17.666/.368/.803	20.505/.460/.704	21.966/.524/.625	23.746/.605/.526	26.312/.709/.384	28.805/.788/.263
GIDC	10.704/.084/.860	11.788/.135/1.000	13.958/.219/.868	17.036/.394/.618	19.537/.571/.427	20.109/.637/.358
PnP-DRUNet	14.209/.279/.921	15.515/.262/.894	17.476/.339/.760	22.238/.581/.421	33.182/.926/.065	38.235/.971/.040
ProxUnroll	14.140/.239/.811	14.606/.188/.790	15.106/.196/.756	16.044/.243/.718	17.970/.349/.641	20.223/.471/.499
FIXED-CG2	17.004/.326/.543	21.456/.573/.402	23.228/.652/.348	25.096/.719/.292	25.977/.764/.247	26.519/.789/.225
Ours	17.772/.311/.531	23.169/.573/.391	26.307/.718/.252	29.915/.835/.126	34.257/.924/.051	36.604/.951/.031

and 128,455 proposals. Splitting by base-image identity gives 44/9/10 train/validation/test images with 88,182/19,007/21,266 proposal samples, respectively, preventing cross-image leakage across SR, timestep, or CG depth. The held-out regression metrics reported in Table 1 are computed on this 10-image test partition.

The model uses 23 causal scalar features from CG phase, diffusion phase, sampling ratio, correction geometry, recursive CG state, measurement consistency, and candidate stability. The shared network is $23 \to 64 \to 64 \to 32$ with SiLU activations and two scalar heads, for 7,842 trainable parameters.

C.1 TRAINING, NORMALIZATION, AND CALIBRATION

Training uses AdamW with learning rate 10^{-3}, weight decay 10^{-4}, batch size 2048, seed 20260910, at most 160 epochs, and early-stopping patience 20; the released checkpoint is epoch 157. Both heads use PyTorch Smooth-L_1 with $\beta = 1$ and equal head weight. Immediate targets are clipped to $[-12, 3]$ dB and future targets to $[-12, 8]$ dB for training only. The immediate loss weights are

$$w_{t,k} = \begin{cases} 1, & g_{t,k} \geq -0.5, \\ 2, & -3 \leq g_{t,k} < -0.5, \\ 5, & g_{t,k} < -3, \end{cases} \tag{19}$$

while raw unclipped targets are retained for evaluation. Each input feature is normalized by the training-set median and interquartile range and then clipped to $[-20, 20]$.

Controller parameters are selected only on the validation split by a fixed 48-configuration grid: damage midpoint $d_0 \in \{0.5, 1.0\}$ dB, future midpoint $f_0 \in \{0, 0.25\}$ dB, trust threshold $\in \{0.35, 0.5\}$, stop threshold $\in \{0.85, 0.90\}$, and re-trust cumulative floor $\in \{-0.5, -1, -2\}$ dB; both sigmoid temperatures are 1 dB. Configurations with more than 1% false stops on validation trajectories without catastrophes are excluded. The frozen configuration is $d_0 = 1$, $f_0 = 0$, trust threshold 0.35, stop threshold 0.90, re-trust floor -2 dB, unconditional trust through CG3, minimum stopping depth CG5, patience two, damage-trust threshold 0.5, and $K_{\max} = 15$. No test reconstruction is used for calibration.

At inference, $d = \sigma((-\hat{g} - d_0)/1)$, $h = \sigma((-\hat{v} - f_0)/1)$, and $\rho = dh$. A valid proposal always advances the internal CG state. It updates the external trusted correction only during the unconditional establishment phase, when individually safe, or after shadow recovery satisfying the cumulative

predicted-gain floor. Two consecutive high-risk proposals after CG5 stop exploration; convergence, maximum depth, and hard numerical failures can stop earlier. The returned proposal is always the deepest trusted one.

C.2 Controller feature schema

The Reliability Controller receives the following 23 causal scalar features. All quantities are available at test time; no ground-truth quality or future CG state is used. Features are normalized using training-set statistics only.

1. **Normalized CG index:** current inner-solver depth (e.g., CG3, CG8, or CG15).
2. **Normalized DDIM progress:** position within the 25-step reverse diffusion trajectory.
3. **Normalized original diffusion timestep:** the original ADM timestep associated with the current DDIM state.
4. **Diffusion log-SNR:** scalar noise-level context for the current diffusion state.
5. **Sampling ratio:** current measurement ratio (SR).
6. $\log(\|\boldsymbol{\Delta}_{k-1}\|_2 + \epsilon)$: magnitude of the previous CG image-space correction.
7. $\log(\|\boldsymbol{\Delta}_{k}\|_2 + \epsilon)$: magnitude of the current correction.
8. $\log(\|\boldsymbol{\Delta}_{k} - \boldsymbol{\Delta}_{k-1}\|_2 + \epsilon)$: size of the newly added correction component.
9. $\log(R_{\mathbf{acc}} + \epsilon)$: accumulated correction-growth ratio.
10. $\log(R_{\mathbf{inc}} + \epsilon)$: incremental correction-growth ratio.
11. **Increment–correction cosine:** alignment between the new correction increment and the previous correction.
12. **Previous-correction indicator:** whether a valid previous correction exists for comparison.
13. **Log relative CG residual:** remaining residual of the inner linear system.
14. **Log CG residual ratio:** change in CG residual relative to the previous valid depth.
15. **Log measurement-residual ratio:** measurement-residual improvement relative to the initial u_0 state.
16. **Measurement-residual change:** most recent change in measurement residual.
17. **Signed-log normalized curvature:** a normalized $p^\top Kp$-type curvature statistic that exposes anomalous or near-nonpositive curvature.
18. **Signed-log CG α:** current CG step-size coefficient.
19. **Signed-log CG β:** current CG search-direction recurrence coefficient.
20. **Clipping ratio:** fraction of candidate pixels outside the valid image range before clipping.
21. **Clipping-ratio jump:** increase in clipping ratio relative to the previous CG proposal.
22. **Pre-clipping candidate standard deviation:** global variation of the candidate image before clipping.
23. **Log increment-to-mean ratio:** magnitude of the new CG correction increment relative to the current clean-image mean.

A damaging current proposal and an unrecoverable trajectory are different conditions. Figure 9 groups held-out proposals offline using their true immediate and future labels and then shows the model predictions within those groups. Large drops correspond to the catastrophic criterion $g < -3$ dB, while useful recovery means future-best gain $v > 0.2$ dB. The rare large-drop-but-recoverable subset remains the most difficult case: the future-value head is pessimistic on average. This is one reason the online controller preserves a trusted fallback and requires persistent risk rather than terminating after a single negative prediction.

D Additional Single-Pixel Imaging Results

The main paper reports representative SR columns for compactness. Tables 6 and 7 provide the complete six-SR results, while Figure 11 shows additional reconstructions. Each entry is reported as PSNR / SSIM / LPIPS-Alex, averaged over the 18 McMaster images at each sampling ratio.

Table 7: Complete color random-mask SPI results. Each cell reports PSNR / SSIM / LPIPS.

Method	SR1	SR5	SR10	SR20	SR40	SR60
APD-Net	11.263/.164/.732	11.543/.136/.766	11.859/.121/.806	12.453/.117/.852	13.656/.142/.834	14.965/.187/.766
D-FDRI	15.276/.310/.837	17.453/.359/.732	18.639/.395/.675	19.967/.444/.598	21.494/.510/.489	22.555/.558/.414
PnP-DRUNet	12.487/.262/.981	12.864/.216/1.024	13.358/.198/.965	14.400/.212/.806	16.609/.303/.527	19.139/.431/.329
ProxUnroll	13.428/.218/.941	13.883/.177/.911	14.367/.188/.852	15.224/.233/.820	17.061/.340/.690	19.157/.456/.514
FIXED-CG2	15.868/.277/.594	18.850/.461/.445	20.636/.531/.384	22.510/.618/.308	23.793/.686/.262	24.093/.690/.254
Ours	15.937/.268/.596	20.931/.481/.407	23.597/.607/.276	26.531/.739/.166	30.015/.852/.079	31.847/.888/.051

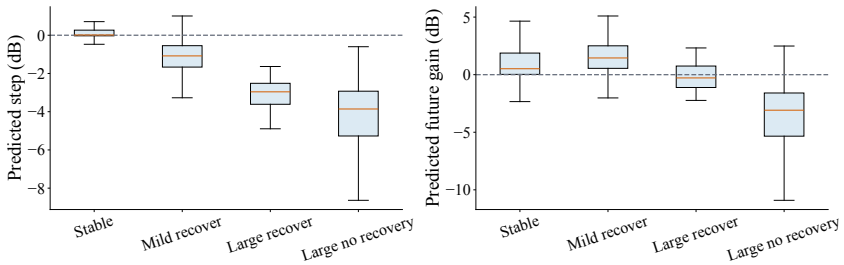


Figure 9: **Recovery-aware held-out predictions.** Groups are defined offline using ground-truth immediate and future labels; the boxplots show the Reliability Controller predictions within those groups. Left: predicted immediate value. Right: predicted future-best gain. The large-drop/recover group is rare and remains the hardest regime for the future-value head.

D.1 CAKE-CUTTING HADAMARD SPI

We additionally evaluate grayscale SPI with cake-cutting ordered Hadamard patterns (Zhang et al., 2025) at SR=5%, 20%, and 60%. This experiment changes the sensing pattern family while reusing the Reliability Controller, feature normalization, and risk calibration learned from random-mask SPI without retraining or recalibration. The quantitative comparison is reported in Table 3; Figure 10 shows representative reconstructions with one sample for each of the three tested SRs.

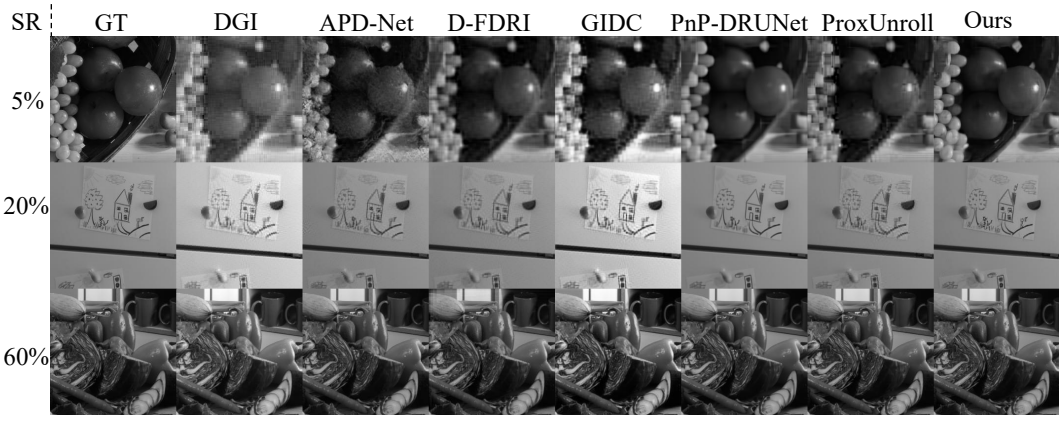


Figure 10: **Cake-cutting Hadamard SPI on McMaster.** Representative grayscale comparisons for three sampling ratios. From top to bottom, rows correspond to SR=5%, 20%, and 60%; each row shows one representative McMaster sample. All methods use identical measurements at each SR. The Reliability Controller is transferred directly from random-mask SPI without retraining or recalibration.

E ADDITIONAL ABLATION OF CG CONTROL

Tables 8 and 9 report the complete six-SR ablation. Across the full sampling range, adaptive control remains consistently competitive while avoiding the pronounced degradations that appear for unrestricted deep CG at some operating points. The Reliability Controller therefore improves the overall quality–stability tradeoff rather than relying on a gain confined to a single sampling-ratio regime.

Table 8: Complete grayscale ablation. Each cell reports PSNR / SSIM / LPIPS.

Method	SR1	SR5	SR10	SR20	SR40	SR60
FIXED-CG2	17.004/.326/.543	21.456/.573/.402	23.228/.652/.348	25.096/.719/.292	25.977/.764/.247	26.519/.789/.225
FIXED-CG10	17.165/.312/.535	21.591/.572/.377	24.357/.696/.279	27.904/.829/.146	32.044/.907/.073	34.761/.941/.042
FIXED-CG12	17.015/.307/.545	20.538/.543/.412	23.592/.677/.274	25.092/.724/.242	28.930/.831/.145	31.050/.854/.124
ResidualStop	18.130/.340/.528	22.769/.559/.384	25.024/.666/.301	26.848/.728/.239	29.790/.802/.175	31.565/.835/.146
Min3Hard	17.081/.312/.535	22.030/.569/.400	25.324/.719/.254	28.811/.839/.139	32.771/.918/.059	34.144/.932/.053
PerCGHard	17.283/.312/.536	22.076/.572/.384	25.293/.718/.254	28.883/.834/.141	31.583/.899/.079	34.983/.944/.040
Ours	17.772/.311/.531	23.169/.573/.391	26.307/.718/.252	29.915/.835/.126	34.257/.924/.051	36.604/.951/.031

Table 9: Complete color ablation. Each cell reports PSNR / SSIM / LPIPS.

Method	SR1	SR5	SR10	SR20	SR40	SR60
FIXED-CG2	15.868/.277/.594	18.850/.461/.445	20.636/.531/.384	22.510/.618/.308	23.793/.686/.262	24.093/.690/.254
FIXED-CG10	15.219/.268/.595	19.023/.469/.410	22.036/.599/.300	24.540/.702/.201	27.030/.790/.121	29.271/.854/.076
FIXED-CG12	14.927/.268/.594	18.349/.454/.427	20.875/.574/.301	23.041/.667/.229	23.797/.666/.238	23.037/.648/.269
ResidualStop	16.176/.280/.590	20.824/.481/.414	22.012/.546/.346	24.593/.662/.233	26.998/.745/.169	23.305/.615/.183
Min3Hard	15.035/.269/.594	19.912/.482/.406	22.601/.612/.293	26.019/.757/.156	28.829/.842/.094	28.854/.840/.088
PerCGHard	15.138/.269/.593	19.930/.484/.404	22.819/.628/.268	25.947/.757/.150	28.663/.837/.093	28.580/.863/.065
Ours	15.937/.268/.596	20.931/.481/.407	23.597/.607/.276	26.531/.739/.166	30.015/.852/.079	31.847/.888/.051

F ADDITIONAL ZERO-SHOT TRANSFER RESULTS

Figure 13 provides additional qualitative results for the Reliability Controller across super-resolution and Gaussian deblurring on McMaster. The Reliability Controller and its feature normalization are used without task-specific retraining or recalibration.

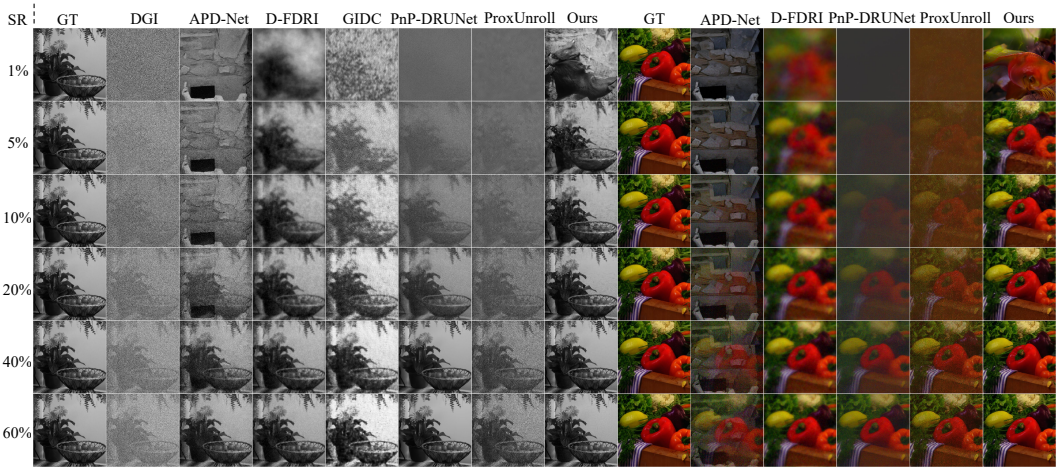


Figure 11: **Additional random-mask SPI results.** Grayscale and color reconstructions from SR 1% to 60%.

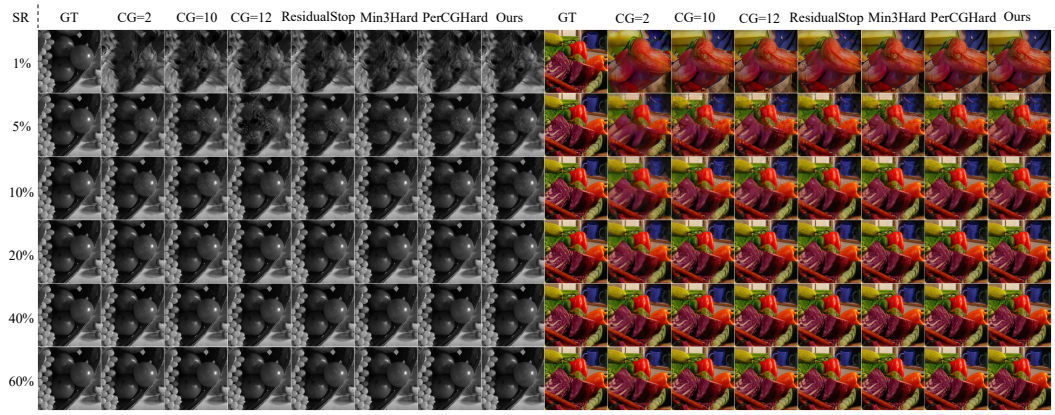


Figure 12: **Additional CG-control ablation.** Fixed-depth CG, hand-crafted controls, and Ours across SR 1%–60%.

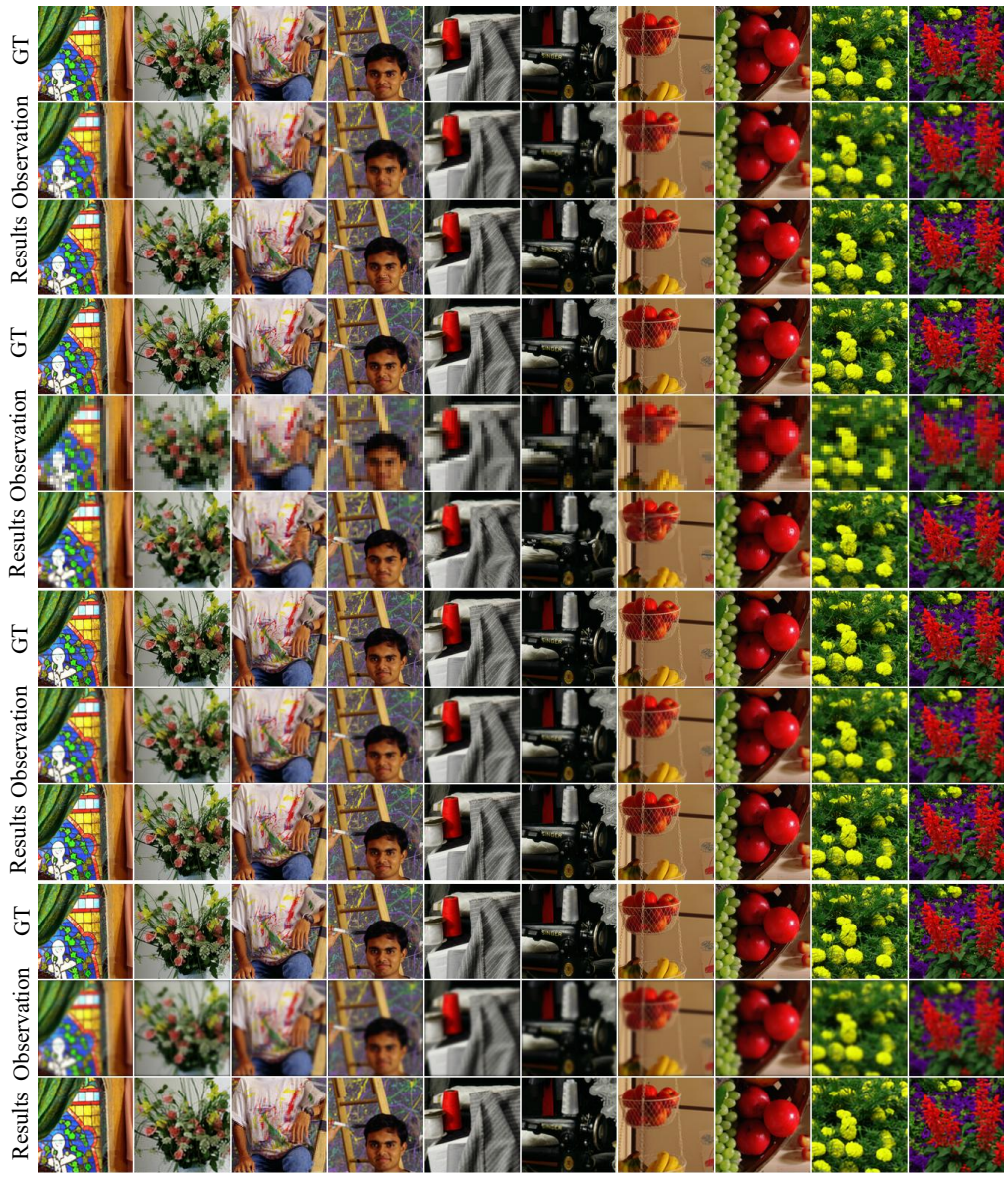


Figure 13: **Additional zero-shot transfer results.** Task groups are $\times 4$ SR, $\times 8$ SR, moderate Gaussian deblurring (31×31, $\sigma = 1.5$), and heavy Gaussian deblurring (61×61, $\sigma = 3.0$); rows show GT, observation, and Ours.