SplitIR: Operator-Splitting Reconstruction of Hidden Screen Content from Indirect Radiance

Anonymous authors
Paper under double-blind review

Abstract

Can an ordinary camera recover the content of a hidden screen from the light it casts on a visible wall? Diffuse scattering weakens and mixes structural cues, obscuring their correspondence with the target layout. Successive refinements can accumulate incompatible corrections in an unrestricted correction space and suppress weak high-frequency cues. We propose Operator-Splitting Reconstruction from Indirect Radiance (SplitIR), which separates reconstruction into potential assimilation and resolvent correction. The Potential Assimilation Module (PAM) uses a learned positive metric and bounded potential-guided transport to organize wall evidence into a coherent layout state. The Resolvent Correction Module (RCM) then analytically adjusts candidate corrections through a metric-aware proximal step and projects them into stage-conditioned feasible correction sets for local refinement. Across Screen, Chart, Password, and Website, SplitIR achieves the best PSNR, SSIM, and LPIPS among 22 representative baselines and reconstructs screen layouts, chart structures, and character patterns from diffuse wall observations. Ablations support the two-operator design, while controlled luminance and Gaussian-noise experiments show a persistent comparative advantage as observation quality degrades. Together, these results establish the initial feasibility of recovering hidden screen content with an ordinary camera observing wall radiance.

1 Introduction

Imagine a screen hidden behind an occluding wall, outside the camera's direct field of view. Fig. 1 shows the acquisition scenario and indirect light paths. A camera records a visible diffuse surface indirectly illuminated by the hidden display. Contributions from distant screen regions overlap into a weak, spatially mixed radiance pattern with strongly attenuated fine detail. Yet this pattern retains cues about the concealed content.

Non-line-of-sight imaging has shown that hidden scenes can be reconstructed from indirect measurements Faccio et al. (2020). Liu et al. Liu et al. (2023) reconstruct hidden-object albedo and normals from time-resolved photon measurements, while Virtual Scanning Cui et al. (2024a) recovers hidden-scene volumes from irregularly undersampled transients. Complementing these transient approaches, computational periscopy demonstrates that an ordinary camera can recover hidden two-dimensional scenes from occluder-induced penumbrae Saunders et al. (2019). This leads to our central question: using only ordinary-camera RGB observations of wall radiance, can a reconstruction model recover screen layouts, chart structures, and character patterns across four content categories?

Recovering hidden content from such observations presents two coupled difficulties. First, diffuse scattering disrupts the spatial organization of observed evidence, obscuring the correspondence between wall evidence and the evolving layout representation. Concatenation or residual fusion leaves the discrepancy metric, evidence-transfer direction, and stage-dependent update implicit, so long-range structure can blur or distort. Second, successive decoder branches propose complementary corrections whose direct accumulation occurs in an unrestricted correction space, allowing residual interactions to overwrite recovered evidence and suppress weak high-frequency cues. Mixed wall evidence must first form a coherent screen layout; subsequent refinement must then preserve the weak structural cues that survive diffuse transport.

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.

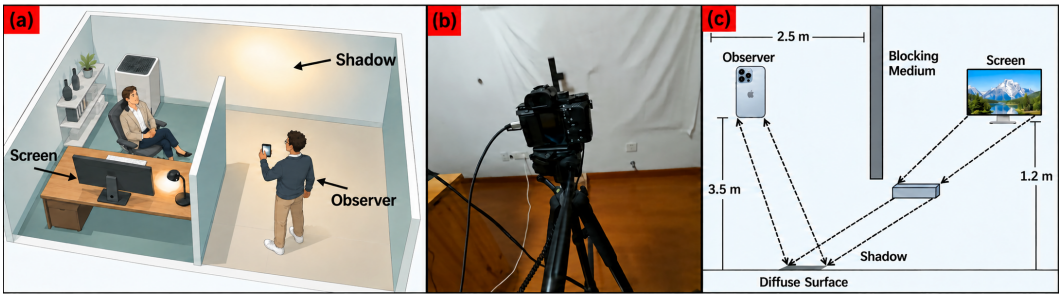


Figure 1: Experimental setup for hidden-screen reconstruction: (a) scenario illustration, (b) acquisition setup, and (c) scene geometry and indirect light paths.

To address these challenges, we propose Operator-Splitting Reconstruction from Indirect Radiance (SplitIR), which separates reconstruction into potential assimilation and resolvent correction. The Potential Assimilation Module (PAM) uses learned discrepancy geometry and bounded potential-guided transport to assimilate mixed wall evidence into a shared layout state. The decoded base reconstruction is then refined by the Resolvent Correction Module (RCM), which analytically adjusts each candidate correction through a metric-aware proximal step and projects it into a learned stage-conditioned feasible set before write-back. This sequence connects mixed wall radiance to coherent layout formation and regulated image refinement. Across four screen-content categories, SplitIR leads 22 representative baselines on PSNR, SSIM, and LPIPS while recovering layouts, chart structures, and character patterns. Luminance, Gaussian-noise, and real-acquisition tests further examine robustness to observation changes. Our contributions are:

- We demonstrate the initial feasibility of recovering structured hidden-screen content from ordinary-camera observations of diffuse wall radiance, including screen layouts, chart structures, and character patterns.
- We introduce SplitIR, which couples potential-guided evidence assimilation with analytic correction within learned feasible sets: PAM forms a shared layout state through learned discrepancy geometry and bounded transport, while RCM regulates stage-wise refinements through metric-aware proximal adjustment and feasible-set projection.
- We evaluate this reconstruction capability across four screen-content categories and 22 representative baselines, and connect the comparative results to luminance and noise changes, real-acquisition robustness, spatial interventions, and stage-wise operator analyses.

2 Related Work

Generic Image Restoration Models Generic image restoration has developed strong data-driven models for diverse degradations. MPRNet Zamir et al. (2021) progressively refines intermediate estimates, Restormer Zamir et al. (2022) strengthens long-range interactions, and AdaIR Cui et al. (2025) introduces adaptive frequency processing for all-in-one restoration. Model-based restoration provides complementary structured priors through plug-and-play priors Venkatakrishnan et al. (2013), regularization by denoising Reehorst & Schniter (2019), consensus equilibrium Buzzard et al. (2018), and diffusion posterior sampling Chung et al. (2023). These approaches organize image recovery through learned restoration models and prior-guided updates. SplitIR first assimilates mixed wall evidence into a shared layout state, then computes image refinements through learned, stage-conditioned correction geometry.

Hidden-Scene Reconstruction Hidden-scene reconstruction recovers occluded content from indirect measurements formed after interaction with relay surfaces. Early transient systems use ultrafast time-of-flight measurements and backprojection Velten et al. (2012), followed by model-based formulations such as the light-cone transform O'Toole et al. (2018) and frequency-wavenumber migration Lindell et al. (2019). Learned transient approaches extend this line through feature embeddings Chen et al. (2020), neural transient fields Shen et al. (2021), transformer-based reconstruction Li et al. (2023a), and Virtual Scanning, which reconstructs hidden scenes from irregularly undersampled transients Cui et al. (2024a). Ordinary-camera hidden-scene imaging has also developed

through penumbra-based computational periscopy Saunders et al. (2019) and two-edge-resolved imaging Czajkowski & Murray-Bruce (2024). Learning-based passive reconstruction organizes indirect observations through latent optimal-transport mappings Geng et al. (2022) or light-transport-conditioned reconstruction and reprojection Zhang et al. (2025). SplitIR makes structured screen recovery explicit at two successive levels: PAM forms a shared screen-layout state from mixed wall evidence, and RCM analytically regulates the corrections that refine the decoded screen image.

3 METHOD

3.1 OVERALL ARCHITECTURE

Two PAM calls organize wall evidence into a shared layout state decoded as a base reconstruction, followed by four RCM stages that adjust and project candidate corrections. Fig. 2 summarizes this wall-to-screen reconstruction path.

We express spatial mixing in indirect observations as

$$x(\xi) = \int_{\Omega} K_{\vartheta}(\xi, \zeta) y(\zeta)\, d\zeta + \varepsilon(\xi), \qquad K_{\vartheta}(\xi, \zeta) \geq 0, \tag{1}$$

where x denotes the wall observation, y the hidden screen, K_{ϑ} the entrywise non-negative diffuse light-transport operator, and ε measurement noise. The mixing couples distant screen regions in the observation domain, producing the observation–layout mismatch addressed by PAM.

SplitIR constructs a 19-channel online evidence field from four RGB terms and seven single-channel structural statistics. The evidence construction is shared: PAM uses $\mathcal{E}(x_i, b_i)$, while RCM uses $\mathcal{E}(f_k, a_k)$. In RCM, concatenating this 19-channel field with the RGB wall observation x and hint h produces the 25-channel conditioning input.

Multi-scale inputs form a shared layout state, while base and side readouts construct an auxiliary hint:

$$\begin{aligned} x_s &= D_s x, \quad s \in \mathcal{S}, & F_0 &= \mathcal{I}_{\theta}(x, \{x_s\}_{s\in\mathcal{S}}), \\ h &= \Pi_{[0,1]}\left(\omega_0 b + \omega_1 x + \omega_2 \uparrow b_2 + \omega_3 \uparrow b_4\right), & \omega_j &= \frac{\exp(\alpha_j)}{\sum_{\ell=0}^{3} \exp(\alpha_\ell)}. \end{aligned} \tag{2}$$

Here, D_s forms the multi-scale inputs and $\mathcal{I}_{\theta}$ constructs the shared layout state. The half-scale readout b_2 and quarter-scale readout b_4 are bilinearly aligned to the base reconstruction before hint fusion, and the four aligned RGB terms are combined through learned softmax weights to form h.

Two PAM stages successively assimilate indirect-radiance evidence into the shared layout state at different scales. Four ordered RCM stages then refine the base reconstruction through frequency, morphology, scanline, and directional-line candidate corrections, each regulated by a stage-conditioned feasible correction mechanism.

3.2 POTENTIAL ASSIMILATION MODULE

PAM addresses the mismatch between diffusely mixed wall evidence and the current layout representation. It maps both to a shared coordinate space, measures their discrepancy with a learned positive metric, and converts the descent direction into bounded stage-conditioned transport.

At PAM call i, $x_i = D_i x$ and side readout b_i are aligned to the current scale. Parameters are shared across calls, while e_i identifies the assimilation stage:

$$\begin{aligned} s_i &= \mathcal{S}(F_i), \qquad o_i = \mathcal{O}\big(\mathcal{E}(x_i, b_i)\big) + e_i, \\ d_i &= s_i - o_i, \qquad F_{i+1} = F_i^{+}, \qquad i \in \mathcal{I} = \{0, 1\}. \end{aligned} \tag{3}$$

Here, $\mathcal{S}$ and $\mathcal{O}$ map the layout state and observation evidence into a shared coordinate space, and d_i is their discrepancy. The stage embedding e_i provides an explicit identifier for each assimilation call within the shared conditioning map.

PAM weights representation dimensions with a positive diagonal metric at each spatial location:

$$M_i^p(\xi) = \mathrm{Diag}\big(m_i^p(\xi)\big), \qquad m_i^p = \mu_p \mathbf{1} + \sigma\big(\Psi(s_i, o_i, |d_i|)\big), \qquad \mu_p > 0, \quad m_i^p(\xi) > 0. \tag{4}$$

The learned metric M_i^p reweights discrepancy dimensions before assimilation.

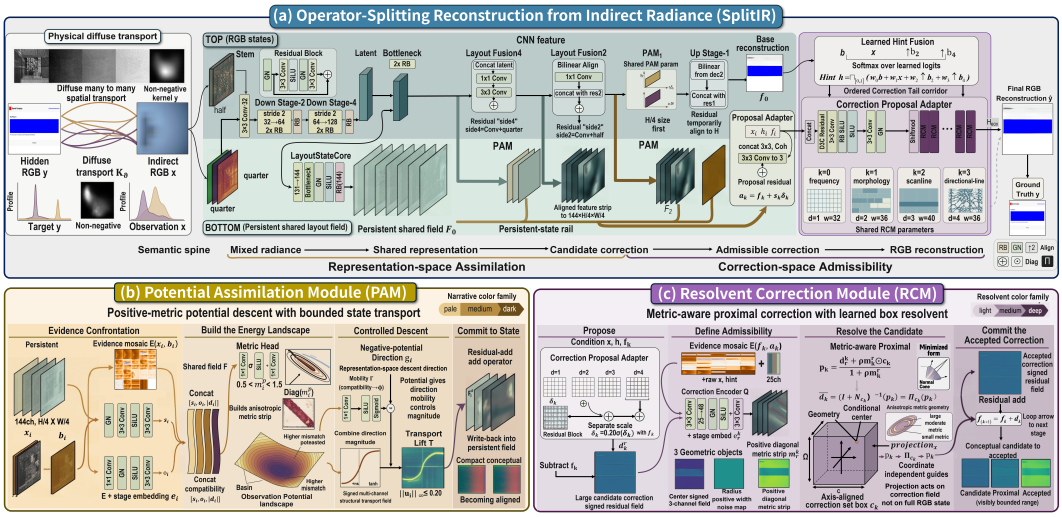


Figure 2: Overall architecture of SplitIR with PAM and RCM.

PAM defines a metric-weighted local energy

$$\widetilde{\phi}_i(\xi) = \frac{1}{2} d_i(\xi)^\top M_i^p(\xi) d_i(\xi), \qquad \phi_i(\xi) = \frac{1}{C_p}\widetilde{\phi}_i(\xi) = \frac{1}{2C_p}\sum_{\alpha=1}^{C_p} m_{i,\alpha}^p(\xi)\big(d_i^\alpha(\xi)\big)^2. \tag{5}$$

The normalized density ϕ_i is the quantity used by the mobility branch and the stage diagnostics, while $\widetilde{\phi}_i$ defines the local quadratic geometry. At each assimilation step, the current metric defines this local discrepancy geometry and is held fixed when computing its descent direction.

PAM converts the local-energy descent direction into a bounded, stage-conditioned transport update:

$$\begin{aligned} g_i &= -\nabla_{d_i}\widetilde{\phi}_i = -M_i^p d_i, \qquad \eta_i = \eta_0 \sigma\big(\Gamma(s_i, o_i, |d_i|, \phi_i)\big), \\ u_i &= \eta_i \odot \tanh\big(\mathcal{T} g_i\big), \qquad F_i^+ = F_i + u_i. \end{aligned} \tag{6}$$

Here, $\mathcal{T}$ maps the representation-space descent direction to the layout state, while Γ predicts its stage-conditioned mobility. Since $0 \le \eta_i \le \eta_0$ and $\tanh(\cdot)$ is bounded, $\|u_i\|_\infty \le \eta_0$. Two PAM calls successively update the shared layout state.

3.3 Resolvent Correction Module

After PAM, RCM regulates decoder-generated corrections before image refinement. At each stage, it predicts the correction geometry, computes a metric-aware proximal adjustment, and projects the result into a learned coordinate-wise feasible set.

At RCM stage $k \in \mathcal{K} = \{0, 1, 2, 3\}$, proposal a_k defines candidate correction d_k^r relative to reconstruction f_k. A shared conditioning map combines the candidate residual, input evidence, and auxiliary hint into a correction representation:

$$a_k = \mathcal{P}_k(x, h, f_k), \qquad d_k^r = a_k - f_k, \qquad q_k = \mathcal{Q}\big(\mathrm{concat}\big(\mathcal{E}(f_k, a_k), x, h\big)\big) + e_k^r. \tag{7}$$

Here, $\mathcal{P}_k$ generates the stage proposal, while $\mathcal{Q}$ maps the 25-channel concatenation of correction evidence, wall observation, and hint into correction representation q_k. The stage embedding e_k^r explicitly conditions the shared correction map on its position in the refinement sequence.

From q_k, RCM predicts the correction geometry through a center, coordinate-wise radius, and positive diagonal metric:

$$\begin{aligned} c_k &= c_0 \tanh\big(\mathcal{H}_c(q_k)\big), \qquad \varrho_k = \varrho_{\min} + \mathrm{softplus}\big(\mathcal{H}_\varrho(q_k) + b_\varrho\big), \\ M_k^r &= \mathrm{Diag}\big(\mu_r \mathbf{1} + \sigma\big(\mathcal{H}_m(q_k)\big)\big), \qquad C_k(\xi) = \big\{z \in \mathbb{R}^{C_r} : |z_\alpha - c_{k,\alpha}(\xi)| \le \varrho_{k,\alpha}(\xi),\ \forall \alpha\big\}. \end{aligned} \tag{8}$$

The center c_k and radius ϱ_k define the stage-conditioned feasible set C_k, while M_k^r weights the proximal adjustment. Each correction coordinate has its own learned admissible interval.

Before write-back, RCM balances candidate correction d_k^r with learned center c_k through a metric-aware proximal adjustment:

$$p_k = \underset{z}{\operatorname{argmin}} \left\{ \frac{1}{2}\|z - d_k^r\|_2^2 + \frac{\rho}{2}\|z - c_k\|_{M_k^r}^2 \right\}, \qquad \|v\|_M^2 := \langle v, Mv \rangle_F, \quad \rho > 0. \tag{9}$$

Here, ρ is the learned positive resolvent strength. Because M_k^r is diagonal, the proximal update has the closed form

$$p_k = (I + \rho M_k^r)^{-1} (d_k^r + \rho M_k^r c_k) = \frac{d_k^r + \rho\, m_k^r \odot c_k}{1 + \rho\, m_k^r}. \tag{10}$$

This metric-aware interpolation gives different correction coordinates different adjustment strengths between the candidate and learned center.

The proximal correction is projected into the learned feasible set through the resolvent:

$$\bar{d}_k = (I + N_{C_k})^{-1} (p_k) = \Pi_{C_k}(p_k) = \min\{\max\{p_k, c_k - \varrho_k\}, c_k + \varrho_k\}. \tag{11}$$

Here, N_{C_k} denotes the normal cone of C_k. With the predicted geometry fixed, the positive diagonal quadratic and coordinate-wise box make the constrained correction problem separable, giving

$$\bar{d}_k = \underset{z \in C_k}{\operatorname{argmin}} \left\{ \frac{1}{2}\|z - d_k^r\|_2^2 + \frac{\rho}{2}\|z - c_k\|_{M_k^r}^2 \right\}. \tag{12}$$

Each coordinate is obtained by clipping its unconstrained proximal solution to the corresponding admissible interval. Uformer forms its restored image by adding a learned residual to the degraded input Wang et al. (2022); with the predicted geometry fixed, RCM instead converts each proposed residual into the unique closed-form solution of the stage-conditioned constrained quadratic through proximal adjustment and box projection Parikh & Boyd (2014). The projected correction is then written back and mapped to the valid image range:

$$\widetilde{f}_{k+1} = f_k + \bar{d}_k, \qquad f_{k+1} = \Pi_{[0,1]}\left(\widetilde{f}_{k+1}\right). \tag{13}$$

Starting from the base reconstruction $f_0 = b$, four RCM stages follow frequency, morphology, scanline, and directional-line refinement. The corresponding proposal adapters use frequency-band operations, local morphological envelopes, bidirectional scanline memory, and fixed directional-line filtering, respectively. Each stage regenerates its proposal and correction geometry from the updated reconstruction, then computes the proximal and feasible corrections before write-back. After four stages, f_4 is the reconstructed hidden screen.

4 Experiment

4.1 Experimental Setup

We evaluate four paired indirect-radiance datasets: Screen, Chart, Password and Website. They contain 21,237 paired samples in total. Each pair contains a wall RGB observation and clear target resized to 256×256.

SplitIR is implemented in PyTorch and optimized with AdamW Loshchilov & Hutter (2019). All four content subsets use the same training configuration: 150 epochs, batch size 6, learning rate 1×10^{-5}, a five-times learning-rate multiplier for PAM and RCM parameters, $\beta = (0.9, 0.999)$, zero weight decay, automatic mixed precision, and seed 3407. Full and ablation models share initialization, optimization, training budget, and checkpoint selection; each ablation changes only its target component. A dataset-specific checkpoint is selected by SSIM on the evaluation partition and is then fixed for comparative evaluation, robustness tests, interventions, and diagnostic analyses.

4.2 Comparative Experiments

We compare SplitIR with 22 methods spanning hidden-scene reconstruction, general image restoration, and restoration under related degradations. All trainable baselines follow the same data split, input resolution, training budget, optimization protocol, and checkpoint-selection rule used for the comparison. Methods requiring task-specific input or output adaptation retain their own network

Table 1: Quantitative comparison on Screen, Chart, Password, and Website using PSNR↑, SSIM↑, and LPIPS↓. Best results are shown in **red**, and second-best results are shown in **blue**.

Method	Screen			Chart			Password			Website		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NLOS Reconstruction Methods												
DRS Shim et al. (2024)	20.44	0.6933	0.4040	13.11	0.6516	0.6510	8.76	0.7949	0.4282	18.38	0.7920	0.3800
DynamicNLOS Li et al. (2024)	21.42	0.7155	0.3793	16.57	0.7029	0.5045	14.05	0.8740	0.1374	18.95	0.7957	0.3548
Flatland Peña et al. (2025)	12.23	0.4734	0.6121	13.12	0.6516	0.6512	8.76	0.7949	0.4282	14.76	0.7703	0.4236
LIKAM Yu et al. (2023)	21.69	0.7267	0.3392	16.28	0.6991	0.5169	13.69	0.8735	0.1320	19.52	0.7971	0.3437
LMS-NLOS Chen et al. (2025)	21.96	0.7066	0.3501	16.68	0.7011	0.4914	13.56	0.8582	0.1432	19.71	0.7746	0.3297
NLOS-LPP Sun et al. (2025)	21.56	0.7143	0.3620	15.54	0.6874	0.5679	10.64	0.8147	0.2959	17.19	0.7788	0.4042
NLOS-LTM Zhang et al. (2025)	21.84	0.7262	0.3350	16.57	0.7038	0.5035	11.76	0.8406	0.2536	18.21	0.7805	0.3719
NLOS-OT Geng et al. (2022)	**24.10**	**0.7933**	**0.2619**	16.61	0.7057	**0.4632**	8.76	0.7949	0.4282	19.02	0.7935	0.3372
NLOST Li et al. (2023a)	14.78	0.5202	0.5412	14.62	0.6130	0.6138	8.89	0.7863	0.4419	15.73	0.7453	0.4418
NLOS-NeuS Fujimura et al. (2023)	12.32	0.4829	0.6445	13.20	0.6458	0.6518	8.76	0.7935	0.4287	15.09	0.7633	0.4186
TransiT Li et al. (2025)	20.53	0.6889	0.3688	13.10	0.6517	0.6511	8.76	0.7949	0.4282	18.73	0.7849	0.3434
UnderScanning Li et al. (2023b)	22.67	0.7567	0.3055	16.66	0.7051	0.4851	14.12	0.8819	0.1416	19.54	0.7965	0.3422
Virtual Scanning Cui et al. (2024a)	23.11	0.7690	0.2973	16.48	0.7039	0.4959	9.90	0.8078	0.3030	19.63	0.7975	0.3361
General-Purpose Image Restoration												
U-Net Ronneberger et al. (2015)	20.06	0.6691	0.3953	16.20	0.6836	0.5183	12.77	0.8563	0.1564	18.36	0.7650	0.3641
NAFNet Chen et al. (2022)	22.81	0.7513	0.3023	16.60	0.6985	0.4949	12.84	0.8470	0.1628	15.74	0.7400	0.4199
Uformer Wang et al. (2022)	21.41	0.7148	0.3546	**17.02**	**0.7179**	0.4894	14.58	0.8829	0.1176	20.07	0.8075	0.3101
ConvIR-Original Cui et al. (2024b)	22.84	0.7565	0.3242	15.73	0.6728	0.5801	**15.13**	**0.8910**	**0.1009**	**20.46**	**0.8179**	**0.2913**
AST Zhou et al. (2024)	22.19	0.7356	0.3182	16.94	0.7173	0.4712	14.70	0.8790	0.1222	15.43	0.7429	0.4239
Related-Degradation Restoration Methods												
PromptIR Potlapalli et al. (2023)	23.27	0.7689	0.2921	13.11	0.6517	0.6513	8.76	0.7949	0.4282	19.57	0.8020	0.3371
InstructIR Conde et al. (2024)	10.18	0.4273	0.6598	9.88	0.5649	0.7489	8.48	0.6920	0.5844	12.57	0.7279	0.5067
DarkIR Feijoo et al. (2025)	20.83	0.6791	0.3790	16.25	0.6876	0.5512	13.76	0.8572	0.1626	19.67	0.7762	0.3333
CIDNet Yan et al. (2025)	21.14	0.7141	0.3728	16.40	0.6982	0.5105	13.17	0.8605	0.1804	19.26	0.7882	0.3550
SplitIR (Ours)	**27.16**	**0.8506**	**0.1781**	**17.36**	**0.7408**	**0.4552**	**15.38**	**0.9000**	**0.0924**	**20.74**	**0.8313**	**0.2756**

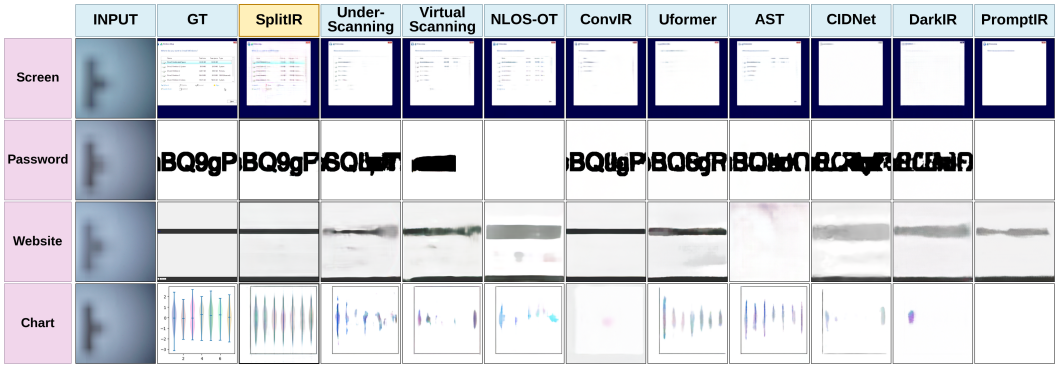


Figure 3: Qualitative comparison on Screen, Password, Website, and Chart. Columns show the wall observation (INPUT), ground truth (GT), SplitIR, and comparison methods.

structure while using the same paired RGB reconstruction setting; the selected checkpoint for each method and dataset is fixed across subsequent evaluations.

SplitIR ranks first on PSNR, SSIM, and LPIPS across all four content categories in Table 1. Geng et al. Geng et al. (2022) reconstruct hidden images by mapping wall observations into an autoencoder-learned target latent space through optimal transport. SplitIR reaches 27.16 dB PSNR on Screen versus 24.10 dB for NLOS-OT in Table 1, a 3.06-dB gain, by first assimilating wall evidence into a shared layout state and then analytically correcting the decoded image. Fig. 3 connects the quantitative advantage to the recovered content: broad Screen regions retain their organization, Password characters preserve localized strokes and spacing, Website reconstructions maintain page-level organization and local interface boundaries, and Chart components preserve their geometric arrangement. Together, these results show that ordinary-camera wall observations support reconstruction at both layout and local-structure levels.

4.3 Ablation Studies

We examine how the two operators contribute to reconstruction through five PAM variants and five RCM variants, with quantitative results in Table 2 and qualitative comparisons in Fig. 4. Full SplitIR

Table 2: Ablation study on Screen, Chart, Password, and Website using PSNR↑, SSIM↑, and LPIPS↓.

Method	Model	Screen			Chart			Password			Website		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Potential Assimilation Module													
w/o PAM	-	22.648	0.769	0.318	16.671	0.724	0.506	13.648	0.870	0.145	20.271	0.818	0.302
w/o PAM Stage Embedding	-	**22.745**	0.768	**0.317**	16.701	0.724	0.504	13.645	0.869	0.142	20.318	0.818	0.302
PAM Metric → Identity	Identity	22.172	0.754	0.334	16.626	0.723	0.508	13.088	0.861	0.177	16.524	0.787	0.408
PAM → Residual Fusion	ResNet	22.671	0.769	**0.317**	16.740	0.725	0.505	13.713	0.870	0.142	20.221	0.818	0.310
PAM Mobility → Constant	Constant	22.733	**0.773**	0.318	16.731	0.725	0.504	13.189	0.862	0.173	20.088	0.819	0.316
Resolvent Correction Module													
w/o RCM	-	22.543	0.770	0.321	16.727	0.726	0.501	**14.228**	**0.878**	**0.129**	**20.425**	**0.822**	**0.300**
w/o RCM Stage Embedding	-	22.305	0.758	0.331	16.664	0.722	0.507	13.323	0.864	0.168	20.081	0.815	0.317
w/o RCM Box Projection	-	21.884	0.747	0.346	**16.813**	0.727	0.500	13.549	0.866	0.156	20.276	0.821	0.313
RCM → Gated Residual	GResNet	22.546	0.768	0.322	16.807	**0.728**	**0.497**	14.012	0.876	0.132	20.283	0.820	0.306
RCM Metric → Identity	Identity	22.189	0.753	0.334	16.642	0.723	0.509	13.223	0.863	0.171	20.290	0.817	0.304
SplitIR (Full)	SplitIR	**27.163**	**0.851**	**0.178**	**17.358**	**0.741**	**0.455**	**15.384**	**0.900**	**0.092**	**20.767**	**0.827**	**0.275**

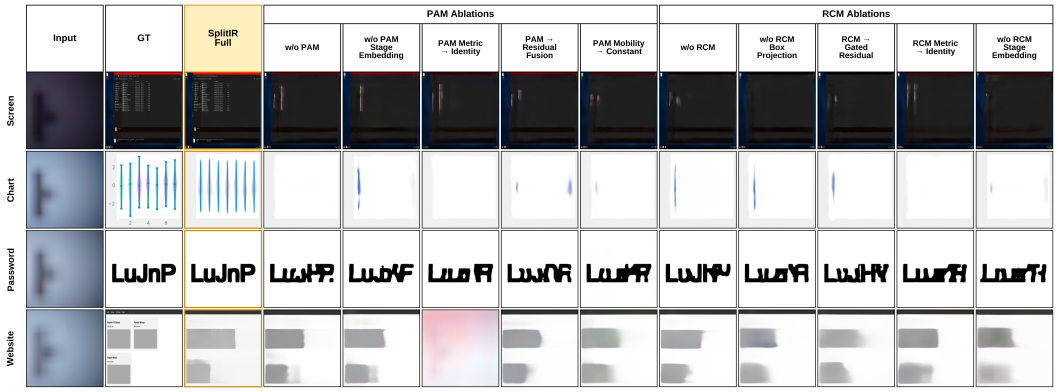


Figure 4: Qualitative ablation comparison of SplitIR across Screen, Chart, Password, and Website, including the full model and PAM and RCM variants.

and all ablations use the same initialization, training budget, optimization settings, data split, and model-selection rule; each variant changes only the target component. PAM tests module or stage-embedding removal, identity-metric and residual-fusion replacement, and constant mobility, while RCM tests module, stage-embedding, or box-projection removal, gated-residual replacement, and identity-metric replacement.

4.3.1 PAM Ablation Analysis

The strongest contrast among the PAM variants concerns the learned weighting of observation–layout differences. Replacing the learned metric with the identity yields an SSIM of 0.754 on Screen, compared with 0.851 for the full model. This contrast highlights the value of assigning different weights to the mixed evidence before it enters the layout state. The stage embedding identifies the assimilation call, while adaptive mobility controls the strength of the evidence-driven update.

4.3.2 RCM Ablation Analysis

Across all four datasets, the full SplitIR model outperforms all five RCM variants across the reported metrics. On Screen, for example, removing box projection produces the weakest RCM result across all three metrics, with 21.884 dB PSNR, 0.747 SSIM, and 0.346 LPIPS, compared with 27.163 dB, 0.851, and 0.178 for the full model. The learned metric and stage embedding shape each correction according to the current reconstruction state and refinement stage. RCM thus converts each proposed refinement into an explicit feasible correction before updating the screen image. With both modules retained, replacing PAM's metric with the identity and removing RCM's box projection reduce Screen SSIM by 0.097 and 0.104, respectively. These losses support the contributions of evidence weighting during assimilation and feasibility constraints during correction to structural recovery. Consistent with the quantitative results, Fig. 4 shows that the full model recovers more complete screen layouts and chart structures, whereas many ablated variants preserve only fragments of these structures.

Table 3: Robustness comparison on the Website dataset under progressive luminance reduction and Gaussian noise at different SNR levels, evaluated using LPIPS↓.

Method	(a) Brightness Reduction ΔL (nit)									(b) Gaussian Noise: SNR (dB)						
	0	10	20	30	40	50	60	70	80	60	55	50	45	40	35	30
NLOS Reconstruction Methods																
DRS	0.464	0.470	0.477	0.486	0.501	0.527	0.556	0.590	0.636	0.473	0.473	0.473	0.473	0.473	0.473	0.474
Flatland	0.522	0.523	0.524	0.525	0.527	0.530	0.533	0.537	0.542	0.529	0.529	0.529	0.529	0.529	0.529	0.530
LIKAM	0.416	0.421	0.427	0.435	0.448	0.467	0.506	0.570	0.665	0.427	0.427	0.427	0.427	0.427	0.427	0.431
LMS-NLOS	0.403	0.414	0.424	0.437	0.466	0.500	0.539	0.574	0.602	0.417	0.417	0.418	0.418	0.419	0.424	0.440
NLOS-LPP	0.491	0.500	0.508	0.523	0.547	0.593	0.633	0.692	0.746	0.500	0.500	0.500	0.500	0.500	0.500	0.500
NLOS-LTM	0.448	0.455	0.473	0.510	0.555	0.605	0.659	0.710	0.763	0.457	0.457	0.457	0.457	0.457	0.457	0.457
NLOS-NeuS	0.522	0.525	0.528	0.531	0.534	0.538	0.543	0.548	0.553	0.532	0.532	0.532	0.532	0.533	0.535	0.541
NLOS-OT	0.411	0.423	0.447	0.478	0.511	0.550	0.595	0.633	0.668	0.421	0.422	0.422	0.421	0.421	0.422	0.423
NLOST	0.558	0.569	0.584	0.613	0.656	0.709	0.766	0.805	0.829	0.962	0.962	0.962	0.962	0.962	0.962	0.962
TransiT	0.415	0.419	0.421	0.429	0.444	0.469	0.513	0.568	0.623	0.423	0.423	0.423	0.423	0.423	0.423	0.424
Under-Scanning	0.417	0.420	0.423	0.435	0.462	0.492	0.532	0.594	0.678	0.427	0.427	0.427	0.427	0.427	0.428	0.431
Virtual-Scanning	0.407	0.411	0.418	0.431	0.450	0.476	0.519	0.576	0.659	0.418	0.418	0.418	0.418	0.419	0.421	0.428
General-Purpose Image Restoration																
NAFNet	0.402	0.406	0.410	0.419	0.435	0.458	0.501	0.568	0.636	**0.412**	**0.412**	**0.412**	**0.412**	**0.412**	**0.413**	**0.416**
U-Net	0.440	0.458	0.489	0.528	0.572	0.618	0.653	0.690	0.720	0.453	0.453	0.453	0.454	0.455	0.459	0.476
Related-Degradation Restoration Methods																
CIDNet	0.418	0.423	0.430	0.435	0.446	0.464	**0.487**	**0.509**	**0.539**	0.428	0.428	0.429	0.429	0.429	0.429	0.431
DarkIR	**0.387**	**0.391**	**0.398**	**0.409**	**0.428**	**0.454**	0.494	0.544	0.602	0.962	0.962	0.962	0.962	0.962	0.962	0.962
InstructIR	0.751	0.747	0.753	0.755	0.759	0.764	0.764	0.767	0.769	0.751	0.751	0.751	0.750	0.749	0.747	0.747
SplitIR (Ours)	**0.328**	**0.334**	**0.340**	**0.344**	**0.351**	**0.361**	**0.374**	**0.393**	**0.420**	**0.356**	**0.356**	**0.356**	**0.357**	**0.358**	**0.363**	**0.383**

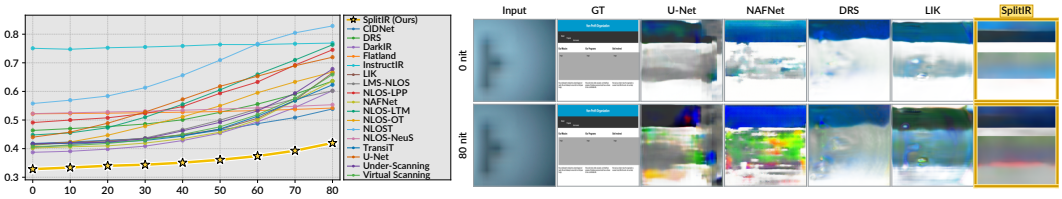


Figure 5: Luminance robustness on Website, showing LPIPS trends and representative qualitative comparisons under progressive brightness reduction.

4.4 Robustness Evaluation

We evaluate Website robustness under luminance reduction and Gaussian noise using SplitIR and the same 17 baselines. All methods use the same prepared observations at each degradation level and fixed dataset-specific checkpoints.

4.4.1 Luminance Robustness

We evaluate luminance robustness on Website by progressively reducing the nominal luminance of the wall observations from 300 to 220 nit-equivalent in 10-nit-equivalent steps. As shown in Table 3(a) and Fig. 5, SplitIR retains the lowest LPIPS throughout the luminance sweep. Yan et al. Yan et al. (2025) introduce HVI and CIDNet to enhance low-light images through separate color and intensity processing. At the strongest 80-nit-equivalent reduction in our Website evaluation, SplitIR reaches 0.420 LPIPS versus 0.539 for CIDNet, while retaining 15.455 dB PSNR and 0.791 SSIM at the nominal 220-nit-equivalent level.

Reducing luminance weakens the contrast that carries the hidden layout while leaving the target arrangement unchanged. SplitIR preserves the main page regions as the wall signal darkens. The remaining evidence first supports layout formation through PAM, and RCM subsequently refines the decoded image through corrections constrained by the learned stage-conditioned geometry.

4.4.2 Noise Robustness

Gaussian noise is evaluated from 60 to 30 dB SNR in 5-dB steps. As shown in Table 3(b) and Fig. 6, SplitIR retains the lowest LPIPS throughout the sweep, reaching 0.383 at 30 dB versus 0.416 for NAFNet Chen et al. (2022), with a 0.633-dB PSNR advantage.

Brightness reduction weakens radiometric evidence, whereas Gaussian perturbation adds spatially distributed uncertainty to that evidence. Across the tested range, broader layout cues remain comparatively persistent while local wall evidence becomes less reliable. The observed response is

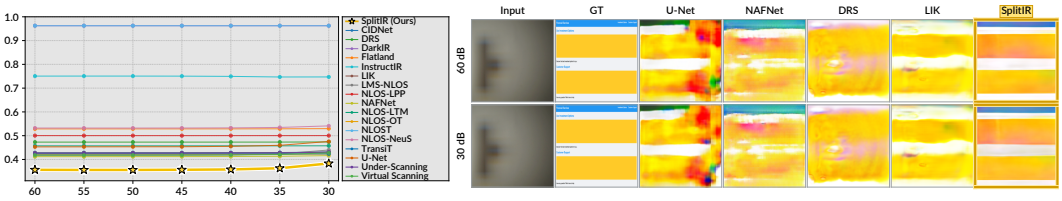


Figure 6: Gaussian-noise robustness on Website, showing LPIPS trends and representative qualitative comparisons under decreasing SNR.

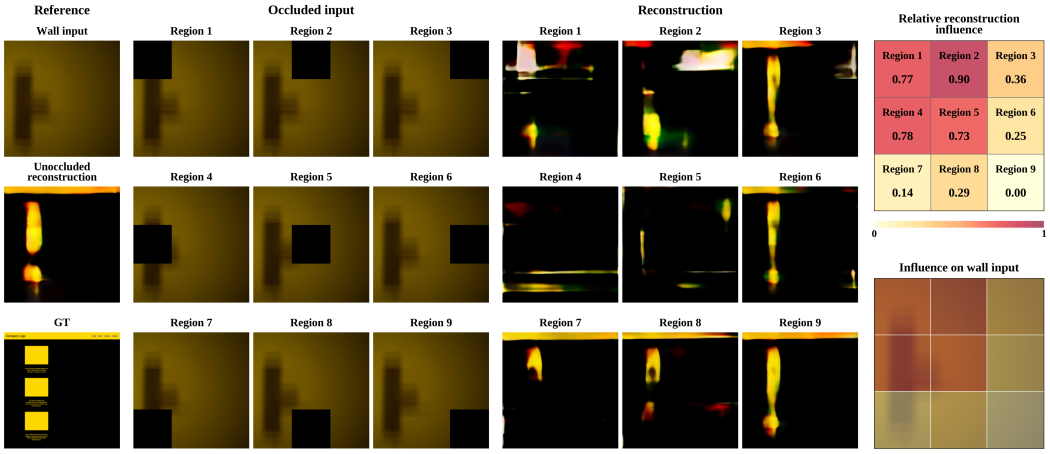


Figure 7: Occlusion analysis of SplitIR on Website. Individually masking the clean wall observation's 3×3 regions yields reconstructions and a relative influence map revealing uneven sensitivity.

consistent with first organizing these cues into a layout state and then regulating the uncertain refinements generated from it.

4.5 Spatial Evidence Analysis

Fig. 7 reveals spatially uneven reconstruction sensitivity to regional occlusion. In the Website example, masking Regions 1, 2, 4, and 5 disrupts the recovered header and vertical foreground arrangement, with relative reconstruction influence scores of 0.73–0.90, whereas masking Region 9 leaves the coarse reconstruction largely unchanged. The strong response in Region 2 further extends this sensitivity beyond the most prominent shadow area. Moreover, one local mask can alter several separated output structures, indicating nonlocal dependencies consistent with diffuse transport mixing screen contributions across the wall. This behavior aligns with SplitIR's formulation: PAM organizes mixed evidence into a shared layout state, while RCM regulates subsequent refinements. These observations confirm that diffuse wall radiance retains recoverable information about concealed screen content, enabling reconstruction using only ordinary-camera observations.

5 Conclusion

This work demonstrates the initial feasibility of recovering hidden screen content from diffuse wall radiance outside the display's direct view. SplitIR addresses two coupled reconstruction difficulties through operator splitting: PAM assimilates spatially mixed evidence into a consistent layout representation, while RCM maps candidate corrections into learned stage-conditioned feasible geometry. Across four screen-content datasets, SplitIR achieves the best PSNR, SSIM, and LPIPS among 22 representative baselines and reconstructs clearer text, chart structures, and interface layouts. Ablations support the distinct roles of PAM and RCM, while controlled luminance and Gaussian-noise tests show that the advantage persists as observations degrade. Together, the results show that diffuse wall radiance retains recoverable information about concealed screen content and supports hidden-display reconstruction with ordinary camera observations.

REFERENCES

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

Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *Computer Vision – ECCV 2022*, pp. 17–33. Springer Nature Switzerland, 2022. doi: 10.1007/978-3-031-20071-7_2.

Pengyun Chen, Shuang Cui, Ning Cao, Wenhao Zhang, Pengfei Wang, Shaohui Jin, and Mingliang Xu. Lightweight multi-scale feature fusion with attention guidance for passive non-line-of-sight imaging. *The Visual Computer*, 41(10):7767–7780, 2025. doi: 10.1007/s00371-025-03837-5.

Wenzheng Chen, Fangyin Wei, Kiriakos N. Kutulakos, Szymon Rusinkiewicz, and Felix Heide. Learned feature embeddings for non-line-of-sight imaging and recognition. *ACM Transactions on Graphics*, 39(6), 2020. doi: 10.1145/3414685.3417825.

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

Marcos V. Conde, Gregor Geigle, and Radu Timofte. InstructIR: High-quality image restoration following human instructions. In *Computer Vision – ECCV 2024*, pp. 1–21. Springer Nature Switzerland, 2024. doi: 10.1007/978-3-031-72764-1_1.

Xingyu Cui, Huanjing Yue, Song Li, Xiangjun Yin, Yusen Hou, Yun Meng, Kai Zou, Xiaolong Hu, and Jingyu Yang. Virtual scanning: Unsupervised non-line-of-sight imaging from irregularly undersampled transients. In *Advances in Neural Information Processing Systems*, volume 37, pp. 109381–109406, 2024a. doi: 10.52202/079017-3472.

Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Revitalizing convolutional network for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12): 9423–9438, 2024b. doi: 10.1109/TPAMI.2024.3419007.

Yuning Cui, Syed Waqas Zamir, Salman H. Khan, Alois Knoll, Mubarak Shah, and Fahad Shahbaz Khan. AdaIR: Adaptive all-in-one image restoration via frequency mining and modulation. In *The Thirteenth International Conference on Learning Representations*, 2025.

Robinson Czajkowski and John Murray-Bruce. Two-edge-resolved three-dimensional non-line-of-sight imaging with an ordinary camera. *Nature Communications*, 15:1162, 2024. doi: 10.1038/s41467-024-45397-7.

Daniele Faccio, Andreas Velten, and Gordon Wetzstein. Non-line-of-sight imaging. *Nature Reviews Physics*, 2(6):318–327, 2020. doi: 10.1038/s42254-020-0174-8.

Daniel Feijoo, Juan C. Benito, Alvaro Garcia, and Marcos V. Conde. DarkIR: Robust low-light image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10879–10889, 2025. doi: 10.1109/CVPR52734.2025.01016.

Yuki Fujimura, Takahiro Kushida, Takuya Funatomi, and Yasuhiro Mukaigawa. NLOS-NeuS: Non-line-of-sight neural implicit surface. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10498–10507, 2023. doi: 10.1109/ICCV51070.2023.00966.

Ruixu Geng, Yang Hu, Zhi Lu, Cong Yu, Houqiang Li, Hengyu Zhang, and Yan Chen. Passive non-line-of-sight imaging using optimal transport. *IEEE Transactions on Image Processing*, 31: 110–124, 2022. doi: 10.1109/TIP.2021.3128312.

Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey Hinton. Similarity of neural network representations revisited. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3519–3529, 2019.

Ruiqian Li, Siyuan Shen, Suan Xia, Ziheng Wang, Xingyue Peng, Chengxuan Song, Yingsheng Zhu, Tao Wu, Shiying Li, and Jingyi Yu. TransiT: Transient transformer for non-line-of-sight videography. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 27542–27551, 2025. doi: 10.1109/ICCV51701.2025.02557.

Yue Li, Jiayong Peng, Juntian Ye, Yueyi Zhang, Feihu Xu, and Zhiwei Xiong. NLOST: Non-line-of-sight imaging with transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13313–13322, 2023a. doi: 10.1109/CVPR52729.2023.01279.

Yue Li, Yueyi Zhang, Juntian Ye, Feihu Xu, and Zhiwei Xiong. Deep non-line-of-sight imaging from under-scanning measurements. In *Advances in Neural Information Processing Systems*, volume 36, pp. 59095–59106, 2023b. doi: 10.52202/075280-2578.

Yue Li, Yi Sun, Shida Sun, Juntian Ye, Yueyi Zhang, Feihu Xu, and Zhiwei Xiong. Toward dynamic non-line-of-sight imaging with mamba enforced temporal consistency. In *Advances in Neural Information Processing Systems*, volume 37, pp. 126452–126473, 2024. doi: 10.52202/079017-4016.

David B. Lindell, Gordon Wetzstein, and Matthew O'Toole. Wave-based non-line-of-sight imaging using fast f-k migration. *ACM Transactions on Graphics*, 38(4):1–13, 2019. doi: 10.1145/3306346.3322937.

Xintong Liu, Jianyu Wang, Leping Xiao, Zuoqiang Shi, Xing Fu, and Lingyun Qiu. Non-line-of-sight imaging with arbitrary illumination and detection pattern. *Nature Communications*, 14(1): 3230, 2023. doi: 10.1038/s41467-023-38898-4.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

Matthew O'Toole, David B. Lindell, and Gordon Wetzstein. Confocal non-line-of-sight imaging based on the light-cone transform. *Nature*, 555(7696):338–341, 2018. doi: 10.1038/nature25489.

Neal Parikh and Stephen Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1 (3):123–231, 2014.

María Peña, Diego Gutierrez, and Julio Marco. Looking around flatland: End-to-end 2D real-time NLOS imaging. *IEEE Transactions on Computational Imaging*, 11:189–200, 2025. doi: 10.1109/TCI.2025.3536092.

Vaishnav Potlapalli, Syed Waqas Zamir, Salman Khan, and Fahad Shahbaz Khan. PromptIR: Prompting for all-in-one image restoration. In *Advances in Neural Information Processing Systems*, volume 36, pp. 71275–71293, 2023. doi: 10.52202/075280-3121.

Edward T. Reehorst and Philip Schniter. Regularization by denoising: Clarifications and new interpretations. *IEEE Transactions on Computational Imaging*, 5(1):52–67, 2019. doi: 10.1109/TCI.2018.2880326.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241. Springer International Publishing, 2015. doi: 10.1007/978-3-319-24574-4_28.

Charles Saunders, John Murray-Bruce, and Vivek K. Goyal. Computational periscopy with an ordinary digital camera. *Nature*, 565:472–475, 2019. doi: 10.1038/s41586-018-0868-6.

Siyuan Shen, Zi Wang, Ping Liu, Zhengqing Pan, Ruiqian Li, Tian Gao, Shiying Li, and Jingyi Yu. Non-line-of-sight imaging via neural transient fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(7):2257–2268, 2021. doi: 10.1109/TPAMI.2021.3076062.

Hyunbo Shim, In Cho, Daekyu Kwon, and Seon Joo Kim. Domain reduction strategy for non-line-of-sight imaging. In *Computer Vision – ECCV 2024*, pp. 75–92. Springer Nature Switzerland, 2024. doi: 10.1007/978-3-031-72751-1_5.

Shida Sun, Yue Li, Yueyi Zhang, and Zhiwei Xiong. Generalizable non-line-of-sight imaging with learnable physical priors. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 25040–25049, 2025. doi: 10.1109/ICCV51701.2025.02322.

Andreas Velten, Thomas Willwacher, Otkrist Gupta, Ashok Veeraraghavan, Moungi G. Bawendi, and Ramesh Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Communications*, 3:745, 2012. doi: 10.1038/ncomms1747.

Singanallur V. Venkatakrishnan, Charles A. Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *IEEE Global Conference on Signal and Information Processing*, pp. 945–948, 2013. doi: 10.1109/GlobalSIP.2013.6737048.

Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17683–17693, 2022. doi: 10.1109/CVPR52688.2022.01716.

Qingsen Yan, Yixu Feng, Cheng Zhang, Guansong Pang, Kangbiao Shi, Peng Wu, Wei Dong, Jinqiu Sun, and Yanning Zhang. HVI: A new color space for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5678–5687, 2025. doi: 10.1109/CVPR52734.2025.00533.

Yanhua Yu, Siyuan Shen, Zi Wang, Binbin Huang, Yuehan Wang, Xingyue Peng, Suan Xia, Ping Liu, Ruiqian Li, and Shiying Li. Enhancing non-line-of-sight imaging via learnable inverse kernel and attention mechanisms. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10529–10539, 2023. doi: 10.1109/ICCV51070.2023.00969.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14816–14826, 2021. doi: 10.1109/CVPR46437.2021.01458.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5718–5729, 2022. doi: 10.1109/CVPR52688.2022.00564.

Jiarui Zhang, Ruixu Geng, Xiaolong Du, Yan Chen, Houqiang Li, and Yang Hu. Passive non-line-of-sight imaging with light transport modulation. *IEEE Transactions on Image Processing*, 34: 410–424, 2025. doi: 10.1109/TIP.2024.3518097.

Shihao Zhou, Duosheng Chen, Jinshan Pan, Jinglei Shi, and Jufeng Yang. Adapt or perish: Adaptive sparse transformer with attentive feature refinement for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2952–2963, 2024. doi: 10.1109/CVPR52733.2024.00285.

APPENDIX FOR SPLITIR: OPERATOR-SPLITTING RECONSTRUCTION OF HIDDEN SCREEN CONTENT FROM INDIRECT RADIANCE

A SCENE DESCRIPTION

A.1 SCENARIO AND RECONSTRUCTION TASK

The physical scenario considered in this work is summarized in Fig. A1. A display presents RGB content on one side of an opaque blocking structure, while the camera is positioned on the opposite side. The blocker occludes both the screen surface and its displayed content from the camera. A diffuse wall remains visible to the camera and is illuminated by light emitted from the hidden display.

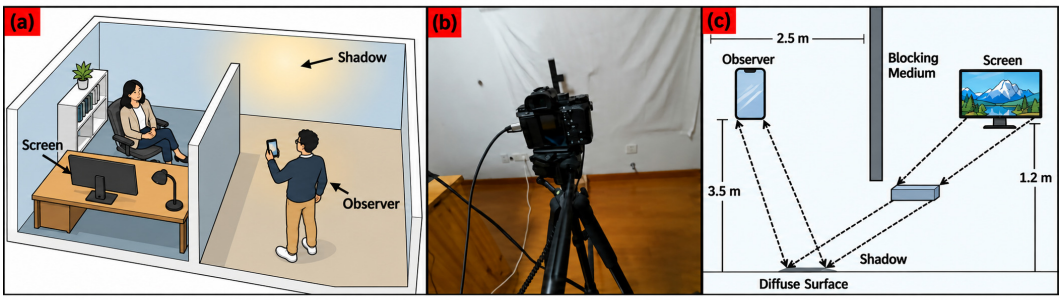


Figure A1: Experimental setup for hidden-screen reconstruction. (a) Hidden-screen acquisition scenario, where the screen is occluded from the camera and illuminates a visible diffuse wall. (b) Real acquisition setup used to capture the wall-radiance observations. (c) Scene geometry showing the relative placement of the camera, blocking medium, screen, and diffuse surface, together with the indirect light paths and the 2.5 m, 3.5 m, and 1.2 m acquisition distances.

This geometry defines the reconstruction task studied in the paper. The camera observes only the RGB appearance of the relay wall, whereas the desired output is the RGB content shown on the occluded screen. The screen image is retained as the paired reference during dataset construction and is used for supervision and quantitative evaluation; during reconstruction, the indirect wall image constitutes the observable input. The task can therefore be stated as recovering an occluded screen image from the content-dependent radiance distribution formed on a visible diffuse surface.

Fig. A1 provides the conceptual representation of this setting. It separates the hidden-screen region from the camera side and illustrates the indirect optical path connecting the display, relay wall, and camera. The figure abstracts the acquisition process into its essential components and defines the scene geometry used throughout the paper.

A.2 REAL-WORLD ACQUISITION CONFIGURATION

The corresponding physical acquisition environment is shown in Fig. A2. The four photographs provide complementary views of the actual setup and show the spatial relationship between the hidden display, blocking structure, relay wall, and capture camera. In the experimental configuration, the display is positioned behind the blocker and remains outside the camera's direct field of view. The camera is instead oriented toward the wall region receiving indirect illumination from the screen.

The acquisition geometry follows the distances marked in Fig. A1(c): the camera-side horizontal offset toward the blocking medium is 2.5 m, the camera is 3.5 m from the diffuse surface along the indicated measurement direction, and the screen is 1.2 m from the diffuse surface. The blocking structure separates the display from the observation side and removes direct visibility of the screen. Consequently, the camera's view is confined to the diffuse wall response induced by the hidden display.

The different viewpoints in Fig. A2 provide a physical counterpart to the abstract configuration in Fig. A1. The camera-facing views show the relay surface used for image acquisition, while the side views expose the relative placement of the blocker and hidden display. Together, these views clarify



Figure A2: Real-world acquisition setup from multiple viewpoints, showing the relative configuration of the hidden display, blocking structure, relay wall, and capture camera.

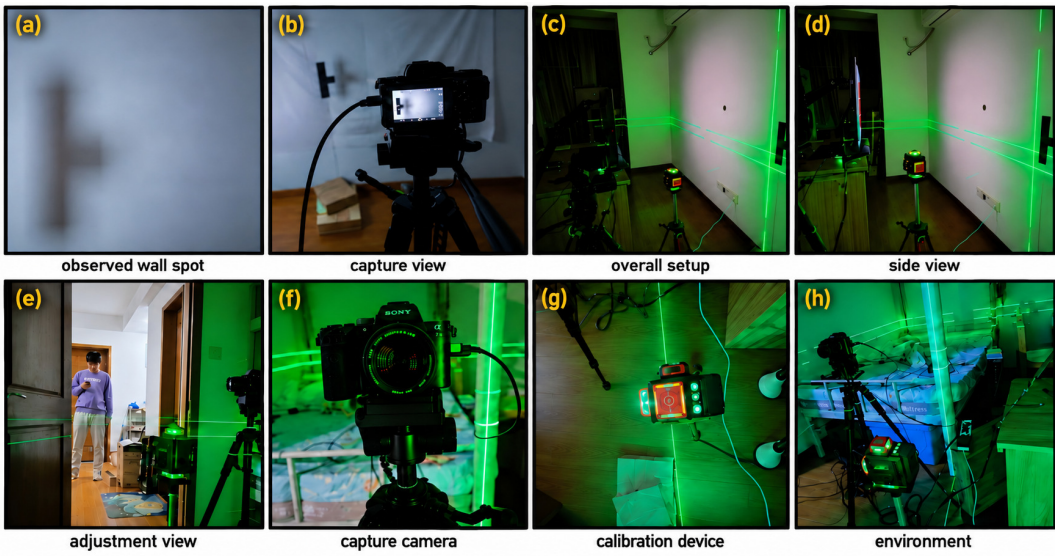


Figure A3: Representative setup-adjustment and geometric-alignment views used to establish the acquisition configuration, including the observed wall region, capture geometry, camera position, alignment device, and surrounding environment.

that the observable signal originates from the wall region accessible to the camera while the source image remains occluded.

For each displayed image, the camera records the corresponding wall appearance under the established acquisition geometry. The displayed RGB image and its wall observation are stored as a paired sample. The screen image provides the reconstruction reference, whereas the wall image forms the indirect-radiance input supplied to the model. The same acquisition procedure is applied to the different content categories considered in the experiments, including general screen content, charts, character-dominated password images, and webpages.

A.3 Setup Adjustment and Geometric Alignment

Before recording the paired data, the physical configuration is inspected and adjusted using the procedure illustrated in Fig. A3. The figure documents representative views used during setup preparation, including the observed wall region, capture view, overall configuration, side view, camera position, alignment device, and surrounding acquisition environment. These views provide a more detailed record of how the conceptual geometry in Fig. A1 is instantiated in the physical scene.

The observed wall spot is first inspected from the camera viewpoint to identify the region containing the indirect response of the hidden display. The camera position and viewing direction are then adjusted with respect to this region. The overall and side views are used to inspect the relative arrangement of the wall, camera, blocker, and hidden side of the scene, while the close-up views document the camera and alignment equipment used during this preparation stage.

Laser references visible in Fig. A3 provide geometric cues during setup adjustment. They support the inspection of relative positions and viewing directions before the final RGB observations are

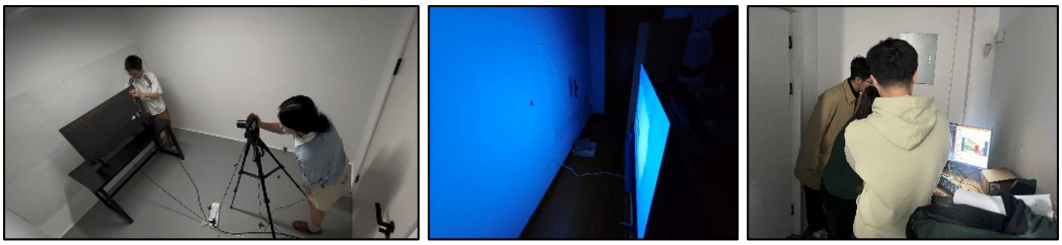


Figure A4: Setup preparation showing camera adjustment, inspection of the display-side arrangement, and coordination of camera placement and viewing direction.

recorded. Once the acquisition geometry is established, the camera is fixed toward the selected relay-wall region for paired data collection.

The preparation process is further documented in Fig. A4. The photographs show camera adjustment, inspection of the display-side arrangement, and coordination of camera placement and viewing direction before paired observations are recorded.

A.4 Physical Formation of the Indirect Observation

The wall observation is formed by indirect light transport between the hidden display, the relay surface, and the camera. The displayed image determines the spatial distribution of emitted RGB radiance, which propagates toward the visible wall. The separation between the display and relay surface spreads each local screen region's contribution over a broader wall area, creating a distributed spatial correspondence. After reaching the relay surface, the incident light is diffusely reflected, and a portion of the reflected radiance is subsequently recorded by the camera.

As a result, the radiance measured at each wall location generally contains contributions from multiple regions of the hidden display, while the radiance originating from one screen region may affect multiple wall locations. These overlapping contributions transform the original spatial organization of the displayed image before it is captured. The resulting RGB wall image retains content-dependent radiometric variations that encode local screen structures as distributed, partially superimposed responses.

This transport process is particularly influential for characters, thin boundaries, local textures, and other fine structures that occupy limited spatial support in the original display. Their responses are attenuated and mixed with neighboring contributions, whereas broader intensity and layout variations remain more readily observable over the relay surface. The camera consequently records an indirect radiance field whose spatial organization differs substantially from that of the hidden image. The following section develops a mathematical formulation of this image-formation process and its corresponding forward model.

B Theoretical Formulation

B.1 Physical Process Derivation

We derive the forward observation model from the physical acquisition geometry described in Sec. A. Let $\zeta \in \Omega_s$ denote a spatial location on the hidden screen, $\eta \in \Omega_w$ a location on the diffuse relay wall, and $\xi \in \Omega_x$ a location in the camera observation. Their corresponding 3D positions are denoted by $\mathbf{p}_s(\zeta)$, $\mathbf{p}_w(\eta)$, and the camera center $\mathbf{c}$, respectively. The following derivation is applied to each RGB channel; the channel index is omitted for compactness.

For a fixed display configuration, the radiometrically linearized screen value $y(\zeta)$ represents the emitted radiance associated with screen location ζ. Consider its propagation toward wall location η. Define

$$d_{sw}(\eta, \zeta) = \|\mathbf{p}_w(\eta) - \mathbf{p}_s(\zeta)\|_2 , \tag{A1}$$

and the corresponding propagation directions

$$\boldsymbol{\omega}_{sw} = \frac{\mathbf{p}_w(\eta) - \mathbf{p}_s(\zeta)}{d_{sw}(\eta, \zeta)}, \qquad \boldsymbol{\omega}_{ws} = -\boldsymbol{\omega}_{sw}. \tag{A2}$$

Let $\mathbf{n}_s(\zeta)$ and $\mathbf{n}_w(\eta)$ denote the surface normals of the display and relay wall. The geometric coupling between the two locations can then be written as

$$G_{sw}(\eta, \zeta) = V_{sw}(\eta, \zeta) \frac{\left[\mathbf{n}_s(\zeta)^\mathsf{T} \boldsymbol{\omega}_{sw}\right]_+ \left[\mathbf{n}_w(\eta)^\mathsf{T} \boldsymbol{\omega}_{ws}\right]_+}{d_{sw}^2(\eta, \zeta)}, \tag{A3}$$

where $[a]_+ = \max(a, 0)$, and $V_{sw}(\eta, \zeta) \in [0, 1]$ represents the visibility and transmission between the screen and wall locations. The two angular terms describe the orientation dependence of emitted and incident radiance, while the inverse-square term accounts for free-space geometric attenuation.

The irradiance received at wall location η is obtained by integrating the contributions from all screen locations:

$$E_w(\eta) = \int_{\Omega_s} G_{sw}(\eta, \zeta)\, y(\zeta)\, d\zeta. \tag{A4}$$

Eq. equation A4 already shows that a wall location generally receives radiance from multiple regions of the hidden display. Conversely, the contribution of a single screen location is distributed over multiple wall locations through the spatial variation of G_{sw}.

The relay surface is matte and is modeled by diffuse reflection. Under a Lambertian approximation, its outgoing radiance is

$$L_w(\eta) = \frac{\rho_w(\eta)}{\pi} E_w(\eta), \tag{A5}$$

where $\rho_w(\eta) \geq 0$ denotes the local diffuse reflectance of the wall. Substituting Eq. equation A4 gives

$$L_w(\eta) = \frac{\rho_w(\eta)}{\pi} \int_{\Omega_s} G_{sw}(\eta, \zeta)\, y(\zeta)\, d\zeta. \tag{A6}$$

The camera records L_w through spatial integration over the visible relay surface, combining finite-aperture, pixel-footprint, viewing, optical, and spatial-sampling effects. We represent this wall-to-camera mapping by a non-negative response kernel $C_\vartheta(\xi, \eta)$, yielding

$$x(\xi) = \int_{\Omega_w} C_\vartheta(\xi, \eta) L_w(\eta)\, d\eta + \varepsilon(\xi), \tag{A7}$$

where $\varepsilon(\xi)$ collects residual measurement noise and unmodeled acquisition perturbations. The parameter set ϑ contains the scene geometry, relay-surface response, camera viewing configuration, and sensor characteristics that affect the observation process.

Substituting Eq. equation A6 into Eq. equation A7 gives

$$x(\xi) = \int_{\Omega_w} C_\vartheta(\xi, \eta) \frac{\rho_w(\eta)}{\pi} \left[\int_{\Omega_s} G_{sw}(\eta, \zeta) y(\zeta)\, d\zeta \right] d\eta + \varepsilon(\xi). \tag{A8}$$

Reordering the integration isolates the contribution from each hidden screen location:

$$x(\xi) = \int_{\Omega_s} \left[\int_{\Omega_w} C_\vartheta(\xi, \eta) \frac{\rho_w(\eta)}{\pi} G_{sw}(\eta, \zeta)\, d\eta \right] y(\zeta)\, d\zeta + \varepsilon(\xi). \tag{A9}$$

We therefore define the effective diffuse-transport kernel as

$$K_\vartheta(\xi,\zeta) = \int_{\Omega_w} C_\vartheta(\xi,\eta)\frac{\rho_w(\eta)}{\pi}G_{sw}(\eta,\zeta)\,d\eta. \tag{A10}$$

Since the geometric transport, wall reflectance, and camera response are non-negative,

$$K_\vartheta(\xi,\zeta) \geq 0. \tag{A11}$$

The complete observation model consequently reduces to

$$x(\xi) = \int_{\Omega_s} K_\vartheta(\xi,\zeta)y(\zeta)\,d\zeta + \varepsilon(\xi), \tag{A12}$$

which is the continuous forward model used in the main paper. Here, K_ϑ jointly represents free-space propagation, diffuse-wall interaction, wall-to-camera transport, and spatial integration within the imaging system.

The integral structure of Eq. equation A12 explains the spatial mixing observed in the acquired wall images. For a fixed screen location ζ, the kernel $K_\vartheta(\xi,\zeta)$ generally has support over multiple observation locations ξ. Similarly, each observation location aggregates contributions from multiple screen locations through the same kernel. Fine structures in y are therefore distributed and attenuated before measurement, while broader radiometric variations remain represented over larger spatial regions. After spatial discretization, the same process can be written compactly as

$$\mathbf{x} = \mathbf{K}_\vartheta\mathbf{y} + \boldsymbol{\varepsilon}, \qquad \big(\mathbf{K}_\vartheta\big)_{u,v} \geq 0 \quad \text{for all } (u,v), \tag{A13}$$

where $\mathbf{K}_\vartheta$ is the entrywise non-negative discrete light-transport operator associated with the physical acquisition geometry.

B.2 Mechanistic Formulation of PAM and RCM

The forward formulation in Sec. B.1 shows that indirect transport replaces the direct spatial correspondence of the hidden image with a distributed observation field. This creates two coupled reconstruction difficulties. First, the spatial organization of the indirect observation is inconsistent with the evolving structural representation used by the decoder, so observation evidence must be transferred into the current layout coordinates before subsequent reconstruction. Second, decoder refinement is performed sequentially, and each newly generated residual is conditioned on an already updated state; direct residual accumulation leaves their admissible range implicit. SplitIR separates the two operations through Potential Assimilation Module (PAM) and Resolvent Correction Module (RCM): PAM converts the discrepancy between observation evidence and the current layout state into bounded representation transport, while RCM transforms each stage-wise candidate residual into a metric-conditioned feasible correction before write-back.

Let F_i denote the shared layout state at PAM stage $i \in \mathcal{I} = \{0,1\}$, with scale-aligned indirect observation $x_i = D_i x$ and side readout b_i. PAM first constructs an online evidence field from their radiometric and structural discrepancy and maps both the observation evidence and layout state into a common representation space. A positive diagonal metric defines the local geometry of their discrepancy, from which an observation potential and its descent field are constructed. The resulting direction is lifted back to the layout coordinates and bounded by a stage-conditioned mobility field. After the two PAM calls, RCM operates on the image-domain reconstruction. At RCM stage $k \in \mathcal{K} = \{0,1,2,3\}$, the current reconstruction, candidate proposal, raw observation, and auxiliary hint jointly parameterize a stage-dependent correction geometry. The candidate residual is first adjusted by a metric-aware proximal operation and is then projected onto the corresponding feasible set before being written into the current reconstruction.

B.2.1 POTENTIAL-GUIDED EVIDENCE ASSIMILATION

For PAM stage i, define the observation–prediction residual as

$$\Delta_i(\xi) = x_i(\xi) - b_i(\xi). \tag{A14}$$

PAM combines Δ_i with complementary radiometric, differential, and directional statistics in the online evidence field

$$\begin{aligned} r_i(\xi) = \mathrm{Cat}\Big[& x_i(\xi),\ b_i(\xi),\ \Delta_i(\xi),\ |\Delta_i(\xi)|, \\ & \mathcal{G}[x_i](\xi),\ \mathcal{G}[b_i](\xi),\ \mathcal{G}[\Delta_i](\xi),\ \mathcal{L}[\Delta_i](\xi), \\ & \mathcal{A}[|\Delta_i|](\xi),\ \mathcal{R}[\Delta_i](\xi),\ \mathcal{C}[\Delta_i](\xi)\Big] \in \mathbb{R}^{19}. \end{aligned} \tag{A15}$$

Here, $\mathcal{G}[\cdot]$ denotes the local gradient response, $\mathcal{L}[\cdot]$ the channel-aggregated Laplacian response, $\mathcal{A}[\cdot]$ a local average operator, and $\mathcal{R}[\cdot]$ and $\mathcal{C}[\cdot]$ the row-wise and column-wise directional statistics, respectively. The four RGB terms contribute 12 channels, and the seven structural statistics contribute one channel each; the row-wise and column-wise statistics are broadcast to the spatial grid. This gives the 19-channel evidence field used by both PAM and RCM.

The structural state and observation evidence are then expressed in a shared representation space:

$$\begin{aligned} s_i(\xi) &= \mathcal{S}(F_i)\,(\xi), \\ \widetilde{o}_i(\xi) &= \mathcal{O}(r_i)\,(\xi), \\ o_i(\xi) &= \widetilde{o}_i(\xi) + e_i^p, \\ d_i(\xi) &= s_i(\xi) - o_i(\xi), \qquad i \in \{0, 1\}. \end{aligned} \tag{A16}$$

The stage embedding e_i^p provides an explicit identifier for the current assimilation call within the shared conditioning map, while the discrepancy field d_i measures the remaining offset between the two representations.

PAM assigns a positive diagonal metric to this discrepancy:

$$\begin{aligned} M_i^p(\xi) &= \mathrm{Diag}\left(m_{i,1}^p(\xi), \ldots, m_{i,C_p}^p(\xi)\right), \\ m_i^p(\xi) &= \mu_p \mathbf{1} + \sigma\left(\Psi\left(s_i(\xi), o_i(\xi), |d_i(\xi)|\right)\right), \qquad \mu_p > 0, \\ m_{i,\alpha}^p(\xi) &> 0, \qquad \alpha = 1, \ldots, C_p. \end{aligned} \tag{A17}$$

The local quadratic energy and the normalized potential density used by the implementation are

$$\begin{aligned} \widetilde{\phi}_i(\xi) &= \frac{1}{2} d_i(\xi)^{\mathsf{T}} M_i^p(\xi) d_i(\xi) = \frac{1}{2} \sum_{\alpha=1}^{C_p} m_{i,\alpha}^p(\xi)\left[d_i^\alpha(\xi)\right]^2, \\ \phi_i(\xi) &= \frac{1}{C_p} \widetilde{\phi}_i(\xi), \qquad \widetilde{\Phi}_i = \int_{\Omega_i} \widetilde{\phi}_i(\xi)\, d\xi. \end{aligned} \tag{A18}$$

The normalized density ϕ_i is passed to the mobility branch and recorded in the PAM diagnostics. To obtain the assimilation direction, M_i^p is evaluated at the current state and held fixed during the local discrepancy variation. For an admissible perturbation δd_i, the first variation of $\widetilde{\Phi}_i$ is

$$\begin{aligned} \delta\widetilde{\Phi}_i\left[d_i; \delta d_i\right] &= \lim_{\epsilon \to 0} \frac{\widetilde{\Phi}_i\left(d_i + \epsilon \delta d_i\right) - \widetilde{\Phi}_i(d_i)}{\epsilon} \\ &= \int_{\Omega_i} \sum_{\alpha=1}^{C_p} m_{i,\alpha}^p(\xi) d_i^\alpha(\xi) \delta d_i^\alpha(\xi)\, d\xi \\ &= \int_{\Omega_i} \left\langle M_i^p(\xi) d_i(\xi), \delta d_i(\xi) \right\rangle d\xi. \end{aligned} \tag{A19}$$

The associated negative variation defines the representation-space assimilation direction

$$g_i(\xi) = -\frac{\delta\widetilde{\Phi}_i}{\delta d_i(\xi)} = -M_i^p(\xi) d_i(\xi). \tag{A20}$$

Because g_i is expressed in the PAM representation coordinates, it is transformed back to the layout-state coordinates before write-back. The transport magnitude is controlled by the stage-conditioned mobility

$$\begin{aligned} \eta_i(\xi) &= \eta_0 \sigma\left(\Gamma\left(s_i(\xi), o_i(\xi), |d_i(\xi)|, \phi_i(\xi)\right)\right), \\ u_i(\xi) &= \eta_i(\xi) \odot \tanh\left(\mathcal{T}\left[g_i\right](\xi)\right) \\ &= \eta_0 \sigma\left(\Gamma\left(s_i, o_i, |d_i|, \phi_i\right)\right)(\xi) \odot \tanh\left(\mathcal{T}\left[-M_i^p d_i\right](\xi)\right), \\ F_i^+(\xi) &= F_i(\xi) + u_i(\xi), \qquad F_{i+1} = F_i^+. \end{aligned} \tag{A21}$$

Here, $\mathcal{T}$ denotes the learned pointwise lifting from the PAM representation space to the shared layout-state space. The sigmoid-modulated mobility and hyperbolic tangent constrain the transport written into F_i, while both PAM calls update the same shared layout state sequentially.

B.2.2 Stage-Conditioned Resolvent Correction

After PAM assimilation and decoder reconstruction, RCM processes an ordered sequence of image-domain candidate corrections. At stage k, let f_k denote the current reconstruction and let a_k denote the candidate proposal generated by the corresponding proposal adapter. Their residual correction is

$$d_k^r(\xi) = a_k(\xi) - f_k(\xi), \qquad k \in \{0, 1, 2, 3\}. \tag{A22}$$

RCM conditions this residual on the current proposal relation, the original indirect observation, and the auxiliary reconstruction hint. The composite conditioning field is

$$\begin{aligned} z_k(\xi) &= \mathrm{Cat}\left[\mathcal{E}\left(f_k, a_k\right)(\xi), x(\xi), h(\xi)\right], \\ z_k &\in \mathbb{R}^{25 \times H \times W}, \\ q_k(\xi) &= \mathcal{Q}\left(z_k\right)(\xi) + e_k^r. \end{aligned} \tag{A23}$$

The evidence term $\mathcal{E}(f_k, a_k)$ contributes 19 channels, while the raw observation x and auxiliary hint h contribute three RGB channels each. The stage embedding e_k^r conditions the shared correction representation on the current position in the ordered refinement sequence.

From q_k, RCM parameterizes the local correction geometry through a center c_k, an admissible radius ϱ_k, and a positive diagonal metric M_k^r:

$$\begin{aligned} c_k(\xi) &= c_0 \tanh\left(\mathcal{H}_c\left[q_k\right](\xi)\right), \\ \varrho_k(\xi) &= \varrho_{\min} + \mathrm{softplus}\left(\mathcal{H}_\varrho\left[q_k\right](\xi) + b_\varrho\right), \qquad \varrho_k(\xi) > 0, \\ M_k^r(\xi) &= \mathrm{Diag}\left(m_{k,1}^r(\xi), \ldots, m_{k,C_r}^r(\xi)\right), \\ m_k^r(\xi) &= \mu_r \mathbf{1} + \sigma\left(\mathcal{H}_m\left[q_k\right](\xi)\right), \qquad m_{k,\alpha}^r(\xi) > 0, \\ C_k(\xi) &= \left\{z \in \mathbb{R}^{C_r} : |z_\alpha - c_{k,\alpha}(\xi)| \le \varrho_{k,\alpha}(\xi),\ \alpha = 1, \ldots, C_r\right\}. \end{aligned} \tag{A24}$$

The center and radius define one learned admissible interval for each correction coordinate, while the positive diagonal metric assigns coordinate-dependent weights to the proximal adjustment. These quantities are regenerated from the updated reconstruction state at every RCM stage.

The candidate residual is then converted into a proximal correction by balancing its original value with the learned correction center. For $\rho > 0$, define

$$\begin{aligned} p_k &= \operatorname*{argmin}_z \left\{ \frac{1}{2}\left\|z - d_k^r\right\|_2^2 + \frac{\rho}{2}\left\|z - c_k\right\|_{M_k^r}^2 \right\}, \\ \left\|z - c_k\right\|_{M_k^r}^2 &= \left\langle z - c_k,\ M_k^r\left(z - c_k\right)\right\rangle_F \\ &= \sum_{\alpha=1}^{C_r} m_{k,\alpha}^r \left(z_\alpha - c_{k,\alpha}\right)^2. \end{aligned} \tag{A25}$$

Expanding the objective channel-wise gives

$$\mathcal{J}_k(z) = \frac{1}{2}\sum_{\alpha=1}^{C_r}\left(z_\alpha - d_{k,\alpha}^r\right)^2 + \frac{\rho}{2}\sum_{\alpha=1}^{C_r} m_{k,\alpha}^r \left(z_\alpha - c_{k,\alpha}\right)^2. \tag{A26}$$

The stationary condition for each correction dimension is

$$\begin{aligned} \frac{\partial \mathcal{J}_k}{\partial z_\alpha} &= z_\alpha - d^r_{k,\alpha} + \rho m^r_{k,\alpha}\left(z_\alpha - c_{k,\alpha}\right) = 0, \\ \left(1 + \rho m^r_{k,\alpha}\right) z_\alpha &= d^r_{k,\alpha} + \rho m^r_{k,\alpha} c_{k,\alpha}. \end{aligned} \tag{A27}$$

Hence the proximal state used by RCM can be written as

$$\begin{aligned} p_k &= \left(I + \rho M^r_k\right)^{-1}\left(d^r_k + \rho M^r_k c_k\right) \\ &= \left[\frac{d^r_{k,1} + \rho m^r_{k,1} c_{k,1}}{1 + \rho m^r_{k,1}}, \ldots, \frac{d^r_{k,C_r} + \rho m^r_{k,C_r} c_{k,C_r}}{1 + \rho m^r_{k,C_r}}\right]^{\mathsf{T}}. \end{aligned} \tag{A28}$$

The proximal state is finally restricted to the stage-conditioned feasible region. Let N_{C_k} denote the normal-cone operator associated with C_k. The corresponding resolvent correction is

$$\begin{aligned} \bar{d}_k &= \left(I + N_{C_k}\right)^{-1}\left(p_k\right) \\ &= \Pi_{C_k}\left(p_k\right) \\ &= \min\left\{\max\left\{p_k,\, c_k - \varrho_k\right\}, c_k + \varrho_k\right\}, \end{aligned} \tag{A29}$$

where the minimum and maximum are evaluated element-wise. Equivalently, the proximal adjustment and box projection solve the stage-conditioned constrained quadratic

$$\bar{d}_k = \underset{z \in C_k}{\operatorname{argmin}} \left\{ \frac{1}{2}\|z - d^r_k\|_2^2 + \frac{\rho}{2}\|z - c_k\|^2_{M^r_k} \right\}. \tag{A30}$$

With the predicted geometry fixed, the positive diagonal quadratic and coordinate-wise box make the problem separable. Each coordinate is minimized by clipping its unconstrained proximal solution to the corresponding interval, yielding the same correction as Eq. equation A29. This gives RCM a direct role in the reconstruction: it converts a proposed refinement into the optimal feasible correction for the current stage's constrained quadratic objective Parikh & Boyd (2014).

Substituting the proximal expression gives the complete correction operator

$$\begin{aligned} \bar{d}_k = \min\Bigg\{ \max\Bigg\{ &\left(I + \rho M^r_k\right)^{-1}\left(d^r_k + \rho M^r_k c_k\right), \\ &c_k - \varrho_k \Bigg\}, c_k + \varrho_k \Bigg\}. \end{aligned} \tag{A31}$$

The admitted correction is then written into the reconstruction and mapped to the valid image-intensity range:

$$\widetilde{f}_{k+1} = f_k + \bar{d}_k, \qquad f_{k+1} = \Pi_{[0,1]}\left(\widetilde{f}_{k+1}\right), \qquad k = 0, 1, 2, 3. \tag{A32}$$

The four stages are applied in the order of frequency, morphology, scanline, and directional-line proposal generation. Since a_k, q_k, c_k, ϱ_k, and M^r_k are recomputed after each write-back, every subsequent correction is conditioned on the state produced by the preceding stage.

The complete reconstruction therefore separates representation-level evidence assimilation from image-domain correction regulation: PAM aligns spatially mixed observation evidence with the evolving shared layout through metric-induced bounded transport, while RCM couples successive decoder proposals to stage-conditioned proximal and feasible-set geometry before their sequential integration.

B.3 Well-Posedness and Operator Properties

This section establishes the positive discrepancy geometry and bounded transport of PAM, together with the existence, uniqueness, and projection properties of the stage-conditioned RCM correction. The proximal and projection formulation follows the standard operator framework of Parikh and Boyd Parikh & Boyd (2014).

Positive discrepancy potential and bounded PAM transport. At PAM stage i, recall the positive diagonal metric

$$M_i^p(\xi) = \mathrm{Diag}\left(m_{i,1}^p(\xi), \ldots, m_{i,C_p}^p(\xi)\right), \qquad m_{i,\alpha}^p(\xi) = \mu_p + \sigma(\cdot), \qquad \mu_p > 0. \tag{A33}$$

Since $0 < \sigma(\cdot) < 1$, every diagonal entry satisfies

$$m_{i,\alpha}^p(\xi) > \mu_p > 0, \tag{A34}$$

and therefore

$$M_i^p(\xi) \succ 0 \qquad \forall \xi \in \Omega_i. \tag{A35}$$

The corresponding PAM potential is

$$\widetilde{\Phi}_i(d_i) = \frac{1}{2} \int_{\Omega_i} d_i(\xi)^\mathsf{T} M_i^p(\xi) d_i(\xi)\, d\xi. \tag{A36}$$

Because $M_i^p(\xi)$ is positive definite pointwise,

$$\widetilde{\Phi}_i(d_i) \geq \frac{\mu_p}{2} \int_{\Omega_i} \|d_i(\xi)\|_2^2 \, d\xi \geq 0. \tag{A37}$$

Moreover,

$$\widetilde{\Phi}_i(d_i) = 0 \quad \Longleftrightarrow \quad d_i(\xi) = 0 \text{ for almost every } \xi \in \Omega_i. \tag{A38}$$

Thus, for the metric evaluated at the current PAM state, $\widetilde{\Phi}_i$ defines a non-negative discrepancy functional whose zero corresponds to coincidence between the structural and observation-derived coordinates.

To characterize the local transport direction, keep M_i^p fixed at the current state and consider the path

$$d_i^{(\tau)} = d_i + \tau g_i, \qquad g_i = -M_i^p d_i. \tag{A39}$$

The directional derivative of the potential at $\tau = 0$ is

$$\begin{aligned} \frac{d}{d\tau} \widetilde{\Phi}_i \left(d_i^{(\tau)}\right)\Big|_{\tau=0} &= \int_{\Omega_i} \langle M_i^p(\xi) d_i(\xi), g_i(\xi) \rangle \, d\xi \\ &= -\int_{\Omega_i} \|M_i^p(\xi) d_i(\xi)\|_2^2 \, d\xi \leq 0. \end{aligned} \tag{A40}$$

The inequality is strict whenever $d_i \neq 0$ on a set of non-zero measure. Hence $-M_i^p d_i$ is a local descent direction of the metric-induced discrepancy potential under the frozen-metric variation used in the PAM update.

The actual PAM write-back applies the bounded transport

$$u_i(\xi) = \eta_i(\xi) \odot \tanh\left(\mathcal{T}[g_i](\xi)\right), \qquad 0 \leq \eta_i(\xi) \leq \eta_0. \tag{A41}$$

Since each component of $\tanh(\cdot)$ lies in $(-1, 1)$,

$$|u_{i,\alpha}(\xi)| \leq \eta_{i,\alpha}(\xi) \leq \eta_0, \tag{A42}$$

which gives

$$\|u_i\|_\infty \le \eta_0. \tag{A43}$$

Therefore, each PAM stage combines a well-defined metric descent direction with a uniformly bounded layout-state transport.

Existence and uniqueness of the RCM proximal state. At RCM stage k, the candidate residual d_k^r and the learned correction geometry define the proximal objective

$$J_k(z) = \frac{1}{2}\|z - d_k^r\|_2^2 + \frac{\rho}{2}\|z - c_k\|_{M_k^r}^2, \qquad \rho > 0, \tag{A44}$$

where

$$M_k^r = \mathrm{Diag}\left(m_{k,1}^r, \ldots, m_{k,C_r}^r\right), \qquad m_{k,\alpha}^r > 0. \tag{A45}$$

Expanding Eq. equation A44 gives

$$J_k(z) = \frac{1}{2}\sum_{\alpha=1}^{C_r}\left(z_\alpha - d_{k,\alpha}^r\right)^2 + \frac{\rho}{2}\sum_{\alpha=1}^{C_r} m_{k,\alpha}^r\left(z_\alpha - c_{k,\alpha}\right)^2. \tag{A46}$$

Its Hessian with respect to z is

$$\nabla_z^2 J_k = I + \rho M_k^r. \tag{A47}$$

For any non-zero vector v,

$$\begin{aligned} v^\mathsf{T}\left(I + \rho M_k^r\right)v &= \|v\|_2^2 + \rho\sum_{\alpha=1}^{C_r} m_{k,\alpha}^r v_\alpha^2 \\ &> 0. \end{aligned} \tag{A48}$$

Hence

$$I + \rho M_k^r \succ 0, \tag{A49}$$

and J_k is strictly, in fact strongly, convex with respect to z. The proximal problem therefore admits a unique minimizer,

$$\exists!\, p_k = \operatorname*{argmin}_z J_k(z). \tag{A50}$$

The first-order optimality condition

$$\nabla_z J_k(p_k) = p_k - d_k^r + \rho M_k^r\left(p_k - c_k\right) = 0 \tag{A51}$$

gives

$$\left(I + \rho M_k^r\right)p_k = d_k^r + \rho M_k^r c_k. \tag{A52}$$

Since $I + \rho M_k^r$ is positive definite and therefore invertible, the unique proximal state is

$$p_k = \left(I + \rho M_k^r\right)^{-1}\left(d_k^r + \rho M_k^r c_k\right), \tag{A53}$$

which coincides with the closed-form update derived in Sec. B.2.

Existence and uniqueness of the resolvent correction. The RCM feasible set is defined as

$$C_k = \left\{ z \in \mathbb{R}^{C_r} : |z_\alpha - c_{k,\alpha}| \le \varrho_{k,\alpha},\ \alpha = 1, \dots, C_r \right\}, \qquad \varrho_{k,\alpha} > 0. \tag{A54}$$

Equivalently,

$$C_k = \prod_{\alpha=1}^{C_r} \left[c_{k,\alpha} - \varrho_{k,\alpha},\ c_{k,\alpha} + \varrho_{k,\alpha} \right]. \tag{A55}$$

Each interval is non-empty because $\varrho_{k,\alpha} > 0$. Their Cartesian product is therefore a non-empty, closed, and convex subset of $\mathbb{R}^{C_r}$. In particular,

$$c_k \in C_k, \tag{A56}$$

so $C_k \neq \varnothing$.

For any proximal state p_k, consider the projection problem

$$\bar{d}_k = \operatorname*{argmin}_{z \in C_k} \frac{1}{2} \|z - p_k\|_2^2. \tag{A57}$$

Because the objective is strictly convex and C_k is non-empty, closed, and convex, the projection exists and is unique:

$$\exists!\, \bar{d}_k = \Pi_{C_k}(p_k). \tag{A58}$$

Let ι_{C_k} denote the indicator functional

$$\iota_{C_k}(z) = \begin{cases} 0, & z \in C_k, \\ +\infty, & z \notin C_k. \end{cases} \tag{A59}$$

Its convex subdifferential is the normal-cone operator,

$$\partial \iota_{C_k} = N_{C_k}. \tag{A60}$$

The proximal operator associated with the indicator functional is therefore

$$\begin{aligned} \operatorname{prox}_{\iota_{C_k}}(p_k) &= \operatorname*{argmin}_z \left\{ \frac{1}{2} \|z - p_k\|_2^2 + \iota_{C_k}(z) \right\} \\ &= \Pi_{C_k}(p_k) \\ &= \left(I + N_{C_k}\right)^{-1}(p_k). \end{aligned} \tag{A61}$$

This establishes the equivalence between the projection used by RCM and the resolvent of the normal-cone operator.

Because C_k is an axis-aligned box, the unique projection is computed component-wise:

$$\bar{d}_{k,\alpha} = \min\left\{ \max\left\{ p_{k,\alpha}, c_{k,\alpha} - \varrho_{k,\alpha} \right\}, c_{k,\alpha} + \varrho_{k,\alpha} \right\}. \tag{A62}$$

For a fixed stage-conditioned set C_k, projection onto the non-empty closed convex box is firmly non-expansive. Hence, for any two proximal states $p_k^{(1)}$ and $p_k^{(2)}$,

$$\left\| \Pi_{C_k}\left(p_k^{(1)}\right) - \Pi_{C_k}\left(p_k^{(2)}\right) \right\|_2^2 \le \left\langle \Pi_{C_k}\left(p_k^{(1)}\right) - \Pi_{C_k}\left(p_k^{(2)}\right),\ p_k^{(1)} - p_k^{(2)} \right\rangle, \tag{A63}$$

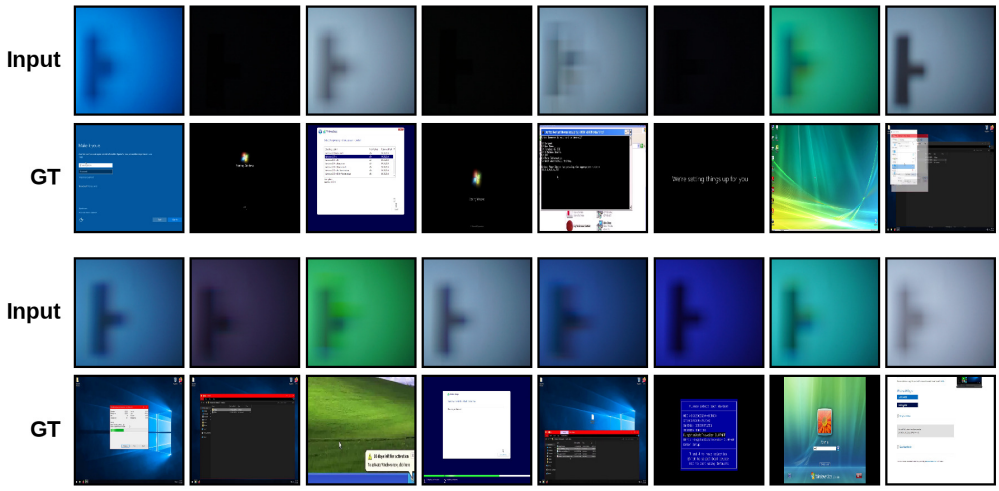


Figure A5: Representative paired samples from the Screen subset, showing indirect relay-wall observations and their corresponding hidden RGB screen content.

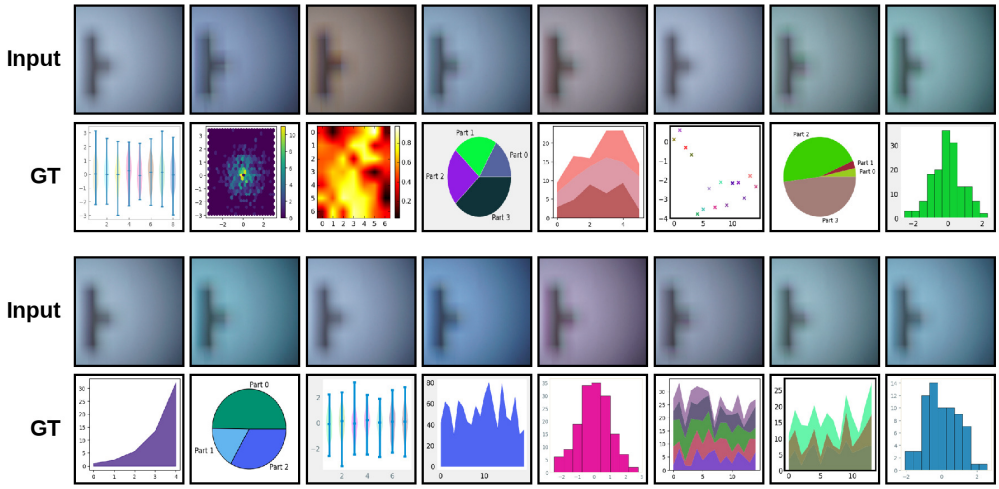


Figure A6: Representative paired samples from the Chart subset, containing structured graphical layouts with axes, curves, bars, annotations, and other geometric elements.

and consequently

$$\left\|\Pi_{C_k}\left(p_k^{(1)}\right)-\Pi_{C_k}\left(p_k^{(2)}\right)\right\|_2 \le \left\|p_k^{(1)}-p_k^{(2)}\right\|_2 . \tag{A64}$$

Thus, the projection step does not increase the Euclidean distance between two proximal correction states.

Well-defined sequential composition. The previous results apply independently at every PAM and RCM stage. For PAM, the learned metric remains positive and the transport is bounded by construction. For RCM, each stage produces a unique proximal state and a unique feasible-set projection whenever the predicted quantities c_k, ϱ_k, and M_k^r are finite.

Accordingly, the stage transition

$$f_{k+1} = \Pi_{[0,1]}\left(f_k + \Pi_{C_k}\left[(I+\rho M_k^r)^{-1}\left(d_k^r + \rho M_k^r c_k\right)\right]\right). \tag{A65}$$

defines a unique next reconstruction state for each input state f_k. Here, Π_{C_k} restricts the correction to the stage-conditioned feasible set, while $\Pi_{[0,1]}$ clips the updated reconstruction element-wise



Figure A7: Representative paired samples from the Password subset, whose hidden targets are dominated by compact character structures and sparse local boundaries.

to the valid image-intensity range. Since both projections are single-valued, the stage transition remains well defined. Applying this relation successively for $k = 0, 1, 2, 3$ yields a well-defined ordered reconstruction sequence

$$f_0 \longmapsto f_1 \longmapsto f_2 \longmapsto f_3 \longmapsto f_4. \tag{A66}$$

Under the learned positive metrics, PAM therefore admits a well-defined discrepancy descent direction with bounded transport, while the RCM subproblem admits a unique proximal state and a unique resolvent correction within its stage-conditioned feasible set.

Training objective. For model output u and target y in RGB $[0, 1]$, the training objective is

$$\mathcal{L} = \mathcal{L}_1 + 0.18\,\mathcal{L}_{\text{edge-SSIM}} + 0.04\,\mathcal{L}_{\text{edge-MSE}}. \tag{A67}$$

We obtain a target-derived edge weight from

$$\begin{aligned} G(y) &= \left[D_x(\text{mean}_c\, y)^2 + D_y(\text{mean}_c\, y)^2 + 10^{-6}\right]^{1/2}, \\ Z(y) &= \text{AvgPool}_9^{\text{reflect}}(G(y)) + \text{mean}_{HW}\, G(y) + 10^{-4}, \\ e &= \text{clip}\left(\frac{G(y)}{Z(y)}, 0, 2\right). \end{aligned} \tag{A68}$$

The edge map is computed from the target and does not introduce an auxiliary PAM or RCM supervision term. The SSIM term uses a 7×7 uniform reflected window; the remaining optional structure, frequency, color, MS-SSIM, and auxiliary losses have zero weight in the reported training configuration.

B.4 Formulation of Evaluation Metrics

The quantitative evaluation uses six full-reference image metrics: MSE, RMSE, PSNR, SSIM, MS-SSIM, and LPIPS. All final comparisons, luminance and noise tests, and real-acquisition robustness experiments use the same preprocessing and metric implementation. Let $y, \hat{y} \in [0, 1]^{3\times H\times W}$ denote the ground-truth and reconstructed RGB images after clipping to the valid intensity range. Each metric is computed independently for every image pair and then averaged with equal weight over the evaluation partition.

Pixel-wise distortion and signal fidelity. For MSE, RMSE, and PSNR, we first rescale the clipped images from $[0, 1]$ to $[0, 255]$ in floating-point precision:

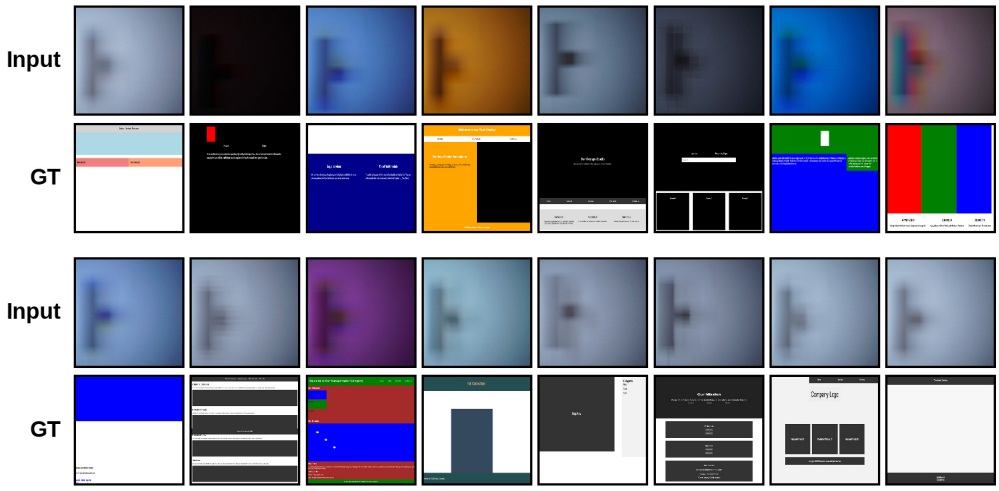


Figure A8: Representative paired samples from the Website subset, containing multi-scale page layouts composed of text, image regions, interface elements, and structured boundaries.

$$\widetilde{y}_{cuv} = 255\, y_{cuv}, \qquad \widetilde{\hat{y}}_{cuv} = 255\, \hat{y}_{cuv}, \qquad c \in \{1, 2, 3\}. \tag{A69}$$

The per-image mean squared error is computed over all color channels and spatial locations,

$$\mathrm{MSE}_{255} = \frac{1}{3HW} \sum_{c=1}^{3} \sum_{u=1}^{H} \sum_{v=1}^{W} \left(\widetilde{\hat{y}}_{cuv} - \widetilde{y}_{cuv} \right)^2 = \frac{255^2}{3HW} \sum_{c=1}^{3} \sum_{u=1}^{H} \sum_{v=1}^{W} \left(\hat{y}_{cuv} - y_{cuv} \right)^2 . \tag{A70}$$

The corresponding root mean squared error is

$$\mathrm{RMSE}_{255} = \sqrt{\mathrm{MSE}_{255}}. \tag{A71}$$

Using the same 255-level dynamic range, PSNR is

$$\mathrm{PSNR} = 10 \log_{10} \left(\frac{255^2}{\mathrm{MSE}_{255}} \right). \tag{A72}$$

Equivalently, defining the normalized-domain error as

$$\mathrm{MSE}_{01} = \frac{1}{3HW} \sum_{c=1}^{3} \sum_{u=1}^{H} \sum_{v=1}^{W} \left(\hat{y}_{cuv} - y_{cuv} \right)^2 , \tag{A73}$$

gives

$$\mathrm{PSNR} = 10 \log_{10} \left(\frac{1}{\mathrm{MSE}_{01}} \right). \tag{A74}$$

The quantitative tables report MSE_{255}, the 255-level form of the normalized-domain error MSE_{01}. Lower MSE and RMSE indicate smaller pixel-wise distortion, while higher PSNR indicates higher signal fidelity.

Structural similarity. SSIM is evaluated independently for the three RGB channels using a 7×7 uniform local window. For a local window $\mathcal{W}_{uv}$ containing $N_w = 49$ samples, the local means are

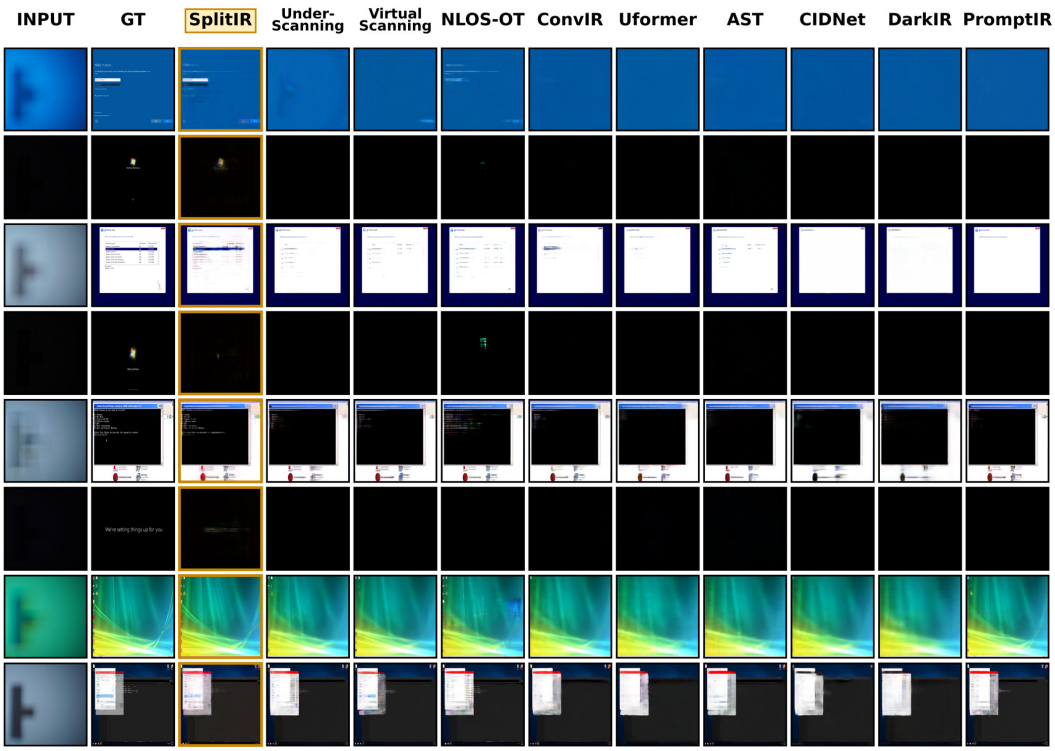


Figure A9: Qualitative comparison on the Screen subset, showing the indirect observation, hidden-screen reference, SplitIR reconstruction, and representative baseline results for heterogeneous screen content.

$$\mu_y = \frac{1}{N_w} \sum_{p \in \mathcal{W}_{uv}} y_p, \qquad \mu_{\hat{y}} = \frac{1}{N_w} \sum_{p \in \mathcal{W}_{uv}} \hat{y}_p. \tag{A75}$$

The implementation uses sample estimates for variance and covariance, giving

$$\begin{aligned} \sigma_y^2 &= \frac{1}{N_w - 1} \sum_{p \in \mathcal{W}_{uv}} (y_p - \mu_y)^2, \\ \sigma_{\hat{y}}^2 &= \frac{1}{N_w - 1} \sum_{p \in \mathcal{W}_{uv}} (\hat{y}_p - \mu_{\hat{y}})^2, \\ \sigma_{y\hat{y}} &= \frac{1}{N_w - 1} \sum_{p \in \mathcal{W}_{uv}} (y_p - \mu_y)(\hat{y}_p - \mu_{\hat{y}}). \end{aligned} \tag{A76}$$

Equivalently, relative to population moments computed with denominator N_w, this corresponds to the sample-covariance correction factor

$$\frac{N_w}{N_w - 1} = \frac{49}{48}. \tag{A77}$$

For each valid spatial location and color channel, the local SSIM index is

$$\mathrm{SSIM}_{uv} = \frac{(2\mu_y \mu_{\hat{y}} + C_1)(2\sigma_{y\hat{y}} + C_2)}{\left(\mu_y^2 + \mu_{\hat{y}}^2 + C_1\right)\left(\sigma_y^2 + \sigma_{\hat{y}}^2 + C_2\right)}, \tag{A78}$$

where the evaluation operates on the 255-level images with

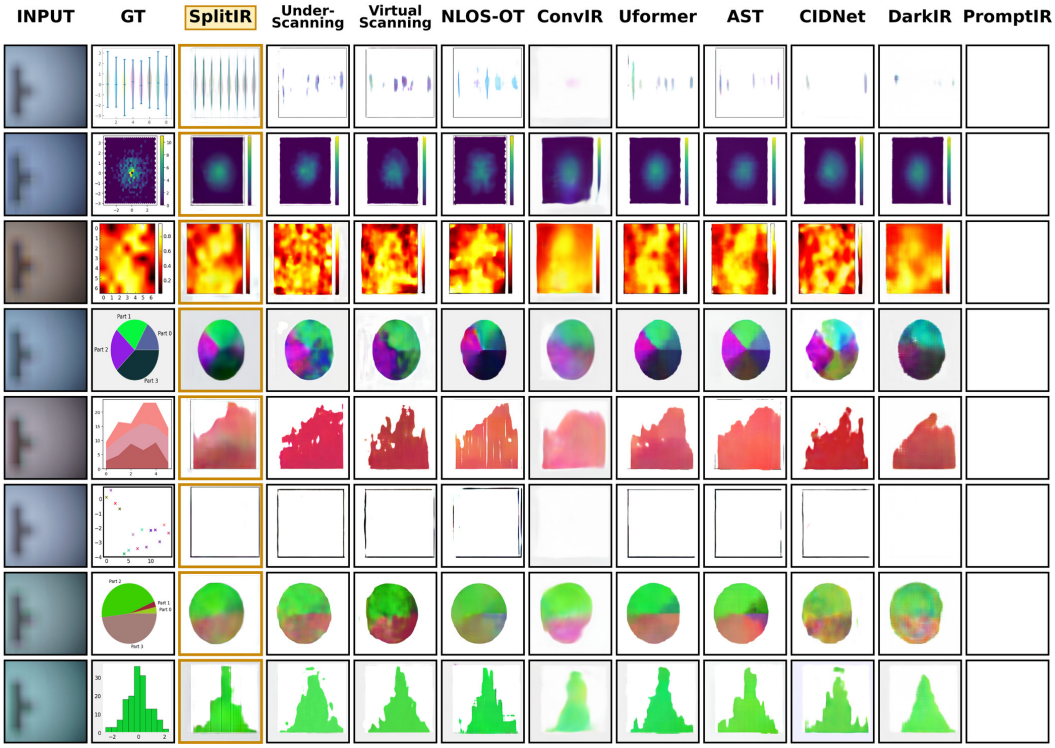


Figure A10: Qualitative comparison on the Chart subset, showing reconstruction of axes, curves, bars, annotations, thin boundaries, and other sparse geometric structures from indirect observations.

$$C_1 = (0.01 \times 255)^2 = 6.5025, \qquad C_2 = (0.03 \times 255)^2 = 58.5225. \tag{A79}$$

The width-three boundary associated with the 7×7 window is excluded before spatial averaging. SSIM is first averaged over valid spatial positions for each RGB channel and is subsequently averaged over the three channels. Higher SSIM indicates greater structural agreement.

Multi-scale structural similarity. MS-SSIM evaluates structural agreement across five progressively coarser resolutions. Starting from $y^{(1)} = y$ and $\hat{y}^{(1)} = \hat{y}$, consecutive scales are generated through 2×2 average pooling,

$$y^{(j+1)} = \mathcal{D}_2\left(y^{(j)}\right), \qquad \hat{y}^{(j+1)} = \mathcal{D}_2\left(\hat{y}^{(j)}\right), \qquad j = 1, \ldots, 4, \tag{A80}$$

where $\mathcal{D}_2$ denotes downsampling by local average pooling. At each scale, local statistics are computed using an 11×11 Gaussian window with standard deviation $\sigma_g = 1.5$. Since the input range is $[0, 1]$, the stability constants are

$$C_1 = 0.01^2, \qquad C_2 = 0.03^2. \tag{A81}$$

For scale j, define the luminance and joint contrast–structure components as

$$l_j = \frac{2\mu_y^{(j)}\mu_{\hat{y}}^{(j)} + C_1}{\left(\mu_y^{(j)}\right)^2 + \left(\mu_{\hat{y}}^{(j)}\right)^2 + C_1}, \tag{A82}$$

and

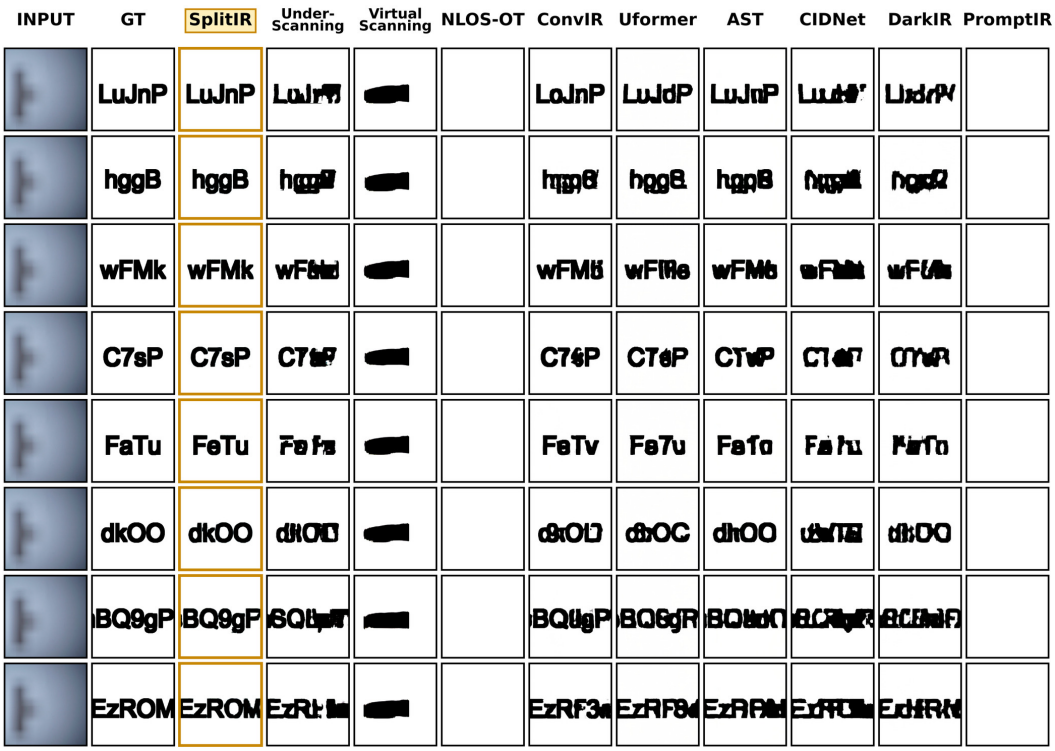


Figure A11: Qualitative comparison on the Password subset, emphasizing compact character strokes, local boundaries, character spacing, and foreground–background transitions.

$$cs_j = \frac{2\sigma_{y\hat{y}}^{(j)} + C_2}{\left(\sigma_y^{(j)}\right)^2 + \left(\sigma_{\hat{y}}^{(j)}\right)^2 + C_2}. \tag{A83}$$

At the finest four scales, the spatial and RGB average of the contrast–structure term is denoted by CS_j, whereas the complete SSIM average at the coarsest scale is

$$S_5 = \langle l_5 \, cs_5 \rangle_{\mathrm{RGB,space}} \, . \tag{A84}$$

With the fixed scale weights

$$\boldsymbol{\beta} = (0.0448,\, 0.2856,\, 0.3001,\, 0.2363,\, 0.1333)\,, \tag{A85}$$

the reported multi-scale score is

$$\mathrm{MS{-}SSIM} = [\max(S_5, 0)]^{\beta_5} \prod_{j=1}^{4} [\max(CS_j, 0)]^{\beta_j} \, . \tag{A86}$$

The implementation applies the default ReLU normalization before the cross-scale product. For CS_j, the width-five boundary associated with the Gaussian window is excluded before averaging; the coarsest full-SSIM term follows the reflection-padding behavior of the evaluation implementation. Higher MS-SSIM indicates greater multi-scale structural similarity.

Learned perceptual distance. LPIPS is evaluated with the AlexNet-based perceptual model through the same evaluation entry used for all final results. We use the metric's prescribed input normalization, network, and feature normalization consistently for every method and experiment.

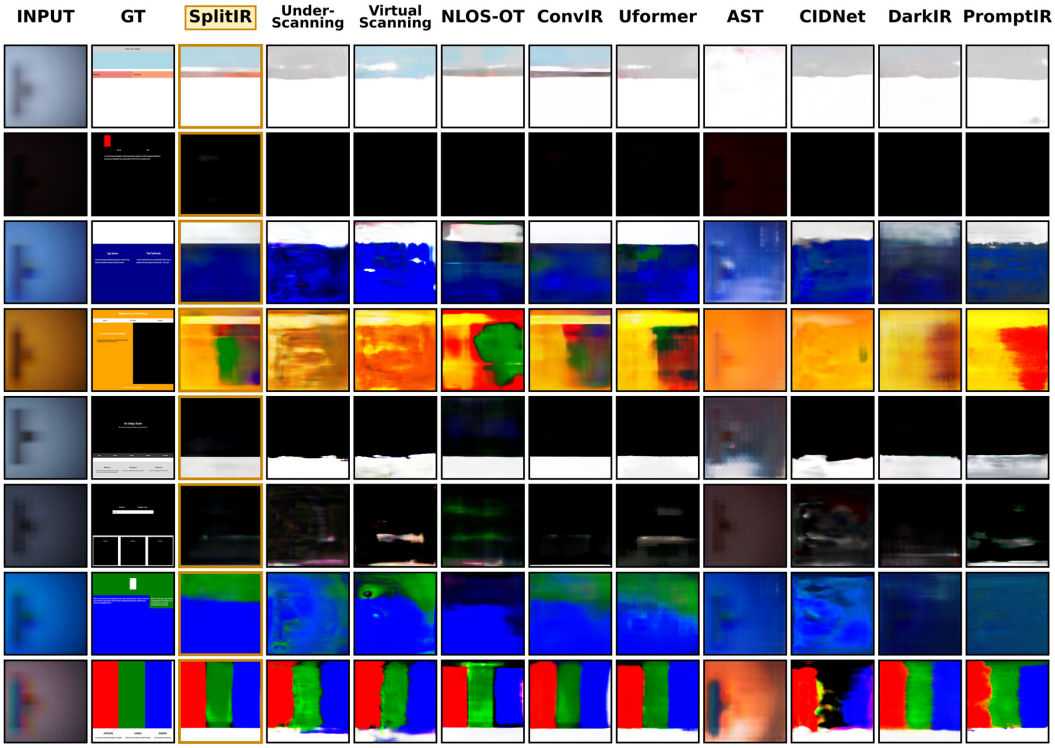


Figure A12: Qualitative comparison on the Website subset, showing page-level organization together with text, image regions, interface components, borders, and fine local structures.

Let $f_{\ell,cuv}(z)$ denote the activation at layer $\ell \in \{1,\dots,5\}$ after the metric's input processing and AlexNet feature extraction. The feature vector at each spatial position is channel-normalized as

$$\bar{f}_{\ell,cuv}(z) = \frac{f_{\ell,cuv}(z)}{\sqrt{\sum_k f_{\ell,kuv}^2(z) + \epsilon_f}}, \tag{A87}$$

where ϵ_f denotes the stabilization constant used by the unified LPIPS implementation.

For layer ℓ, the normalized feature discrepancy between prediction and reference is weighted channel-wise by the learned linear calibration coefficients $a_{\ell c}$. The per-image LPIPS distance is

$$\mathrm{LPIPS}\,(\hat{y}, y) = \sum_{\ell=1}^{5} \frac{1}{H_\ell W_\ell} \sum_{u=1}^{H_\ell} \sum_{v=1}^{W_\ell} \sum_{c} a_{\ell c} \left[\bar{f}_{\ell,cuv}(\hat{y}) - \bar{f}_{\ell,cuv}(y)\right]^2 . \tag{A88}$$

The layer-wise distances are spatially averaged and accumulated across the five AlexNet feature stages. Lower LPIPS indicates a smaller learned perceptual distance between the reconstruction and reference.

Dataset-level aggregation. Let $\mathcal{D}_{\mathrm{eval}} = \{(y_n, \hat{y}_n)\}_{n=1}^{N}$ denote an evaluation set and let $\mathcal{M}(y_n, \hat{y}_n)$ denote any of the six per-image metrics. The value reported in the quantitative tables is the arithmetic mean

$$\overline{\mathcal{M}} = \frac{1}{N} \sum_{n=1}^{N} \mathcal{M}\,(y_n, \hat{y}_n)\,. \tag{A89}$$

Accordingly, RMSE is computed for each image before dataset-level averaging,

Table A1: Comprehensive quantitative comparison on Screen and Password across six evaluation metrics. Best results are shown in **red**, and second-best results are shown in **blue**.

Method	Metrics					
	PSNR↑	SSIM↑	MS-SSIM↑	MSE↓	RMSE↓	LPIPS↓
Dataset I: Screen						
NLOS Reconstruction Methods						
DRS	20.44	0.6933	0.6619	1109.52	28.80	0.4040
DynamicNLOS	21.42	0.7155	0.6974	903.91	25.85	0.3793
Flatland	12.23	0.4734	0.3464	6605.63	73.01	0.6121
LIKAM	21.69	0.7267	0.7120	936.19	25.52	0.3392
LMS-NLOS	21.96	0.7066	0.7228	732.10	23.57	0.3501
NLOS-LPP	21.56	0.7143	0.6988	935.17	25.65	0.3620
NLOS-LTM	21.84	0.7262	0.7209	794.81	24.21	0.3350
NLOS-OT	**24.10**	**0.7933**	**0.8005**	**596.29**	**19.67**	**0.2619**
NLOST	14.78	0.5202	0.4441	3609.17	54.73	0.5412
NLOS-NeuS	12.32	0.4829	0.3220	5905.10	70.77	0.6445
TransiT	20.53	0.6889	0.6734	941.30	27.29	0.3688
UnderScanning	22.67	0.7567	0.7502	797.47	23.07	0.3055
Virtual Scanning	23.11	0.7690	0.7667	751.67	22.08	0.2973
General-Purpose Image Restoration						
U-Net	20.06	0.6691	0.6657	1213.30	30.23	0.3953
NAFNet	22.81	0.7513	0.7638	779.50	22.45	0.3023
Uformer	21.41	0.7148	0.7088	817.29	25.22	0.3546
ConvIR	22.84	0.7565	0.7592	673.48	21.79	0.3242
AST	22.19	0.7356	0.7390	770.48	23.60	0.3182
Related-Degradation Restoration Methods						
PromptIR	23.27	0.7689	0.7710	706.17	21.53	0.2921
InstructIR	10.18	0.4273	0.3156	8007.40	85.25	0.6598
DarkIR	20.83	0.6791	0.6830	991.64	27.23	0.3790
CIDNet	21.14	0.7141	0.6953	978.61	26.43	0.3728
SplitIR (Ours)	**27.16**	**0.8506**	**0.8551**	**480.40**	**15.75**	**0.1781**
Dataset II: Password						
NLOS Reconstruction Methods						
DRS	8.76	0.7949	0.4792	8768.45	93.32	0.4282
DynamicNLOS	14.05	0.8740	0.8705	2652.82	51.06	0.1374
Flatland	8.76	0.7949	0.4792	8768.45	93.32	0.4282
LIKAM	13.69	0.8735	0.8594	2894.70	53.28	0.1320
LMS-NLOS	13.56	0.8582	0.8522	2967.16	54.01	0.1432
NLOS-LPP	10.64	0.8147	0.6886	5718.99	75.26	0.2959
NLOS-LTM	11.76	0.8406	0.7641	4462.55	66.36	0.2536
NLOS-OT	8.76	0.7949	0.4792	8768.45	93.32	0.4282
NLOST	8.89	0.7863	0.4813	8513.17	91.95	0.4419
NLOS-NeuS	8.76	0.7935	0.4790	8768.38	93.32	0.4287
TransiT	8.76	0.7949	0.4792	8768.45	93.32	0.4282
UnderScanning	14.12	0.8819	0.8742	2610.49	50.66	0.1416
Virtual Scanning	9.90	0.8078	0.6450	6773.19	81.94	0.3030
General-Purpose Image Restoration						
U-Net	12.77	0.8563	0.8277	3568.18	59.20	0.1564
NAFNet	12.84	0.8470	0.8231	3479.75	58.58	0.1628
Uformer	14.58	0.8829	0.8876	2390.69	48.27	0.1176
ConvIR	**15.13**	**0.8910**	**0.8982**	**2148.46**	**45.55**	**0.1009**
AST	14.70	0.8790	0.8866	2311.01	47.51	0.1222
Related-Degradation Restoration Methods						
PromptIR	8.76	0.7949	0.4792	8768.45	93.32	0.4282
InstructIR	8.48	0.6920	0.3612	9261.91	96.15	0.5844
DarkIR	13.76	0.8572	0.8527	2831.78	52.76	0.1626
CIDNet	13.17	0.8605	0.8373	3214.01	56.34	0.1804
SplitIR (Ours)	**15.38**	**0.9000**	**0.9059**	**2067.82**	**44.50**	**0.0924**

$$\overline{\mathrm{RMSE}} = \frac{1}{N}\sum_{n=1}^{N}\sqrt{\mathrm{MSE}_{255}^{(n)}}, \tag{A90}$$

rather than as $\sqrt{\overline{\mathrm{MSE}}}$. PSNR is treated analogously:

$$\overline{\mathrm{PSNR}} = \frac{1}{N}\sum_{n=1}^{N} 10\log_{10}\left(\frac{255^2}{\mathrm{MSE}_{255}^{(n)}}\right). \tag{A91}$$

Table A2: Comprehensive quantitative comparison on Chart and Website across six evaluation metrics. Best results are shown in **red**, and second-best results are shown in **blue**.

Method	Metrics					
	PSNR↑	SSIM↑	MS-SSIM↑	MSE↓	RMSE↓	LPIPS↓
Dataset III: Chart						
NLOS Reconstruction Methods						
DRS	13.11	0.6516	0.3620	5555.39	65.55	0.6510
DynamicNLOS	16.57	0.7029	0.5753	1827.11	40.36	0.5045
Flatland	13.12	0.6516	0.3620	5555.36	65.55	0.6512
LIKAM	16.28	0.6991	0.5497	1959.47	41.78	0.5169
LMS-NLOS	16.68	0.7011	0.5838	1754.41	39.69	0.4914
NLOS-LPP	15.54	0.6874	0.4952	2275.18	45.28	0.5679
NLOS-LTM	16.57	0.7038	0.5666	1833.13	40.38	0.5035
NLOS-OT	16.61	0.7057	0.5820	1806.84	40.12	**0.4632**
NLOST	14.62	0.6130	0.4069	2991.66	51.30	0.6138
NLOS-NeuS	13.20	0.6458	0.3600	5357.21	64.62	0.6518
TransiT	13.10	0.6517	0.3622	5555.39	65.55	0.6511
UnderScanning	16.66	0.7051	0.5830	1810.50	40.04	0.4851
Virtual Scanning	16.48	0.7039	0.5655	1900.07	40.96	0.4959
General-Purpose Image Restoration						
U-Net	16.20	0.6836	0.5559	1951.99	41.94	0.5183
NAFNet	16.60	0.6985	0.5682	1818.04	40.25	0.4949
Uformer	**17.02**	**0.7179**	0.6091	**1605.02**	**38.04**	0.4894
ConvIR	15.73	0.6728	0.5040	2206.46	44.49	0.5801
AST	16.94	0.7173	**0.6139**	1612.23	38.22	0.4712
Related-Degradation Restoration Methods						
PromptIR	13.11	0.6517	0.3620	5555.38	65.57	0.6513
InstructIR	9.88	0.5649	0.2262	7167.20	83.15	0.7489
DarkIR	16.25	0.6876	0.5338	1946.73	41.80	0.5512
CIDNet	16.40	0.6982	0.5600	1875.18	40.99	0.5105
SplitIR (Ours)	**17.36**	**0.7408**	**0.6382**	**1527.58**	**36.83**	**0.4552**
Dataset IV: Website						
NLOS Reconstruction Methods						
DRS	18.38	0.7920	0.6680	1942.76	37.16	0.3800
DynamicNLOS	18.95	0.7957	0.6984	1785.05	35.12	0.3548
Flatland	14.76	0.7703	0.5902	5198.03	59.87	0.4236
LIKAM	19.52	0.7971	0.7262	1510.75	32.48	0.3437
LMS-NLOS	19.71	0.7746	0.7239	1317.00	31.01	0.3297
NLOS-LPP	17.19	0.7788	0.6183	2700.45	43.66	0.4042
NLOS-LTM	18.21	0.7805	0.6629	2012.05	37.77	0.3719
NLOS-OT	19.02	0.7935	0.7022	1566.76	33.86	0.3372
NLOST	15.73	0.7453	0.5563	3768.80	52.06	0.4418
NLOS-NeuS	15.09	0.7633	0.5830	4861.41	57.61	0.4186
TransiT	18.73	0.7849	0.6763	1688.44	35.25	0.3434
UnderScanning	19.54	0.7965	0.7237	1467.47	32.22	0.3422
Virtual Scanning	19.63	0.7975	0.7323	1413.85	31.77	0.3361
General-Purpose Image Restoration						
U-Net	18.36	0.7650	0.7004	1808.34	36.50	0.3641
NAFNet	15.74	0.7400	0.5839	3799.58	52.09	0.4199
Uformer	20.07	0.8075	0.7539	1127.51	29.29	0.3101
ConvIR	**20.46**	**0.8179**	**0.7701**	**1016.39**	**27.76**	**0.2913**
AST	15.43	0.7429	0.5738	3894.48	53.19	0.4239
Related-Degradation Restoration Methods						
PromptIR	19.57	0.8020	0.7305	1475.61	32.21	0.3371
InstructIR	12.57	0.7279	0.4777	4872.10	64.68	0.5067
DarkIR	19.67	0.7762	0.7214	1383.43	31.49	0.3333
CIDNet	19.26	0.7882	0.7044	1588.75	33.51	0.3550
SplitIR (Ours)	**20.74**	**0.8313**	**0.7814**	**982.72**	**27.08**	**0.2756**

PSNR, SSIM, and MS-SSIM are reported with higher values indicating closer agreement, whereas MSE, RMSE, and LPIPS are reported with lower values indicating smaller reconstruction error or perceptual distance.

C Supplementary Experiments

C.1 Datasets and Checkpoint Selection

The experiments are conducted on four paired hidden-screen datasets: Screen, Chart, Password, and Website. Each sample contains an indirect RGB observation recorded from the relay wall and its corresponding hidden-screen reference. The wall observation is used as the reconstruction input,

while the screen image provides the paired supervision and full-reference evaluation target. All samples are represented as 256×256 RGB images. The four subsets contain different forms of screen content and therefore exhibit distinct spatial organizations, edge distributions, text densities, and frequency characteristics.

C.1.1 SCREEN

The Screen subset contains general display content with heterogeneous visual composition. The hidden targets combine image regions, text, icons, interface components, colored areas, local textures, and foreground–background transitions within the same screen. Their spatial structures vary substantially across samples: some regions are dominated by smooth intensity variation, while others contain sharp contours, compact text, or irregular high-frequency details. This subset therefore represents general screen content spanning multiple structural patterns.

The Screen subset contains $2{,}121$ training pairs and 375 evaluation pairs. The indirect observations exhibit smooth radiometric variation on the relay wall, whereas the paired targets contain considerably denser spatial organization and local detail. Representative observation, target, and reconstruction results are provided in Fig. A9.

C.1.2 CHART

The Chart subset contains screen content dominated by structured graphical elements. Typical images include coordinate axes, curves, bars, legends, grid structures, text annotations, colored regions, and combinations of horizontal, vertical, and oblique primitives. Compared with general display imagery, a larger portion of the visual organization is encoded by thin boundaries and the relative positions of geometric components. The subset therefore contains both large-scale chart layouts and spatially sparse graphical structures with limited local support.

The Chart subset contains $5{,}000$ training pairs and $2{,}000$ evaluation pairs. The corresponding observation–target pairs expose a substantial difference between the spatially mixed wall measurements and the structured geometric organization of the hidden charts. Representative comparisons are provided in Fig. A10.

C.1.3 PASSWORD

The Password subset is dominated by character-based screen content. Its principal spatial information is concentrated in compact textual regions composed of narrow strokes, local edges, repeated character patterns, and foreground–background intensity transitions. Compared with the other subsets, the global layout is relatively sparse, while the local appearance depends more directly on stroke connectivity, character spacing, boundary localization, and local contrast.

The Password subset contains $5{,}000$ training pairs and $1{,}741$ evaluation pairs. Since the informative structures occupy relatively small spatial regions in the hidden image, their corresponding observations encode character structure through distributed radiometric variations over the relay surface. Representative comparisons are shown in Fig. A11.

C.1.4 WEBSITE

The Website subset contains webpage-style screen content with multi-scale spatial organization. The hidden targets combine text blocks, image regions, navigation elements, interface components, background areas, separators, and other page-level structures. These elements are organized within a common layout, producing a mixture of large structural regions and substantially finer local details. Consequently, the same target may contain low-frequency page organization together with high-frequency characters, icons, and boundaries.

The Website subset contains $4{,}000$ training pairs and $1{,}000$ evaluation pairs. Its paired observations and targets illustrate how global page organization and local interface structures are jointly encoded into the indirect wall observation after diffuse transport. Representative comparisons are provided in Fig. A12.

Table A3: Ablation study on Screen and Password across six evaluation metrics. Best results are shown in **red**, and second-best results are shown in **blue**.

Method	Metrics					
	PSNR↑	SSIM↑	MS-SSIM↑	MSE↓	RMSE↓	LPIPS↓
Dataset I: Screen						
PAM Ablations						
w/o PAM	22.648	0.769	0.756	781.657	22.973	0.318
w/o PAM Stage Embedding	**22.745**	0.768	0.758	764.067	22.763	**0.317**
PAM Metric → Identity	22.172	0.754	0.738	822.581	24.053	0.334
PAM → Residual Fusion	22.671	0.769	0.753	791.674	23.027	0.317
PAM Mobility → Constant	22.733	**0.773**	**0.761**	**760.799**	**22.758**	0.318
RCM Ablations						
w/o RCM	22.543	0.770	0.758	774.771	23.182	0.321
w/o RCM Stage Embedding	22.305	0.758	0.743	814.972	23.783	0.331
w/o RCM Box Projection	21.884	0.747	0.720	857.999	24.712	0.346
RCM → Gated Residual	22.546	0.768	0.751	765.024	23.114	0.322
RCM Metric → Identity	22.189	0.753	0.736	819.116	23.983	0.334
SplitIR (Full)	**27.163**	**0.851**	**0.855**	**480.406**	**15.746**	**0.178**
Dataset II: Password						
PAM Ablations						
w/o PAM	13.648	0.870	0.850	2926.583	53.556	0.145
w/o PAM Stage Embedding	13.645	0.869	0.850	2937.083	53.615	0.142
PAM Metric → Identity	13.088	0.861	0.825	3311.780	57.044	0.177
PAM → Residual Fusion	13.713	0.870	0.852	2886.535	53.174	0.142
PAM Mobility → Constant	13.189	0.862	0.830	3230.982	56.361	0.173
RCM Ablations						
w/o RCM	**14.228**	**0.878**	**0.871**	**2579.629**	**50.198**	**0.129**
w/o RCM Stage Embedding	13.323	0.864	0.836	3136.758	55.519	0.168
w/o RCM Box Projection	13.549	0.866	0.844	2987.010	54.136	0.156
RCM → Gated Residual	14.012	0.876	0.863	2713.185	51.474	0.132
RCM Metric → Identity	13.223	0.863	0.832	3209.566	56.162	0.171
SplitIR (Full)	**15.384**	**0.900**	**0.906**	**2067.832**	**44.503**	**0.092**

C.1.5 Checkpoint Selection

A dataset-specific SplitIR checkpoint is selected independently for each of the four subsets. Training proceeds for 150 epochs, and checkpoint evaluation is performed every five epochs. SSIM on the corresponding evaluation partition is used as the checkpoint-selection criterion. The checkpoint associated with the selected SSIM value is then retained for the final quantitative evaluation on that subset.

The same selection procedure is applied to Screen, Chart, Password, and Website. The evaluation partition used for checkpoint selection is also used for final reporting of PSNR, SSIM, MS-SSIM, MSE, RMSE, and LPIPS. Once selected, the dataset-specific checkpoint is fixed across the main comparison, robustness tests, interventions, internal-feature analysis, and operator diagnostics for that subset.

C.2 Extended Comparison Experiments

The four subsets differ in spatial organization and in the type of structure carried by their hidden targets. We analyze the complete six-metric comparison by content category to connect each result to the hidden target's spatial organization. Screen and Password are reported in Table A1, while Chart and Website are reported in Table A2. The comparison uses PSNR, SSIM, MS-SSIM, MSE, RMSE, and LPIPS under the evaluation definitions introduced in Sec. B.4. The corresponding qualitative results are presented separately for each subset so that the quantitative measurements can be examined together with the reconstructed spatial structures.

C.2.1 Screen Comparison

Screen tests whether heterogeneous content can be recovered within one coherent image. SplitIR leads all six metrics, with NLOS-OT providing the strongest comparison. Fig. A9 shows the corresponding combination of broad-region placement, window boundaries, and localized content. Together, the recovered organization and component boundaries turn the diffuse wall observation into a structured screen image.

Table A4: Ablation study on Chart and Website across six evaluation metrics. Best results are shown in **red**, and second-best results are shown in **blue**.

Method	Metrics					
	PSNR↑	SSIM↑	MS-SSIM↑	MSE↓	RMSE↓	LPIPS↓
Dataset III: Chart						
PAM Ablations						
w/o PAM	16.671	0.724	0.574	1791.382	39.942	0.506
w/o PAM Stage Embedding	16.701	0.724	0.579	1780.828	39.816	0.504
PAM Metric → Identity	16.626	0.723	0.573	1807.126	40.135	0.508
PAM → Residual Fusion	16.740	0.725	0.579	1770.103	39.669	0.505
PAM Mobility → Constant	16.731	0.725	0.578	1769.615	39.692	0.504
RCM Ablations						
w/o RCM	16.727	0.726	0.581	1778.394	39.745	0.501
w/o RCM Stage Embedding	16.664	0.722	0.575	1792.841	39.970	0.507
w/o RCM Box Projection	**16.813**	0.727	0.585	**1738.787**	**39.320**	0.500
RCM → Gated Residual	16.807	**0.728**	**0.587**	1744.098	39.364	**0.497**
RCM Metric → Identity	16.642	0.723	0.573	1802.309	40.078	0.509
SplitIR (Full)	**17.358**	**0.741**	**0.638**	**1527.582**	**36.834**	**0.455**
Dataset IV: Website						
PAM Ablations						
w/o PAM	20.271	0.818	0.759	1122.995	28.921	0.302
w/o PAM Stage Embedding	20.318	0.818	0.758	1103.758	28.700	0.302
PAM Metric → Identity	16.524	0.787	0.623	2897.069	46.075	0.408
PAM → Residual Fusion	20.221	0.818	0.755	1122.167	28.937	0.310
PAM Mobility → Constant	20.088	0.819	0.750	1166.897	29.480	0.316
RCM Ablations						
w/o RCM	**20.425**	**0.822**	**0.762**	**1043.992**	**28.101**	**0.300**
w/o RCM Stage Embedding	20.081	0.815	0.749	1165.254	29.487	0.317
w/o RCM Box Projection	20.276	0.821	0.756	1111.466	28.797	0.313
RCM → Gated Residual	20.283	0.820	0.758	1116.030	28.851	0.306
RCM Metric → Identity	20.290	0.817	0.759	1117.330	28.843	0.304
SplitIR (Full)	**20.767**	**0.827**	**0.782**	**960.744**	**26.889**	**0.275**



Figure A13: Qualitative ablation results on the Screen subset. The visualization compares the complete SplitIR configuration with PAM and RCM variants under heterogeneous screen content containing global regions, text-like structures, boundaries, and local details.

C.2.2 Password Comparison

Password brings the comparison to character-scale structure. ConvIR uses convolutional restoration operators across several image degradation tasks Cui et al. (2024b). SplitIR leads all six Password metrics in Table A1, including an LPIPS of 0.092, compared with 0.101 for ConvIR. Fig. A11 locates the reconstruction differences at stroke connections, character spacing, and foreground–background boundaries. The comparison extends the recovery evidence from broad screen layouts to the compact strokes and spacing that define character patterns.

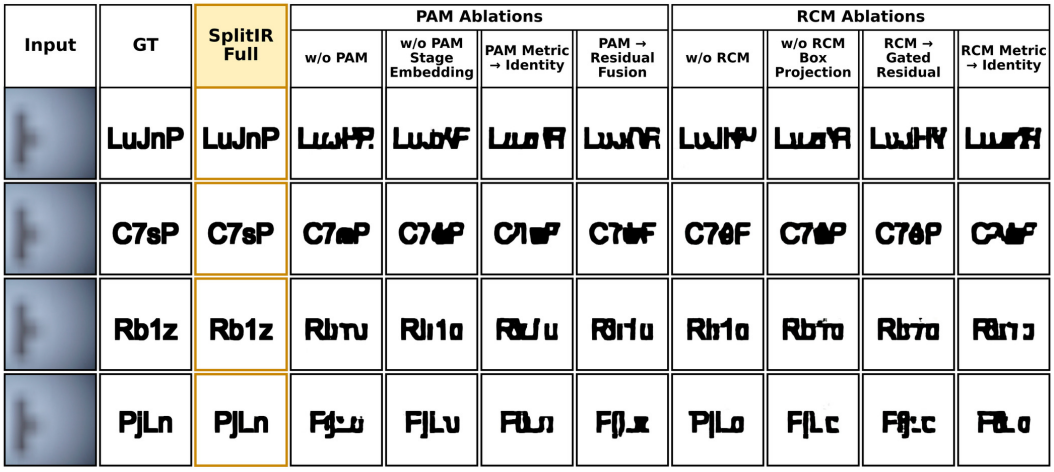


Figure A14: Qualitative ablation results on the Password subset. The comparison focuses on compact character strokes, boundary localization, character spacing, and foreground–background structure under different PAM and RCM configurations.

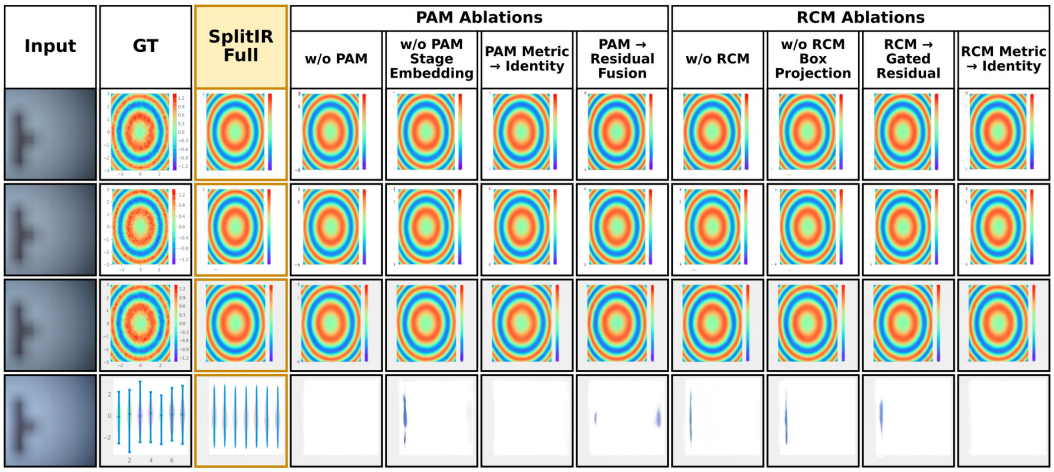


Figure A15: Qualitative ablation results on the Chart subset. The visualization compares the reconstruction of axes, curves, bars, annotations, thin boundaries, and other sparse geometric structures under different PAM and RCM configurations.

C.2.3 Chart Comparison

Chart requires the relative placement of graphical components and the preservation of thin structures within them. The strongest baseline varies with the criterion: Uformer leads the baseline comparison on PSNR, SSIM, MSE, and RMSE; AST on MS-SSIM; and NLOS-OT on LPIPS. AST combines adaptive sparse attention with attentive feature refinement Zhou et al. (2024); SplitIR reaches 0.638 MS-SSIM, compared with 0.614 for AST, and leads all six reconstruction metrics. Fig. A10 connects this result to the joint recovery of chart organization and sparse graphical structure, consistent with forming a shared layout before refining local components.

C.2.4 Website Comparison

Website combines the demands exposed separately by the preceding categories: page regions establish the layout, while local boundaries separate neighboring interface components. SplitIR leads all six metrics, with ConvIR ranking second. PromptIR conditions restoration on learned degradation prompts Potlapalli et al. (2023), while NLOS-LTM conditions hidden-image reconstruction and reprojection on a learned transport representation Zhang et al. (2025). SplitIR reaches 0.831 SSIM

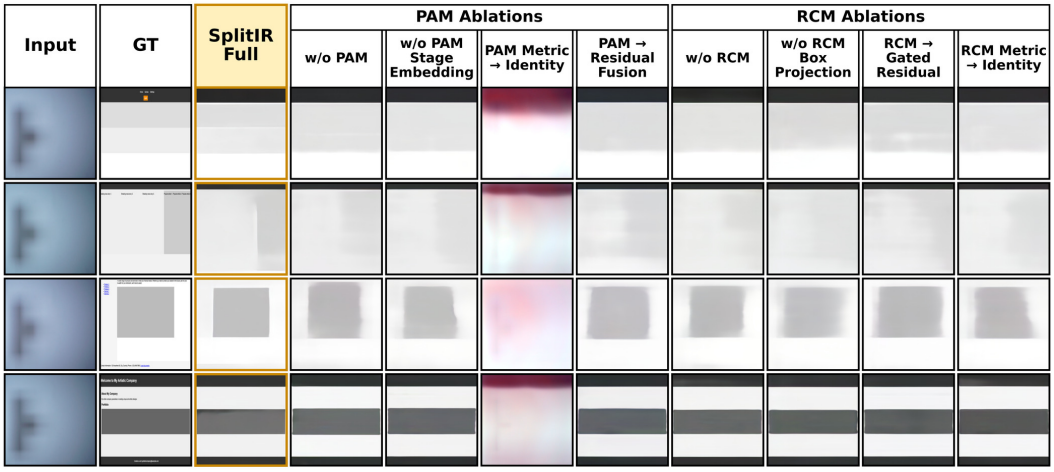


Figure A16: Qualitative ablation results on the Website subset. The comparison shows the effect of different PAM and RCM configurations on page-level organization together with text, interface components, borders, and fine local structures.

on Website, compared with 0.802 for PromptIR and 0.781 for NLOS-LTM. Fig. A12 shows this advantage across both scales, linking page-level organization to the separation of local regions.

C.3 Supplementary Ablation Studies

We conduct a structured ablation study to examine the two operators that form SplitIR. For each dataset, ten configurations are evaluated: five variants modify the Potential Assimilation Module (PAM), and five variants modify the Resolvent Correction Module (RCM). The PAM ablations cover complete module removal, removal of stage embedding, replacement of the learned metric by an identity metric, replacement of PAM by residual fusion, and replacement of adaptive mobility by a constant value. The RCM ablations similarly cover complete module removal, removal of stage embedding, removal of the box-constrained projection, replacement by gated residual refinement, and replacement of the learned correction metric by an identity metric.

All configurations are evaluated using the same six reconstruction criteria adopted in the comparative experiments: PSNR, SSIM, MS-SSIM, MSE, RMSE, and LPIPS. To preserve the content-dependent analysis used throughout the appendix, Screen and Password are grouped in Table A3, while Chart and Website are grouped in Table A4. Within each dataset, the PAM and RCM variants are shown as separate blocks, followed by the complete SplitIR configuration. The corresponding qualitative ablations are presented separately in Figs. A13–A16, allowing the metric variations to be examined together with their image-domain reconstruction behavior.

C.3.1 PAM Ablation Analysis

The PAM variants examine three aspects of observation assimilation: whether an explicit assimilation operator is required, whether the observation–layout discrepancy requires a learned representation geometry, and whether the resulting transport should depend on the current reconstruction stage and local evidence. Accordingly, the experiments remove PAM, remove its stage embedding, replace the learned metric by an identity metric, replace the complete operator by residual fusion, or replace adaptive mobility by a constant value. The same five variants are evaluated on all four content subsets in Tables A3 and A4.

The value of the learned discrepancy geometry varies with the spatial organization of the screen content. The complete configuration gives the best performance across all six metrics on all four subsets, while the internal variants exhibit content-dependent ordering. On Screen, constant mobility retains the second-best SSIM, MS-SSIM, MSE, and RMSE, whereas removing the PAM stage embedding gives the second-best PSNR and LPIPS. On Chart, residual fusion and constant mobility remain relatively close to one another, but both remain below the complete formulation. On Password, replacing the learned metric or adaptive mobility produces a larger reduction in structural

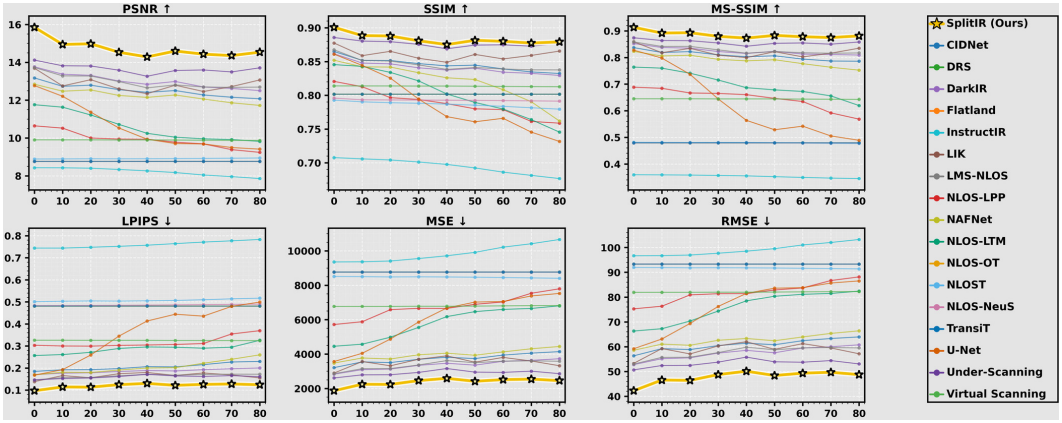


Figure A17: Six-metric robustness curves under brightness reduction on the Password subset. PSNR, SSIM, MS-SSIM, MSE, RMSE, and LPIPS are reported for $\Delta L \in \{0, 10, 20, 30, 40, 50, 60, 70, 80\}$ nit-equivalent.

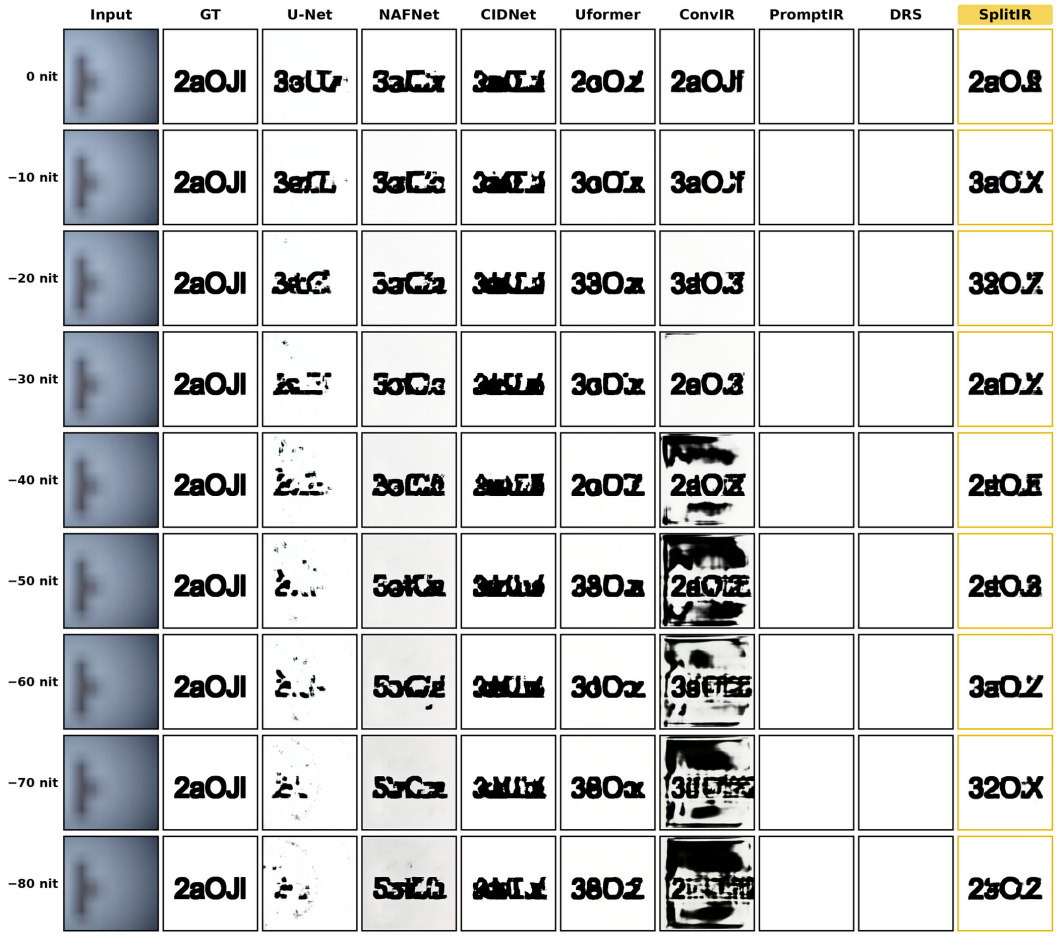


Figure A18: Representative Password reconstructions under increasing brightness reduction. The examples visualize the preservation of character strokes, local boundaries, and spatial arrangement as the indirect wall observation becomes progressively weaker.

similarity than removing PAM itself. The most pronounced PAM change occurs on Website when the learned metric is replaced by an identity metric: PSNR decreases from 20.767 to 16.524 dB, while MS-SSIM decreases from 0.782 to 0.623. These results indicate that the contribution of indi-

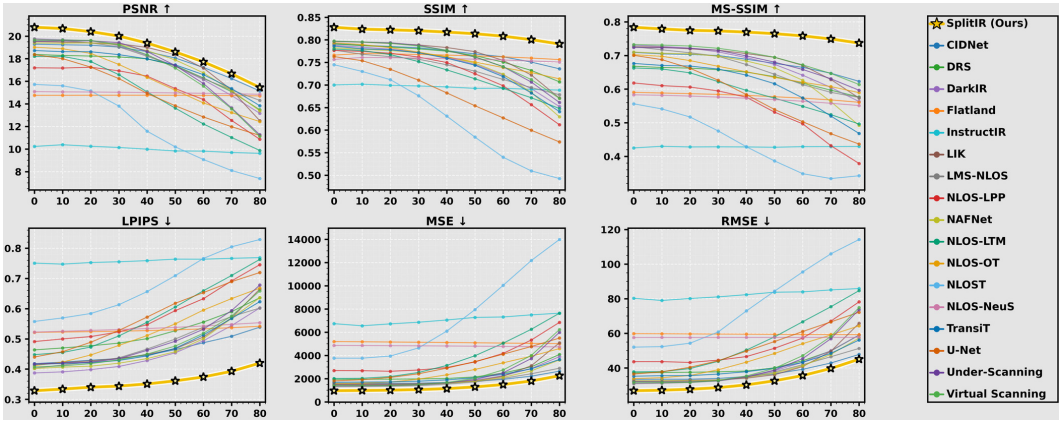


Figure A19: Six-metric robustness curves under brightness reduction on the Website subset. The curves describe reconstruction quality from $\Delta L = 0$ to 80 nit-equivalent across pixel-domain, structural, and perceptual criteria.

vidual PAM components depends on how the target content distributes spatial evidence across the indirect observation.

The qualitative ablations in Figs. A13–A16 provide the image-domain counterpart to these measurements. Across the four content categories, modifying PAM changes the preservation of spatial organization and local structure after the indirect evidence has been mapped into the reconstruction representation. The visual differences are especially informative for content whose identifying structure is spatially compact, such as character strokes in Password and thin graphical elements in Chart, as well as for Website images that combine global page organization with local interface details.

This behavior is consistent with the physical formation process and the operator formulation in Sec. B.2. Diffuse wall transport distributes the contribution of a local screen region over a broader relay area, so the observation evidence and the evolving layout representation do not retain direct spatial correspondence. PAM evaluates their discrepancy under the learned positive metric M_i^p, converts the resulting potential variation into a transport direction, and regulates its write-back through stage-conditioned mobility. Replacing M_i^p with an identity metric removes the adaptive geometry of this discrepancy. The stage embedding provides an explicit identifier for each assimilation call within the shared conditioning map, while adaptive mobility sets the magnitude of the evidence-driven update. The quantitative and qualitative ablations therefore connect the learned discrepancy geometry and bounded transport to layout-state formation before image refinement.

C.3.2 RCM Ablation Analysis

The RCM variants examine how stage-wise decoder proposals are regulated before they are written into the current reconstruction. The evaluated configurations remove RCM, remove its stage embedding, remove the box-constrained projection while retaining the preceding correction operation, replace RCM by gated residual refinement, or replace the learned correction metric with an identity metric. These modifications separately alter the presence of the correction operator, its stage-dependent representation, its feasible-set constraint, and the metric used to reconcile candidate corrections.

The RCM variants expose content-dependent demands on correction regulation. The complete formulation gives the best performance across all six metrics on all four datasets, while the simplified variants show different responses across content categories. On Screen, removing the box projection gives the weakest performance among the five RCM variants across all six metrics, including 21.884 dB PSNR, 0.720 MS-SSIM, and 0.346 LPIPS. On Chart, the two closest RCM variants divide the second-best measurements: the projection-free configuration gives the second-best PSNR, MSE, and RMSE, while gated residual refinement gives the second-best SSIM, MS-SSIM, and LPIPS. On Password and Website, removing the complete RCM produces the strongest result among the simplified RCM configurations, but remains below the complete model on all six criteria. The varying

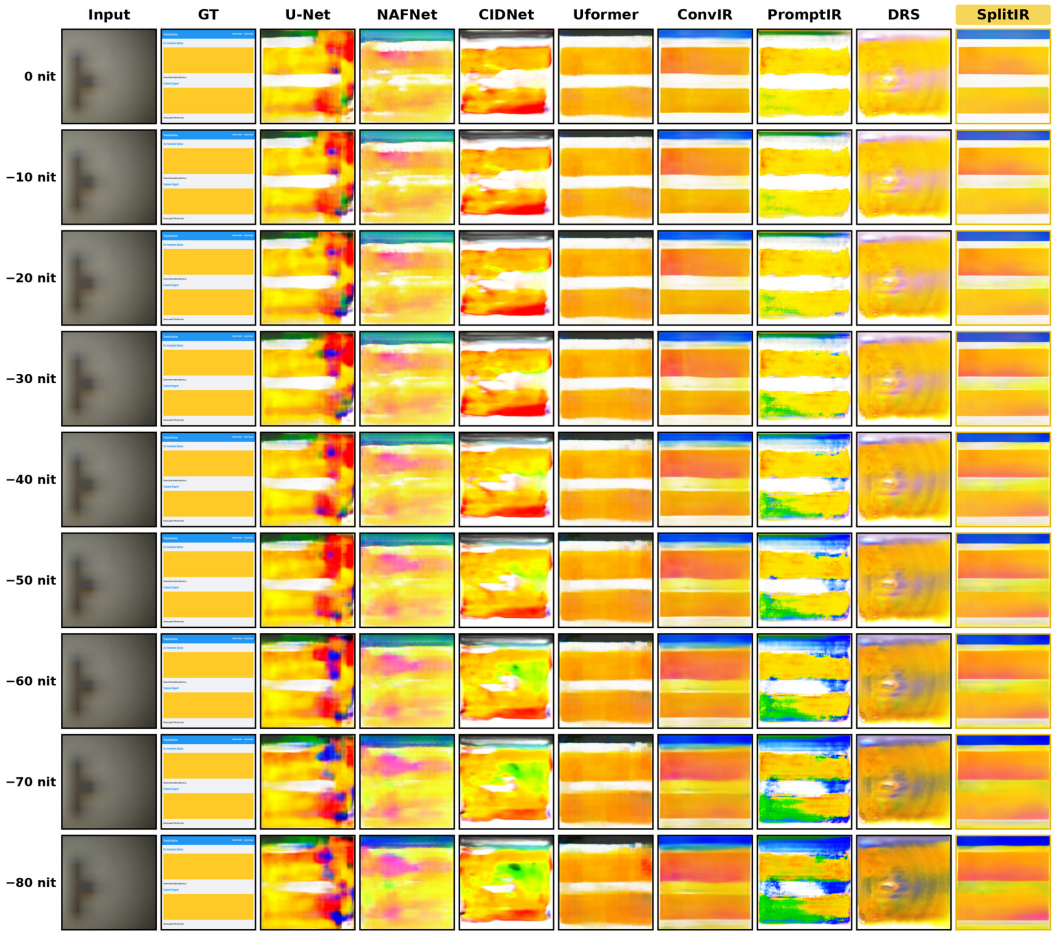


Figure A20: Representative Website reconstructions under increasing brightness reduction, showing the evolution of page-level organization, text regions, boundaries, and local interface structures.

ordering indicates that the effect of correction regulation depends on the spatial form of the hidden target rather than on a single pixel-domain error characteristic.

The qualitative results in Figs. A13–A16 further expose the effect of modifying the sequential correction process. RCM variants operate after the shared layout representation has already been decoded, so the visual differences are concentrated in the refinement of structures that remain ambiguous or incomplete after the base reconstruction. These differences can be inspected in the continuity of local boundaries, compact character structures, geometric primitives, and fine page elements across the four datasets. The complete configuration retains both the proposal adaptation and the stage-dependent admissibility constraint throughout the refinement sequence.

The corresponding mechanism follows the RCM construction in Sec. B.2. Each decoder stage generates a candidate correction from a reconstruction state that already contains the effects of preceding updates. RCM embeds this candidate in a stage-dependent correction representation, evaluates it under the learned positive metric M_k^r, forms a metric-aware proximal state, and restricts the admitted correction through projection onto the stage-conditioned feasible set C_k. The RCM stage embedding explicitly conditions the shared correction map on its position in the refinement sequence; replacing the metric by the identity removes dimension-dependent correction geometry; and removing the box projection removes the final feasible-set restriction. The quantitative and qualitative ablation responses are therefore consistent with RCM acting as an image-domain regulation operator for sequential decoder corrections.

Across the four content categories, the two ablation groups expose different stages of the SplitIR reconstruction process. PAM variants primarily alter the conversion of spatially mixed wall evidence into the shared layout state, whereas RCM variants alter the admission of successive corrections

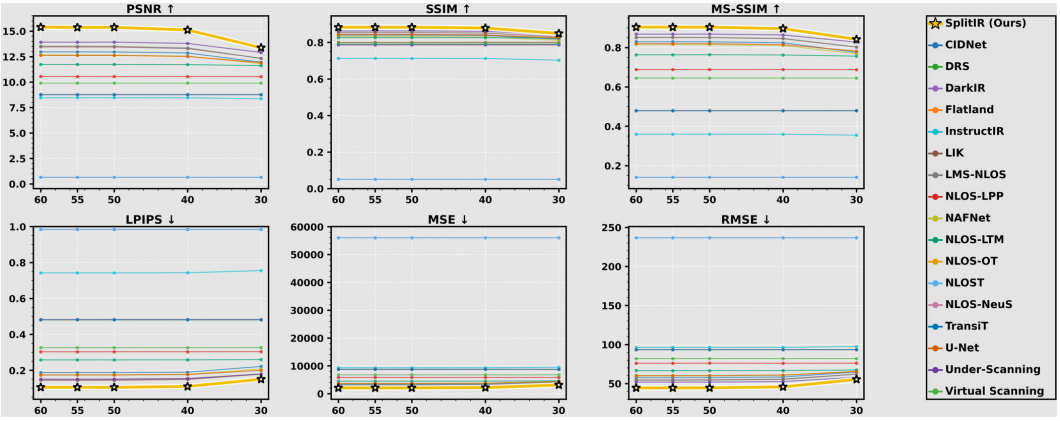


Figure A21: Six-metric robustness curves under additive Gaussian noise on the Password subset at SNR levels of 60, 55, 50, 45, 40, 35, and 30 dB.

after decoding has begun. This distinction is visible both in the six quantitative criteria and in the corresponding qualitative reconstructions in Figs. A13–A16. The complete configuration retains the learned metric, stage dependence, bounded assimilation, and feasible-set correction mechanisms simultaneously, providing the reference configuration for the subsequent sensitivity analysis.

C.4 Supplementary Robustness Experiments

We further evaluate SplitIR as the wall observations become darker or noisier. Brightness reduction weakens radiometric evidence, whereas Gaussian perturbation adds spatially distributed uncertainty to that evidence. These two settings test complementary demands on wall-to-screen recovery: extracting structure from progressively weaker contrast and refining it from increasingly uncertain local evidence.

The experiments are conducted on Password and Website because the two subsets exhibit substantially different spatial organizations. Password concentrates most of its informative content in compact character strokes and local boundaries, whereas Website combines page-level layout with text, image regions, separators, and interface components. For each perturbation type, we report the six reconstruction metrics used throughout the paper (PSNR, SSIM, MS-SSIM, MSE, RMSE, and LPIPS) and visualize their evolution across degradation levels. Representative reconstructions are further provided for each dataset to connect the metric trajectories with the corresponding image-domain behavior.

C.4.1 Robustness to Brightness Reduction

Brightness reduction tests how much screen structure remains recoverable as the wall signal weakens at fixed hidden content and acquisition geometry. We apply controlled reductions of $\Delta L \in \{0, 10, 20, 30, 40, 50, 60, 70, 80\}$ nit-equivalent relative to the nominal reference level. The perturbation is applied to the indirect observation, while the paired hidden-screen target remains fixed. Consequently, increasing ΔL reduces the available radiometric contrast without changing the target structure.

This setting directly stresses the evidence-assimilation stage of the reconstruction problem. Diffuse transport has already distributed screen-dependent information over a smooth wall region; further attenuating the observation decreases the contrast from which the hidden spatial organization must be inferred. The complete six-metric trajectories for Password and Website are shown in Figs. A17 and A19, respectively.

Password On Password, the structural metrics change gradually as the wall signal darkens (Fig. A17). Across the full brightness range, SplitIR changes from 15.852 to 14.537 dB in PSNR, from 0.901 to 0.879 in SSIM, and from 0.097 to 0.124 in LPIPS between $\Delta L = 0$ and 80. The structural measurements therefore remain comparatively stable despite the reduction in the radiometric strength of the observation. This behavior is relevant to Password because most of its semantic

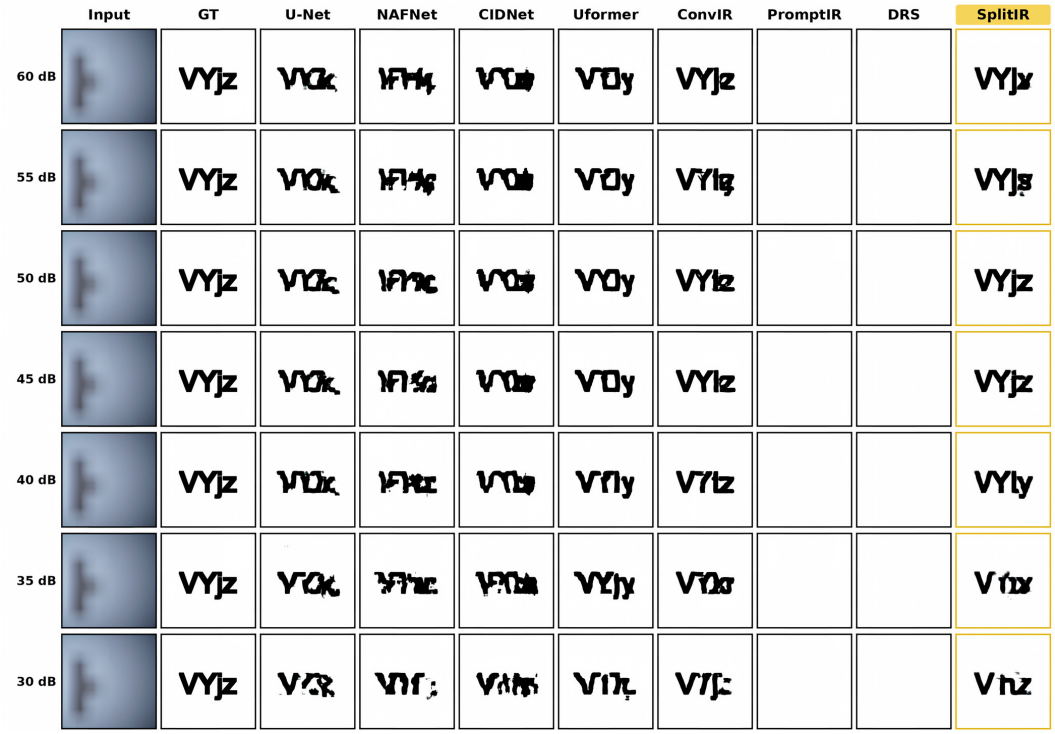


Figure A22: Representative Password reconstructions at seven SNR levels from 60 to 30 dB in steps of 5 dB, illustrating changes in character strokes, local boundaries, and compact spatial structures.

structure is concentrated in narrow strokes, local junctions, and foreground–background boundaries rather than in densely textured regions.

The qualitative examples in Fig. A18 provide the corresponding image-domain view. As the observation becomes darker, the available evidence for small character components is reduced, making stroke continuity and boundary localization increasingly difficult to infer. The reconstructed examples nevertheless preserve the dominant character arrangement over the tested range, which is consistent with the relatively gradual changes observed in SSIM, MS-SSIM, and LPIPS.

Website Website exhibits a different response because relevant information is distributed over multiple spatial scales. As shown in Fig. A19, SplitIR changes from 20.783 to 15.455 dB in PSNR and from 0.828 to 0.791 in SSIM over the same 0–80 reduction range, while LPIPS increases from 0.328 to 0.420. The larger pixel-domain change relative to Password is consistent with the more spatially distributed composition of Website targets: page regions, text blocks, borders, and local interface elements contribute jointly to the reconstruction error.

Fig. A20 shows the corresponding visual progression. Reduced observation intensity first limits the contrast available for local page elements and subsequently affects the separation between larger layout regions. The qualitative and quantitative results therefore describe the same degradation process from complementary perspectives: the global webpage organization remains observable over the tested range, while fine-scale evidence becomes progressively less distinct as the relay-wall signal weakens.

The two datasets also reveal how target structure interacts with radiometric attenuation. Password concentrates information in a small number of high-contrast local structures and therefore retains a relatively stable structural profile, whereas Website distributes relevant evidence over larger and more heterogeneous spatial regions. The same reduction of indirect intensity consequently produces different metric trajectories even though the perturbation protocol is identical.

Under reduced brightness, PAM evaluates the observation–layout discrepancy using the remaining wall evidence and regulates its transport into the shared layout state. RCM then adjusts and projects the corrections generated from the decoded state. The brightness experiment therefore follows layout

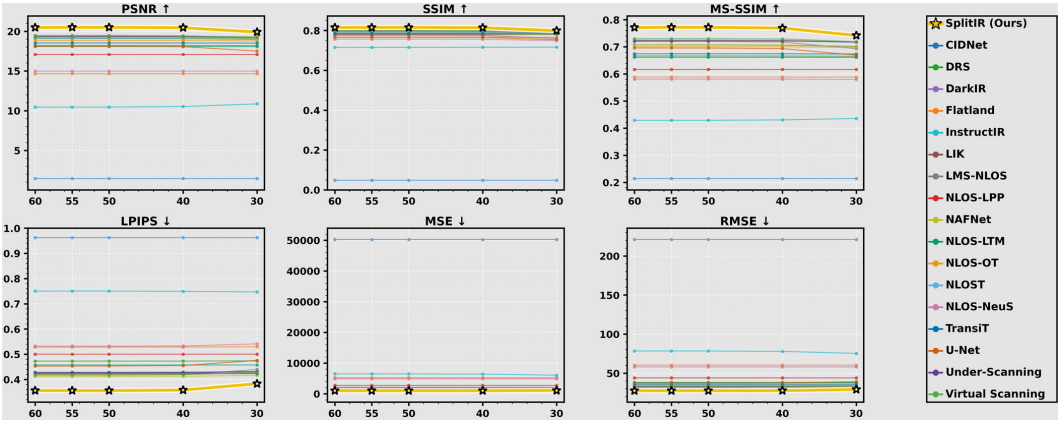


Figure A23: Six-metric robustness curves under additive Gaussian noise on the Website subset at SNR levels of 60, 55, 50, 45, 40, 35, and 30 dB.

formation and feasible correction as observable radiometric evidence decreases for a fixed hidden target.

C.4.2 Robustness to Additive Gaussian Noise

Gaussian noise is evaluated at 60, 55, 50, 40, and 30 dB SNR using the same prepared noisy observations for all compared methods at each level. Lower SNR denotes stronger noise, while the paired hidden-screen targets remain unchanged. Brightness reduction weakens radiometric evidence; Gaussian perturbation adds spatially distributed uncertainty to that evidence. The complete Password and Website trajectories are shown in Figs. A21 and A23. The quantitative comparison uses these five SNR levels, while the qualitative examples additionally show intermediate and stronger perturbations where indicated.

Password The Password curves in Fig. A21 exhibit a relatively flat region under weak and moderate noise, followed by a clearer change at the strongest tested condition. SplitIR records 15.399 dB PSNR, 0.883 SSIM, 0.904 MS-SSIM, and 0.105 LPIPS at 60 dB. At 30 dB, these values become 13.379 dB, 0.849, 0.842, and 0.150, respectively, while MSE and RMSE increase from 2052.7 and 44.38 to 3111.5 and 55.24.

The intermediate conditions show a comparatively stable higher-SNR response followed by a larger change at 30 dB. SSIM remains close to 0.883 from 60 to 50 dB and reaches 0.879 at 40 dB before decreasing more clearly at 30 dB. This pattern indicates that moderate perturbations preserve much of the evidence required for character organization, whereas stronger noise increasingly disrupts the local cues associated with narrow strokes and character boundaries.

The qualitative examples in Fig. A22 complement this trend. As noise increases, uncertainty becomes increasingly visible in the recovery of compact local structures. Character strokes provide a particularly sensitive visual indicator because small deviations in local evidence can alter continuity, thickness, and separation between neighboring components. The visual progression therefore provides a direct counterpart to the reductions in SSIM and MS-SSIM and the increase in LPIPS.

Website Fig. A23 shows a similar but content-dependent pattern on Website. SplitIR records 20.474 dB PSNR, 0.815 SSIM, 0.771 MS-SSIM, 0.356 LPIPS, 1000.8 MSE, and 27.59 RMSE at 60 dB. At 30 dB, the corresponding values are 19.903 dB, 0.799, 0.742, 0.383, 1089.3, and 29.10.

The higher-SNR region remains comparatively stable: PSNR stays near 20.5 dB from 60 through 40 dB before decreasing at 30 dB, while LPIPS remains near 0.356–0.358 over the same higher-SNR interval and then increases to 0.383. Website therefore retains its principal page organization under moderate Gaussian perturbation, while the strongest tested noise level produces a clearer loss of fine evidence. The behavior differs from Password in absolute scale because Website jointly contains broad page regions and high-frequency local elements.

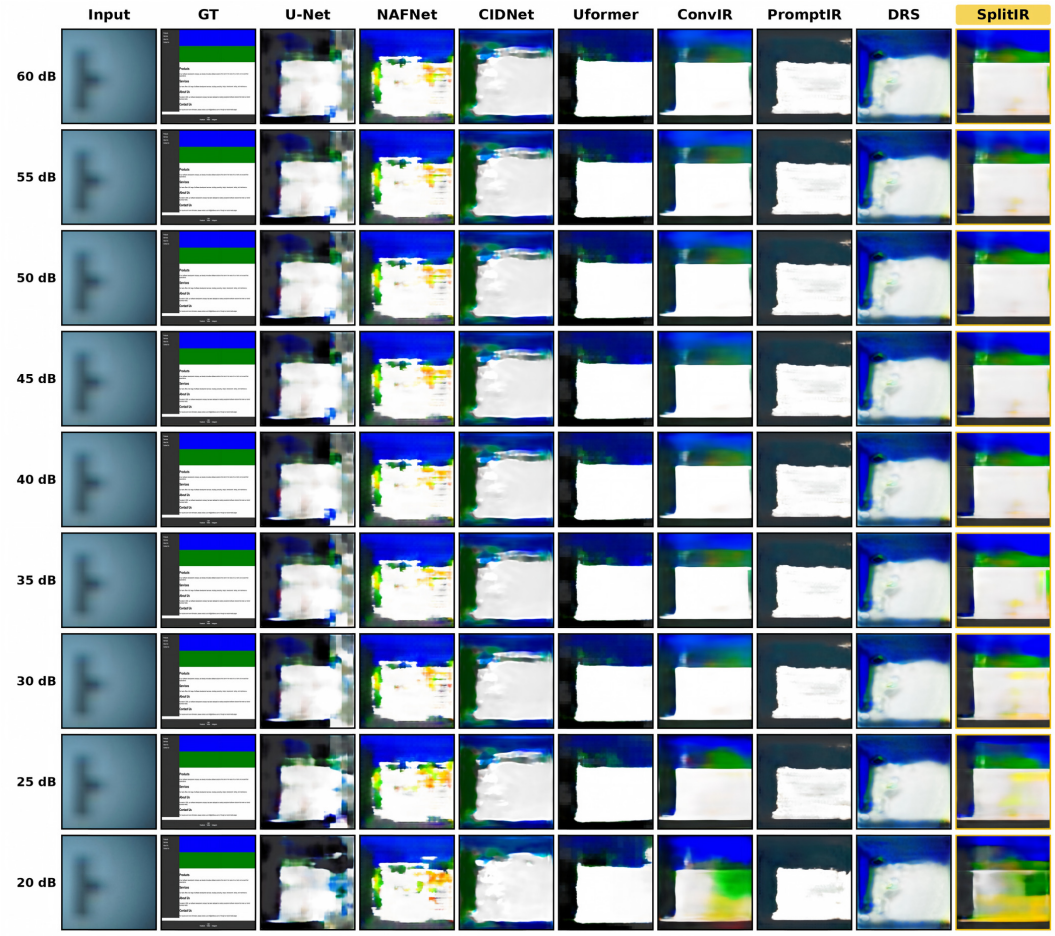


Figure A24: Representative Website reconstructions at nine SNR levels from 60 to 20 dB in steps of 5 dB, showing the evolution of page-level regions, text, boundaries, and local interface structures.

The qualitative comparison in Fig. A24 makes this distinction visible. Large layout regions provide relatively persistent spatial cues, whereas text, separators, borders, and compact interface elements are more sensitive to the increasing uncertainty of the indirect observation. The six quantitative metrics and the visual examples therefore describe complementary aspects of the same degradation: global structure can remain identifiable while local reconstruction fidelity progressively decreases.

The Gaussian-noise experiment probes a different aspect of the reconstruction pipeline from brightness reduction. Brightness attenuation decreases the strength of the available evidence, whereas Gaussian perturbation changes its local reliability. Under the latter condition, PAM must assimilate an observation in which evidence and noise coexist within the same spatial field. Its learned metric and bounded transport determine how this perturbed evidence enters the shared layout representation. RCM then operates on decoder proposals derived from that state and constrains their sequential incorporation through stage-conditioned correction geometry.

Across both datasets, the relatively stable higher-SNR region and the larger change at 30 dB indicate that reconstruction quality degrades as uncertainty begins to dominate increasingly fine observation cues. The qualitative results further show that this transition depends on target structure: compact characters expose local stroke degradation, whereas webpage content exposes the joint degradation of local details and large-scale organization. Together, the brightness and Gaussian experiments evaluate SplitIR under two complementary perturbations of the wall evidence while preserving the same hidden reconstruction objective.

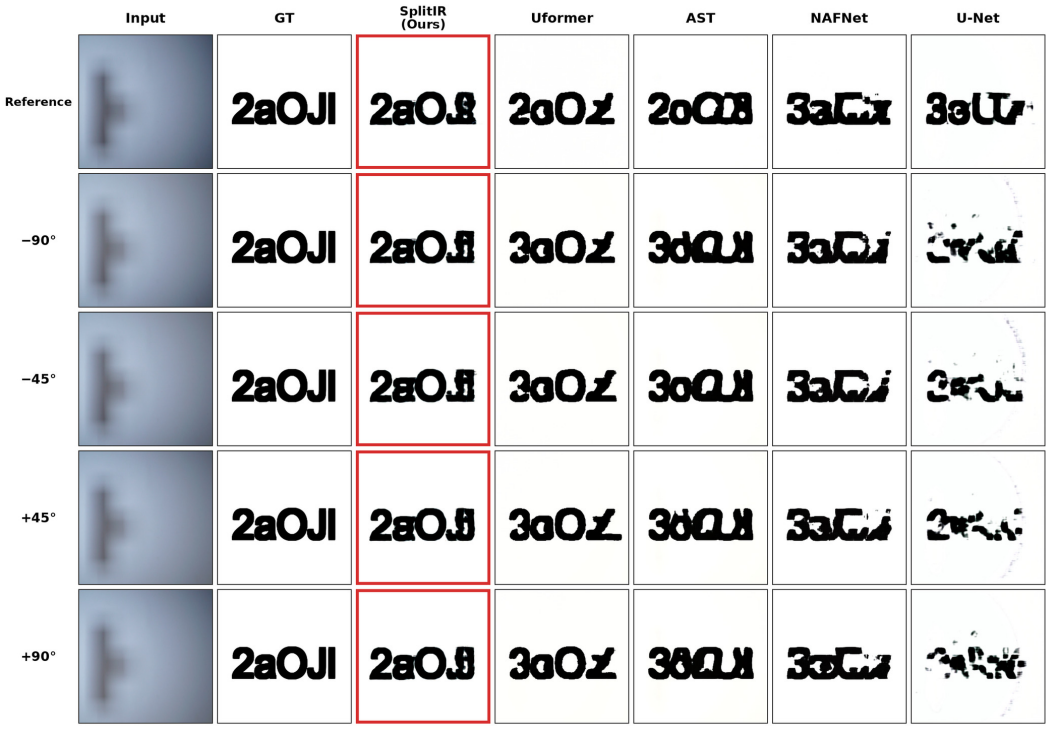


Figure A25: Representative Password reconstructions under different ambient-light directions. Each row uses the same hidden target, while the wall observation is reacquired under the corresponding physical illumination direction.

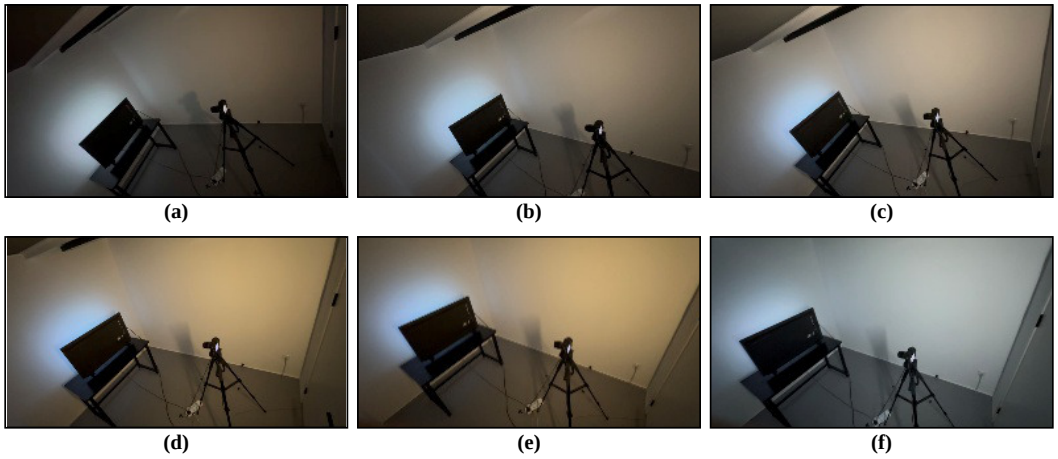


Figure A26: Representative acquisition views under different ambient illumination conditions, showing the display, camera, and illuminated relay wall.

C.4.3 Additional Environmental, Geometric, and Imaging Robustness

We further evaluate SplitIR under real acquisition changes in ambient illumination, camera motion, camera shake, relay-wall material, viewing distance, viewing pitch, and lens configuration. For each condition, the corresponding physical factor is changed and the wall RGB observation is reacquired while the hidden target, reconstruction model, and evaluation protocol remain fixed.

Figure A26 documents representative ambient-illumination conditions in the acquisition environment. The photographs show changes in the appearance of the relay wall and surrounding scene, providing physical context for the illumination experiments discussed below.

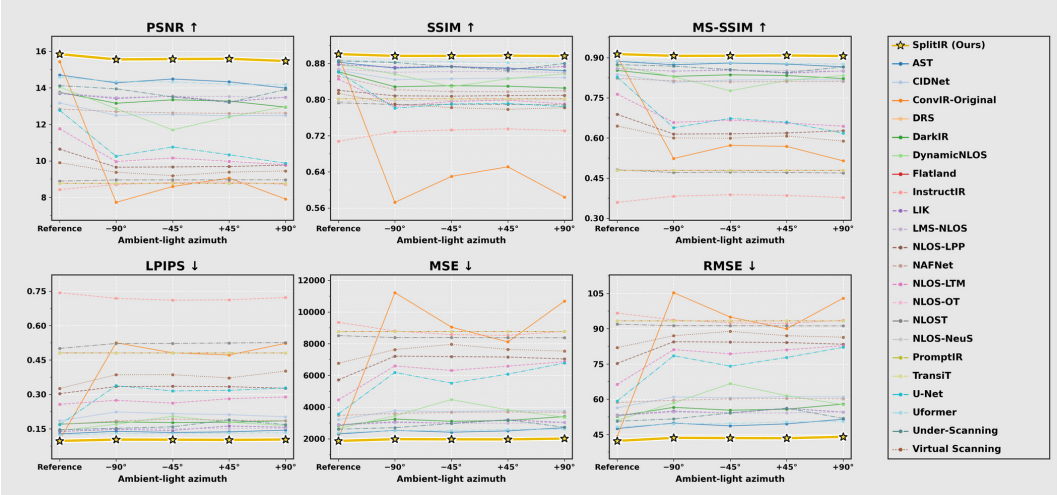


Figure A27: Six-metric robustness curves under ambient-light direction changes on the Password subset. The displayed conditions are the reference observation and directional fields at $-90°$, $-45°$, $+45°$, and $+90°$.

These experiments probe complementary changes along the acquisition path. Ambient illumination redistributes radiometric energy over the relay surface; motion and camera shake alter the recorded wall pattern through real camera–scene motion; relay-wall material changes the surface response; viewing distance and pitch change the geometric correspondence between the relay wall and camera; and the lens experiment changes the optical formation of the recorded observation. Together, the seven families examine illumination, surface response, spatial bandwidth, viewing geometry, and camera characteristics under real acquisition conditions.

For the six-metric experiments, PSNR, SSIM, MS-SSIM, MSE, RMSE, and LPIPS are evaluated over the complete corresponding evaluation set. The wall-material experiment is visualized using MS-SSIM, LPIPS, and MSE, while the viewing-distance, vertical-angle, and camera-lens experiments use PSNR, MS-SSIM, and LPIPS. For every experiment, the curve figure is followed by a fixed-scene qualitative comparison so that aggregate metric variation and spatial reconstruction behavior can be examined separately.

Ambient-light direction on Password. The first experiment changes the direction of the physical ambient illumination and reacquires the wall observation. The reference condition uses the reference illumination, while the four additional conditions use azimuths of $-90°$, $-45°$, $+45°$, and $+90°$. The hidden Password target and scene geometry remain fixed, so the comparison isolates how a direction-dependent illumination field redistributes radiometric evidence over the relay surface.

This perturbation is particularly relevant to Password because its target information is concentrated in compact foreground structures. The hidden characters modulate the smooth wall response recorded by the camera. Directional illumination therefore changes the local contrast of this response without changing which hidden characters generated it. The experiment tests whether the reconstruction depends on a narrowly fixed spatial illumination profile or can retain the target structure when that profile changes.

Fig. A27 shows that the metric variation remains limited across the four non-reference directions. SplitIR gives 15.852 dB PSNR, 0.901 SSIM, 0.914 MS-SSIM, and 0.097 LPIPS under the reference condition. Across the four directional perturbations, PSNR remains between 15.470 and 15.593 dB, SSIM between 0.897 and 0.897, MS-SSIM between 0.906 and 0.908, and LPIPS between 0.102 and 0.103. MSE and RMSE follow the same narrow variation. SplitIR retains the first position on all six displayed metrics across all five conditions within the comparison set.

The comparatively small change is consistent with the spatial statistics of Password. Although the illumination gradient modifies the wall image, the dominant target structure is compact and foreground–background separation remains encoded in the perturbed observation. The experiment

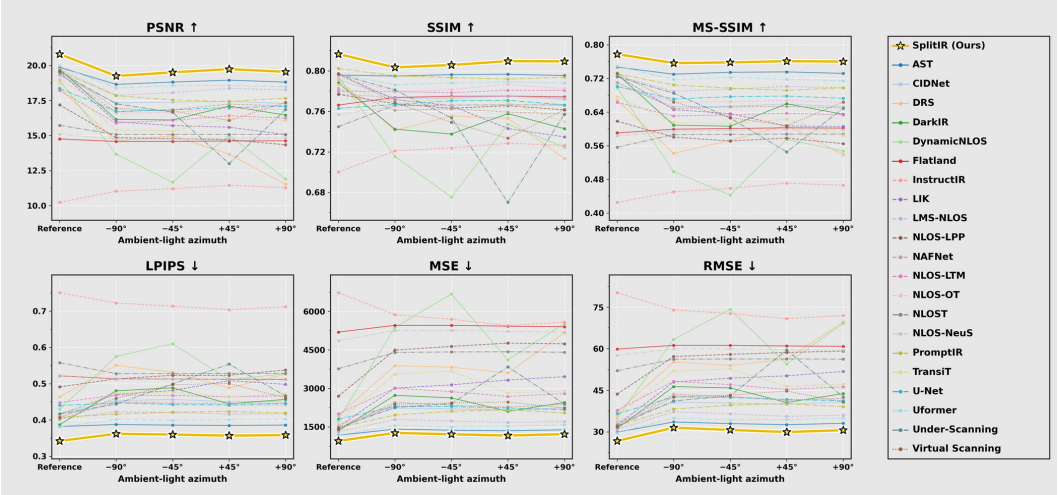


Figure A28: Six-metric robustness curves under ambient-light direction changes on the Website subset using the reference, $-90°$, $-45°$, $+45°$, and $+90°$ conditions.

therefore differs from severe geometric mismatch: the evidence is radiometrically redistributed, but its underlying coordinate system remains unchanged.

The qualitative comparison in Fig. A25 provides the image-domain counterpart of this result. The SplitIR reconstruction maintains the dominant character arrangement across the reference and four directional conditions. Changes are mainly concentrated around stroke thickness, terminal regions, and local character boundaries. Several competing reconstructions show larger changes in character shape or fragmentation when the illumination direction changes, whereas the main arrangement produced by SplitIR remains more stable across rows.

From the operator perspective, directional illumination modifies the observation evidence before layout reconstruction. PAM therefore receives a spatially varying radiometric change in the discrepancy between observation-derived features and the evolving layout state. Because the perturbation does not change the target coordinate system, the assimilation process can still organize this evidence into a common layout representation. RCM subsequently acts on spatially varying residual errors associated with the illumination field under fixed hidden-content geometry.

Ambient-light direction on Website. The same directional protocol is applied to Website to examine whether the response changes when target information is distributed over multiple spatial scales. Website contains large page regions together with text blocks, separators, navigation elements, and local interface details. The additional directional component therefore overlaps with both low-frequency page organization and finer local structures.

Compared with Password, this setting provides a more spatially distributed test. A local change in the wall illumination can affect several target components simultaneously because the reconstructed page distributes its content across multiple regions. The same five illumination conditions are retained so that the difference between Password and Website reflects content composition rather than a change in the perturbation protocol.

As shown in Fig. A28, the Website trajectory exhibits a larger response than Password. SplitIR reaches 20.817 dB PSNR, 0.817 SSIM, 0.778 MS-SSIM, and 0.342 LPIPS under the reference condition. Across the four directional fields, PSNR ranges from 19.253 to 19.743 dB, SSIM from 0.803 to 0.810, MS-SSIM from 0.756 to 0.761, and LPIPS from 0.357 to 0.362. The largest PSNR difference from the reference is approximately 1.56 dB.

SplitIR retains the leading result on all six metrics across the five illumination conditions. Fig. A29 shows that the dominant upper and lower page regions remain identifiable as the directional illumination changes. This persistence connects the comparative advantage to the recovery of page organization under spatially redistributed illumination.

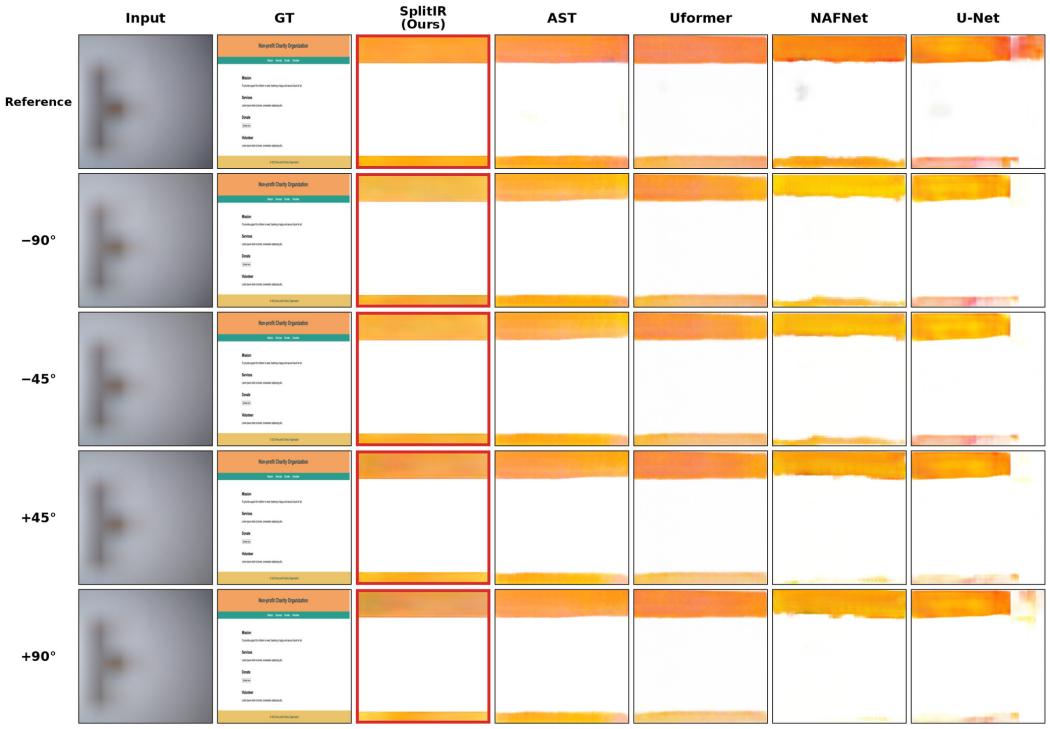


Figure A29: Representative Website reconstructions under different ambient-light directions. The comparison shows the response of large page regions and local interface structures to spatially varying illumination.

Fig. A29 shows that the directional field primarily changes the reconstructed intensity and local consistency of the page bands while leaving the coarse vertical organization identifiable. In the displayed example, SplitIR preserves the dominant upper and lower page regions across the five conditions. The remaining discrepancy is concentrated in local page content, which is more weakly represented in the diffuse observation.

The difference between Password and Website illustrates the dependence of robustness on target statistics. Compact characters require preservation of a relatively localized structural pattern, whereas webpage reconstruction must jointly recover several spatial scales. A directional radiometric field therefore interacts with a larger fraction of the Website layout. This explains why the Website curves show a wider metric range even though the same perturbation is used.

Motion blur on Screen. Motion robustness is evaluated by reacquiring the wall observation under controlled horizontal motion. The six conditions correspond to nominal wall-image path lengths of 0, 1, 3, 5, 9, and 13 pixels, with 0 px as the reference. The hidden target remains fixed across the sequence.

Increasing motion progressively reduces the local spatial bandwidth of the recorded wall evidence while preserving its broad support. Screen is well suited to this test because each target combines smooth regions, text, window boundaries, and heterogeneous foreground–background transitions.

The curves in Fig. A30 contain a broad stable region for short motion paths. SplitIR obtains 24.566 dB PSNR and 0.821 MS-SSIM at 0 px. At 1 and 3 px, PSNR is 24.598 and 24.603 dB, respectively, while MS-SSIM remains above 0.822. PSNR remains within 24.566–24.603 dB over the 0–3 px range, showing a stable response to short motion paths.

A clearer degradation appears once the path becomes longer. At 9 px, PSNR decreases to 24.419 dB and MS-SSIM to 0.817; at 13 px they decrease further to 23.972 dB and 0.804. LPIPS increases from 0.278 at the reference condition to 0.292 at 13 px. SplitIR remains first on all six metrics across the six displayed conditions.

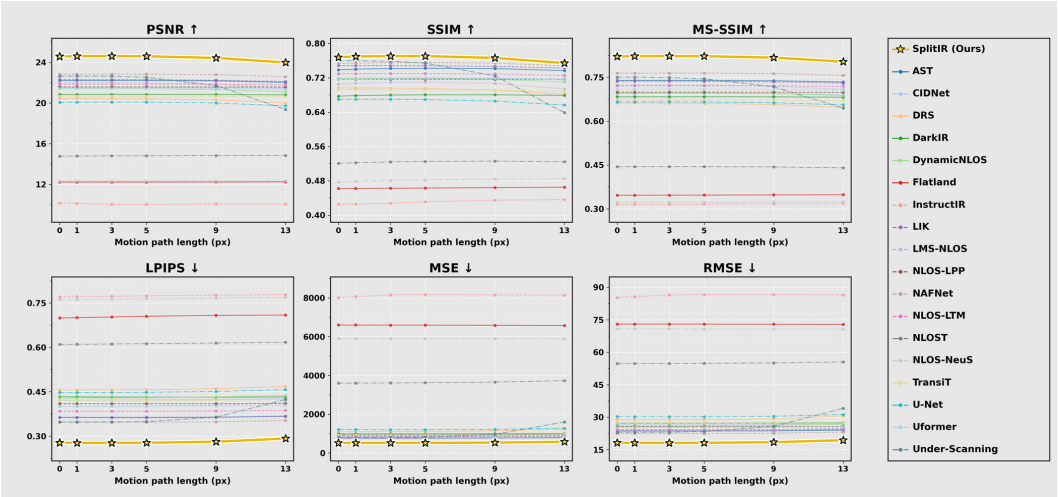


Figure A30: Six-metric robustness curves for motion blur on the Screen subset at path lengths of 0, 1, 3, 5, 9, and 13 pixels.

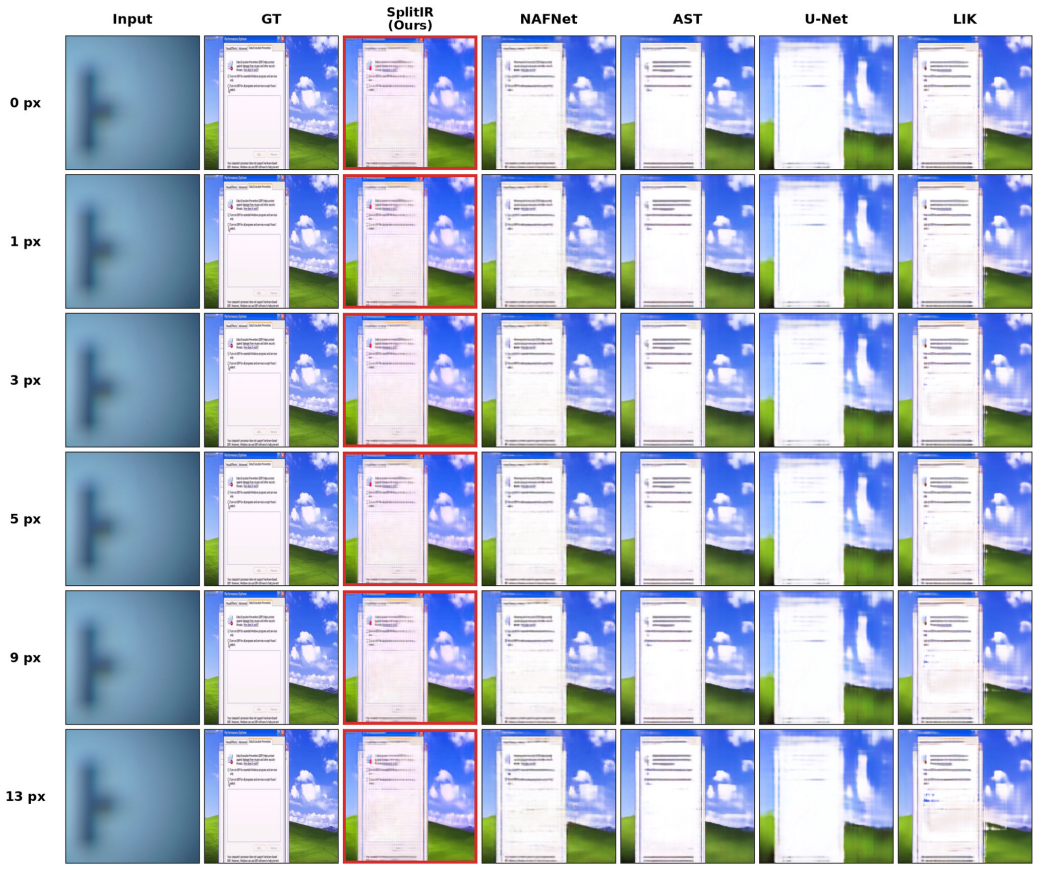


Figure A31: Representative Screen reconstructions under increasing horizontal motion blur. The rows follow the same hidden target from 0 to 13 px motion-path length.

The qualitative comparison in Fig. A31 shows why the metric change remains small initially. The large blue and green regions of the screen, the dominant window geometry, and the major foreground boundaries remain recognizable under short and intermediate motion paths. At larger path lengths,

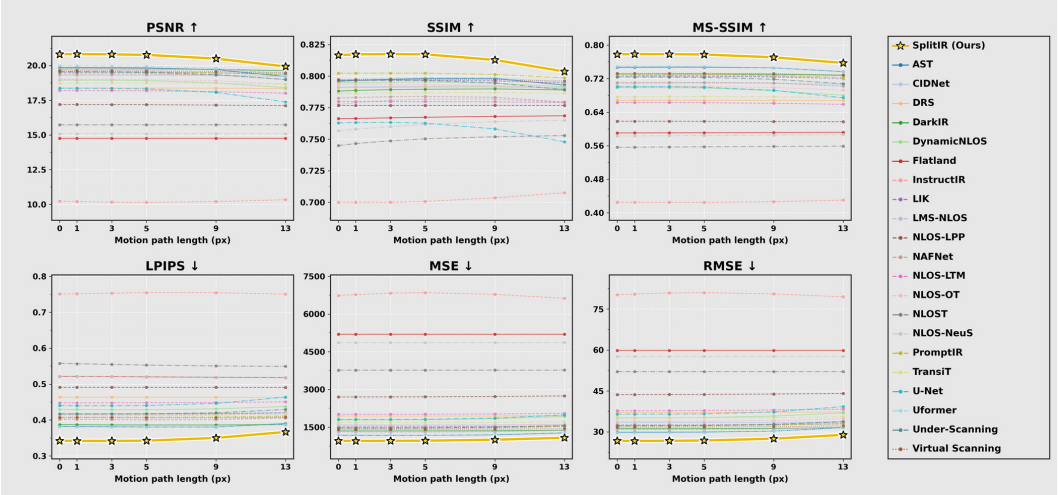


Figure A32: Six-metric robustness curves for motion blur on the Website subset at path lengths of 0, 1, 3, 5, 9, and 13 pixels.

the smallest text-like structures and window details become less distinct, but the overall spatial layout is still reconstructed.

The broad spatial support of the diffuse wall signal explains this stable initial region. At longer paths, motion increasingly suppresses the local modulations associated with fine screen structure, while the dominant layout remains recoverable and SplitIR retains its lead across all six metrics.

Motion blur on Website. The same real horizontal-motion acquisition protocol is applied to Website. This dataset contains a different mixture of information: broad page blocks provide low-frequency layout cues, while text, separators, and local interface components contribute higher-frequency structure. The experiment therefore tests whether the stable low-blur region observed on Screen also appears for a multi-scale page distribution.

Fig. A32 again shows limited variation for the first several conditions. SplitIR records 20.817 dB PSNR at 0 px, 20.829 dB at 1 px, 20.812 dB at 3 px, and 20.771 dB at 5 px. MS-SSIM remains close to 0.778 over the same interval, and LPIPS remains close to 0.342.

At 9 and 13 px, the effect becomes more visible. PSNR decreases to 20.502 and 19.934 dB, while MS-SSIM decreases to 0.771 and 0.757. LPIPS reaches 0.350 and 0.367, respectively. As on Screen, SplitIR remains first on the six metrics over all displayed blur levels.

The qualitative results in Fig. A33 provide a complementary interpretation. The longer persistence of large page regions is consistent with broad spatial variations remaining in the blurred wall observation. Finer structures are affected earlier as their weaker local modulations are progressively smoothed in the wall observation. The reconstruction therefore exhibits a scale-dependent response: page organization remains comparatively stable while local detail gradually weakens.

The agreement between Screen and Website also separates motion blur from the geometric perturbations considered later. Here, the spatial coordinate system remains fixed and only local bandwidth is reduced. The reconstruction changes gradually under local bandwidth reduction; the later geometry acquisitions produce larger departures from their reference conditions.

Camera shake on Screen. Camera shake extends the one-dimensional motion test to a two-dimensional camera-motion trajectory. We reacquire the wall observation at RMS displacement levels of 0, 0.25, 0.5, 1, 2, and 3 pixels in the recorded wall image, with the same hidden Screen target retained across conditions. The sequence therefore tests how increasing real camera jitter redistributes local wall evidence while preserving the overall scene configuration.

Fig. A34 shows that the response is small over the lower displacement range. PSNR changes from 24.566 dB at the reference to 24.606, 24.623, and 24.615 dB at 0.25, 0.5, and 1 px, respectively. MS-SSIM remains around 0.821–0.823, and LPIPS remains close to 0.277.

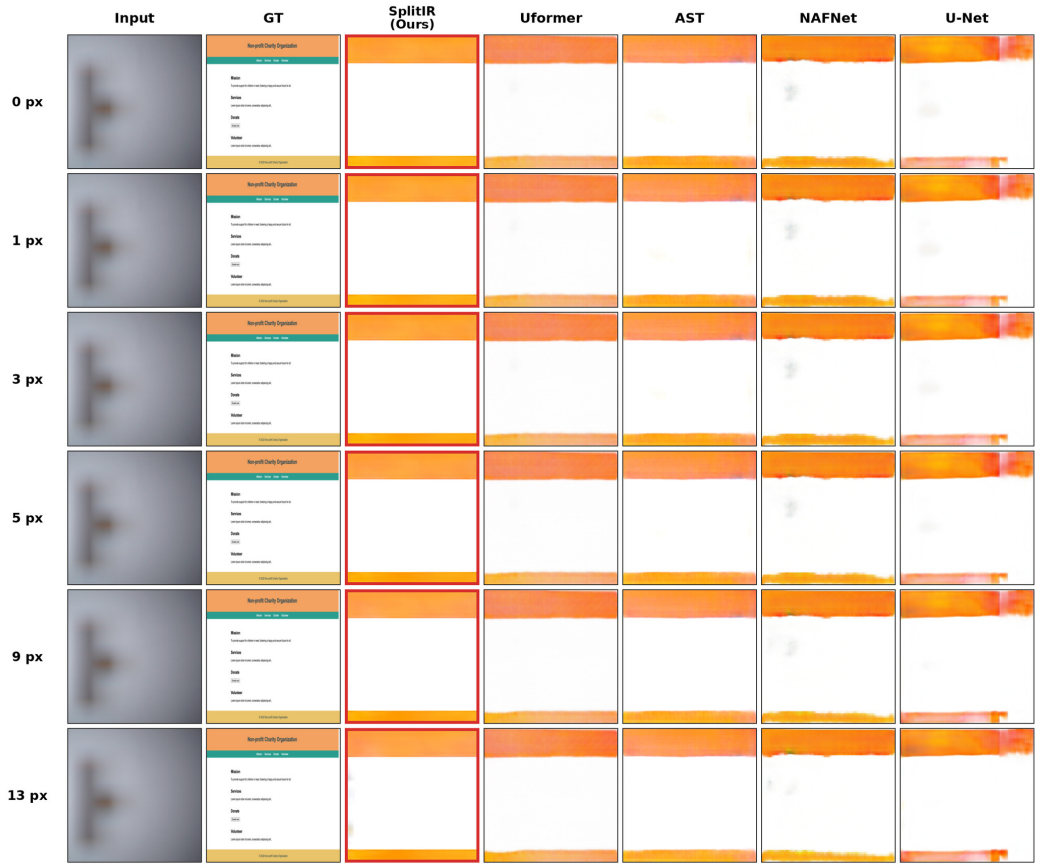


Figure A33: Representative Website reconstructions under increasing motion blur, showing the response of page-level regions, separators, and local interface structures as the motion path increases.

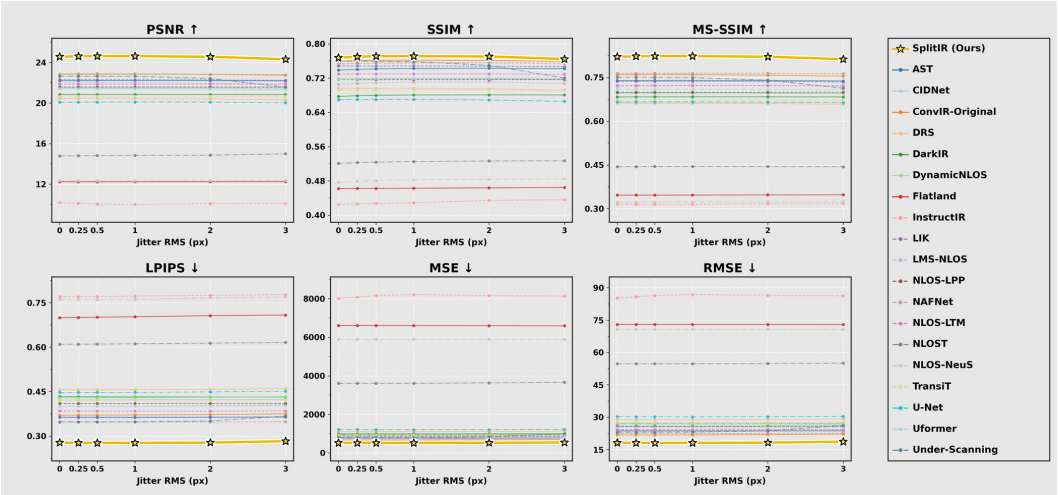


Figure A34: Six-metric robustness curves for camera shake on the Screen subset at RMS displacements of 0, 0.25, 0.5, 1, 2, and 3 pixels.

The first clearer degradation occurs at larger displacement. At 3 px RMS, PSNR is 24.310 dB, MS-SSIM is 0.813, and LPIPS is 0.283. The total PSNR change from the unperturbed condition is only about 0.26 dB, which is smaller than the change produced by the strongest one-dimensional

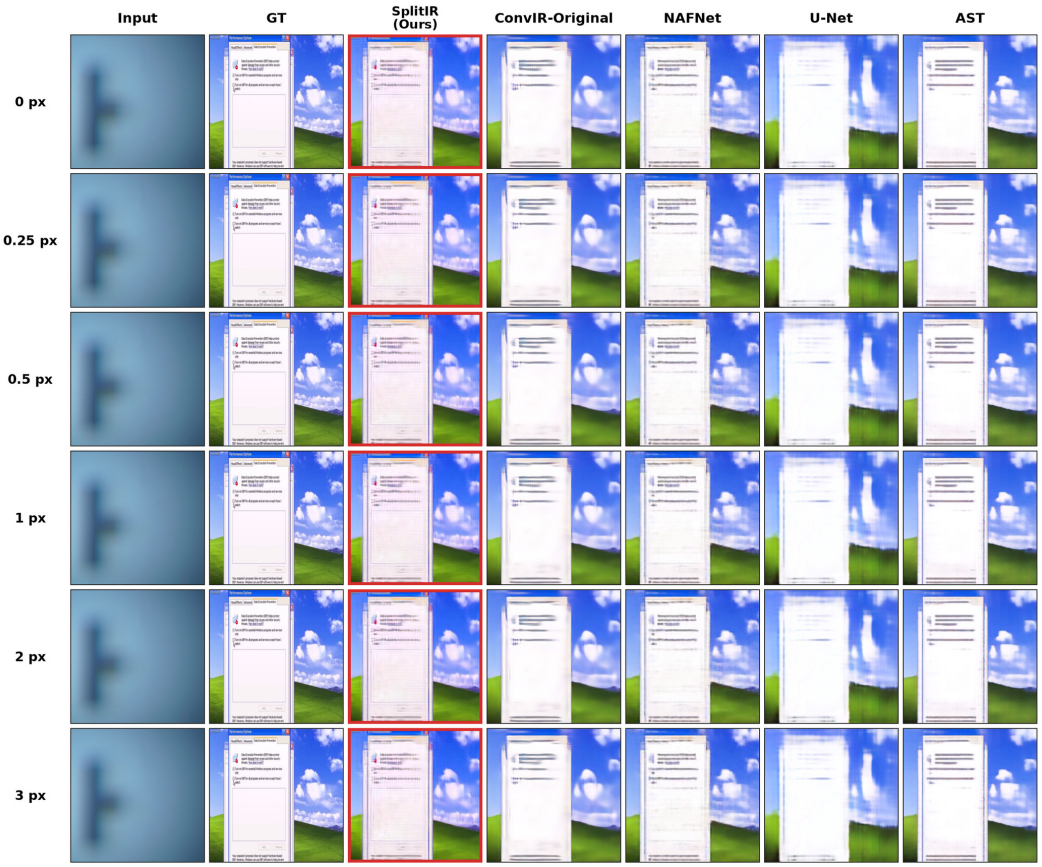


Figure A35: Representative Screen reconstructions under increasing real camera shake. The same hidden scene is retained while the RMS displacement of the jitter trajectory is varied.

motion-blur condition. SplitIR remains first on all six metrics throughout the displayed camera-shake sweep.

The visual comparison in Fig. A35 is consistent with the curve. The major window region, background, and dominant screen boundaries remain stable for small and moderate jitter. At the largest displacement, fine transitions become softer, but the overall structure remains aligned with the target.

The camera records a diffuse wall response with broad spatial support, so subpixel and near-pixel motion produces relatively small changes in the already smooth radiance field. Larger jitter progressively redistributes the residual spatial modulation that encodes the hidden content.

Relay-wall material protocol. Relay-wall robustness is measured by reacquiring the wall observation with three physical relay surfaces: the reference white wall, matte white paint, and glazed tile. The hidden target, camera-side reconstruction protocol, and evaluation remain fixed while the surface material is changed. These materials jointly vary attenuation, spatial smoothing, color response, and texture in the observed wall radiance before it is processed by the reconstruction network.

Figure A36 presents the three relay surfaces used in our real-capture experiments: the reference white wall, a surface coated with matte white paint, and a glazed tile surface. For each surface, a close-up highlights its local texture, while a wider view from the screen-facing side shows its spatial arrangement relative to the display.

Relay-wall material on Screen. On Screen, the quantitative comparison in Fig. A37 shows that the reference wall gives 0.821 MS-SSIM, 0.278 LPIPS, and 522.76 MSE. Glazed tile remains relatively close, with 0.811 MS-SSIM, 0.289 LPIPS, and 535.74 MSE. Matte paint produces a larger shift: MS-SSIM decreases to 0.734, LPIPS increases to 0.357, and MSE increases to 756.32.



Figure A36: Relay surfaces used in the real-capture experiments. From left to right, the three image pairs show the reference white wall, the surface coated with matte white paint, and the glazed tile surface. Each pair contains a surface close-up (left) and a wider scene view from the screen-facing side (right).

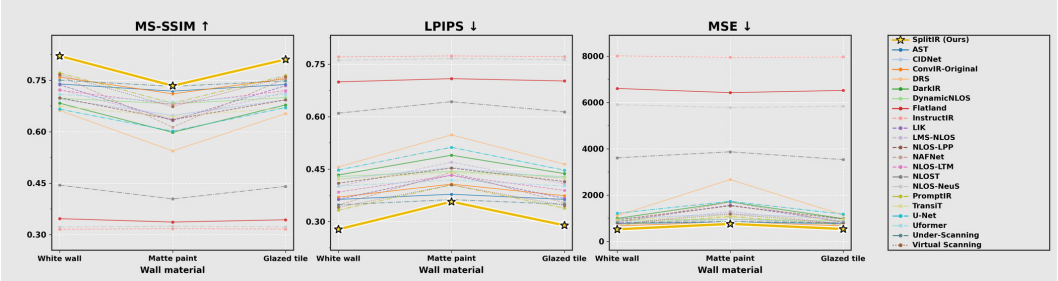


Figure A37: Relay-wall material robustness on Screen using MS-SSIM, LPIPS, and MSE for the reference white-wall, matte-paint, and glazed-tile relay surfaces.



Figure A38: Representative Screen reconstructions under the reference white wall, matte-paint, and glazed-tile relay surfaces.

The two non-reference surfaces produce distinct radiometric and spatial responses in the reacquired wall observations. Glazed tile remains closer to the reference behavior, whereas matte paint produces a larger change in attenuation, smoothing, and local texture, matching the wider metric separation.

The qualitative comparison in Fig. A38 shows that the surface change affects both broad image appearance and local boundaries. Under glazed tile, the dominant Screen layout remains close to the reference reconstruction. The matte-paint condition introduces a larger change in the recovered local content, consistent with the larger structural and perceptual metric differences.

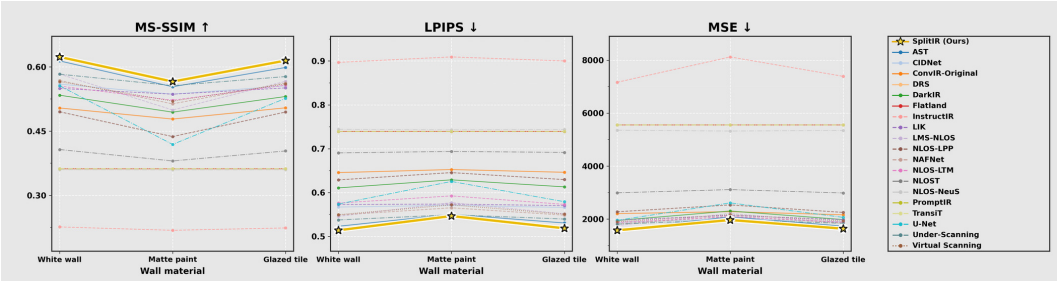


Figure A39: Relay-wall material robustness on Chart using MS-SSIM, LPIPS, and MSE under the reference white-wall, matte-paint, and glazed-tile relay surfaces.

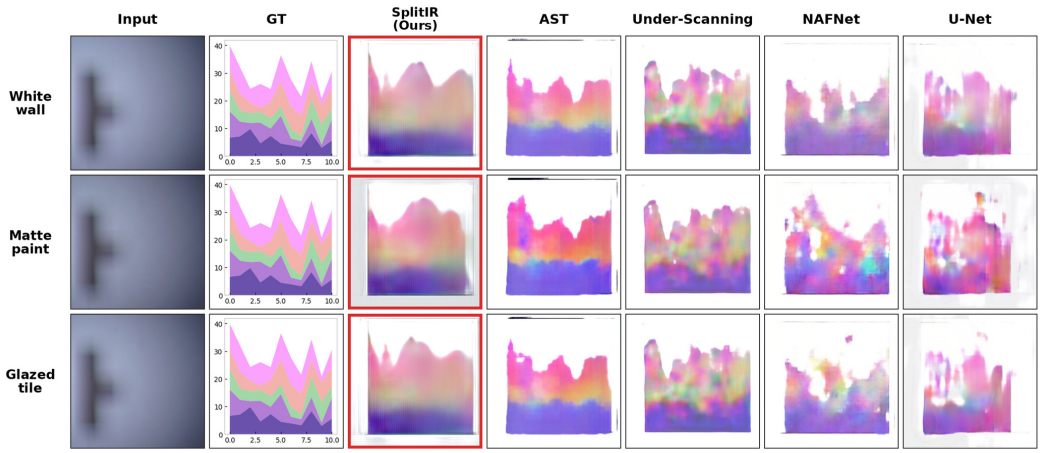


Figure A40: Representative Chart reconstructions under different relay-wall materials, emphasizing axes, thin boundaries, graphical primitives, and their relative spatial organization.

For Screen, the wall-material comparison shows how layout recovery responds to combined surface changes in attenuation, spatial mixing, color response, and texture. These changes alter the evidence presented to PAM before the screen layout is formed.

Relay-wall material on Chart. As shown in Fig. A39, Chart exhibits the same ordering but with different absolute metric values. The reference condition gives 0.623 MS-SSIM, 0.514 LPIPS, and 1575.92 MSE. Glazed tile changes these to 0.615, 0.519, and 1637.20, respectively. Matte paint produces the larger response, with 0.566 MS-SSIM, 0.547 LPIPS, and 1964.94 MSE.

The sensitivity of Chart is associated with its sparse geometric structure. Thin curves, bars, axis boundaries, and annotations occupy limited spatial support in the hidden target. Additional smoothing in the relay response therefore reduces the contrast of the already weak observation components from which these structures are inferred.

The qualitative samples in Fig. A40 show that the principal difference appears in the recovery of thin and spatially sparse structures rather than in large uniform regions. This agrees with the reduction in MS-SSIM: surface-induced smoothing changes the continuity and relative placement of the geometric elements that dominate the Chart distribution.

The same physical wall change produces content-dependent numerical responses across subsets because the hidden targets distribute informative structure differently across spatial scales. For Chart, the change is most visible in thin and spatially sparse graphical components.

Relay-wall material on Password. As shown in Fig. A41, Password shows a larger structural change under wall variation. The reference wall gives 0.914 MS-SSIM, 0.097 LPIPS, and 1869.00 MSE. Glazed tile gives 0.849 MS-SSIM, 0.148 LPIPS, and 3038.57 MSE, while matte paint gives 0.813, 0.173, and 3666.58, respectively.

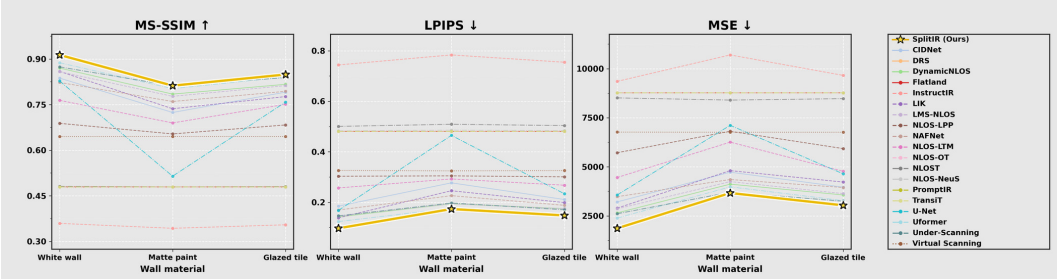


Figure A41: Relay-wall material robustness on Password using MS-SSIM, LPIPS, and MSE under the reference white-wall, matte-paint, and glazed-tile relay surfaces.

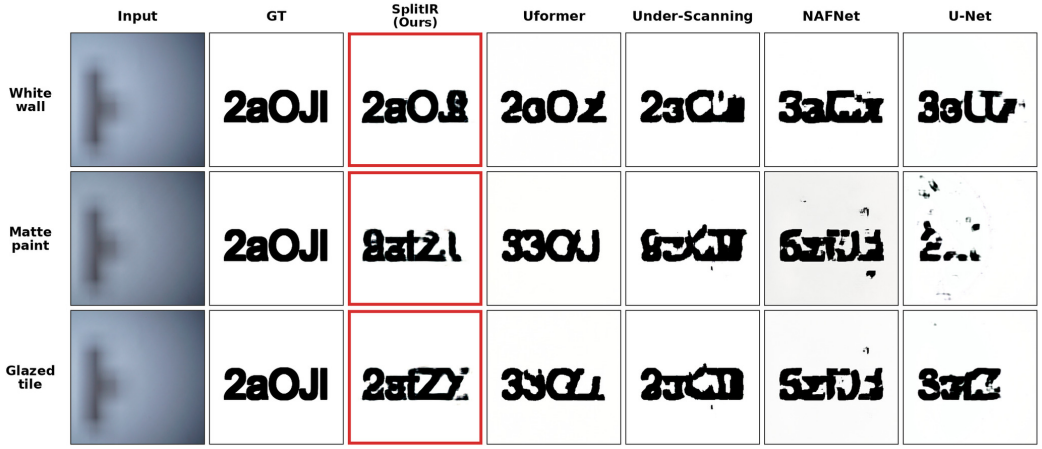


Figure A42: Representative Password reconstructions under different relay-wall materials, showing changes in compact character strokes, local boundaries, and foreground–background separation.

These differences are larger than those observed for moderate illumination-direction changes on the same dataset. The distinction is physical: directional illumination modifies the distribution of an additional component, whereas wall-material variation modifies the effective transport of the screen-dependent signal itself. Compact character evidence is therefore affected before it reaches the learned inverse mapping.

The qualitative comparison in Fig. A42 shows that the reference condition retains the clearest correspondence to the target character sequence. Under the two non-reference wall materials, character strokes become more distorted and neighboring components are more likely to merge or lose their original shape. The matte-paint condition shows the largest visible departure among the three wall materials.

Surface-response changes act on the evidence presented to PAM before the screen layout is formed. Their effect is especially visible in compact character structures, where small changes in the wall response influence stroke separation. RCM then regulates the corrections generated from the resulting layout state.

Relay-wall material on Website. For Website, Fig. A43 shows that the reference condition produces 0.778 MS-SSIM, 0.342 LPIPS, and 953.95 MSE. Glazed tile changes these values to 0.769, 0.354, and 981.05, respectively. Matte paint again produces the larger change, with 0.749 MS-SSIM, 0.384 LPIPS, and 1208.67 MSE.

The Website response is smaller than Password in relative structural terms but remains consistent across all three reported metrics. Large page regions provide persistent low-frequency cues, while local text, borders, and interface details are more sensitive to the altered surface response. The two spatial scales therefore contribute differently to the final metric trajectory.

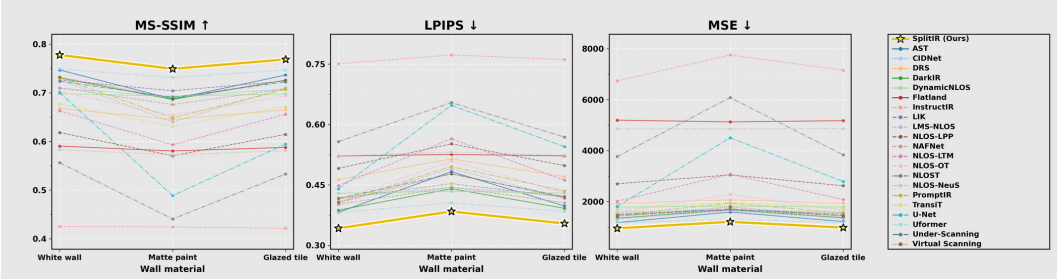


Figure A43: Relay-wall material robustness on Website using MS-SSIM, LPIPS, and MSE under the reference white-wall, matte-paint, and glazed-tile relay surfaces.

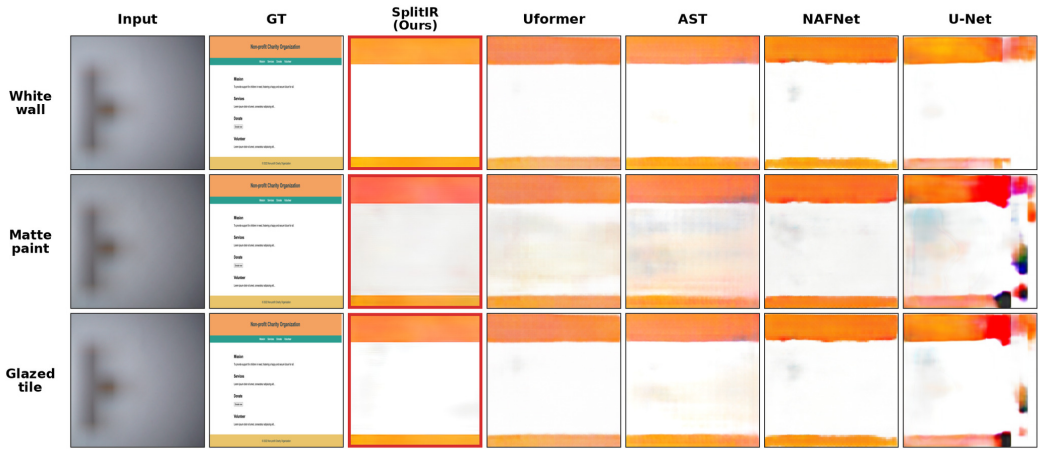


Figure A44: Representative Website reconstructions under different relay-wall materials, showing page-level organization together with text, borders, and local interface elements.

The qualitative comparison in Fig. A44 confirms this scale-dependent behavior. The broad page organization remains identifiable across the displayed relay surfaces, while smaller internal structures undergo more visible changes. The difference between matte paint and glazed tile is also consistent with their corresponding curve separation.

Across Screen, Chart, Password, and Website, glazed tile remains closer to the reference response than matte paint in the plotted metrics. This repeated ordering links the reconstruction response to surface-perturbation strength across all four content categories, while the magnitude of the change remains content dependent because each dataset places informative structure at different spatial scales.

Viewing-distance variation on Password. We next change the physical viewing distance around a 100 cm reference and reacquire the wall observation at 80, 90, 100, 110, and 120 cm. Distance variation changes the scale and spatial support of the recorded wall pattern relative to the reference acquisition, directly probing how layout recovery responds when the same hidden content is observed under different sampling geometry.

The physical placement references are shown in Fig. A45. The black floor markers identify locations used to position the camera for different observation positions and viewing directions. Their arrangement relative to the display and relay wall provides a physical reference for camera placement during the geometric acquisition experiments.

The reference geometry forms a clear maximum in Fig. A46. SplitIR reaches 15.852 dB PSNR, 0.914 MS-SSIM, and 0.097 LPIPS at 100 cm. At 90 and 110 cm, PSNR decreases to 10.871 and 10.928 dB, MS-SSIM to 0.715 and 0.706, and LPIPS increases to 0.208 and 0.220.

The two outer conditions produce a larger change. At 80 cm, SplitIR gives 9.967 dB PSNR, 0.653 MS-SSIM, and 0.242 LPIPS. At 120 cm, the corresponding values are 9.847 dB, 0.638, and 0.270.



Figure A45: Floor-marker layout for camera placement. The black marks indicate placement locations for different camera positions and viewing directions relative to the display and relay wall.

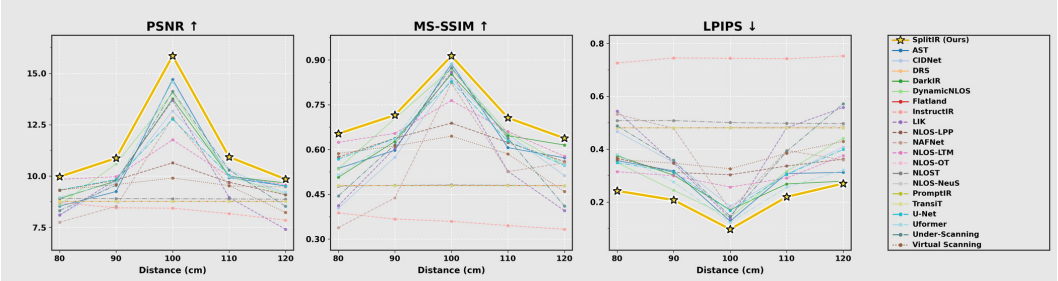


Figure A46: Robustness to viewing-distance variation on Password using PSNR, MS-SSIM, and LPIPS at 80, 90, 100, 110, and 120 cm. The 100 cm condition is the reference geometry.

Reconstruction quality is highest at the 100 cm reference and decreases as the acquisition departs from that geometry in either direction. SplitIR retains the best PSNR, MS-SSIM, and LPIPS across all five distance conditions.

The qualitative comparison in Fig. A47 makes the geometric mismatch visible. At the reference distance, the recovered character sequence retains the closest spatial organization to the target. Moving the observation plane in either direction changes the apparent support of the diffuse pattern, and the recovered characters exhibit larger deformation and misalignment.

Changing distance alters where the wall evidence appears and the spatial support over which it is distributed. The comparison therefore extends the reconstruction analysis from changes in signal strength to changes in spatial correspondence: PAM reorganizes the reacquired wall evidence into the shared layout state, and RCM regulates the corrections generated from that state.

Vertical viewing-angle variation on Password. Vertical-angle robustness is evaluated by physically changing the camera pitch to $-20°$, $-10°$, $0°$, $+10°$, and $+20°$ and reacquiring the wall observation at each setting. The $0°$ condition is the reference. Distance changes primarily alter global scale, whereas pitch changes produce position-dependent changes in the observed wall geometry and therefore shift evidence locations unevenly across the image.

The adjustment procedure is further illustrated in Fig. A48. The first photograph shows manual tripod adjustment, while the remaining views document the camera-support configuration with and without horizontal reference lines. These lines provide visual guides for the support arrangement rather than calibrated height measurements.

The reference angle yields the strongest reconstruction within the five-angle comparison in Fig. A49. SplitIR reaches 15.852 dB PSNR, 0.914 MS-SSIM, and 0.097 LPIPS at $0°$. At $-20°$ and $-10°$, PSNR decreases to 9.248 and 9.286 dB, while MS-SSIM becomes 0.584 and 0.577. At $+10°$ and $+20°$, PSNR is 9.074 and 8.975 dB, and MS-SSIM further decreases to 0.552 and 0.546.

LPIPS follows the same pattern, increasing from 0.097 at the reference view to 0.299–0.342 at the four non-reference angles. SplitIR achieves the best PSNR, MS-SSIM, and LPIPS at each of the five viewing angles.

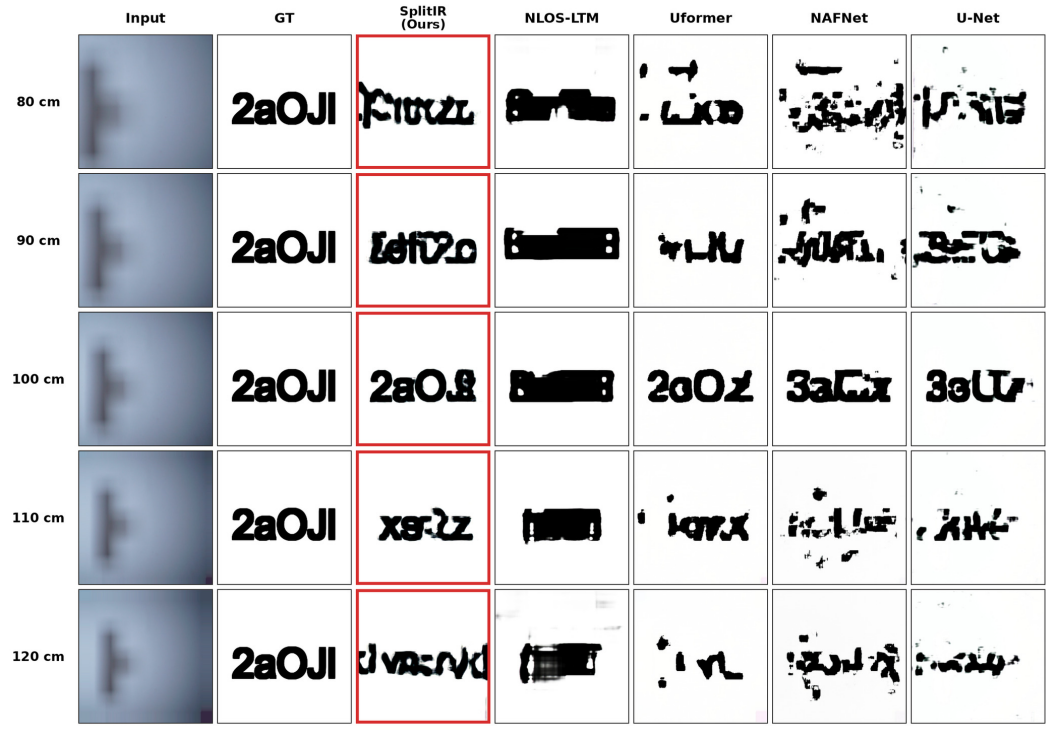


Figure A47: Representative Password reconstructions under viewing-distance variation around the reference configuration.



Figure A48: Tripod adjustment: (a) manual adjustment of the tripod, (b) an annotated view with horizontal reference lines, and (c) the corresponding unannotated view.

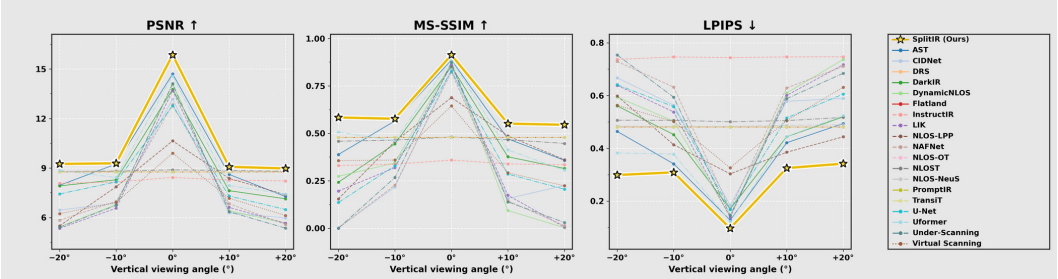


Figure A49: Robustness to vertical viewing-angle variation on Password using PSNR, MS-SSIM, and LPIPS for $-20°$, $-10°$, $0°$, $+10°$, and $+20°$.

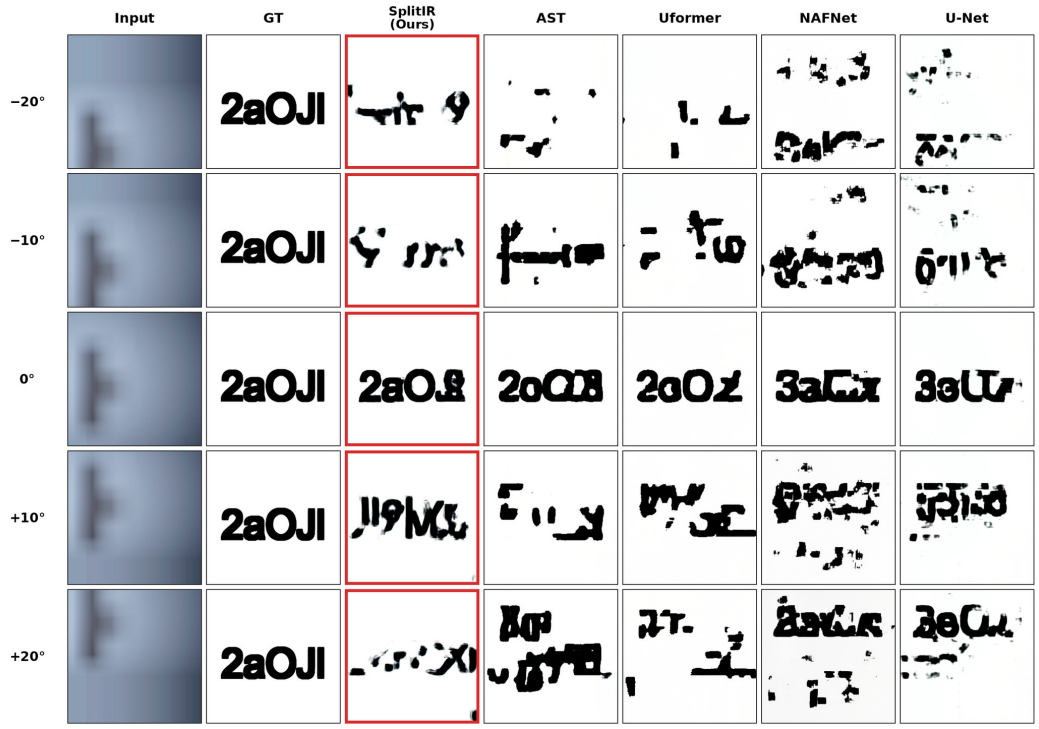


Figure A50: Representative Password reconstructions under vertical viewing-angle changes around the reference view, showing viewpoint-induced changes in character geometry.

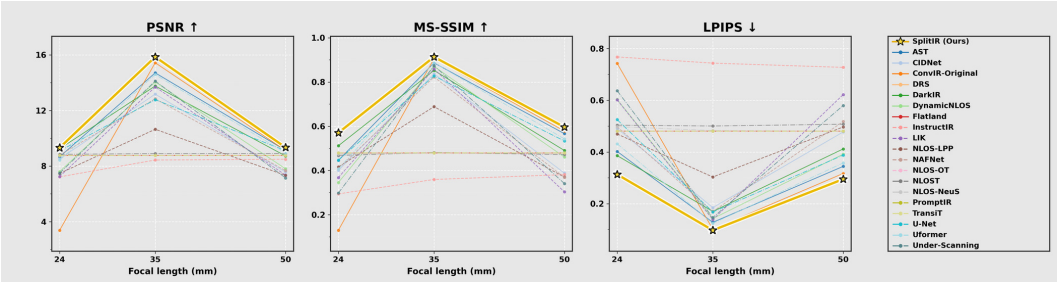


Figure A51: Robustness to camera-lens variation on Password using PSNR, MS-SSIM, and LPIPS for the 24, 35, and 50 mm lens configurations. The 35 mm configuration is the reference.

The qualitative results in Fig. A50 show that angle variation changes more than stroke sharpness. The reconstructed character structures are displaced and deformed because the underlying wall evidence is sampled under a different projective configuration. This distinguishes pitch from blur, noise, and illumination perturbations, all of which preserve the reference spatial mapping.

Pitch changes the wall evidence through position-dependent viewing geometry. The resulting comparison tests how screen structure is reconstructed when evidence locations shift unevenly across the image, extending the global scale changes studied in the distance experiment.

Camera-lens variation on Password. The final experiment evaluates real 24, 35, and 50 mm lens configurations, with 35 mm as the reference. For each focal length, the wall observation is reacquired with the corresponding optical configuration while the hidden target and reconstruction model remain fixed. The three settings jointly probe changes in field of view, geometric projection, optical blur, and spatial intensity response introduced by the physical lens configuration.

The reference 35 mm configuration again produces a clear maximum in Fig. A51. SplitIR reaches 15.852 dB PSNR, 0.914 MS-SSIM, and 0.097 LPIPS at 35 mm. The 24 mm configuration gives

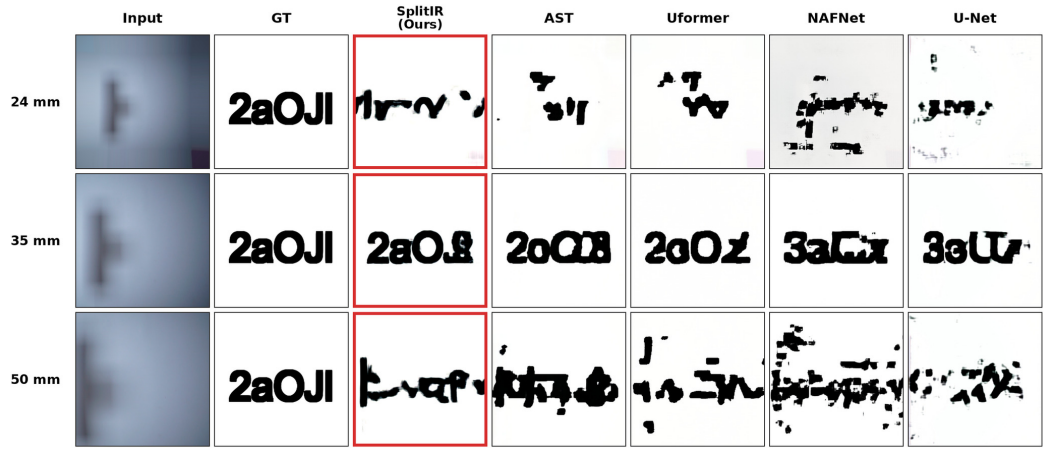


Figure A52: Representative Password reconstructions under the 24, 35, and 50 mm real camera-lens configurations, showing the reconstruction response to the corresponding optical acquisitions.

9.315 dB PSNR, 0.571 MS-SSIM, and 0.313 LPIPS, while the 50 mm configuration gives 9.335 dB, 0.596, and 0.295, respectively. SplitIR remains first on each of the three displayed metrics for all lens conditions.

The 24 and 50 mm configurations yield similar PSNR values but different MS-SSIM and LPIPS responses. The three physical lens settings therefore probe distinct geometric and optical changes in the wall evidence. SplitIR retains the best result on all three metrics across the lens configurations.

The qualitative comparison in Fig. A52 shows that the non-reference lens configurations produce substantial changes in the reconstructed character geometry. The 35 mm condition preserves the character sequence and local spacing most consistently with the target, whereas the 24 and 50 mm conditions introduce stronger deformation and merging of local structures.

Together with the distance and pitch experiments, the lens comparison extends the robustness analysis to joint geometric and optical changes. Across these acquisition conditions, PAM organizes the reacquired wall evidence into the shared layout state, and RCM regulates the stage-conditioned corrections generated from it.

Cross-condition analysis. The additional experiments separate three forms of robustness that are otherwise mixed when only a single degradation family is considered. Ambient-light direction changes the spatial distribution of radiometric interference while leaving the observation coordinate system intact. Motion blur and camera shake preserve the same scene geometry but reduce or redistribute spatial bandwidth. Their curves therefore exhibit comparatively gradual changes, particularly in the low-perturbation regime.

Relay-wall material changes the physical intermediate surface. The consistent ordering between the reference wall, glazed tile, and matte paint across Screen, Chart, Password, and Website links the reconstruction response to changes in attenuation and spatial mixing, with content-dependent magnitude across the four target categories.

Viewing distance, vertical angle, and camera-lens variation form a third group. These conditions alter the geometric relation between the camera observation and the reference acquisition. Their metric curves show substantially larger departures from the reference condition than the moderate motion or illumination perturbations. The corresponding qualitative examples also show geometric deformation of reconstructed content rather than only reduced sharpness or contrast.

Light-transport modulation adapts hidden-image reconstruction and reprojection across transport conditions through a learned condition representation Zhang et al. (2025). Our analysis follows the screen-reconstruction pathway by distinguishing changes in evidence strength, local bandwidth, and spatial correspondence. Across these real acquisition variations, PAM organizes the available wall evidence into a layout state, and RCM regulates the corrections generated from it. The experiments

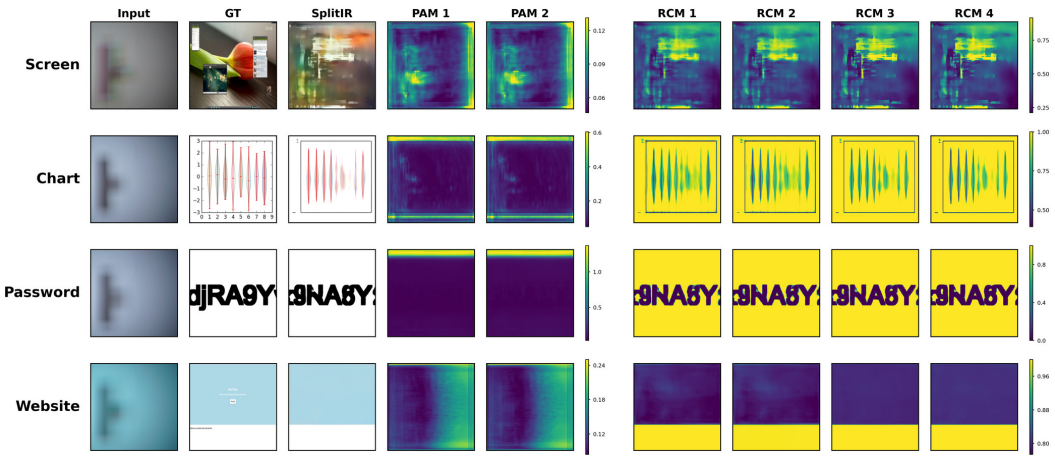


Figure A53: Core-stage visualization across Screen, Chart, Password, and Website. Each row shows the wall observation, ground truth, SplitIR reconstruction, the two PAM responses, and the four RCM-stage responses for a fixed representative example. Each heat map displays the channel-averaged absolute response with independent display normalization.

therefore connect the comparative results to different stages of the same wall-to-screen recovery process.

C.5 Supplementary Feature and Operator Analysis

We next trace how wall evidence becomes screen structure. Internal responses reveal where content-aligned organization emerges, spatial interventions show which observation regions affect reconstruction, and stage-wise analyses describe how PAM and RCM transform the evolving state.

The visualization and diagnostic protocols use a predetermined sample set selected independently of reconstruction quality. Each heat map displays the channel-averaged absolute response, $A(h, w) = \mathrm{mean}_c\,|F(c, h, w)|$, with independent display normalization. The maps therefore expose the spatial support and organization of the intermediate representation. Intervention metrics are computed per image and averaged over the complete evaluation partition of each dataset. Stage diagnostics use 32 clean images per dataset and record every PAM and RCM call.

Each content category uses its fixed dataset-specific checkpoint across the comparative, intervention, internal-feature, and operator-diagnostic analyses.

C.5.1 Internal Representation Visualization

We inspect 15 locations along the forward path using four fixed examples per content category, giving 16 visualization examples in total. The two PAM calls and four RCM stages are recorded separately, together with representative layout, multi-scale, decoding, proposal, and feedback features. Quantitative intervention analyses use the complete evaluation partitions, comprising 5,116 image pairs across the four categories.

Recording the intermediate responses preserves the predictions for all 16 visualization examples: the maximum absolute difference between the clipped floating-point outputs with and without the recording hooks is zero.

Fig. A53 provides the most direct view of the two operator families. The PAM responses remain relatively coarse and are expressed in the shared representation domain, whereas the RCM responses already exhibit structures aligned with the evolving image-domain reconstruction. This distinction is especially visible in Chart and Password: the PAM maps remain broad and low-resolution, while the RCM sequence progressively contains line-like or character-like patterns that correspond to the decoded target layout. Website shows a similar transition at a larger spatial scale, with the RCM stages reflecting the dominant horizontal page partition.

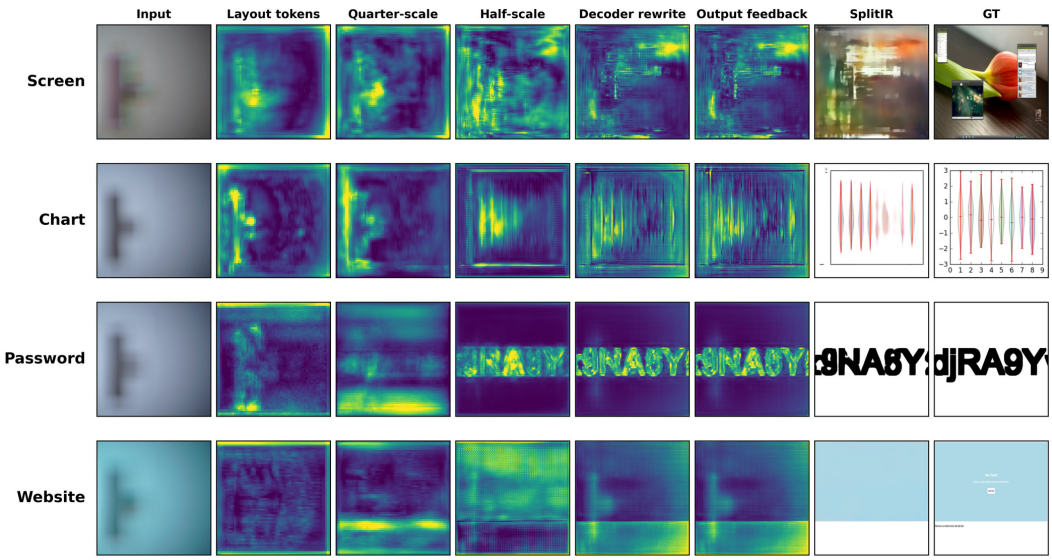


Figure A54: Spatial responses at five representative reconstruction locations: initial shared layout state, quarter-scale features, half-scale features, decoder-stage feature, and output-feedback feature. Each map shows the channel-averaged absolute response with independent display normalization.

The heat maps reveal where spatial structure appears and how its organization changes along the reconstruction path. Later responses follow content-dependent layouts, from heterogeneous Screen regions to broad Website partitions.

The pathway view in Fig. A54 makes the coarse-to-fine transition more explicit. Responses in the initial shared layout state are spatially broad, preceding the content-aligned structure of the final reconstruction. The quarter-scale and half-scale stages reorganize this response into more localized structures. By the decoder-stage and output-feedback positions, the maps become increasingly aligned with the dominant organization of the final reconstruction. Password makes this transition explicit: compact character-shaped responses become visible as the shared layout passes through the later decoding stages. Chart exhibits a related transition from distributed wall evidence to a more regular set of directional primitives.

The same qualitative progression is present for Screen and Website, but the recovered structures are less sparse. Screen contains mixed textures, windows, and image regions, so later responses remain spatially heterogeneous. Website is dominated by broad page regions; its later responses therefore emphasize large horizontal or vertical partitions rather than isolated high-frequency strokes. The difference across content classes is consistent with the reconstruction problem itself: the network is not recovering one universal spatial template, but mapping different diffuse observations into content-dependent layouts.

Screen examples. The four fixed Screen examples provide a wider view of how the same internal locations respond to heterogeneous content. Because Screen contains text, photographs, interface regions, and sharp boundaries in the same subset, the intermediate maps vary substantially from one example to another even though their forward locations are identical.

Fig. A55 shows that the later representation follows the dominant structure of each target rather than converging to one repeated activation pattern. Examples containing large rectangular interfaces produce extended boundary responses, whereas examples containing more diffuse image content generate broader spatial organization. The decoder-stage and output-feedback maps are also visibly closer to one another than either is to the initial layout tokens, which is consistent with both operating near the end of the shared decoding path.

Chart examples. Chart provides a complementary case in which informative content is concentrated in sparse geometric primitives. The four examples include thin line patterns, segmented

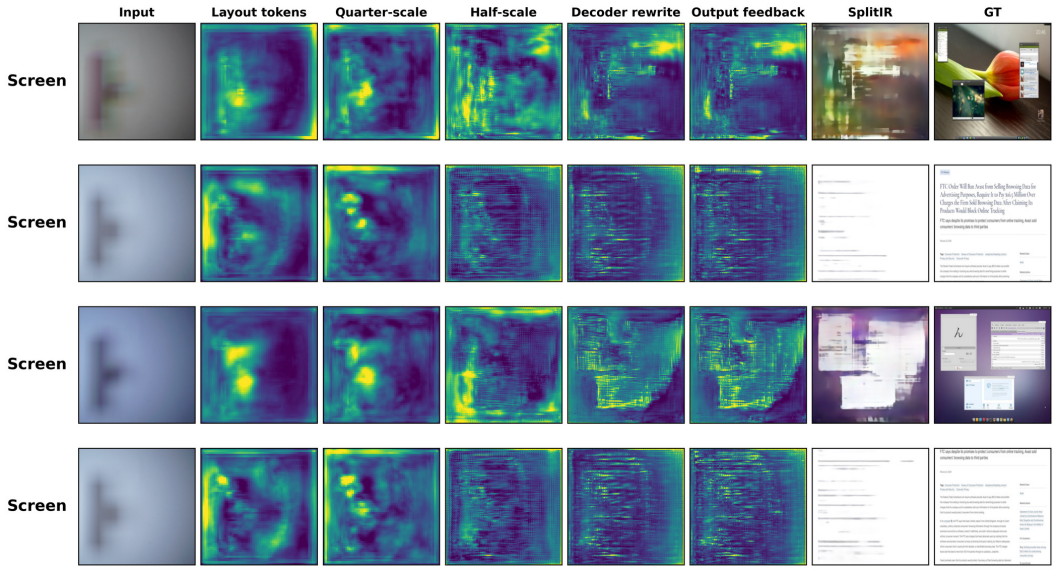


Figure A55: Internal spatial responses for the four fixed Screen examples along the same five reconstruction locations used in Fig. A54. Each row is one example and uses independently normalized display maps.

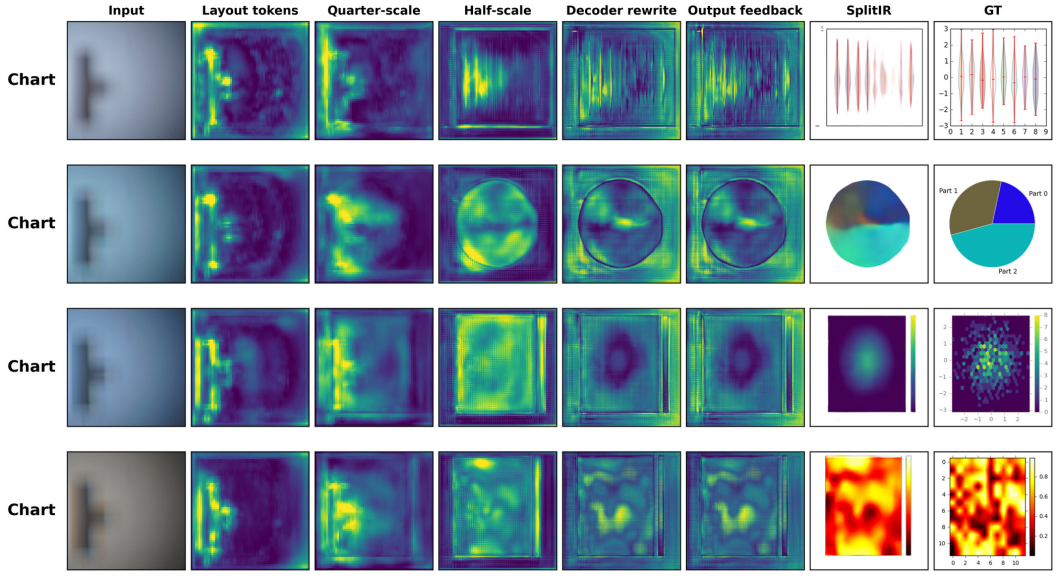


Figure A56: Internal spatial responses for four fixed Chart examples. The transition from broad layout evidence to more directional and localized structure is shown at the same five forward locations.

graphic regions, and spatially localized chart structures. Their wall observations are smooth, yet the later internal maps recover increasingly anisotropic responses.

In Fig. A56, the quarter- and half-scale maps begin to resolve directional or localized responses from the spatially mixed wall input. The decoder-stage and feedback stages further concentrate those structures. Because every map is normalized independently for display, the relevant evidence is the spatial shape and orientation of the response rather than its color magnitude. The examples develop distinct response shapes and orientations that follow their graphical organization.

Password examples. Password is dominated by compact character strokes, making it useful for examining when high-frequency target organization becomes visible inside the network. The input

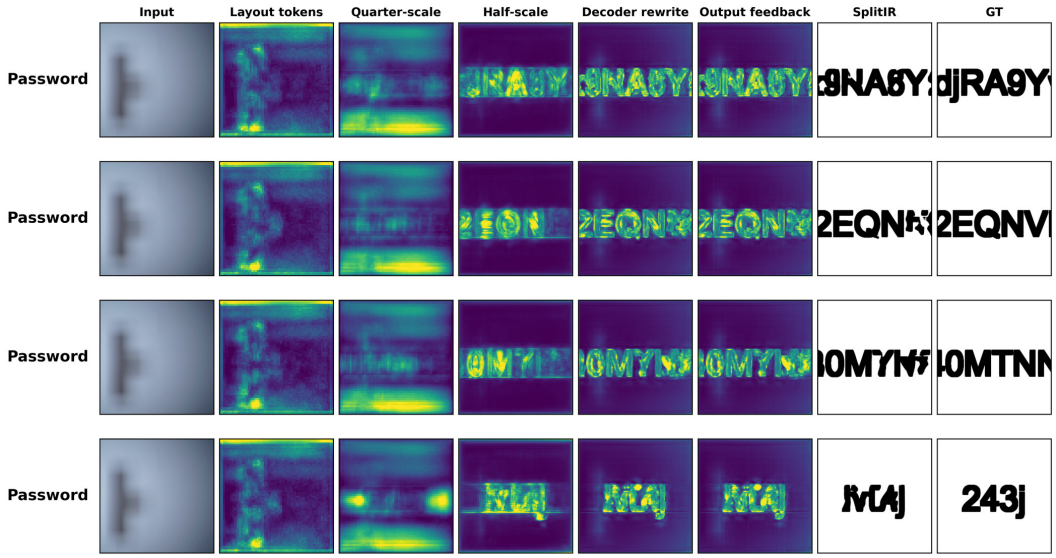


Figure A57: Internal spatial responses for four fixed Password examples. Later decoder and feedback locations exhibit character-aligned spatial patterns that are not directly visible in the diffuse input.

maps contain no directly readable characters, while the final reconstruction is strongly structured by character shape and spacing.

Fig. A57 shows that the early layout and quarter-scale maps remain broad, whereas the half-scale and later maps increasingly reflect the support of the hidden characters. The effect is repeated across the four fixed samples despite changes in character sequence and spacing. This shows character-scale organization becoming more explicit in the half-scale and later responses, connecting the broad wall-derived representation to localized reconstructed-character structure.

Website examples. Website combines large page-level regions with local content blocks and text-like elements. Compared with Password and Chart, much of the target energy is organized at a larger spatial scale, so the internal responses remain correspondingly broad.

The four rows in Fig. A58 show that the later maps follow the coarse page geometry of the corresponding GT: large header, body, side, or lower-page regions are represented as extended spatial bands. The result is consistent with the multi-scale nature of the Website subset and complements the sharper character-aligned responses observed for Password. Taken together, Figs. A55– A58 show content-dependent internal organization across the same fixed forward locations.

C.5.2 Spatial Sensitivity and Representation Response

We next perturb the input while keeping the trained weights fixed. Each 256×256 observation is divided into a 3×3 grid using boundaries $[0, 85, 170, 256]$, and one region is set to zero at a time. For each base condition, the occlusion value shown below is the mean SSIM of the unoccluded condition minus the mean SSIM after occluding the specified region. Positive values therefore indicate a reduction in reconstruction SSIM caused by the additional spatial occlusion. Each of the nine masks is evaluated over the complete evaluation partition of its dataset.

The five base conditions are clean input, brightness reductions of 60 and 130 nit-equivalent, and horizontal motion blur of 5 and 13 pixels. The occlusion protocol is identical across conditions: brightness images are read from the corresponding prepared inputs, motion blur is applied before the spatial mask, and the GT remains unchanged. This design separates the additional effect of removing a spatial region from the base degradation already present in the observation.

The clean-input heat maps in Fig. A59 show a common positional pattern across all four datasets: the central region produces the largest average SSIM decrease. The central decrease is 0.638 for Screen, 0.453 for Chart, 0.432 for Password, and 0.754 for Website. The result is especially strong

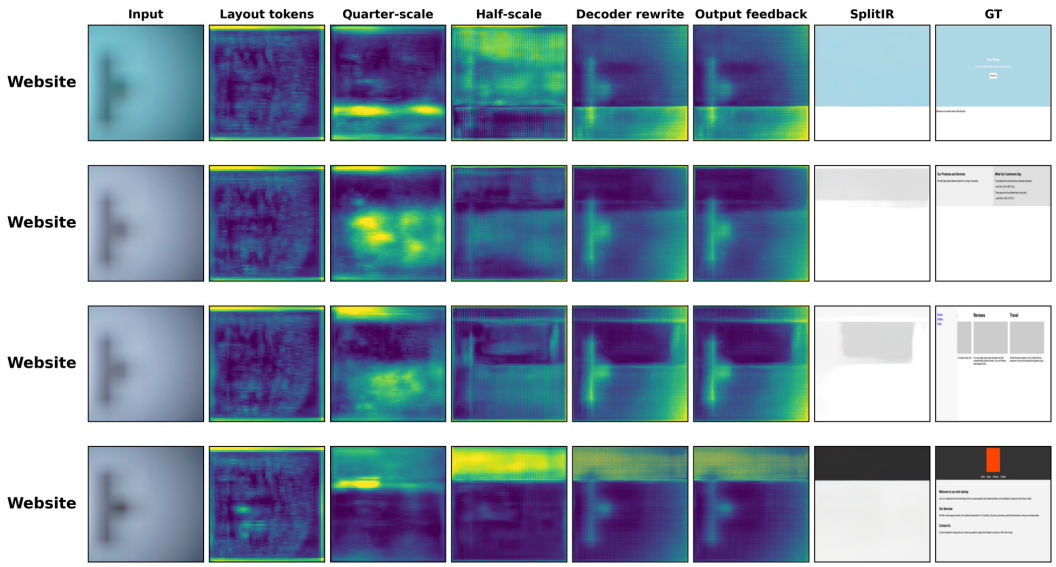


Figure A58: Internal spatial responses for four fixed Website examples. Broad page partitions and large layout regions become increasingly organized toward the decoder-stage and output-feedback stages.

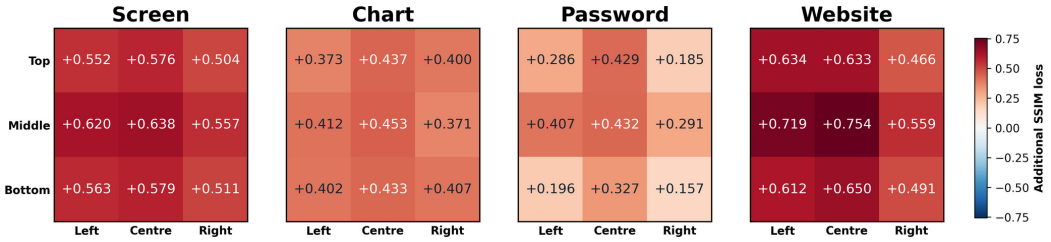


Figure A59: Nine-region occlusion sensitivity under the clean input condition. Each cell reports the mean SSIM decrease caused by zeroing the corresponding input region relative to the same clean condition without occlusion. Regions are arranged from top-left to bottom-right.

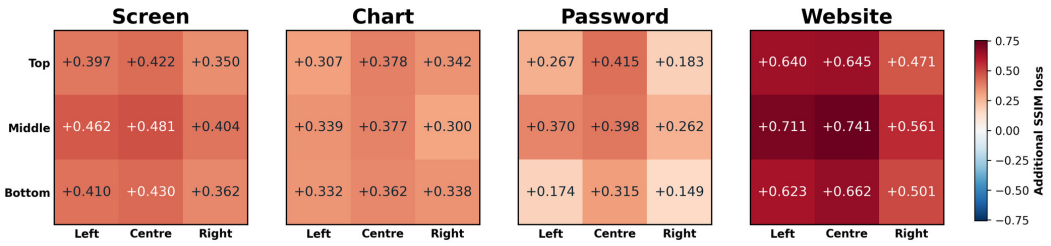


Figure A60: Nine-region occlusion sensitivity after a 60-nit-equivalent brightness reduction. Values measure the additional SSIM decrease caused by each zeroed region relative to the same brightness-reduced input without occlusion.

for Website, where removing the center of the wall observation reduces mean SSIM by more than 0.75 relative to the corresponding unoccluded condition.

Under clean-input region masking, the central cell produces the largest average SSIM decrease in all four content categories. Reconstruction is therefore spatially sensitive to the wall observation, with a concentrated response within the acquisition crop. This intervention connects the recovered screen structure to the observed wall evidence.

After a 60-nit-equivalent reduction, the strongest sensitivity remains concentrated along the central column across all four content categories (Fig. A60). The persistence of this spatial pattern shows

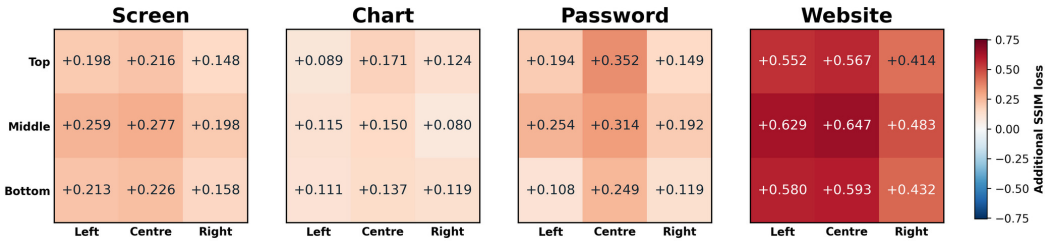


Figure A61: Nine-region occlusion sensitivity after a 130-nit-equivalent brightness reduction. The same spatial mask and full-test-set averaging are used as in Fig. A59.

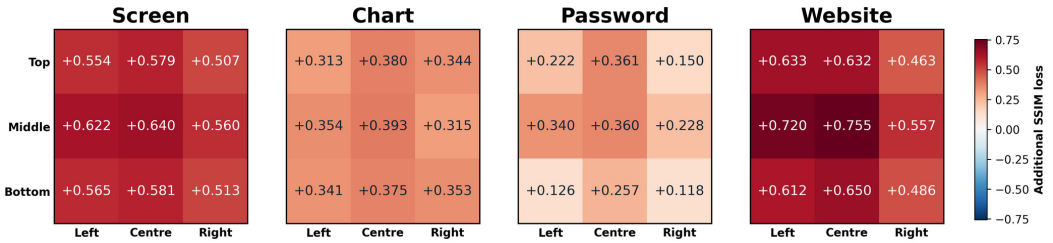


Figure A62: Nine-region occlusion sensitivity with a 5-pixel horizontal motion blur applied before the spatial mask. Each value is relative to the unoccluded 5-pixel-blur condition.

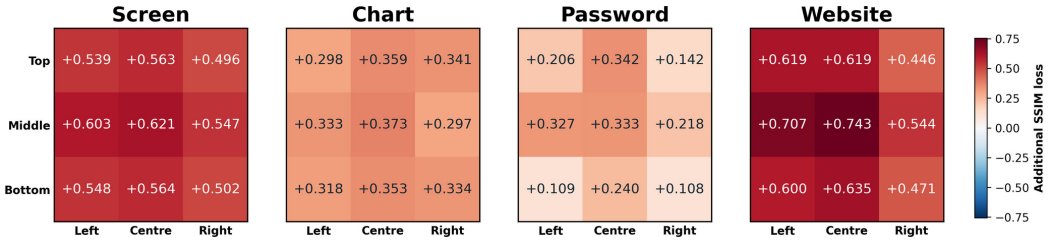


Figure A63: Nine-region occlusion sensitivity with a 13-pixel horizontal motion blur applied before masking. Values represent the additional SSIM decrease relative to the unoccluded 13-pixel-blur condition.

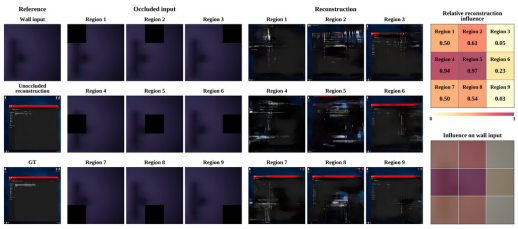
that reconstruction continues to depend on an organized region of the wall observation as the signal weakens.

The similarly sized masks produce different SSIM changes, revealing a spatially uneven sensitivity within the wall crop. The middle and upper-middle regions remain more influential than several corners despite comparable masking area.

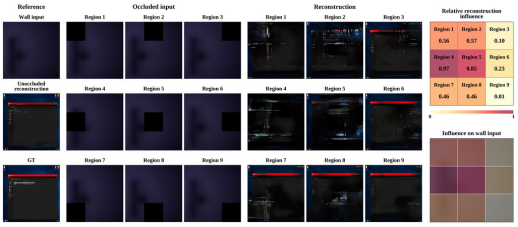
At the stronger 130-nit-equivalent reduction, Fig. A61 shows a lower absolute occlusion penalty for several datasets because the degraded input has already lost part of the evidence available in the clean condition. The Website map still exhibits the largest additional losses near the center, whereas Password shows a more compressed range. The result highlights an important distinction between base degradation and spatial importance: a strongly degraded base observation can show both lower reconstruction quality and a smaller additional loss when one more region is removed.

Each heat-map value is the additional SSIM decrease caused by masking relative to the corresponding unmasked base condition.

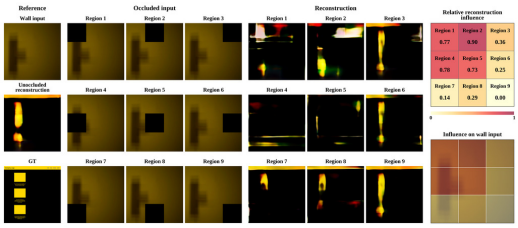
The 5-pixel motion condition in Fig. A62 preserves the same broad position dependence. For Screen and Website, the central and adjacent middle regions remain the most sensitive locations. Chart and Password also retain a center-biased response, although the relative differences between cells change compared with clean input. The blur redistributes local evidence horizontally but does not change the global crop geometry, so the principal sensitive region remains spatially similar.



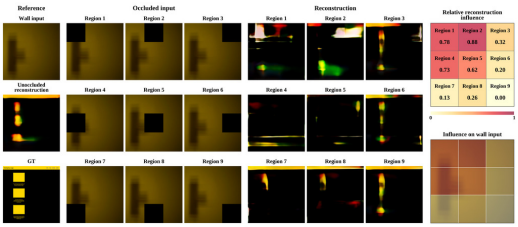
(a) Screen, 60-nit brightness reduction



(b) Screen, 130-nit brightness reduction



(c) Website, clean input



(d) Website, 13-pixel horizontal motion blur

Figure A64: Representative nine-region occlusion examples under different observation conditions. Each panel shows the reference wall input, nine region-wise occluded observations, the corresponding reconstructions, and the relative reconstruction influence across the 3×3 spatial partition.

With 13-pixel blur, the largest masking-induced losses remain within the central column across the four content categories (Fig. A63). This recurring spatial concentration connects reconstruction to wall evidence across both radiometric attenuation and spatial smoothing. The loss magnitude still depends on the base condition because the available evidence changes before masking.

Representation similarity and output response. To relate input perturbations to internal and output changes, we compare pooled, column-centered features using linear centered kernel alignment (CKA; Kornblith et al., 2019). Screen and Password use Decoder.0.layers.3.main.1.convs.0, while Chart and Website use Encoder.2.layers.3.main.1.convs.0; each location produces a 128-channel 64×64 feature map for a 256×256 input. Features are adaptively pooled to 8×8 before linear CKA. The output-change measure is the mean absolute difference between perturbed and clean predictions after 16×16 average pooling. Here, CKA measures agreement between perturbed and clean features at the same selected layer of each trained model.

Fig. A64 illustrates how removing different wall regions affects the reconstructed screen structure. In the clean Website example, occluding Regions 1, 2, 4, and 5 substantially disrupts the recovered header and vertically arranged foreground structures, with relative reconstruction influence scores ranging from 0.73 to 0.90. Occluding Region 9 leaves the coarse reconstruction largely unchanged. These differences reveal a spatially uneven dependence on the wall observation despite the similar masking areas. The strong response in Region 2 further shows that sensitivity extends beyond the most visually prominent shadow area into the surrounding radiance field. Moreover, removing a single wall region can alter several separated structures in the reconstructed image, indicating nonlocal dependencies between the observed evidence and the recovered layout. This behavior is consistent with diffuse transport mixing contributions from different screen locations across the wall. Within SplitIR, PAM organizes this mixed evidence into a shared layout state, while RCM regulates subsequent image refinements. The occlusion results connect the recovered structure to the available wall observation, complementing the module ablations and stage-wise analyses.

Fig. A65 shows two related but distinct responses. Conditions that move far from the clean representation often appear at lower CKA and larger SSIM loss, while larger output changes generally co-occur with larger reconstruction loss. The relationship is clearest in the pooled-output plots, where the four datasets show an approximately increasing trend between output L1 change and SSIM loss across the intervention set. The CKA plots are more dataset dependent because a high same-layer similarity can coexist with substantial output degradation when the relevant output change is small in feature-space orientation but important for reconstruction.

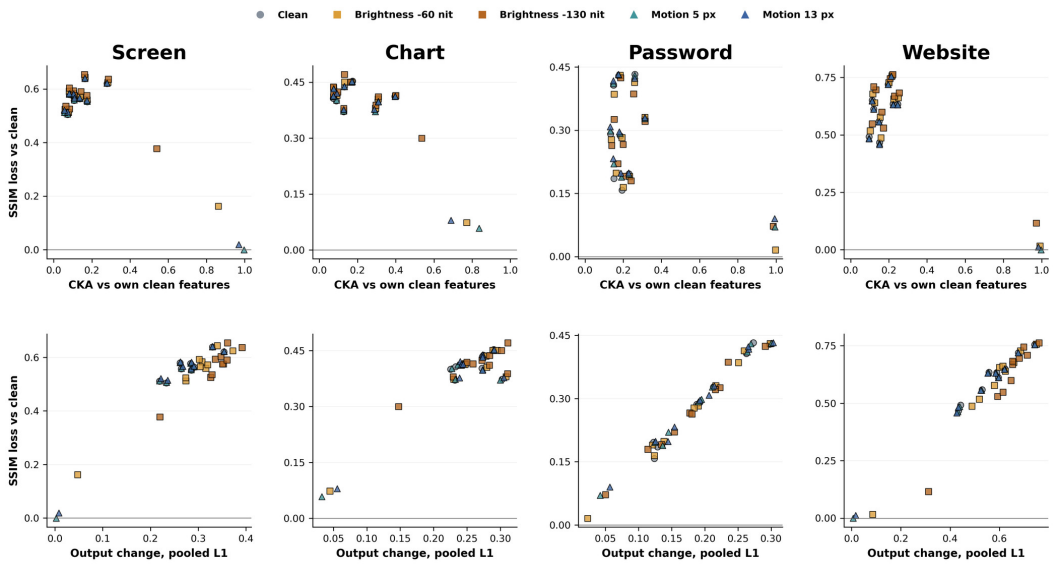


Figure A65: Relationship between representation change, output change, and reconstruction degradation for SplitIR. The upper row plots SSIM loss against same-layer CKA and the lower row against pooled output L1 change. The 49 conditions comprise 45 masked conditions across five base settings and four unmasked degraded settings; the unmasked clean input provides the reference.

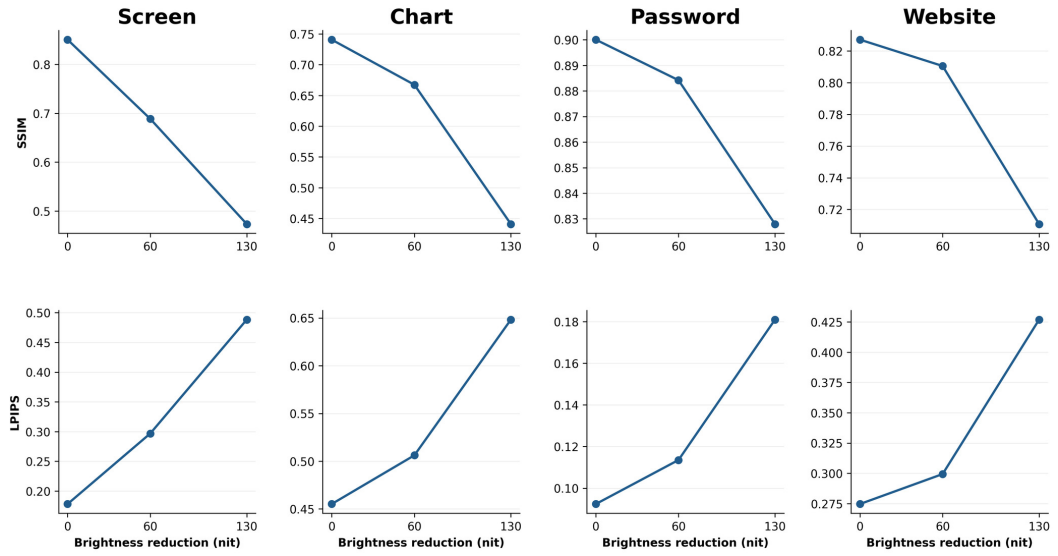


Figure A66: Full-model response to the three brightness conditions used by the feature-intervention experiment: 0, 60, and 130 nit-equivalent of brightness reduction. Curves report full-test-set SSIM and LPIPS for Screen, Chart, Password, and Website.

The paired measurements trace associated changes in the selected internal representation and reconstructed image as the wall observation is perturbed.

The brightness-response curves in Fig. A66 provide the base conditions underlying two of the occlusion groups. Screen SSIM changes from 0.851 under clean input to 0.688 at 60 nit-equivalent and 0.473 at 130 nit-equivalent. Chart changes from 0.741 to 0.667 and 0.441, while Password changes more gradually from 0.900 to 0.884 and 0.828. Website changes from 0.827 to 0.810 and 0.711. LPIPS moves in the opposite direction for all four datasets.

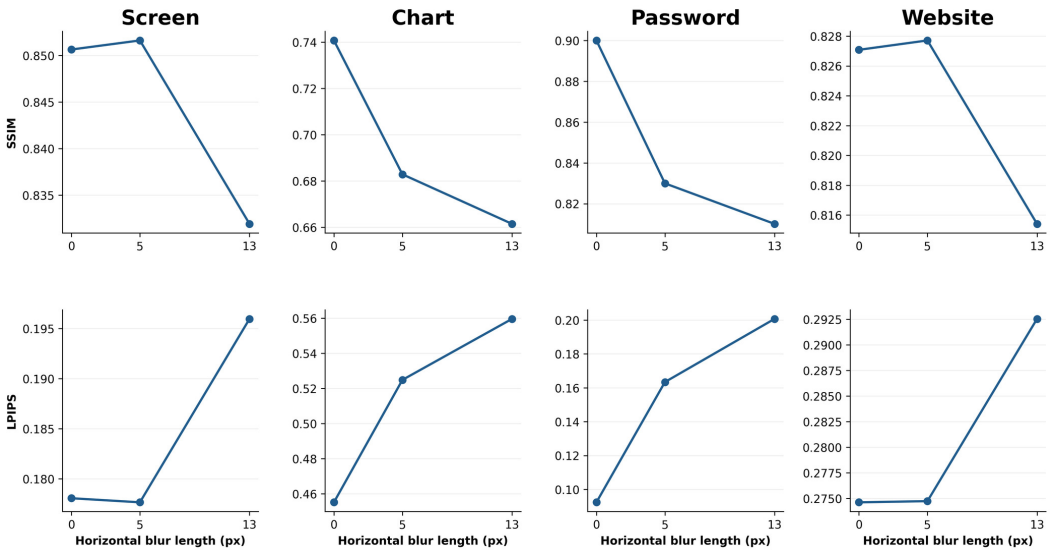


Figure A67: Full-model response to the horizontal motion-blur conditions used in the intervention analysis: 0, 5, and 13 pixels. Curves report full-test-set SSIM and LPIPS for the four datasets.

The three brightness levels anchor the response analysis: the 130-nit-equivalent reduction produces larger output changes and reconstruction losses than the 60-nit-equivalent reduction, with distinct responses across content categories.

Fig. A67 shows a different pattern for motion blur. Screen is nearly unchanged at 5 px ($0.851 \rightarrow 0.852$ SSIM) and decreases to 0.832 at 13 px. Website similarly changes only slightly at 5 px and reaches 0.815 at 13 px. Chart and Password are more sensitive: Chart decreases from 0.741 to 0.683 and 0.661, while Password decreases from 0.900 to 0.830 and 0.810. These differences reflect the amount of useful information carried by thin directional or character-scale structures in each target category.

The CKA values also distinguish brightness and motion perturbations. For example, Password retains very high same-layer CKA under both 5 and 13 px blur (0.995 and 0.993) even though its SSIM drops by approximately 0.07 and 0.09. High feature agreement can coexist with visible changes in image structure, showing that representation agreement and reconstruction fidelity capture complementary aspects of the response.

The examples in Fig. A68 give a visual counterpart to the aggregate curves. The 130-nit-equivalent condition changes both contrast and recovered appearance, while the 13-pixel blur mainly suppresses fine spatial structure. The effect differs by content: Chart loses thin geometric components, Password loses stroke fidelity, and Website preserves coarse page partitioning longer than fine local detail. Screen combines both behaviors because it contains heterogeneous image and interface structures.

Together, the occlusion and response experiments show that spatial sensitivity and representation stability are related but not interchangeable. The nine-region masks probe where evidence enters the model, whereas CKA and pooled-output changes describe how the trained representation and prediction move away from clean input under a broad set of conditions. The two analyses therefore provide complementary views of the same reconstruction process.

C.5.3 PAM Stage Diagnostics

We inspect both PAM calls on 32 clean evaluation images per dataset, yielding 64 stage records per dataset and 256 records in total. For layout state F_i and updated state F_i^+, the diagnostic records $100\|F_i^+ - F_i\|_2 / \|F_i\|_2$, mean mobility, and the conditional potential change. The potential is recomputed after the update while holding the observation encoding and metric fixed, giving a forward-pass diagnostic of the local update under fixed observation and metric conditioning.

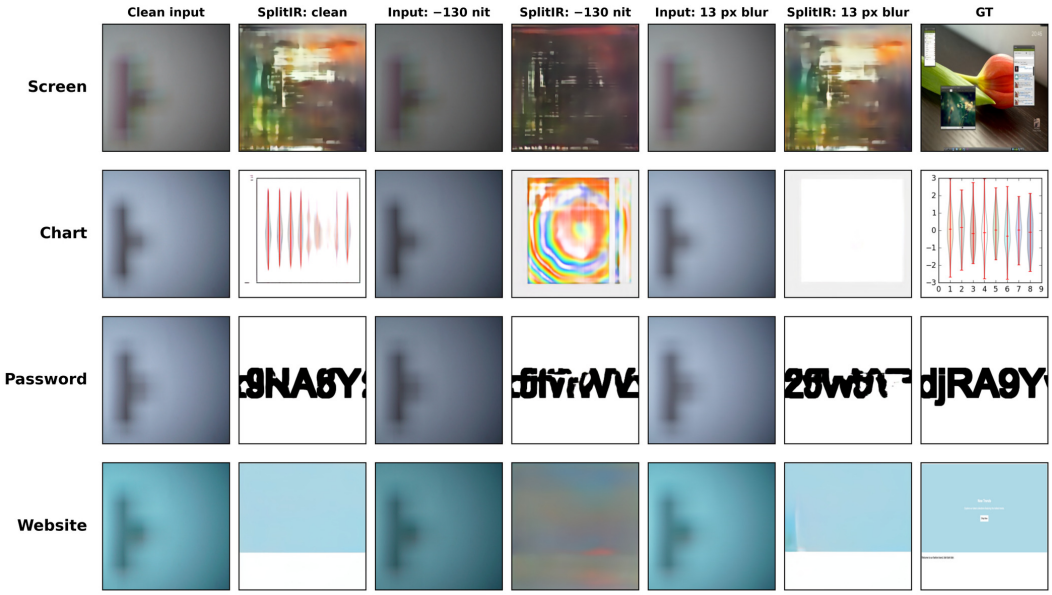


Figure A68: Representative input and reconstruction examples for clean input, 130-nit-equivalent brightness reduction, and 13-pixel horizontal motion blur. Each row uses the fixed first example of Screen, Chart, Password, or Website and shows the corresponding SplitIR output together with GT.

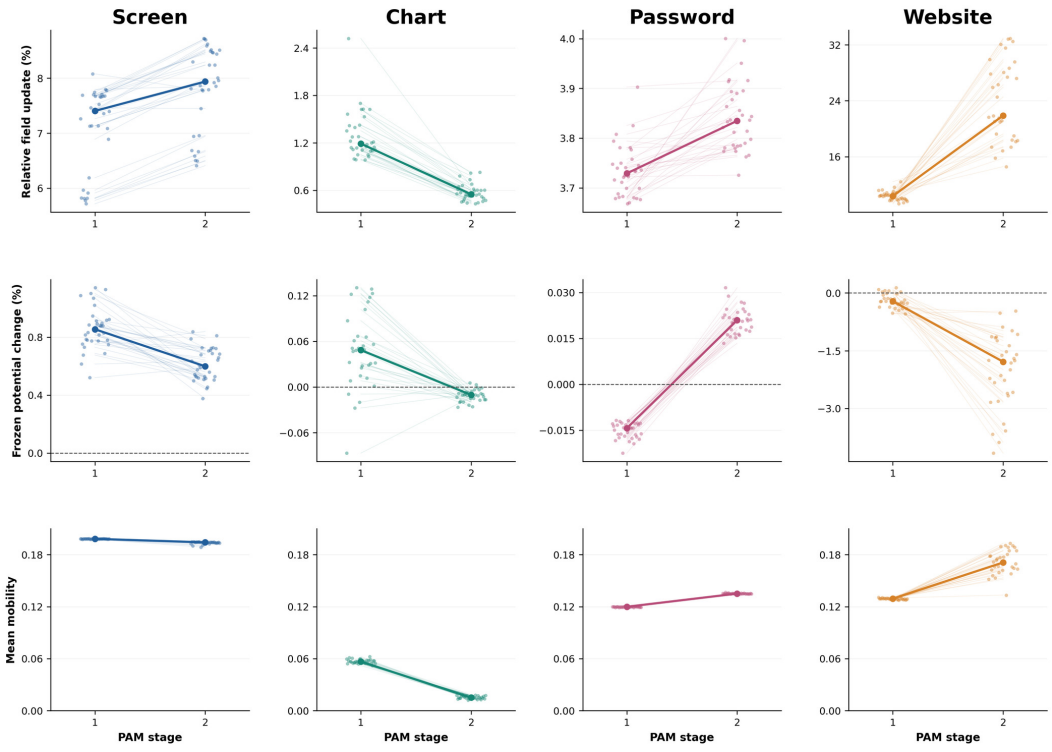


Figure A69: PAM stage diagnostics over 32 clean images per dataset. The first row reports relative layout-state update magnitude, the second row reports the conditional potential change with observation and metric held fixed, and the third row reports mean mobility. Pale lines connect the same image across the two PAM calls; the solid line connects stage medians.

Fig. A69 shows content-dependent update scales and mobilities across the four dataset-specific models. The text reports arithmetic means, while the solid line connects stage medians. Screen changes by an average of 7.10% and 7.83% in stages 1 and 2. Chart is much smaller at 1.29% and

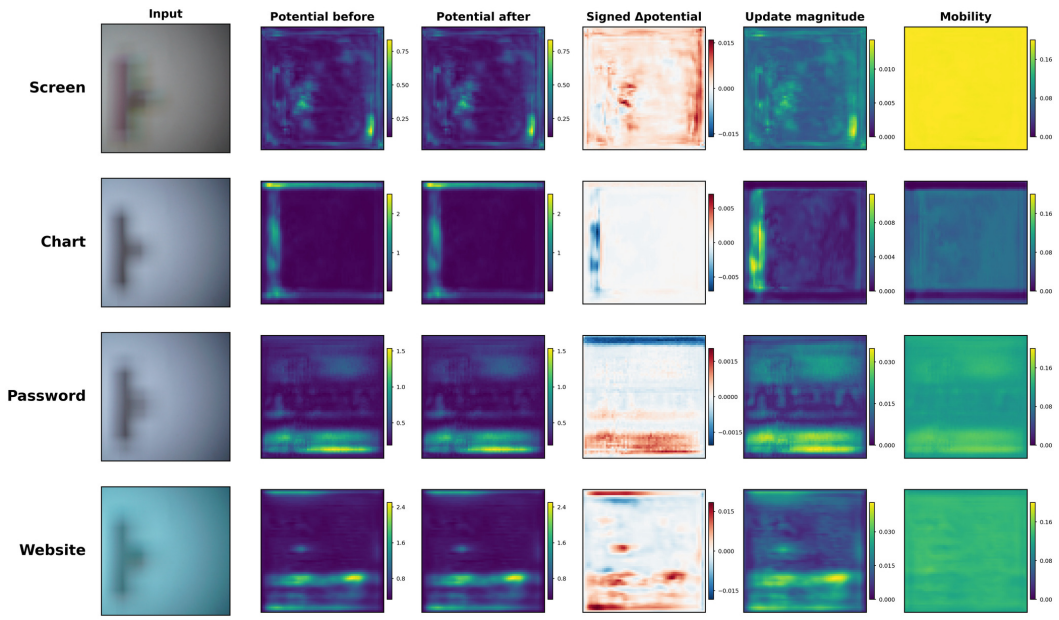


Figure A70: Spatial diagnostics for PAM stage 1 on the fixed representative example of each dataset. Columns show the input, conditional potential before and after the update, signed potential difference, update magnitude, and mobility. Potential-before and potential-after maps share a row-wise scale; mobility uses the common $[0, 0.2]$ display range.

0.57%. Password remains near 3.73% and 3.84%, while Website increases from 10.37% in stage 1 to 23.83% in stage 2. The learned mobility follows a related but not identical pattern: Screen is near its upper range at approximately 0.198 and 0.194, Chart decreases from about 0.057 to 0.015, Password moves from 0.120 to 0.135, and Website from 0.129 to 0.172.

The conditional-potential response varies across content categories and between the two assimilation calls. The recorded decrease counts are 0/64 for Screen, 34/64 for Chart, 32/64 for Password, and 58/64 for Website; the stage means likewise differ in sign across categories and calls. Together with the update and mobility measurements, these responses describe how learned bounded transport reorganizes the current layout state.

The stage-1 maps in Fig. A70 localize where the update acts in the shared layout state. Screen shows broad distributed changes, Chart concentrates its update around the region carrying line structure, Password exhibits character-aligned responses, and Website emphasizes broad page regions. The signed potential difference contains both positive and negative values, matching the aggregate observation that the diagnostic potential is not constrained to decrease for every sample and stage.

The mobility field also differs by content. Screen is close to the upper displayed range over much of the field, whereas Chart uses a much smaller mobility in the selected example. Password and Website lie between these cases. This agrees with the stage statistics and supports an interpretation of mobility as a learned, content-dependent scale on the transported update rather than a fixed global step size.

The second PAM call in Fig. A71 operates on a state that already includes the first assimilation update. The spatial patterns therefore need not repeat stage 1. Website shows the clearest increase in update magnitude, consistent with its average relative update rising to 23.83%. Chart shows a weaker second-stage response and a lower mobility field, consistent with the drop in both its mean update magnitude and mobility. Password maintains a similar overall update scale across the two stages but reverses the sign of the mean conditional potential change.

The two assimilation calls adapt both the spatial support and magnitude of the layout update. Website places a stronger update in the second call, while Chart uses a smaller second-stage response. These content-dependent trajectories show how the potential-guided direction and learned mobility work together within bounded layout-state transport.

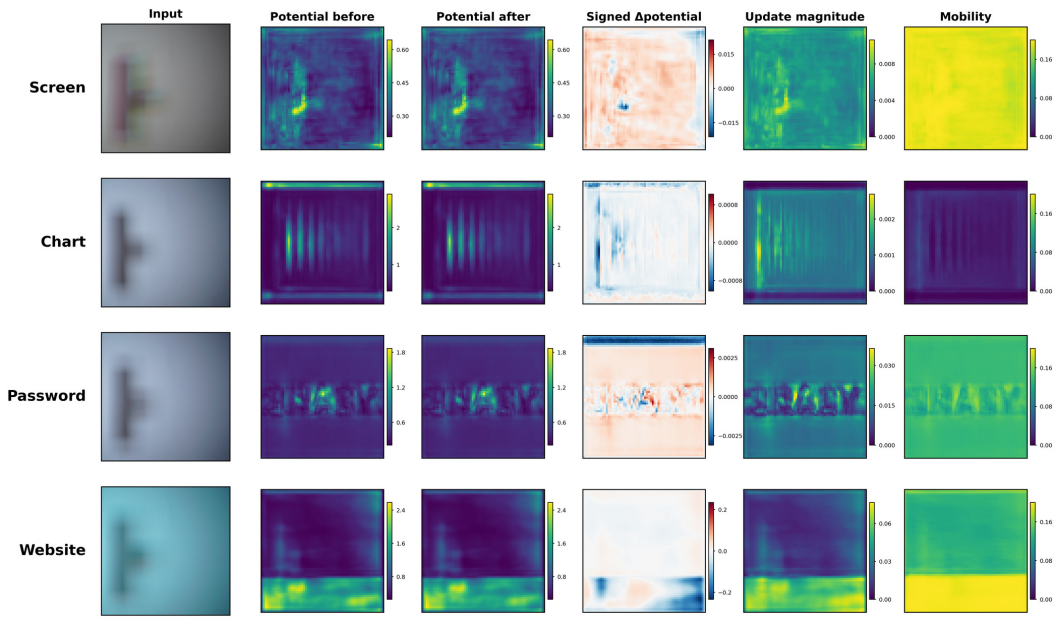


Figure A71: Spatial diagnostics for PAM stage 2 using the same quantities as Fig. A70. The maps show how the second assimilation call redistributes update magnitude and mobility after the shared state has already been modified once.

C.5.4 RCM Stage Diagnostics

RCM is analyzed on the same 32 clean evaluation images per dataset. Each image produces four RCM calls, yielding 128 stage records per dataset and 512 records in total. For current reconstruction f_k and proposal a_k, the candidate correction $d_k^r = a_k - f_k$ is transformed into proximal correction p_k and feasible correction $\bar{d}_k$. The diagnostics record correction magnitude, projection-active coordinates, and a fixed-condition non-expansiveness residual.

Fig. A72 reveals a strongly stage- and content-dependent correction schedule. Screen has mean relative corrections of 3.44%, 1.62%, 4.52%, and 1.67% across stages 1–4. Chart follows 6.80%, 1.16%, 3.74%, and 1.47%. Password applies its largest update at the first RCM call, 18.18%, followed by 1.90%, 3.73%, and 2.37%. Website uses smaller mean corrections of 0.66%, 0.33%, 1.35%, and 0.60%.

RCM combines continuous proximal adjustment with content-dependent feasible-set projection. The Screen, Chart, and Password proximal corrections lie within their learned intervals across the sampled calls, while Website additionally activates projection across the refinement sequence. Website triggers projection in 6/32, 32/32, 32/32, and 30/32 records over stages 1–4, with mean active-coordinate fractions of below 0.01%, 2.75%, 27.14%, and 2.01%, respectively. The largest projection activity therefore occurs at stage 3, connecting the learned correction geometry to the evolving reconstruction.

RCM stage 1. The first RCM stage acts directly on the earliest full-resolution proposal in the sequential correction chain. Password has the largest mean relative correction at this stage, while Website has the smallest. The spatial maps below show the candidate correction, projected correction, learned center magnitude, radius, and fraction of active projection channels for the fixed representative sample of each dataset.

Fig. A73 shows structured correction fields that follow the reconstructed content, including chart-line organization and character support. For Screen, Chart, and Password, the proximal corrections already lie inside their learned intervals in the sampled calls; Website additionally activates a small projection fraction at this stage. The learned correction geometry is spatially content dependent.

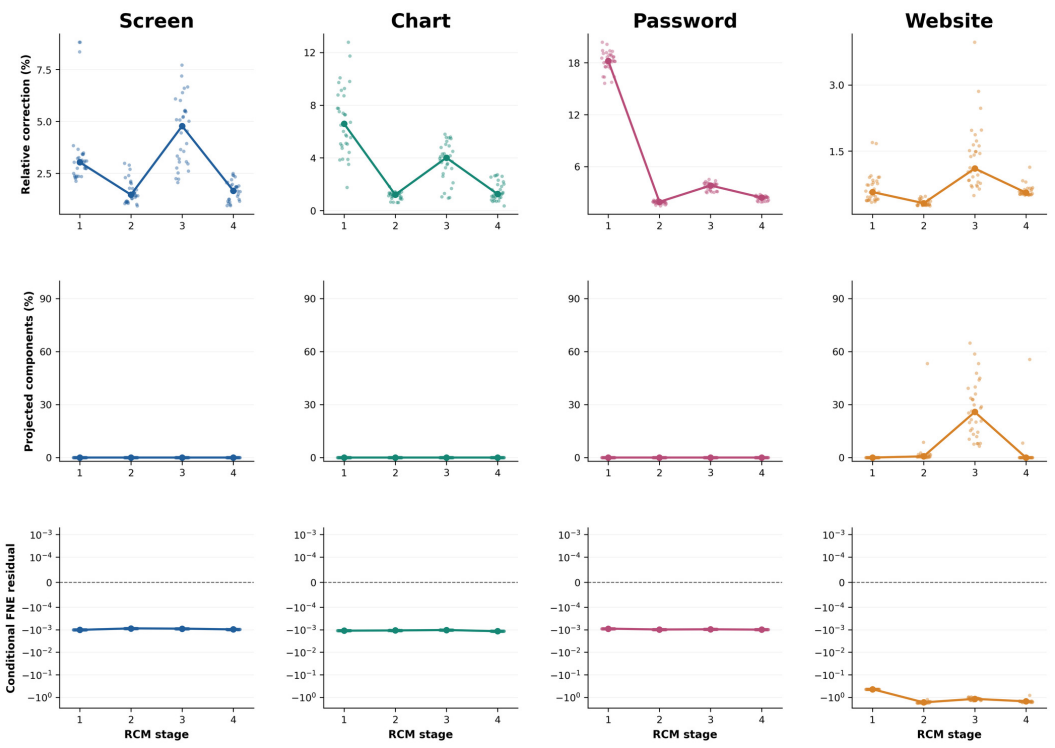


Figure A72: RCM stage diagnostics over 32 clean images per dataset. The first row reports relative reconstruction-correction magnitude, the second row reports the fraction of RGB coordinates changed by box projection, and the third row reports the fixed-condition residual $\|\Delta T\|_2^2 - \langle \Delta T, \Delta d \rangle$. Pale lines connect the same image across stages; the solid line connects stage medians, while the text reports arithmetic means.

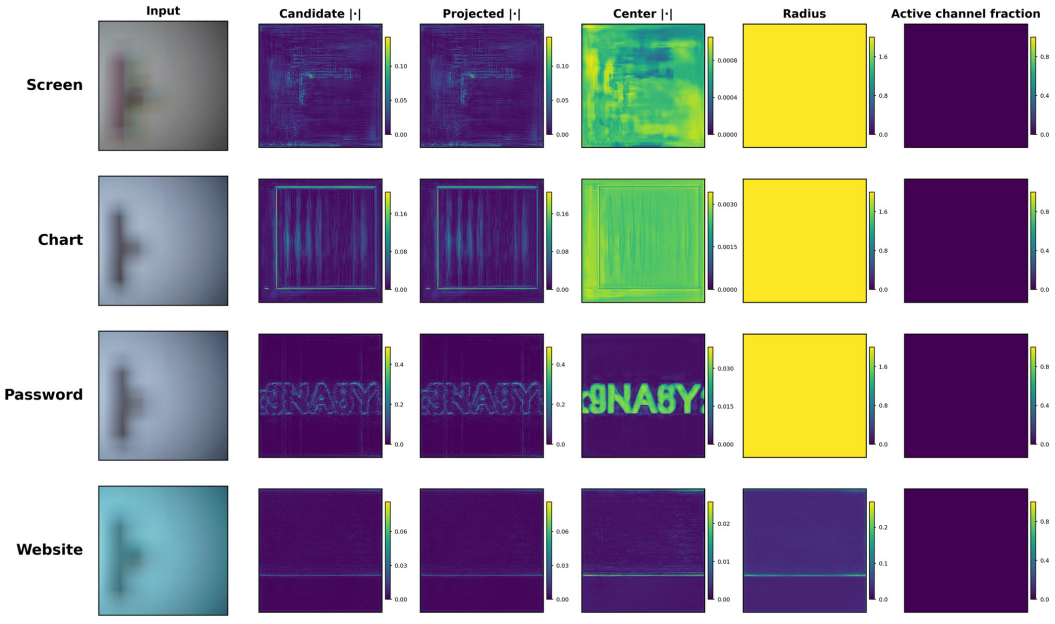


Figure A73: Spatial diagnostics for RCM stage 1. Columns show input, candidate correction magnitude, projected correction magnitude, center magnitude, radius, and active projection-channel fraction for the fixed representative example of each dataset.

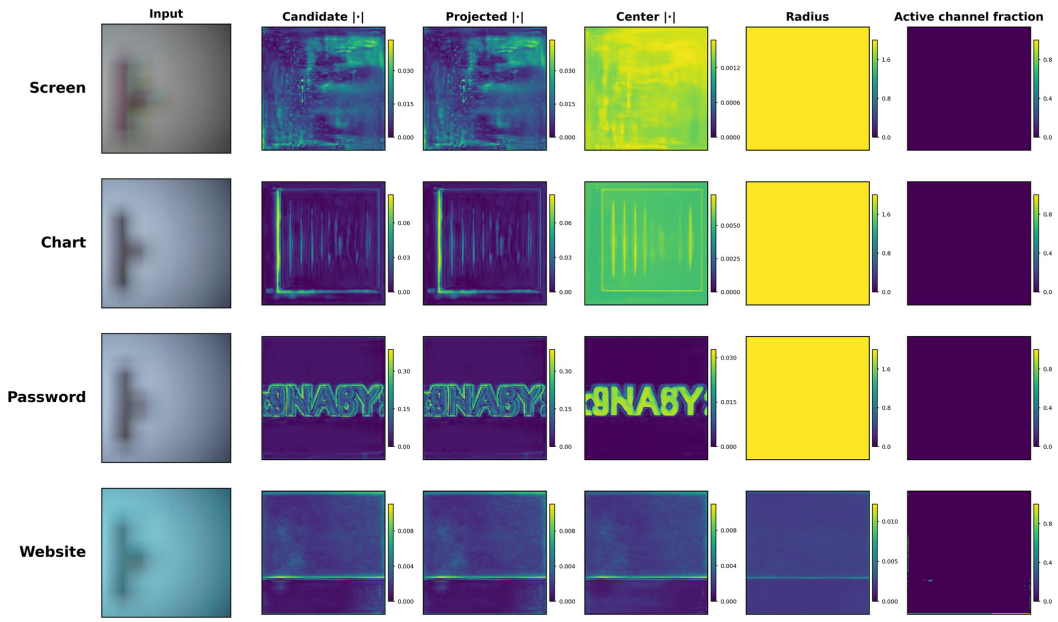


Figure A74: Spatial diagnostics for RCM stage 2 using the same quantities as Fig. A73. Website activates the learned box on all 32 sampled records at this stage, whereas the other three datasets remain inside their boxes for the sampled clean inputs.

RCM stage 2. The second stage receives the reconstruction already modified by stage 1. Aggregate correction magnitudes decrease for all four datasets, while Website becomes the first dataset in which every sampled record activates the projection at this stage.

The selected example in Fig. A74 reflects this difference. Screen, Chart, and Password retain nearly zero active-channel maps, while Website exhibits a visible band of projected coordinates. The correction magnitude remains spatially organized even in the non-projecting datasets, again separating the role of the proximal candidate update from the additional role of the box projection.

RCM stage 3. Stage 3 produces the largest average Website correction and the largest projection activity. It also forms the largest mean correction for Screen after stage 1 and 2 have already modified the reconstruction.

For Website, the aggregate active-coordinate fraction reaches 27.14% at stage 3, and all 32 sampled records activate projection. The corresponding spatial map in Fig. A75 shows concentrated projection activity. The other three datasets retain non-zero candidate and projected corrections, with their proximal corrections already inside the learned boxes. This stage shows how proximal adjustment and feasible-set projection divide their roles across content regimes.

RCM stage 4. The final RCM call produces a smaller mean correction after the stage-3 update but preserves content-dependent spatial structure in the candidate and projected fields.

In Fig. A76, Website still shows localized projection activity, while Screen, Chart, and Password have proximal corrections within their learned boxes for the selected example. The candidate maps retain content structure, including character support in Password and page partitioning in Website. Across all four RCM figures, the spatial correction pattern changes between calls, consistent with the sequential stage-conditioned formulation in Sec. B.2.

Constraint and fixed-condition checks. For every recorded RCM call, we verify the coordinate-wise box and a fixed-condition correction mapping. With the stage geometry fixed, define

$$T_k(d) = \Pi_{C_k}\left(\frac{d + \rho\, m_k^r \odot c_k}{1 + \rho\, m_k^r}\right), \qquad \Delta T = T_k(d + \Delta d) - T_k(d).$$

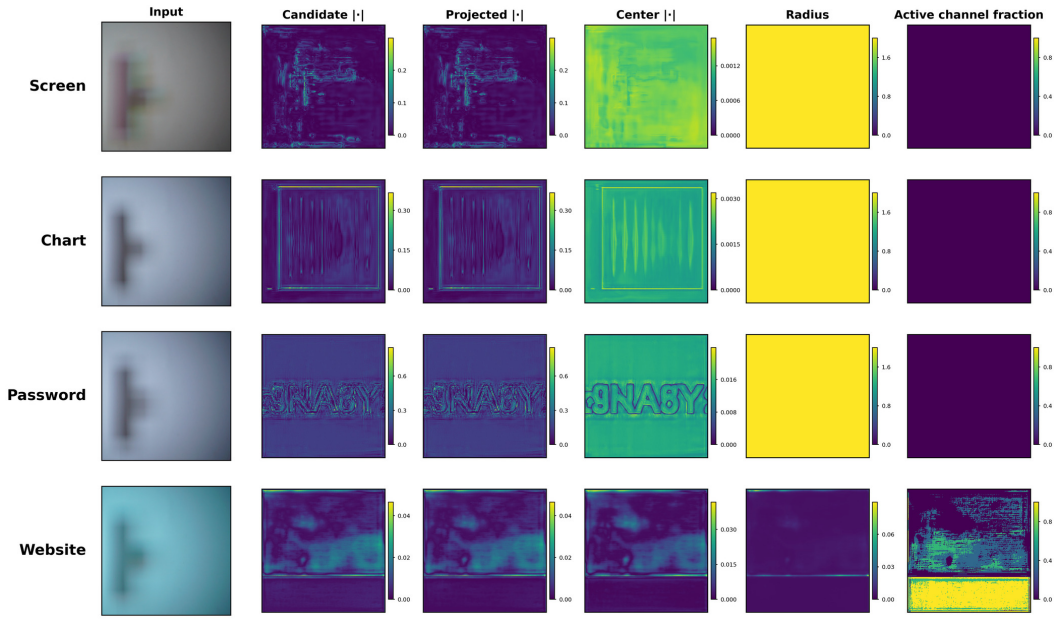


Figure A75: Spatial diagnostics for RCM stage 3. This stage exhibits the largest mean projection-channel fraction on Website and a renewed increase in correction magnitude for Screen, Chart, and Password relative to stage 2.

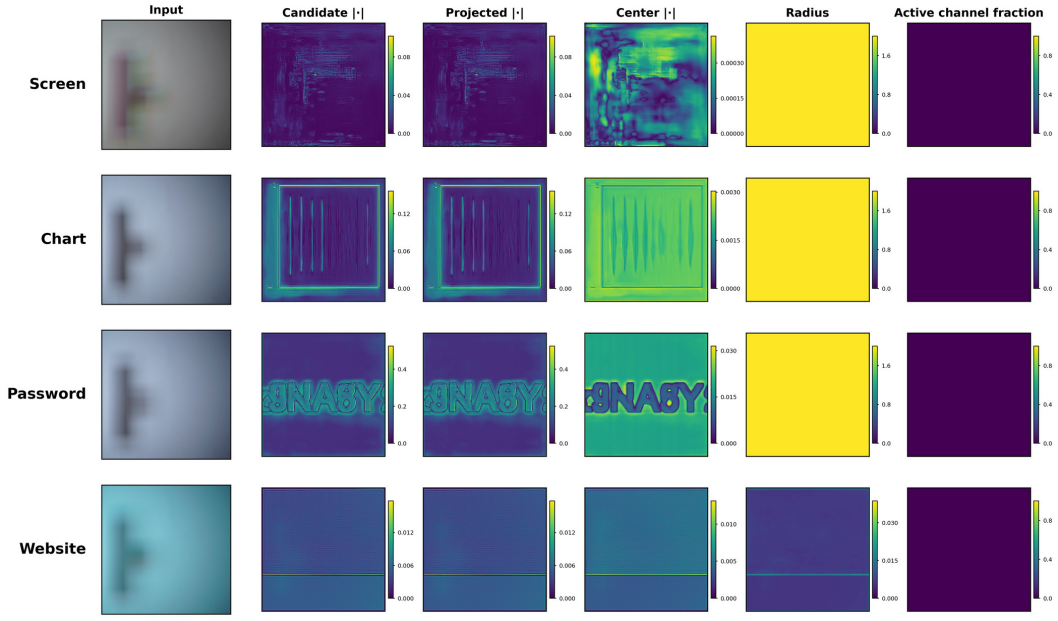


Figure A76: Spatial diagnostics for RCM stage 4. Candidate and projected corrections, center magnitude, radius, and active-channel fraction are shown for the same representative examples used in the previous RCM-stage figures.

The deterministic perturbation is $\Delta d = 0.01 \sin(j)$, where j is the flattened integer index over the BCHW candidate tensor reshaped back to its original shape. Across all 512 recorded calls, $\|\Delta T\|_2^2 - \langle \Delta T, \Delta d \rangle \leq 0$ with the box, center, metric, and strength held fixed.

The projected corrections satisfy their learned boxes in all recorded calls. Image-range mapping can further change the correction realized in the reconstruction; two Password image-stage records exceed the 10^{-6} plotting tolerance after this mapping, while Website deviations remain at floating-

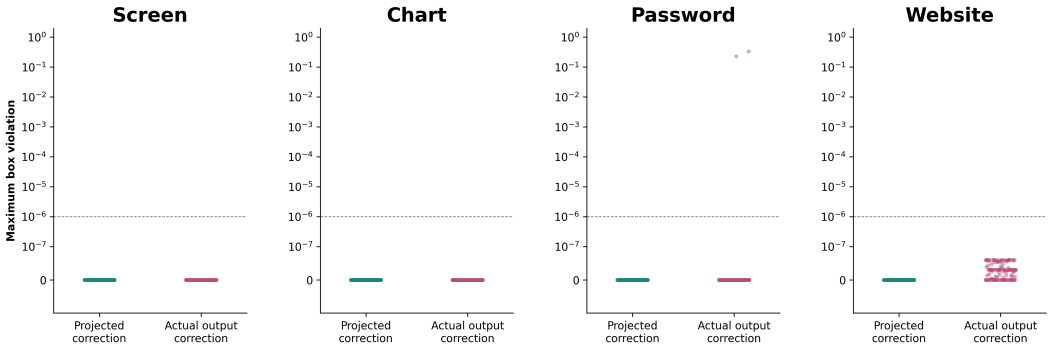


Figure A77: RCM constraint-membership diagnostics. Each point is one image-stage record. The left group checks the projected correction against its learned local box; the right group shows the image-range-adjusted correction after mapping the reconstruction to the valid intensity range. The horizontal dashed line marks the 10^{-6} plotting tolerance.

point scale. Across all 512 recorded calls, the prescribed candidate perturbation yields a non-positive fixed-condition residual, consistent with the analytical correction structure in Sec. B.2.

Summary of the feature analysis. Together, the analyses trace a continuous reconstruction process: spatially distributed wall evidence develops into content-aligned layout structure, and stage-conditioned corrections refine that structure into the screen image. Internal responses reveal where this organization emerges, spatial interventions connect it to the observed wall regions, and operator measurements show how the updates adapt across content categories. The ablation comparisons provide the corresponding reconstruction-level evidence for the two-operator design. Together, these results support the initial feasibility of recovering concealed screen layouts, chart structures, and character patterns from ordinary-camera observations of wall radiance.