WEAK LIGHT AS STRUCTURAL EVIDENCE: RELAY-STATE HIERARCHICAL REPRESENTATION LEARNING FOR HIDDEN-SCREEN RECONSTRUCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

We investigate whether an ordinary camera can recover hidden display content from the weak light that the display projects onto a visible relay wall while direct viewing is blocked. Diffuse reflection, spatial spread, and propagation attenuation disperse and mix the brightness, edges, and spatial relations of screen structures, so the wall response carries these cues in a rearranged form. We present Relay-State Hierarchical Representation Learning (RSHRL), a two-pass framework that first organizes this structural evidence through Relay-Wall Invertibility Representation (RWIR) and then repairs residual damage through Damage-State Learning (DSL). RWIR rebuilds wall-to-screen structural correspondence by combining fine evidence, regional relationships, and coarse screen context. DSL estimates a scalar damage severity from the preliminary reconstruction; the second-pass consumer uses this condition through its internal DACA, DASA, and SDR operations to refine the affected structures. Experiments on four paired screen-content datasets show that RSHRL leads the evaluated restoration baselines across all 12 PSNR, SSIM, and LPIPS comparisons. Removal and replacement studies demonstrate the complementary value of hierarchical representation and damage-guided repair. Across two acquisition setups and varied observation conditions, the reconstructions recover principal brightness distributions, text-bearing regions, icon contours, window boundaries, and overall interface layouts. These results establish the initial feasibility of recovering hidden screen content with an ordinary camera by observing only wall-reflected display light.

1 INTRODUCTION

An occluded display still illuminates a visible wall, creating a measurable relay signal. Kuhn recovered readable CRT text from diffusely reflected optical emission using a fast photosensor and deconvolution under low background illumination Kuhn (2002). Here, an ordinary camera records the spatial wall-light pattern, from which we reconstruct principal screen regions, icon contours, window boundaries, and interface layouts. We study passive hidden-screen reconstruction when the display and camera share the same space while the camera's direct view of the display is blocked. The camera observes naturally emitted display light after it reaches a rough wall and undergoes diffuse reflection, spatial spread, and energy attenuation. This screen-to-wall-to-camera path transforms text strokes, icon contours, window boundaries, and interface regions into low-contrast, overlapping weak-light blobs. Our central question is whether this wall response retains enough structural evidence to rebuild correspondence with the original screen and recover its principal content.

Existing passive NLOS methods recover hidden motion, texture, or coarse appearance from naturally available illumination and occlusion cues Bouman et al. (2017); Saunders et al. (2019), while modern image restoration networks achieve strong performance in denoising, deblurring, and low-light enhancement Zamir et al. (2022); Chen et al. (2022); Cai et al. (2023). Passive hidden-screen reconstruction couples two structural transitions. First, relay-wall propagation mixes distant display regions into the same smooth observation: evidence from one text line or window may be dispersed over a broad wall region, and one observed region may contain contributions from several screen locations. Recovering the screen therefore requires reorganizing distributed weak-light evidence and rebuilding wall-to-screen structural correspondence. Second, the remaining reconstruction errors are

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.

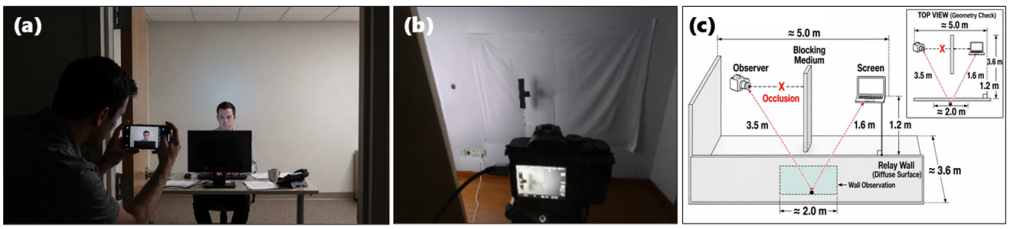

Figure 1: Hidden-screen observation through wall-reflected display light. (a) An occluded-display acquisition scene. (b) Camera observation of wall illumination with a T-shaped blocker. (c) The screen-to-wall-to-camera geometry. In both acquisition configurations, the display remains outside the camera's direct view.

unevenly distributed: some regions fade, some boundaries break, and some details become missing or misplaced. Recovering the screen therefore requires both rebuilding its spatial organization and directing further repair to the structures that need it.

To answer this question, we propose Relay-State Hierarchical Representation Learning (RSHRL), a unified two-pass framework that progresses from relay-evidence organization to damage-guided screen restoration. Relay-Wall Invertibility Representation (RWIR) combines fine evidence, regional relationships, and coarse context to reorganize the dispersed wall response into the preliminary screen reconstruction Y_{pre}. Damage-State Learning (DSL) then estimates a scalar severity $s \in [0, 1]$, which conditions the second-pass consumer and its internal channel, spatial, and state-driven repair operations. This progression connects dispersed wall illumination to recoverable screen organization. Ordinary-camera reconstruction has used an occluder's penumbra Saunders et al. (2019) and two orthogonal occluding edges Czajkowski & Murray-Bruce (2024) to recover hidden scenes. RSHRL recovers principal brightness distributions, icon contours, window boundaries, and interface layouts from both blocker-assisted observations and smooth behind-screen wall reflections, outperforming the evaluated restoration methods across four content datasets and establishing initial screen recovery from ordinary-camera observations (Figure 3; Table 1). Figure 1 illustrates the acquisition setups and reconstruction objective.

The main contributions are summarized as follows:

(1) We introduce RSHRL, a unified two-pass framework that turns weak relay-wall illumination into a screen reconstruction through hierarchical evidence organization and severity-conditioned restoration.
(2) We develop RWIR to rebuild wall-to-screen structural correspondence through fine-, regional-, and coarse-scale representations, and DSL to estimate a bounded scalar severity for second-pass refinement. The two components connect structural organization with severity-conditioned repair.
(3) We conduct extensive experiments on four paired screen-content datasets, including comparisons with state-of-the-art restoration methods, component ablations, robustness evaluations, and qualitative analyses. The results establish recovery across blocker-assisted and behind-screen wall-reflection configurations.

2 Related Work

Passive non-line-of-sight imaging. Passive non-line-of-sight imaging recovers hidden-scene information from naturally available illumination or occlusion cues. Accidental pinhole and pinspeck imaging exploits naturally occurring occluders to reveal hidden visual information Torralba & Freeman (2014), while computational periscopy uses occlusion-induced variations in light transport to reconstruct hidden two-dimensional scenes Saunders et al. (2019). Optimal-transport-based approaches further demonstrate that diffuse relay measurements contain recoverable information about hidden appearance Geng et al. (2022). Aittala et al. Aittala et al. (2019) recovered hidden video by jointly estimating the hidden sequence and its light transport through deep matrix factorization. RSHRL instead learns a paired wall-to-screen mapping within each content set and reuses the trained reconstruction path to recover brightness regions, icon contours, and window layouts in held-out desktop, lock-screen, and game images (Figure 3).

Image restoration under structured degradation. Direct image restoration methods learn mappings from degraded observations to aligned clean targets. Progressive and multiscale architectures such as MPRNet and Restormer combine contextual information with spatial detail Zamir et al. (2021; 2022), while NAFNet and Uformer provide strong convolutional and hierarchical transformer operators for general image restoration Chen et al. (2022); Wang et al. (2022). Retinexformer uses illumination guidance for low-light enhancement Cai et al. (2023), AdaIR adapts restoration through frequency mining and modulation Cui et al. (2025), and MODEM estimates global and local degradation priors for adverse-weather recovery Wang et al. (2025). RSHRL connects hierarchical relay representation with a scalar severity estimated from the preliminary reconstruction, directing refinement toward screen structures affected by wall propagation. This task-oriented progression links recovery of spatial correspondence to repair of the resulting image.

3 METHOD

3.1 OVERALL ARCHITECTURE

RSHRL resolves two linked structural transitions in passive hidden-screen reconstruction: relay transport mixes local screen evidence, and the resulting preliminary reconstruction retains heterogeneous residual damage. The framework comprises Relay-Wall Invertibility Representation (RWIR) and Damage-State Learning (DSL). Pass 1 uses the shared restoration backbone and RWIR to form a preliminary screen reconstruction Y_{pre}, from which DSL estimates a bounded scalar severity s. The fixed context $(Y_{\mathrm{pre}}, F_{\mathrm{fine}}, F_{\mathrm{layout}}, s)$ conditions Pass 2: the RWIR fields enter through the context adapter, whereas s controls the internal DACA, DASA, and SDR operations. The downstream decoder and RGB head then produce the final reconstruction. Figure 2 summarizes this dependency.

Let $t_c(s)$ be the screen radiance at $s \in \Omega_s$, $r \in \Omega_w$ a relay-wall point, and $q \in \Omega_o$ a camera sample. The steady-state screen-to-wall-to-camera path is represented by

$$\begin{aligned} x_c(q) &= (\mathcal{X}_c t_c)(q) + \zeta_c(q) + \eta_c(q), \\ (\mathcal{X}_c t_c)(q) &= \int_{\Omega_s} \kappa_c(q,s) t_c(s)\,\mathrm{d}s, \\ \kappa_c(q,s) &= \int_{\Omega_w} h_c(q,r)\rho_c(r)V(s,r)g(s,r)\,\mathrm{d}r. \end{aligned} \tag{1}$$

Here, $x_c \in L^2(\Omega_o)$ is the observed wall-light field, while the unknown nonnegative kernel κ_c combines camera footprint h_c, wall response ρ_c, visibility V, and propagation g. The fields ζ_c and η_c denote ambient response and signed acquisition noise. Spatial spreading and attenuation make the inverse ill-conditioned; the complete operator analysis is provided in Appendix B.

Let $\Omega = \{1,\dots,H\}\times\{1,\dots,W\}$ and let the normalized wall observation be $x \in [0,1]^{B\times 3\times 384\times 512}$. With shared restoration parameters θ, representation parameters ϕ, inverse-decoder parameters δ, and damage-state parameters ψ, the forward dependency is

$$\begin{aligned} (Y^{(1)}, \widehat{F}_{\mathrm{early}}, \widehat{F}_{\mathrm{up}}) &= \mathcal{F}_\theta^{(1)}(x), \\ F_{\mathrm{early}} = \mathrm{stopgrad}(\widehat{F}_{\mathrm{early}}), &\qquad F_{\mathrm{up}} = \mathrm{stopgrad}(\widehat{F}_{\mathrm{up}}), \\ (Z_2, Z_3, Z_4) &= \mathcal{R}_\phi(x, F_{\mathrm{early}}, F_{\mathrm{up}}), \\ (F_{\mathrm{layout}}, F_{\mathrm{fine}}, Y_{\mathrm{layout}}, Y_{\mathrm{pre}}) &= \mathcal{D}_\delta(Z_2, Z_3, Z_4), \\ s = \sigma(\mathcal{E}_\psi(Y_{\mathrm{pre}}))\,, &\qquad s \in [0,1]^{B\times 1}, \\ \mathcal{C} &= \mathrm{stopgrad}(Y_{\mathrm{pre}}, F_{\mathrm{fine}}, F_{\mathrm{layout}}, s), \\ y &= \mathcal{F}_\theta^{(2)}(x;\mathcal{C}). \end{aligned} \tag{2}$$

The two auxiliary RWIR tensors are generated internally from the wall observation during Pass 1. The full-resolution observation-embedding feature $\widehat{F}_{\mathrm{early}} \in \mathbb{R}^{B\times 24\times 384\times 512}$ and the coarse decoder feature $\widehat{F}_{\mathrm{up}} \in \mathbb{R}^{B\times 192\times 48\times 64}$ provide complementary spatial scales and are held fixed when they enter RWIR. Thus, inference requires only the relay-wall observation as an external image input. RWIR produces $Z_2 \in \mathbb{R}^{B\times 32\times 192\times 256}$, $Z_3 \in \mathbb{R}^{B\times 48\times 96\times 128}$, and $Z_4 \in \mathbb{R}^{B\times 64\times 48\times 64}$. The inverse decoder returns $F_{\mathrm{layout}} \in \mathbb{R}^{B\times 64\times 96\times 128}$, $F_{\mathrm{fine}} \in \mathbb{R}^{B\times 48\times 192\times 256}$, and $Y_{\mathrm{pre}} \in [0,1]^{B\times 3\times 384\times 512}$.

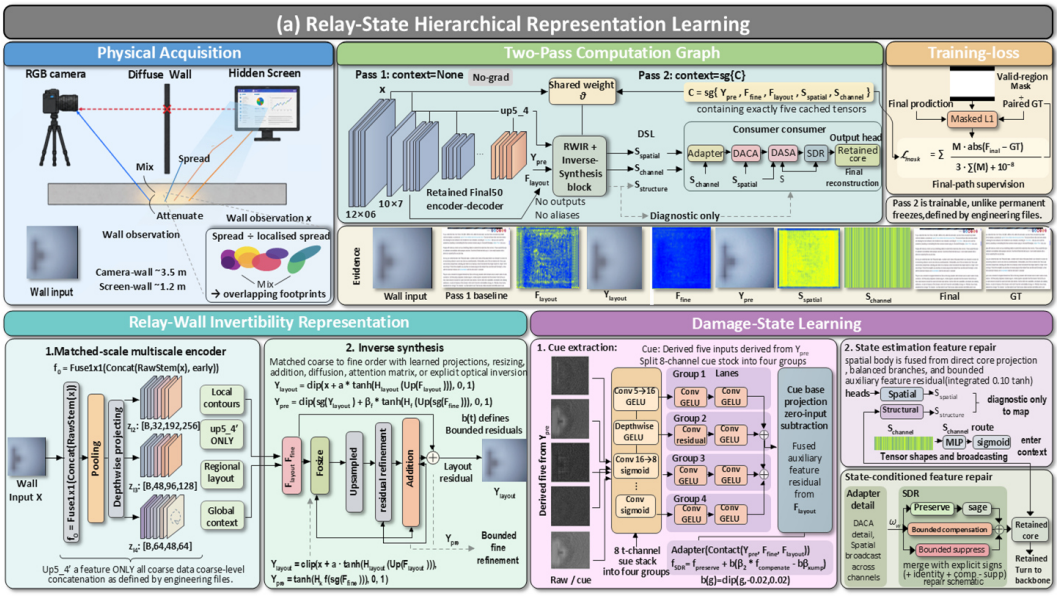

Figure 2: Two-pass RSHRL architecture. Pass 1 uses the shared restoration backbone and RWIR to organize the wall observation into hierarchical representations, the fine and layout fields, and the preliminary reconstruction Y_{pre}. DSL estimates one bounded scalar severity $s \in [0, 1]$ from Y_{pre}. The fixed RWIR fields enter the Pass-2 context adapter, while s conditions the consumer's internal DACA, DASA, and SDR operations before the decoder and RGB head produce the final output.

The first-pass RGB output $Y^{(1)}$ is the shared-backbone image readout, whereas Y_{pre} is formed by the RWIR inverse decoder and is the image input to DSL. DSL compresses the remaining reconstruction difficulty into one bounded severity value per sample. Optimization updates the shared Pass-2 backbone, the severity-conditioned consumer, the copied core block, and the downstream decoder and RGB head.

3.2 Relay-Wall Invertibility Representation

RWIR rebuilds relay-to-screen correspondence through the three hierarchy levels Z_2, Z_3, Z_4. Its inputs combine the wall observation with the fixed full-resolution observation-embedding feature and the fixed coarse decoder feature:

$$(Z_2, Z_3, Z_4) = \mathcal{R}_\phi(x, F_{\mathrm{early}}, F_{\mathrm{up}}). \tag{3}$$

The inverse decoder aligns these scales and produces the layout field, fine field, layout image, and preliminary reconstruction. The preliminary readout is

$$Y_{\mathrm{pre}} = \mathrm{clip}_{[0,1]}[\mathrm{stopgrad}(Y_{\mathrm{layout}}) + \beta \tanh(H_f(\mathrm{Up}(\mathrm{stopgrad}(F_{\mathrm{fine}}))))] \,, \qquad \beta = 0.35\,\sigma(\widetilde{\beta}). \tag{4}$$

Here, σ denotes the logistic sigmoid function.

Thus, Y_{pre} is a bounded refinement of the layout prediction rather than a residual added directly to the wall observation. The hierarchy carries coarse screen context into the layout and fine fields before image readout, so the image received by DSL already integrates multiscale RWIR evidence. The fine and layout fields remain separate RWIR inputs to the Pass-2 context adapter; they are not additional DSL outputs.

Document restoration methods such as DocRes recover geometry and appearance through task-specific visual prompts Zhang et al. (2024). RWIR addresses the different correspondence problem created by spatially mixed wall evidence: its reverse hierarchy carries coarse context into regional and fine representations before damage-guided refinement. This organization yields 0.8540 SSIM on Screen, compared with 0.7490 for DocRes in Table 1.

3.3 DAMAGE-STATE LEARNING

DSL reads Y_{pre} and produces one scalar severity per sample:

$$\begin{aligned} s &= \sigma(\mathcal{E}_{\psi}(Y_{\mathrm{pre}}))\,, \\ s &\in [0,1]^{B\times 1}. \end{aligned} \tag{5}$$

The preliminary reconstruction records the residual damage left after hierarchical correspondence formation. DSL summarizes its overall extent as the bounded scalar s, which controls the strength of the channel, spatial, and repair operations inside the Pass-2 consumer without introducing additional state maps. At the fixed conditioning position, $\mathcal{F}_{\theta}^{u}$ and $\mathcal{F}_{\theta}^{d}$ denote the upstream and downstream portions of the second backbone evaluation. The consumer sequence is

$$\begin{aligned} \omega &= \mathcal{F}_{\theta}^{u}(x), \\ Z_r &= \mathrm{Adapter}(Y_{\mathrm{pre}}, F_{\mathrm{fine}}, F_{\mathrm{layout}}), \\ \omega^r &= \omega + Z_r, \\ \omega^c &= \mathrm{DACA}(\omega^r; s), \\ \omega^p &= \mathrm{DASA}(\omega^c; s), \\ \omega^a &= \mathrm{SDR}(\omega^p; s), \\ y &= \mathcal{F}_{\theta}^{d}(\mathrm{Core}(\omega^a))\,. \end{aligned} \tag{6}$$

The consumer modulates a 192-channel intermediate decoder feature. DACA applies severity-controlled channel attention, DASA applies severity-controlled spatial attention, and SDR performs severity-driven preserve, compensate, and suppress refinement. The copied core block, remaining decoder, and RGB head transform this conditioned feature into the final Pass-2 RGB prediction. Training uses the valid-region masked ℓ_1 loss:

$$J(y,t;M) = \frac{\sum |y-t| \odot M}{C_y \sum M + 10^{-8}}. \tag{7}$$

Here, M is the valid-region mask aligned with the paired target t. Complete tensor definitions, the fixed context boundary, and the expanded masked objective are provided in Appendix B.

4 EXPERIMENTS AND RESULTS

4.1 EXPERIMENTAL SETUP

We evaluate passive hidden-screen reconstruction on four paired datasets grouped by acquisition setup. The blocker-assisted setup is represented by Screen, which contains 2,496 paired samples captured with a T-shaped blocker between the display and camera. Its wall observations combine a diffuse light field with a smooth blocker shadow. The behind-screen wall-reflection setup contains OSHome, OSLock, and Game, with 3,632, 1,399, and 1,250 paired samples, respectively. In this setup, the display faces a white wall and the camera is behind the display while facing the wall, observing the smooth mixed reflection. These datasets emphasize daily desktop and multi-window content, sparse lock-screen content, and dark game content. Separate dataset columns preserve the distinction between the two acquisition setups.

Each dataset is trained, tested, and reported separately with paired weak wall-light inputs and ground-truth screen images. All images are converted to three channels at 512 by 384 resolution while preserving aspect ratio through central padding. For each dataset, all methods use the same image preprocessing, paired training targets, and valid evaluation region. Training uses the paired screen image as supervision; inference uses the weak wall-light image and internally derived features.

We report peak signal-to-noise ratio (PSNR) Huynh-Thu & Ghanbari (2008), structural similarity (SSIM) Wang et al. (2004), and Learned Perceptual Image Patch Similarity (LPIPS) Zhang et al. (2018). Higher PSNR and SSIM are better, while lower LPIPS is better. The baseline pool covers Adaptive Sparse Transformer Zhou et al. (2024), AdaIR Cui et al. (2025), ConvIR Cui et al. (2024), DGUNet Mou et al. (2022), DeepSN-Net Deng et al. (2025), IRNeXt Cui et al. (2023a), MIRNetv2



Figure 3: Qualitative comparison of RSHRL and 13 restoration baselines on four test samples. From top to bottom, the content groups come from the OSHome, OSLock, Screen, and Game test splits. Each content set occupies a two-row group containing the wall observation, ground truth, and reconstruction outputs. The comparison highlights recovery of brightness regions, window boundaries, and screen layout from indirect wall illumination.

Zamir et al. (2023), SFHformer Jiang et al. (2024), SFNet Cui et al. (2023b), Uformer Wang et al. (2022), Ada-Deblur Wu et al. (2023), DEA-Net Chen et al. (2024), DarkIR Feijoo et al. (2025), DehazeFormer Song et al. (2023), DnCNN Zhang et al. (2017), DocRes Zhang et al. (2024), HVI-CIDNet Yan et al. (2025), Retinexformer Cai et al. (2023), pix2pix Isola et al. (2017), and U-Net Ronneberger et al. (2015). They are evaluated with the common resolution, preprocessing, valid region, per-dataset protocol, and metrics described above.

The preliminary readout refines the fixed layout prediction with a bounded fine-feature term whose scale is $0.35\sigma(\widetilde{\beta})$. During training, Pass 1, RWIR, the inverse decoder, and DSL provide a fixed context, while the Pass-2 backbone and severity-conditioned reconstruction path are optimized by the valid-region masked ℓ_1 loss. Evaluation uses the corresponding valid bounding box for each sample.

4.2 Comparison with State of the Art

Figure 3 provides a representative qualitative comparison between our method and the compared baselines on four selected test samples.

Table 1 compares RSHRL with 20 restoration baselines on four paired screen-content datasets, covering blocker-assisted observation and behind-screen wall reflection.

RSHRL leads all 12 dataset–metric comparisons in Table 1. AdaIR adapts restoration to different degradations by mining and modulating low- and high-frequency features Cui et al. (2025). RSHRL rebuilds wall-to-screen correspondence before damage-guided refinement and achieves 25.95 dB PSNR on Screen, compared with 21.71 dB for AdaIR, together with higher SSIM and lower LPIPS. Figure 3 makes this advantage concrete through the recovery of the main content regions and their enclosing window structure. The same favorable ranking across OSHome, OSLock, and Game connects this result to different screen contents and both acquisition setups.

Table 1: Quantitative comparison on OSHome, OSLock, Screen, and Game. Best values are shown in red and second-best values are shown in blue.

Method	Source	OSHome			OSLock			Screen			Game		
		PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
General Purpose Image Restoration Frameworks													
AST	CVPR'24	12.90	0.4960	0.7672	15.07	0.5366	0.6920	19.39	0.7050	0.4668	17.45	0.6453	0.6158
AdaIR	ICLR'25	13.95	0.5338	**0.7324**	15.16	0.5729	0.6455	**21.71**	**0.7621**	**0.3873**	17.43	0.6394	0.5754
ConvIR	TPAMI'24	12.66	0.5168	0.7872	14.91	0.5925	0.6867	20.27	0.7283	0.4495	17.64	0.6515	0.6013
DGUNet	CVPR'22	13.98	0.5416	0.7459	**16.54**	0.6236	0.6514	18.96	0.6960	0.4818	17.88	0.6596	0.5950
DeepSN-Net	TPAMI'25	10.66	0.3751	0.8343	12.83	0.4228	0.7711	10.48	0.2954	0.7787	10.86	0.4440	0.7390
IRNeXt	ICML'23	12.43	0.5070	0.7693	16.13	0.6082	0.6785	19.95	0.7234	0.4542	18.13	0.6681	0.5735
MIRNetv2	TPAMI'23	12.79	0.5342	0.7598	15.21	0.6042	0.6575	18.23	0.6917	0.4923	17.67	0.6559	0.6120
SFHformer	ECCV'24	10.02	0.4480	0.8530	12.46	0.5095	0.7404	8.817	0.3940	0.8066	10.01	0.4620	0.7598
SFNet	ICLR'23	14.02	**0.5421**	0.7484	16.37	**0.6237**	**0.6374**	19.62	0.7164	0.4618	**18.15**	**0.6682**	**0.5699**
Uformer	CVPR'22	**14.34**	0.5356	0.7392	15.14	0.5652	0.6743	19.20	0.7068	0.4639	17.54	0.6470	0.6142
Degradation-Specific / Domain-Specific Restoration Models													
Ada-Deblur	TIP'23	14.05	0.5358	0.7522	16.27	0.6086	0.6608	19.42	0.7067	0.4637	17.75	0.6577	0.5878
DEA-Net	TIP'24	13.69	0.5334	0.7610	15.96	0.6069	0.6742	20.65	0.7500	0.4240	17.72	0.6579	0.5879
DarkIR	CVPR'25	6.570	0.06650	0.8897	8.780	0.09490	0.8040	9.220	0.1749	0.8673	4.460	0.04090	0.9361
DehazeFormer	TIP'23	12.79	0.4704	0.7511	14.87	0.5375	0.6687	19.36	0.6931	0.4794	17.77	0.6494	0.6074
DnCNN	TIP'17	9.019	0.3979	0.7900	11.81	0.4787	0.7221	8.883	0.3466	0.7245	8.737	0.4275	0.7258
DocRes	CVPR'24	12.69	0.5153	0.7378	16.35	0.6094	0.6467	21.14	0.7490	0.3999	17.52	0.6472	0.5795
HVI-CIDNet	CVPR'25	13.84	0.5331	0.7420	15.60	0.5758	0.6714	19.69	0.7178	0.4724	17.77	0.6555	0.5979
Retinexformer	ICCV'23	12.66	0.5078	0.7343	14.60	0.5586	0.6506	19.56	0.6732	0.4554	17.92	0.6556	0.5924
Generic Encoder-Decoder / Image-to-Image Baselines													
pix2pix	CVPR'17	10.42	0.3787	0.7669	12.40	0.5178	0.7303	10.32	0.4816	0.6637	11.14	0.4502	0.6917
U-Net	MICCAI'15	8.720	0.3726	0.8223	12.17	0.4874	0.7618	9.184	0.2401	0.8061	10.28	0.4715	0.7581
RSHRL (Ours)	-	**14.52**	**0.5489**	**0.7239**	**16.74**	**0.6240**	**0.6286**	**25.95**	**0.8540**	**0.2048**	**18.45**	**0.6769**	**0.5339**

4.3 Ablation Study

Table 2 tests the two links in the RSHRL evidence chain using internally defined replacement configurations. LAER (Linear Autoencoder Representation) maps the relay-wall image to layout and fine features and a preliminary reconstruction through a linear convolutional encoder with scale-specific readout heads. SS-CAER (Single-Scale Convolutional Autoencoder Representation) uses a shallow nonlinear single-scale latent with convolutional readout heads for the same interface. Both replace the complete RWIR context-generation path while retaining DSL and the original severity-conditioned consumer. PWRL (Pixel-Wise Regression Learning) replaces DSL and its consumer with a direct bounded pixel-wise residual update, whereas GRL (Global Residual Learning) replaces them with a globally pooled residual-control pathway. Both retain the same RWIR output interface, so the comparison changes the post-RWIR learning mechanism rather than the representation source.

Table 2: Removal and replacement ablation of RWIR and DSL on OSHome, OSLock, Screen, and Game.

Method	**Source**	**Metrics**											
		OSHome			OSLock			Screen			Game		
		PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
RWIR Variants													
w/o RWIR	—	12.80	0.4827	0.7757	15.17	0.5880	0.6614	16.73	0.5994	0.5866	16.38	0.5723	0.6556
RWIR → LAER	Internal	12.82	0.5258	0.7531	15.21	0.5915	0.6549	17.80	0.6818	0.5160	16.99	0.6386	0.6253
RWIR → SS-CAER	Internal	12.88	0.5333	0.7554	15.34	0.5967	0.6504	17.91	0.6840	0.5136	17.06	0.6398	0.6229
DSL Variants													
w/o DSL	—	12.59	0.5362	0.7543	15.24	0.5957	0.6441	16.11	0.6001	0.5995	16.75	0.5754	0.6607
DSL → PWRL	Internal	14.25	0.5478	0.7367	16.42	0.6180	0.6417	20.61	0.7399	0.4387	18.35	0.6727	0.5614
DSL → GRL	Internal	13.16	0.5314	0.7445	16.27	0.6115	0.6449	18.21	0.6949	0.4974	17.31	0.6438	0.6144
RSHRL (Ours)	—	**14.52**	**0.5489**	**0.7239**	**16.74**	**0.6240**	**0.6286**	**25.95**	**0.8540**	**0.2048**	**18.45**	**0.6769**	**0.5339**

The two components make complementary contributions to the reconstruction. On Screen, the full model improves PSNR by 9.22 dB over the RWIR-removed configuration and by 9.84 dB over the DSL-removed configuration. Hierarchical correspondence and damage-guided refinement therefore work together to deliver the complete reconstruction advantage: RWIR supplies the structural basis, and DSL directs the subsequent repair.

Component-level ablations of RWIR and DSL are presented in Appendices C.3.2 and C.4.1, respectively.

4.4 Robust Hidden-Screen Reconstruction

Table 3 presents the complete Gaussian-noise robustness comparison for all 20 evaluated restoration baselines and RSHRL. Each entry is averaged over the transform-free clean condition and five Gaussian-noise levels at 40, 35, 30, 25, and 20 dB.

Table 3: Complete Gaussian-noise robustness comparison for all 20 evaluated restoration baselines and RSHRL on OSHome, OSLock, Screen, and Game. Each value is averaged over the transform-free Clean-00 condition and Gaussian-noise levels of 40, 35, 30, 25, and 20 dB. Best values are shown in red and second-best values are shown in blue.

Method	Source	OSHome			OSLock			Screen			Game		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
General Purpose Image Restoration Frameworks													
AST	CVPR'24	12.89	0.4914	0.7699	15.09	0.5325	0.6941	19.28	0.6702	0.4790	17.45	0.6337	0.6189
AdaIR	ICLR'25	6.993	0.1278	0.9559	9.185	0.1512	0.8941	9.308	0.2786	0.8743	4.839	0.1180	0.9406
ConvIR	TPAMI'24	7.365	0.2266	0.8467	9.342	0.2046	0.7771	9.227	0.3963	0.8072	5.319	0.2278	0.9119
DGUNet	CVPR'22	6.921	0.1485	0.8475	9.038	0.1498	0.7826	9.538	0.3853	0.7925	4.804	0.1286	0.8737
DeepSN-Net	TPAMI'25	6.839	0.1091	0.8575	9.032	0.1381	0.7891	9.552	0.1315	0.8469	4.818	0.07090	0.8893
IRNeXt	ICML'23	7.152	0.1925	0.8492	9.411	0.2243	0.7730	9.460	0.3918	0.8091	4.838	0.1518	0.8891
MIRNetv2	TPAMI'23	6.787	0.1265	0.8494	8.957	0.1444	0.7798	8.485	0.3519	0.8224	4.708	0.1243	0.8952
SFHformer	ECCV'24	10.03	0.4466	0.8556	12.46	0.5083	0.7394	8.812	0.3815	0.8168	10.01	0.4584	0.7632
SFNet	ICLR'23	7.094	0.1644	0.8527	9.586	0.2595	0.7870	9.515	0.3943	0.8229	4.934	0.1746	0.8950
Uformer	CVPR'22	14.33	0.5301	0.7419	15.15	0.5625	0.6760	19.09	0.6762	0.4781	17.54	0.6355	0.6173
Degradation-Specific / Domain-Specific Restoration Models													
Ada-Deblur	TIP'23	6.767	0.1249	0.8529	9.018	0.1521	0.7782	8.344	0.3513	0.8121	4.650	0.1097	0.8767
DEA-Net	TIP'24	6.869	0.07920	0.9214	9.127	0.1086	0.8359	6.499	0.1051	0.8727	4.544	0.04550	1.016
DarkIR	CVPR'25	6.578	0.06670	0.8927	8.788	0.09520	0.8057	9.221	0.1697	0.8825	4.467	0.04070	0.9414
DehazeFormer	TIP'23	7.136	0.05180	0.9506	9.611	0.04730	0.9021	8.834	0.04270	0.8851	5.087	0.03190	0.9564
DnCNN	TIP'17	9.013	0.3933	0.7951	11.76	0.4739	0.7239	8.873	0.3239	0.7346	8.442	0.3946	0.7392
DocRes	CVPR'24	6.477	0.08010	0.9975	8.841	0.1147	0.9480	6.269	0.1353	0.9027	4.239	0.05860	1.009
HVI-CIDNet	CVPR'25	6.833	0.1395	0.8548	9.078	0.1704	0.7850	9.681	0.3940	0.8152	4.801	0.1365	0.8766
Retinexformer	ICCV'23	6.995	0.1461	0.8850	9.064	0.1482	0.7969	9.340	0.3487	0.8154	4.888	0.1383	0.8890
Generic Encoder-Decoder Models													
pix2pix	CVPR'17	10.42	0.3546	0.7820	12.40	0.5171	0.7324	9.979	0.4596	0.6758	10.51	0.3732	0.7439
U-Net	MICCAI'15	8.764	0.3604	0.8401	12.15	0.4845	0.7631	9.209	0.2323	0.8096	10.26	0.4648	0.7603
RSHRL (Ours)	This work	**14.53**	**0.5447**	**0.7250**	**16.77**	**0.6214**	**0.6295**	**25.04**	**0.8221**	**0.2280**	**18.44**	**0.6667**	**0.5341**

The advantage extends across all four content datasets, with the highest average PSNR and SSIM and the lowest average LPIPS in Table 3. Figure 4 shows what these averages represent: principal screen regions and window organization remain recoverable as noise is added to the wall observation.

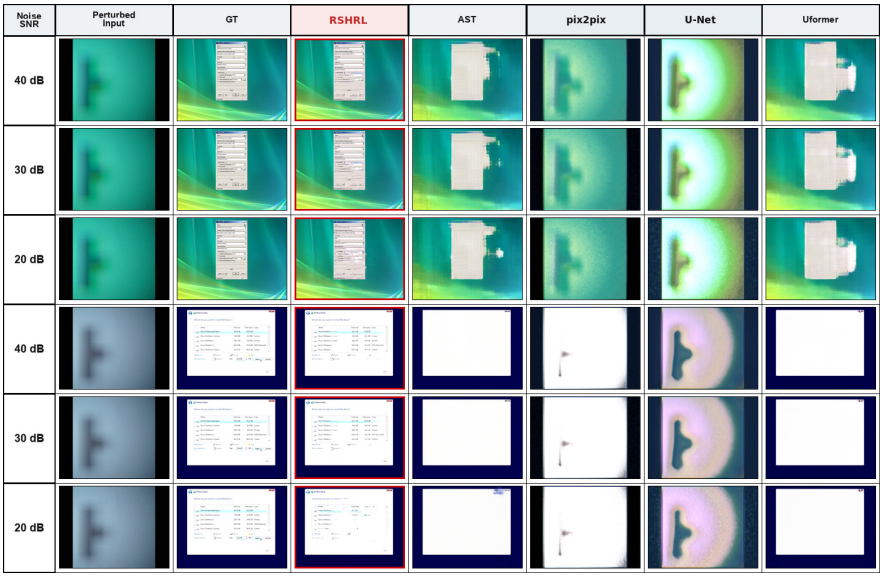

Figure 4: Qualitative comparison on two Screen examples under Gaussian-noise levels of 40, 30, and 20 dB.

Brightness results are presented in Appendix C.6.3, and motion-blur and camera-shake results are presented in Appendix C.6.6.

4.5 Structural Recovery with General-Purpose Restoration Baselines

Table 4 compares RSHRL with RED-Net Mao et al. (2016), NAFNet Chen et al. (2022), SwinIR Liang et al. (2021), and Restormer Zamir et al. (2022). NAFNet develops a simplified restoration architecture, while Restormer efficiently models contextual interactions for high-resolution restoration. For hidden-screen reconstruction, RSHRL achieves the best PSNR, SSIM, and LPIPS across all four datasets; on Screen, its 25.95 dB exceeds the strongest compared value of 14.31 dB. The qualitative results connect this separation to recovery of the screen's dominant regions and window arrangement from the mixed wall response.

Table 4: Quantitative comparison of traditional restoration methods and our task-aligned model on OSHome, OSLock, Screen, and Game. Best values are shown in red and second-best values are shown in blue.

Method	Source	OSHome			OSLock			Screen			Game		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
General-Purpose Restoration Networks													
RED-Net	NIPS'16	10.42	0.3994	0.7897	12.29	0.4749	0.7099	10.58	0.4394	0.6982	9.894	0.3689	0.7593
NAFNet	ECCV'22	12.38	0.5023	0.7419	14.48	0.5552	0.6785	14.31	0.5198	0.6468	14.26	0.5111	0.6813
SwinIR	ICCVW'21	10.61	0.4097	0.7902	11.81	0.5074	0.7098	10.71	0.4624	0.6902	10.21	0.4918	0.7323
Restormer	CVPR'22	12.77	0.5051	0.7505	15.26	0.5579	0.6752	12.78	0.3892	0.7261	16.76	0.6073	0.6292
Hidden-Screen Reconstruction													
RSHRL (Ours)	This work	**14.52**	**0.5489**	**0.7239**	**16.74**	**0.6240**	**0.6286**	**25.95**	**0.8540**	**0.2048**	**18.45**	**0.6769**	**0.5339**

RSHRL achieves the best listed result in every dataset and metric combination reported in Table 4. On Screen, it gains 11.64 dB PSNR and 0.3342 SSIM over the strongest general-purpose baseline while reducing LPIPS by 0.4420. The separation corresponds to recovery of the screen's dominant regions and window arrangement from the mixed wall response.

5 Conclusion

We demonstrate the initial feasibility of recovering hidden screen content with an ordinary camera from light reflected onto a visible wall. RSHRL addresses the loss of spatial correspondence through hierarchical relay representation and directs subsequent refinement through learned damage conditions. Comparisons across four screen-content datasets show consistent advantages over the evaluated restoration methods, while removal and replacement studies establish the complementary contributions of RWIR and DSL. Across the acquisition setups and observation changes, the reconstructions recover principal brightness distributions, text-bearing regions, icon contours, window boundaries, and overall interface layouts. Taken together, these results indicate that the wall response preserves recoverable structural evidence even though direct pixel-wise correspondence with the screen is disrupted by reflection, spatial spreading, attenuation, and observation degradation. The robustness studies further connect this evidence to changing observation quality: as the relay signal is perturbed, the reconstruction continues to retain the dominant screen regions and their spatial organization across the evaluated conditions. This behavior is consistent with the division of roles inside RSHRL. RWIR first reorganizes distributed wall evidence into screen-oriented fine, regional, and coarse structure, while DSL uses the residual state of the preliminary reconstruction to condition the second-pass repair. The experiments therefore connect the physical formation of indirect wall observations, learned relay-to-screen correspondence, and residual-damage refinement within one reconstruction pathway. Building on earlier optical screen recovery and computational periscopy Kuhn (2002); Saunders et al. (2019); Geng et al. (2022), this study advances recovery of everyday screen organization from indirect light. Wall-reflected display light therefore provides usable structural evidence for reconstructing content hidden from direct view.

REFERENCES

Miika Aittala, Prafull Sharma, Lukas Murmann, Adam B. Yedidia, Gregory W. Wornell, William T. Freeman, and Frédo Durand. Computational mirrors: Blind inverse light transport by deep matrix factorization. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Katherine L. Bouman, Vickie Ye, Adam B. Yedidia, Fredo Durand, Gregory W. Wornell, Antonio Torralba, and William T. Freeman. Turning corners into cameras: Principles and methods. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2289–2297, 2017. doi: 10.1109/ICCV.2017.249.

Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage Retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12504–12513, 2023. doi: 10.1109/ICCV51070.2023.01149.

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.

Zixuan Chen, Zewei He, and Zhe-Ming Lu. DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention. *IEEE Transactions on Image Processing*, 33: 1002–1015, 2024. doi: 10.1109/TIP.2024.3354108.

Yuning Cui, Wenqi Ren, Sining Yang, Xiaochun Cao, and Alois Knoll. IRNeXt: Rethinking convolutional network design for image restoration. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 6545–6564. PMLR, 2023a. URL https://proceedings.mlr.press/v202/cui23d.html.

Yuning Cui, Yi Tao, Zhenshan Bing, Wenqi Ren, Xinwei Gao, Xiaochun Cao, Kai Huang, and Alois Knoll. Selective frequency network for image restoration. In *The Eleventh International Conference on Learning Representations*. OpenReview.net, 2023b. URL https://openreview.net/forum?id=tyZ1ChGZIKO.

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, 2024. doi: 10.1109/TPAMI.2024.3419007.

Yuning Cui, Syed Waqas Zamir, Salman 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*. OpenReview.net, 2025. URL https://openreview.net/forum?id=M5t0WvjfCg.

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.

Xin Deng, Chenxiao Zhang, Lai Jiang, Jingyuan Xia, and Mai Xu. DeepSN-Net: Deep semi-smooth Newton driven network for blind image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(4):2632–2646, 2025. doi: 10.1109/TPAMI.2024.3525089.

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.

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.

Quan Huynh-Thu and Mohammed Ghanbari. Scope of validity of PSNR in image/video quality assessment. *Electronics Letters*, 44(13):800–801, 2008. doi: 10.1049/el:20080522.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5967–5976, 2017. doi: 10.1109/CVPR.2017.632.

Xingyu Jiang, Xiuhui Zhang, Ning Gao, and Yue Deng. When fast Fourier transform meets transformer for image restoration. In *Computer Vision – ECCV 2024*, pp. 381–402. Springer Nature Switzerland, 2024. doi: 10.1007/978-3-031-72995-9_22.

Markus G. Kuhn. Optical time-domain eavesdropping risks of CRT displays. In *Proceedings 2002 IEEE Symposium on Security and Privacy*, pp. 3–18. IEEE Computer Society, 2002. doi: 10.1109/SECPRI.2002.1004358.

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image restoration using Swin Transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pp. 1833–1844, 2021. doi: 10.1109/ICCVW54120.2021.00210.

Xuezhe Ma, Chunting Zhou, Xiang Kong, Junxian He, Liangke Gui, Graham Neubig, Jonathan May, and Luke Zettlemoyer. MEGA: Moving average equipped gated attention. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=qNLe3iq2El.

Xiaojiao Mao, Chunhua Shen, and Yu-Bin Yang. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper_files/paper/2016/hash/0ed9422357395a0d4879191c66f4faa2-Abstract.html.

Chong Mou, Qian Wang, and Jian Zhang. Deep generalized unfolding networks for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17399–17410, 2022. doi: 10.1109/CVPR52688.2022.01688.

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(7740):472–475, 2019. doi: 10.1038/s41586-018-0868-6.

Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 32:1927–1941, 2023. doi: 10.1109/TIP.2023.3256763.

Yehui Tang, Kai Han, Jianyuan Guo, Chang Xu, Chao Xu, and Yunhe Wang. GhostNetV2: Enhance cheap operation with long-range attention. In *Advances in Neural Information Processing Systems*, volume 35, pp. 9969–9982, 2022.

Antonio Torralba and William T. Freeman. Accidental pinhole and pinspeck cameras: Revealing the scene outside the picture. *International Journal of Computer Vision*, 110(2):92–112, 2014. doi: 10.1007/s11263-014-0697-5.

Fu-Jen Tsai, Yan-Tsung Peng, Yen-Yu Lin, Chung-Chi Tsai, and Chia-Wen Lin. Stripformer: Strip transformer for fast image deblurring. In *Computer Vision – ECCV 2022*, pp. 146–162. Springer Nature Switzerland, 2022. doi: 10.1007/978-3-031-19800-7_9.

Hainuo Wang, Qiming Hu, and Xiaojie Guo. MODEM: A Morton-Order degradation estimation mechanism for adverse weather image recovery. In *Advances in Neural Information Processing Systems*, volume 38, pp. 47410–47439. Curran Associates, Inc., 2025. doi: 10.52202/085713-1581. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/43d2b7fbee8431f7cef0d0afed51c691-Paper-Conference.pdf.

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.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004. doi: 10.1109/TIP.2003.819861.

Qiucheng Wu, Yifan Jiang, Junru Wu, Victor Kulikov, Vidit Goel, Nikita Orlov, Humphrey Shi, Zhangyang Wang, and Shiyu Chang. Broad spectrum image deblurring via an adaptive supernetwork. *IEEE Transactions on Image Processing*, 32:5270–5282, 2023. doi: 10.1109/TIP.2023.3312912.

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.

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. 14821–14831, 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. 5728–5739, 2022. doi: 10.1109/CVPR52688.2022.00564.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for fast image restoration and enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2):1934–1948, 2023. doi: 10.1109/TPAMI.2022.3167175.

Jiaxin Zhang, Dezhi Peng, Chongyu Liu, Peirong Zhang, and Lianwen Jin. DocRes: A generalist model toward unifying document image restoration tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15654–15664, 2024. doi: 10.1109/CVPR52733.2024.01482.

Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017. doi: 10.1109/TIP.2017.2662206.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018. doi: 10.1109/CVPR.2018.00068.

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

Weak Light as Structural Evidence: Appendix for Relay-State Hierarchical Representation Learning for Hidden-Screen Reconstruction

A SUPPLEMENTARY SCENE INFORMATION

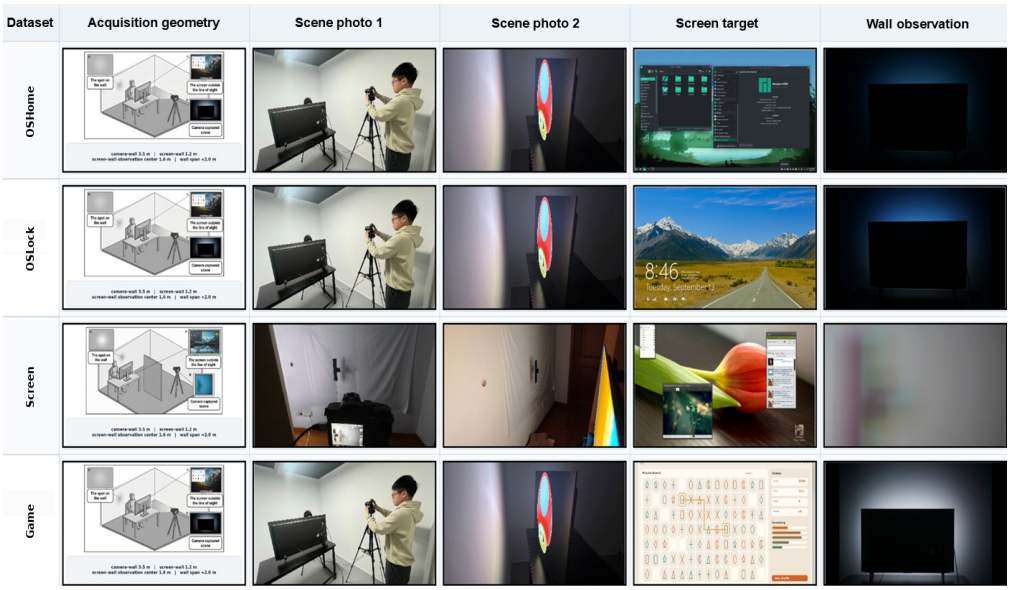

Figure A1: Acquisition setups for the four paired hidden-screen datasets. Each row connects the acquisition geometry and scene photographs with a screen target and its wall observation.

Figure A1 makes the source of the reconstruction evidence concrete. In every setting, the camera receives a weak wall-light pattern after screen emission has undergone reflection, spatial spreading, overlap, and attenuation. Text strokes, icon contours, window boundaries, and region relations therefore appear as dispersed changes in the wall response. The Screen setup adds the smooth transition created by the T-shaped blocker, while the other three datasets capture the mixed reflection behind the display. Together, the four scenes establish the shared challenge that drives RSHRL: screen structure must be recovered after its direct spatial correspondence has been disrupted on the relay wall.

The four content types then test different parts of the same recovery chain. Dense desktop and multiwindow layouts require RWIR to reconnect fine strokes with regional window relations; sparse lock-screen elements require it to preserve weak isolated structures; dark game observations require it to use the surviving regional and coarse context. RWIR organizes these cues into fine, regional, and coarse screen-oriented fields and produces Y_{pre}. DSL subsequently estimates a scalar severity from the preliminary reconstruction, and the second-pass consumer uses this condition through its internal DACA, DASA, and SDR operations to repair faded, broken, missing, or spurious structures.

The formal evidence covers two acquisition setups. In the blocker-assisted setup used by Screen, a T-shaped blocker prevents a direct view of the display while the camera observes a white wall. Display emission forms a blurred wall response, and the blocked path adds a smooth T-shaped transition inside that response. This smooth transition can overlap changes induced by text, icons, and window boundaries. The dataset contains 2,496 paired observations and targets, with desktop pages, settings, file managers, dialog boxes, and dense interface layouts. Its role is to test the combined difficulty of smooth blocker transitions, regional layout, and fine structure.

The behind-screen wall-reflection setup is used by OSHome, OSLock, and Game. The display faces a white wall, and the camera behind the display sees the mixed wall reflection. The datasets retain distinct content roles. OSHome contains 3,632 paired desktop and multiwindow samples; OSLock contains 1,399 paired lock-screen and wallpaper samples with sparse interface elements; Game

contains 1,250 paired game samples whose wall observations are often globally dark. Across the documented acquisition geometry, the camera-to-wall distance is 3.5 m, the perpendicular screen-to-wall distance is 1.2 m, the distance from the screen to the wall observation center is 1.6 m, and the horizontal relay-wall span is approximately 2.0 m. Separate reporting preserves both acquisition setup and content difficulty.

All formal images are represented as three channels at 512 by 384 pixels after aspect-ratio-preserving resizing and central padding. Training, testing, and reporting are performed separately for each dataset, and the metrics are evaluated inside a common valid region. Training uses paired targets; inference uses the weak wall-light image and internally derived features. The first backbone evaluation provides the multiresolution feature fields used by RWIR. RWIR and the inverse decoder produce Y_{pre}, F_{fine}, and F_{layout}; DSL estimates the scalar severity s from Y_{pre}; and the second-pass consumer modulates an intermediate decoder feature before the remaining decoder and RGB head produce the final reconstruction.

B Extended Method Derivation

This appendix expands the compact formulation in Section 3. It uses the same wall observation x, three-level RWIR hierarchy Z_2, Z_3, Z_4, inverse-decoder fields $F_{\mathrm{layout}}, F_{\mathrm{fine}}$, preliminary and final reconstructions Y_{pre}, y, scalar severity s, and objective J as the main text.

B.1 Relay Observation and Ill-Posedness

For each color coordinate c, the screen-to-wall-to-camera model in Equation 1 is induced by the nonnegative kernel

$$\kappa_c(q,s) = \int_{\Omega_w} h_c(q,r)\rho_c(r)V(s,r)g(s,r)\,\mathrm{d}r, \tag{A1}$$

which aggregates propagation and reflection through the relay wall before camera sampling. Here, $t_c \in L^2(\Omega_s;\mathbb{R}_{\geq 0})$ is the hidden screen radiance, $x_c \in L^2(\Omega_o)$ is observed, $h_c : \Omega_o \times \Omega_w \to \mathbb{R}_{\geq 0}$, $\rho_c : \Omega_w \to \mathbb{R}_{\geq 0}$, $V : \Omega_s \times \Omega_w \to [0,1]$, and $g : \Omega_s \times \Omega_w \to \mathbb{R}_{\geq 0}$. The latent factors explain spatial spreading, mixing, attenuation, and occlusion; RSHRL learns from the observed wall-light field.

Fixing c and writing $\mathcal{X} = \mathcal{X}_c$, finite kernel energy makes $\mathcal{X}$ compact on bounded domains. For its singular system (σ_n, u_n, v_n) and an explanatory cutoff $\tau > 0$,

$$\begin{aligned} \mathcal{X}u_n &= \sigma_n v_n, \qquad \sigma_1 \geq \sigma_2 \geq \cdots \downarrow 0, \\ \widehat{t}_\tau &:= \mathcal{X}_\tau^\dagger(x_c - \zeta_c) = \sum_{\sigma_n \geq \tau} \frac{\langle x_c - \zeta_c, v_n\rangle}{\sigma_n} u_n, \\ \widehat{t}_\tau - t_c &= -\Pi_0 t_c - \sum_{0<\sigma_n<\tau} \langle t_c, u_n\rangle u_n + \sum_{\sigma_n \geq \tau} \frac{\langle \eta_c, v_n\rangle}{\sigma_n} u_n. \end{aligned} \tag{A2}$$

The three terms identify unavailable null-space content, cutoff bias, and noise amplification as the principal sources of ambiguity. This decomposition motivates the RSHRL progression from scale-restricted evidence organization to severity-conditioned restoration. In the proposed formulation, $\mathcal{X}_\tau^\dagger$ characterizes the inverse problem, while RWIR learns a hierarchical representation of the observed relay evidence and DSL supplies a bounded severity condition to the subsequent restoration.

For hidden-screen reconstruction, the same decomposition explains why a direct local restoration path easily loses the original screen arrangement. Weak singular components contain precisely the fine strokes and boundaries most strongly attenuated by wall propagation, while the observation noise is amplified when those components are recovered independently. RWIR addresses this conflict by combining fine evidence with regional layout and coarse interface context before image readout. The resulting Y_{pre} recovers the main screen organization and creates a stable basis for the next problem: heterogeneous residual damage across different structures.

B.2 TWO-PASS FORMULATION

On $\Omega = \{1, \dots, H\} \times \{1, \dots, W\}$, the wall observation satisfies $x \in [0,1]^{B\times 3\times 384\times 512}$. Equation 2 evaluates the shared backbone twice. Pass 1 produces the early and coarse backbone features, the RWIR hierarchy, F_{layout}, F_{fine}, and Y_{pre}. DSL maps Y_{pre} to the scalar severity $s \in [0,1]$. The tuple $(Y_{\text{pre}}, F_{\text{fine}}, F_{\text{layout}}, s)$ forms the fixed context for Pass 2: the three RWIR outputs feed the adapter, and the scalar severity conditions the internal repair operations.

B.3 FULL RWIR HIERARCHY

Let $\Omega_0 = \Omega$, $\Omega_f = \{1, \dots, H_f\} \times \{1, \dots, W_f\}$, $\Omega_m = \{1, \dots, H_m\} \times \{1, \dots, W_m\}$, and $\Omega_c = \{1, \dots, H_c\} \times \{1, \dots, W_c\}$. For a field f, deterministic coarse-graining is

$$(\Pi_j^x f)(\upsilon) = \frac{1}{|\Omega_j(\upsilon)|} \sum_{\xi \in \Omega_j(\upsilon)} f(\xi), \qquad j \in \{f, m, c\}. \tag{A3}$$

The indices ξ and υ belong to the preceding and next grids, respectively. Repeated application increases the support summarized by each coarse position, providing evidence fields at the spatial scales used by the subsequent RWIR hierarchy.

The evidence recursion from Equation 3 can be expanded as

$$\begin{aligned} e_0 &= \Phi_0(x, F_{\text{early}}), \\ e_f &= \Phi_f(\Pi_f^x e_0), \\ e_m &= \Phi_m(\Pi_m^x e_f), \\ e_c &= \Phi_c(\Pi_c^x e_m, F_{\text{up}}) \\ &= \Phi_c\big(\Pi_c^x \Phi_m\big(\Pi_m^x \Phi_f\big(\Pi_f^x \Phi_0(x, F_{\text{early}})\big)\big), F_{\text{up}}\big). \end{aligned} \tag{A4}$$

For $j \in \{0, f, m, c\}$, $e_j : \Omega_j \to \mathbb{R}^{C_j^e}$. Here, F_{early} is the fixed full-resolution observation-embedding feature fused with the wall observation in e_0, and F_{up} is the fixed coarse decoder feature entering when e_c is formed. Both tensors are generated internally during the first pass through the shared backbone and provide complementary spatial scales to RWIR. The subsequent hierarchy and inverse decoder produce F_{fine}, F_{layout}, and Y_{pre}.

The reverse hierarchy is

$$\begin{aligned} \widetilde{r}_c &= \Psi_c(e_c), & r_c &= \widetilde{r}_c + \Gamma_c(\widetilde{r}_c), \\ \widetilde{r}_m &= \Psi_{c\to m}(\Pi_{c\to m} r_c) + \Psi_m(e_m), & r_m &= \widetilde{r}_m + \Gamma_m(\widetilde{r}_m), \\ \widetilde{r}_f &= \Psi_{m\to f}(\Pi_{m\to f} r_m) + \Psi_f(e_f), & r_f &= \widetilde{r}_f + \Gamma_f(\widetilde{r}_f). \end{aligned} \tag{A5}$$

For $j \in \{f, m, c\}$, $\widetilde{r}_j, r_j : \Omega_j \to \mathbb{R}^{C_j^r}$. The Π operators are deterministic grid alignments; $\Psi_c, \Psi_{c\to m}, \Psi_m, \Psi_{m\to f}, \Psi_f$ are learned coordinate maps; and $\Gamma_c, \Gamma_m, \Gamma_f$ are learned corrections. Substitution gives the complete fine-field dependency:

$$\begin{aligned} r_f = (I + \Gamma_f)\Big[\Psi_{m\to f}\Big(\Pi_{m\to f}(I + \Gamma_m)\big[\Psi_{c\to m}\big(\Pi_{c\to m}(I + \Gamma_c)\Psi_c(e_c)\big) + \Psi_m(e_m)\big]\Big) \\ + \Psi_f(e_f)\Big]. \end{aligned} \tag{A6}$$

Equation A6 shows the explicit route from e_c through r_c and r_m to r_f, while e_m and e_f enter at their corresponding grids. This dependency implements the central RWIR role: coarse contextual evidence is progressively aligned with regional and fine evidence to form a screen-oriented reconstruction field.

This coarse-to-fine return gives each scale a clear task in rebuilding wall-to-screen correspondence. The coarse branch anchors the overall interface arrangement, the middle branch reconnects text areas,

icons, and windows, and the fine branch restores the strokes and boundaries that distinguish neighboring structures. Their progressive alignment turns dispersed wall evidence into a representation that can be decoded in screen space. Once this hierarchy forms Y_{pre}, the remaining errors appear as local fading, breaks, omissions, and spurious edges, which motivates the DSL construction that follows.

The preliminary readout follows Equation 4. It upsamples the fixed fine feature, predicts a bounded refinement through H_f, adds this refinement to the fixed layout prediction, and clips the result to the valid image range. The learnable factor satisfies $0 < \beta < 0.35$. The resulting Y_{pre} forms the explicit DSL input; the hierarchy fields remain internal to the RWIR reconstruction path.

B.4 Scalar Damage-Severity Estimation

The estimator maps the preliminary reconstruction to one bounded value per sample:

$$s = \sigma(\mathcal{E}_\psi(Y_{\mathrm{pre}})), \qquad s \in [0,1]^{B\times 1}. \tag{A7}$$

The scalar s summarizes the residual reconstruction difficulty after RWIR has formed Y_{pre}. It is the sole DSL state passed to the final reconstruction path:

$$\mathcal{C} = \mathrm{stopgrad}\,(Y_{\mathrm{pre}}, F_{\mathrm{fine}}, F_{\mathrm{layout}}, s)\,. \tag{A8}$$

Pass 1, RWIR, the inverse decoder, and DSL form the fixed context $\mathcal{C}$. The stop-gradient boundary holds these context-producing branches fixed while the final reconstruction loss optimizes the active Pass-2 path.

B.5 Severity-Conditioned Second Pass

Let $\omega \in \mathbb{R}^{B\times 192\times 48\times 64}$ be the decoder feature at the fixed conditioning point. The RWIR adapter first aligns the preliminary reconstruction, fine feature, and layout feature:

$$Z_r = \mathrm{Adapter}\,(Y_{\mathrm{pre}}, F_{\mathrm{fine}}, F_{\mathrm{layout}})\,, \qquad Z_r \in \mathbb{R}^{B\times 192\times 48\times 64}. \tag{A9}$$

The consumer then follows the sequence

$$\begin{aligned} \omega^r &= \omega + Z_r, \\ \omega^c &= \mathrm{DACA}(\omega^r; s), \\ \omega^p &= \mathrm{DASA}(\omega^c; s), \\ \omega^a &= \mathrm{SDR}(\omega^p; s), \\ \omega^{\mathrm{core}} &= \mathrm{Core}(\omega^a), \\ y &= \mathcal{F}^d_\theta(\omega^{\mathrm{core}}). \end{aligned} \tag{A10}$$

DACA performs severity-controlled channel attention, DASA performs severity-controlled spatial attention, and SDR combines preserve, compensate, and suppress branches under the same scalar condition. These are internal operations of the Pass-2 consumer rather than separate DSL outputs. The consumer returns an intermediate 192-channel feature. The downstream decoder and RGB head produce the final reconstruction.

B.6 Expanded Training Objective

Let $M_\iota(p) \in \{0,1\}$ be the valid-region mask for sample ι. Expanding Equation 7 gives

$$J(y,t;M) = \frac{\sum_{\iota=1}^{B}\sum_{c=1}^{C_y}\sum_{p\in\Omega} M_\iota(p)\,|y_{\iota,c}(p) - t_{\iota,c}(p)|}{C_y \sum_{\iota=1}^{B}\sum_{p\in\Omega} M_\iota(p) + 10^{-8}}. \tag{A11}$$

Training is supervised by the valid-region masked ℓ_1 error on the Pass-2 reconstruction, while evaluation metrics are computed within cropped valid bounding boxes. RWIR and the inverse decoder produce $Y_{\text{pre}}, F_{\text{fine}}, F_{\text{layout}}$; DSL produces the scalar severity s from Y_{pre}; and the fixed context conditions the second pass.

C Additional Experimental Results

The experiments form a continuous evidence chain. We first establish the paired wall-observation task, then examine how RWIR reconstructs relay-to-screen correspondence and how DSL repairs heterogeneous residual damage. Joint ablations connect the two stages, while acquisition, degradation, and data-limited studies test the resulting reconstruction pathway across the observation conditions.

C.1 Task and Acquisition Evidence

The paired acquisitions establish the physical evidence available to RSHRL. Across the four datasets, each weak wall observation is linked to its displayed screen target, allowing the following experiments to evaluate recovery of brightness regions, text areas, icon contours, window boundaries, and overall layout from the screen-to-wall-to-camera indirect path.

C.1.1 Acquisition Pairing and Scene Evidence

The paired acquisition material makes the information source and the learning target explicit. Each observation is a wall response produced by an unseen display, and each target is the screen image displayed at the same time. Figure A2 shows two test pairs for each formal dataset. The four datasets cover two acquisition conditions: Screen uses the T-shaped blocker, whereas OSHome, OSLock, and Game share the behind-screen wall-reflection configuration.

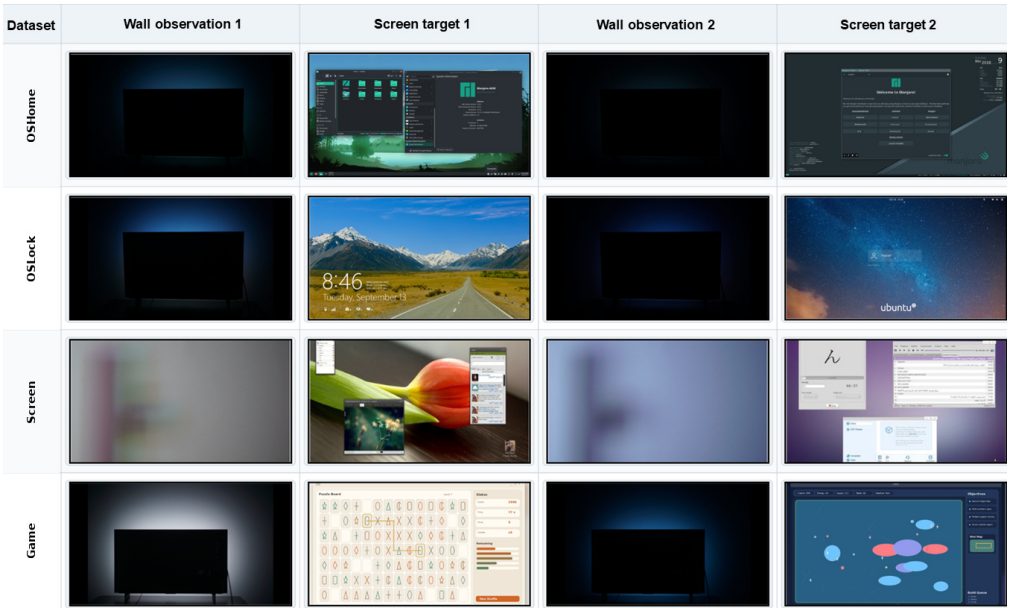

Figure A2: Paired acquisition examples from the four formal datasets. Each row shows two test pairs, with the relay-wall observation followed by the simultaneously displayed screen target. The examples cover the shared behind-screen wall-reflection setup and the T-shaped-blocker setup.

Each pair connects the wall-light observation to the image displayed on the hidden screen at the same time. The contrast between the diffuse observation and the structured target makes the task concrete: recovering screen brightness, regions, and layout from their indirect optical response.

Across the two acquisition setups, diffuse wall patterns correspond to targets with distinct windows, icons, and large display regions. The pairs therefore show the structural change that the reconstruction must reverse rather than treating the input as an ordinary degraded screen image.

Table 5: Dataset scale, acquisition setup, and paired input–target form for formal evaluation. Counts denote the train/validation/test split; all samples are aspect-ratio preserved and centrally padded to 512×384.

Attribute	OSHome	OSLock	Screen	Game
Acquisition setup	Behind-screen	Behind-screen	T-shaped blocker	Behind-screen
Screen content	Desktop/UI	Lock screen	Desktop/Web/UI	Game/HUD
Train	2615	1008	2121	900
Val	291	112	124	100
Test	726	279	251	250
Total	3632	1399	2496	1250
Raw wall observation	1920×1080	1920×1080	256×256	1920×1080
Raw screen target	Variable	Variable	500×375	Three sizes[†]
Model size	512×384	512×384	512×384	512×384

[†] The native Game targets have resolutions 2048×1280, 2304×1440, and 2560×1440.

The four content sets probe complementary aspects of hidden-screen reconstruction. Desktop and multiwindow scenes emphasize window boundaries and regional organization; lock-screen content emphasizes isolated interface elements; and game content tests recovery of large display regions from darker wall observations. Together, they examine whether indirect light can support screen organization across different visual contents and two acquisition setups.

C.2 Recovering Screen Structure from Wall Observations

Diffuse wall propagation spreads each screen structure across neighboring observation regions. The traditional-restoration comparison tests mechanisms designed for locally aligned degradation against the screen-space organization required by this task. Table 4 and the qualitative evidence below establish the starting point for the representation study.

C.2.1 Traditional-Restoration Qualitative Comparison

Appendix Figure A3 visualizes the traditional-method comparison on one selected test example per dataset, ordered OSHome, OSLock, Screen, and Game from top to bottom.

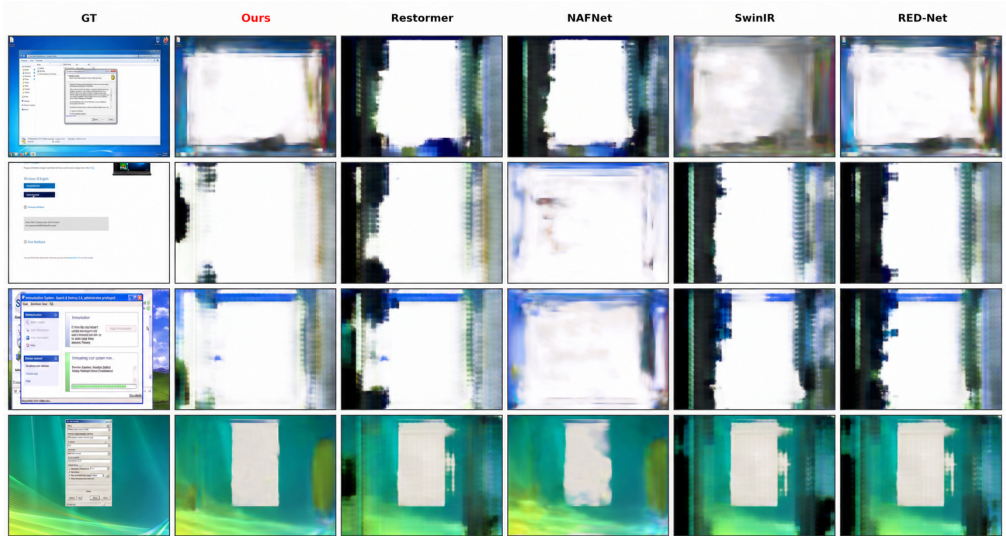

Figure A3: Representative hidden-screen reconstructions from the OSHome, OSLock, Screen, and Game test splits, ordered from top to bottom. Columns show the ground truth, RSHRL, Restormer, NAFNet, SwinIR, and RED-Net.

The examples show that RSHRL recovers dominant brightness regions, window positions, and visible interface boundaries more clearly than the compared restoration outputs. These recovered regions and window arrangements raise the next question: which representation and refinement choices make this reconstruction possible?

C.3 RWIR: Reconstructing Relay-to-Screen Correspondence

This section traces the RWIR evidence chain. Full RWIR is compared with its removal, LAER and SS-CAER replacements, and module-level alternatives based on the DFC mechanism of GhostNetV2 Tang et al. (2022), MEGA Ma et al. (2023), and the intra- and inter-strip operations of Stripformer Tsai et al. (2022). These mechanisms are instantiated at the RWIR representation slot within the shared restoration network, providing controlled comparisons under a common backbone. Internal ablations identify the encoder, structure projector, and inverse decoder contributions, while stage evolution and regional intervention expose how distributed wall evidence becomes a screen-oriented reconstruction.

C.3.1 Representation-Level Comparison

Table 6: Comparison of module-level representation choices with the same backbone. Best values are shown in red and second-best values are shown in blue.

Representation	Source	OSHome			OSLock			Screen			Game		
		PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
GhostNetV2 DFC-Ghost	NeurIPS'22	10.97	**0.4259**	0.8390	11.09	0.3563	0.7987	9.731	0.3984	**0.7846**	10.74	0.3883	0.8235
MEGA	ICLR'23	10.80	0.4204	**0.8330**	10.71	0.3777	**0.7929**	9.663	**0.4227**	0.8211	10.77	**0.3949**	0.8148
Stripformer Intra-Strip	ECCV'22	11.18	0.4189	0.8338	11.47	**0.4153**	0.8297	9.808	0.3309	0.8334	**11.26**	0.3610	0.7986
Stripformer Inter-Strip	ECCV'22	**11.23**	0.4173	0.8332	**11.48**	0.4131	0.8318	**9.830**	0.3398	0.8305	11.22	0.3524	**0.7971**
Full RWIR (Ours)	-	**11.65**	**0.4524**	**0.8001**	**11.73**	**0.4575**	**0.7577**	**10.94**	**0.4467**	**0.7429**	**12.80**	**0.4100**	**0.7351**

Analysis. GhostNetV2 combines local features with long-range DFC attention Tang et al. (2022), MEGA equips gated attention with an exponential moving average Ma et al. (2023), and Stripformer models directional dependencies through intra- and inter-strip attention Tsai et al. (2022). RWIR's wall-to-screen hierarchy leads all 12 dataset–metric comparisons in Table 6, reaching 12.80 dB PSNR on Game compared with 11.26 dB for the strongest competing representation on this metric, the Stripformer Intra-Strip variant. Its coarse-to-fine synthesis carries contextual evidence into regional and fine representations before reconstruction. Table 7 next examines the encoder, structural projection, and decoder within this hierarchy.

C.3.2 Internal Ablation of RWIR

Table 7 compares six internally constructed RWIR replacement configurations. SCE (Shallow Convolutional Encoder) fuses shallow wall features with the first-pass observation embedding and injects the coarse decoder feature at the coarsest level, whereas SSE (Single-Scale Encoder) constructs all hierarchy levels from one wall-only latent. LSP (Linear Structure Projector) uses 1×1 projections and bounded RGB residual heads, whereas LCP (Local Convolutional Projector) uses local 3×3 convolutional heads. BUD (Bilinear Upsampling Decoder) uses a simple cross-scale transmission chain, bilinear upsampling, and 1×1 RGB readouts, whereas SCD (Shallow Convolutional Decoder) uses the same transmission interface followed by a shallow three-layer convolutional decoder. Together, these alternatives probe the encoder, structure-projection, and decoder functions of RWIR.

Table 7: Internal ablation of RWIR components across four datasets.

Method	Metrics											
	OSHome			OSLock			Screen			Game		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Relay-Wall Encoder												
w/o RWIR-RWRE	13.17	0.5227	0.7679	15.57	0.5984	0.6641	15.92	0.6225	0.5750	15.88	0.6160	0.6719
RWIR-RWRE → SCE	13.42	0.5281	0.7787	15.48	0.5957	0.6726	16.17	0.6236	0.5767	16.23	0.6202	0.6643
RWIR-RWRE → SSE	13.16	0.5194	0.7787	15.51	0.5965	0.6646	16.03	0.6248	0.5704	16.30	0.6229	0.6560
Relay-to-Screen Projection												
w/o RWIR-RSSP	13.20	0.5225	0.7743	15.57	0.5984	0.6642	16.06	0.6196	0.5748	16.01	0.6154	0.6733
RWIR-RSSP → LSP	13.06	0.5204	0.7696	15.59	0.5997	0.6611	16.11	0.6182	0.5888	16.23	0.6141	0.6621
RWIR-RSSP → LCP	13.23	0.5292	0.7772	15.47	0.5914	0.6591	16.02	0.6242	0.5778	15.78	0.6095	0.6827
Screen Decoder												
w/o RWIR-RWID	13.22	0.5227	0.7708	15.57	0.5984	0.6641	16.03	0.6213	0.5735	16.03	0.6168	0.6733
RWIR-RWID → BUD	13.26	0.5263	0.7711	15.43	0.5926	0.6663	16.11	0.6267	0.5683	15.82	0.6157	0.6772
RWIR-RWID → SCD	13.33	0.5247	0.7781	15.69	0.5979	0.6688	16.04	0.6206	0.5686	16.25	0.6171	0.6662
RSHRL (Ours)	**14.52**	**0.5489**	**0.7239**	**16.74**	**0.6240**	**0.6286**	**25.95**	**0.8540**	**0.2048**	**18.45**	**0.6769**	**0.5339**

Across the nine modified rows on Screen, PSNR spans 15.92 to 16.17, SSIM spans 0.6182 to 0.6267, and LPIPS spans 0.5683 to 0.5888. Against the strongest modified result for each metric, the full row is higher by 9.778 PSNR and 0.2273 SSIM and lower by 0.3635 LPIPS. The different modifications exhibit complementary metric behavior across the four datasets, while the complete RWIR configuration consistently provides the strongest joint reconstruction performance. The relay-to-screen structural-projector rows are reported as separate alternative configurations for examining different structure-mapping choices, complementing the encoder and inverse-decoder variants evaluated within the active hierarchical pathway.

The complete hierarchy produces the strongest final reconstructions among the internal variants. The encoder gathers evidence across scales, the projection organizes its relationship to the screen, and the decoder brings coarse and regional information into finer reconstruction features. Their combination preserves the full advantage of RWIR and supplies a structured basis for subsequent damage-guided refinement.

C.3.3 Formation of the Screen-Oriented Representation

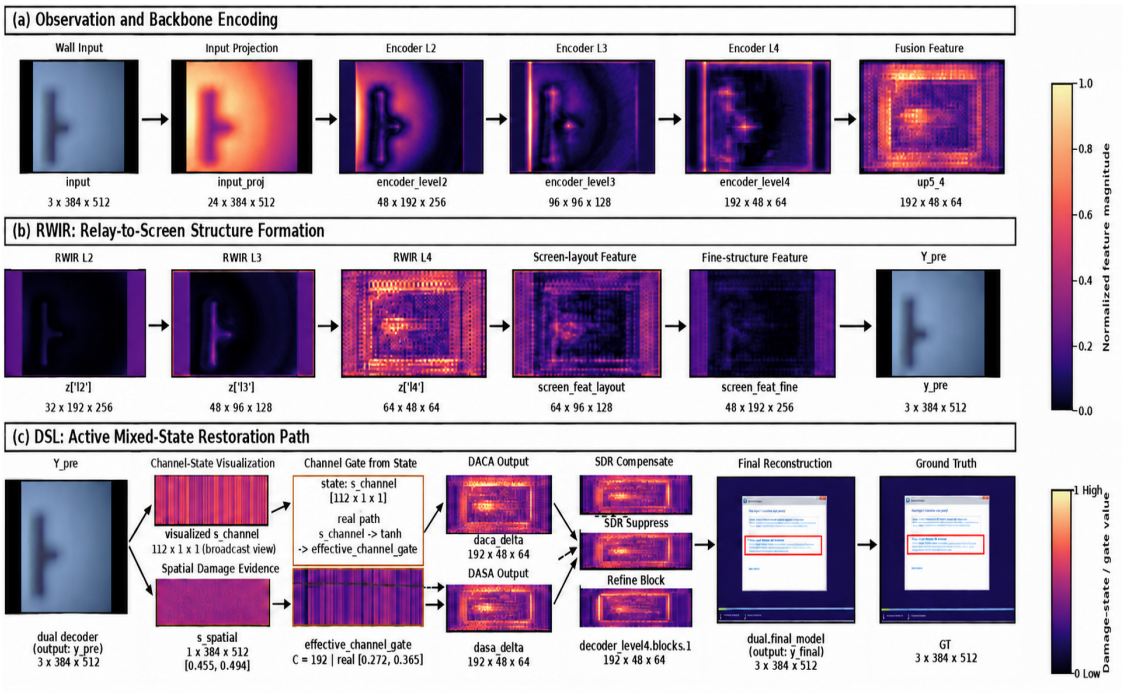

Figure A4: Stage evolution from the wall observation through RWIR and the preliminary reconstruction. DSL emits one scalar severity, while DACA, DASA, and SDR are shown as internal Pass-2 consumer responses before the final reconstruction and target reference.

Figure A4 follows the transformation from the wall-light pattern to the features used for screen reconstruction. RWIR combines responses across scales and produces Y_{pre}; DSL then summarizes its residual difficulty as the scalar s. Conditioned on this value, the internal DACA, DASA, and SDR operations modify the Pass-2 feature before the final image recovers the main window region and its surrounding screen arrangement.

C.3.4 REGION-LEVEL MECHANISM EVIDENCE

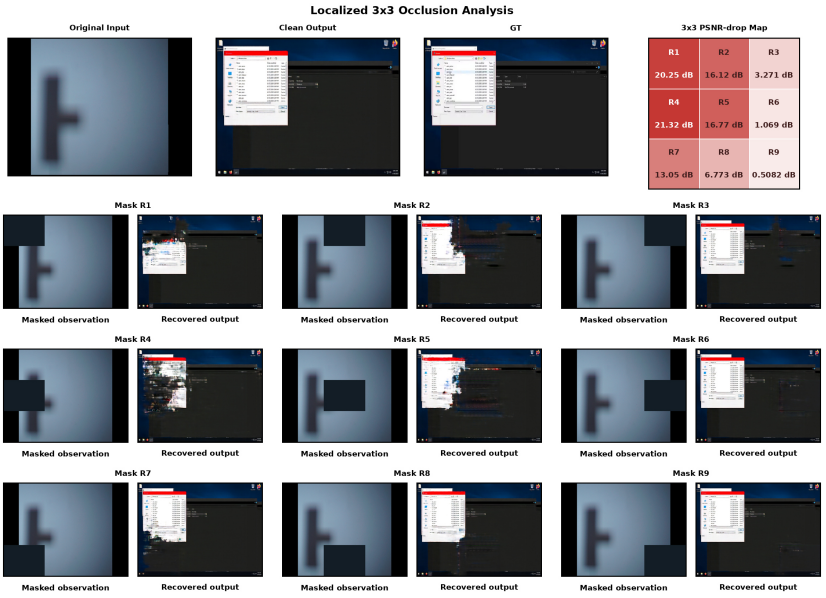

Figure A5: Fixed-canvas regional masking on a 3×3 wall-observation partition.

The 512×384 input is divided by horizontal boundaries $[0, 170, 341, 512]$ and vertical boundaries $[0, 128, 256, 384]$. Each condition replaces one region with neutral RGB $[19, 29, 37]$ while retaining the original canvas size.

Removing one wall region can weaken screen structures whose locations extend beyond that region. This nonlocal effect shows that the relay evidence for a window or content block is distributed rather than confined to a matching image patch.

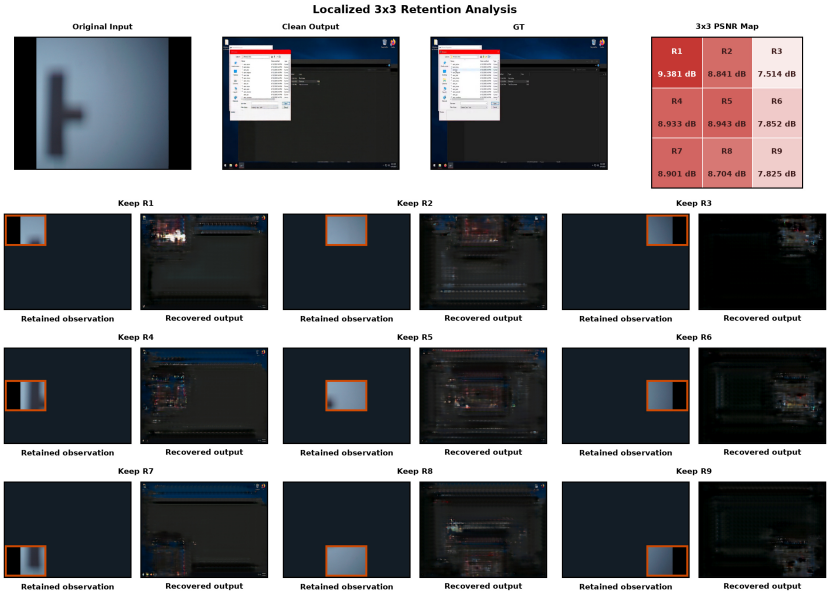

Figure A6: Fixed-canvas reconstruction with one retained region of the 3×3 wall-observation partition.

Each keep-only condition preserves one region at its original position and fills the remaining eight regions with neutral RGB $[19, 29, 37]$. The input, prediction, GT, and valid bounding box keep their original dimensions.

A single retained region can support parts of the screen layout, but it does not preserve all boundaries and relationships. Combining evidence from several wall regions is therefore important for reconstructing the complete interface organization.

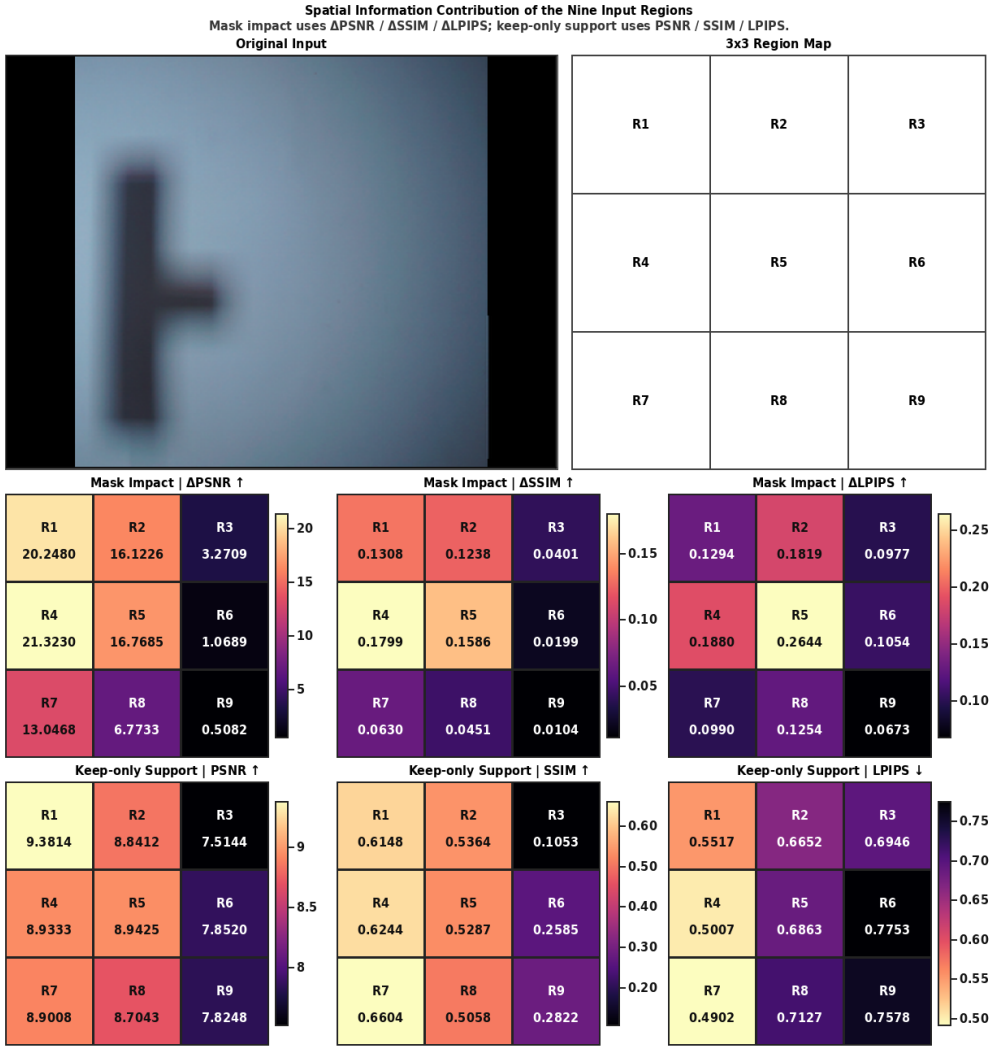

Figure A7: Regional reconstruction sensitivity under fixed-canvas mask-one and keep-one conditions. Upper heatmaps show clean-minus-corrupted changes after masking; lower heatmaps show the absolute PSNR, SSIM, and LPIPS obtained from each retained region.

The Heat Map is generated directly from the three task metrics and does not use a composite score. R8 contributes most strongly to OSHome across the masking measures, whereas R5 has the largest effect on Screen.

The contribution pattern depends on both scene and content: R8 is most influential for OSHome across the masking measures, whereas R5 has the largest effect on Screen. The heatmaps make this content-dependent distribution visible without collapsing the three metrics into one score.

Table 8: Region-wise information contribution across OSHome, OSLock, Screen, and Game. Panel (a) reports the reconstruction change after masking each region, where $\Delta\text{PSNR} = \text{PSNR}_{\text{clean}} - \text{PSNR}_{\text{mask}}$, $\Delta\text{SSIM} = \text{SSIM}_{\text{clean}} - \text{SSIM}_{\text{mask}}$, and $\Delta\text{LPIPS} = \text{LPIPS}_{\text{mask}} - \text{LPIPS}_{\text{clean}}$. Larger changes indicate a greater contribution of the masked region to reconstruction. Panel (b) reports reconstruction quality when the indicated region alone is retained. Best and second-best regional responses within each dataset are shown in red and blue, respectively. Ties at the displayed precision receive the same ranking mark.

(a) Reconstruction change after regional masking

Region	OSHome			OSLock			Screen			Game		
	ΔPSNR↑	ΔSSIM↑	ΔLPIPS↑	ΔPSNR↑	ΔSSIM↑	ΔLPIPS↑	ΔPSNR↑	ΔSSIM↑	ΔLPIPS↑	ΔPSNR↑	ΔSSIM↑	ΔLPIPS↑
R1	1.578	0.04010	0.04170	3.020	0.06740	0.06620	8.395	0.1245	0.1369	0.4744	0.01020	0.01200
R2	2.712	0.06210	0.05620	1.868	0.05400	0.06540	11.31	0.2346	0.2517	**4.513**	**0.1020**	0.07880
R3	1.428	0.03660	0.03860	2.650	0.06260	0.06300	7.328	0.08320	0.09400	0.4064	0.008300	0.01140
R4	2.849	0.06690	0.04880	3.640	**0.08110**	**0.07840**	11.35	0.3155	0.3495	3.630	0.07670	0.07500
R5	1.777	0.04500	0.03640	3.442	0.07460	0.04880	**13.60**	**0.3770**	**0.3948**	3.426	0.08250	0.07480
R6	3.019	0.06630	0.04800	**3.661**	0.07950	0.07670	8.649	0.1124	0.1328	4.490	**0.1020**	**0.1029**
R7	2.004	0.04720	0.04120	3.323	0.07120	0.06860	8.283	0.1363	0.1623	0.3943	0.007700	0.01160
R8	**3.780**	**0.09080**	**0.08020**	3.564	0.07750	0.06690	11.25	0.2546	0.2853	4.025	0.06610	0.07880
R9	2.151	0.05070	0.04210	3.330	0.07240	0.06320	6.355	0.06610	0.08170	0.6256	0.01250	0.02060

(b) Region-alone retention

Region	OSHome			OSLock			Screen			Game		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
R1	6.802	**0.3998**	0.9823	3.994	0.3876	0.8318	7.351	0.2806	0.7975	7.570	0.3055	0.7905
R2	**7.045**	0.3963	0.9552	**4.003**	0.3872	0.8304	7.090	0.2265	0.7927	8.196	**0.3949**	0.7497
R3	6.912	0.3985	0.9608	3.992	0.3875	0.8318	6.267	0.1479	0.8583	7.904	0.2749	0.7764
R4	6.543	0.3895	0.9766	3.976	0.3880	0.8310	**7.416**	0.2861	0.7955	7.373	0.3571	0.7611
R5	6.607	0.3917	0.9848	3.984	0.3879	**0.8288**	7.291	0.2468	**0.7922**	**8.403**	0.3874	**0.7289**
R6	6.876	0.3978	0.9401	3.981	0.3878	0.8330	6.540	0.1753	0.8427	7.415	0.3665	0.7363
R7	6.580	0.3926	**0.9354**	3.980	**0.3881**	0.8332	7.342	**0.2884**	0.7991	7.396	0.2875	0.7832
R8	6.829	0.3958	0.9672	3.989	0.3876	0.8327	7.077	0.2333	0.8005	8.332	0.3336	0.7435
R9	6.631	0.3895	0.9467	3.979	**0.3881**	0.8321	6.442	0.1735	0.8603	7.606	0.2962	0.7722

Analysis. Regional masking measures the reconstruction cost of removing each wall region, while retention measures the screen information available from that region alone. The responses are content dependent: the most influential areas differ across OSHome, OSLock, Screen, and Game, showing that useful screen evidence is distributed differently across scenes.

C.3.5 Representation Stability under Perturbation

We next examine whether the representation remains useful as the wall signal weakens or parts of the observation become unavailable.

The three perturbation families progressively reduce the strength or completeness of the available wall evidence and test how much reconstructable screen organization remains.

RWIR preserves a favorable reconstruction trend across relay-signal attenuation, localized missing observations, and reduced observation area. The three sweeps extend the comparison from spatially mixed evidence to progressively weaker and less complete wall observations.

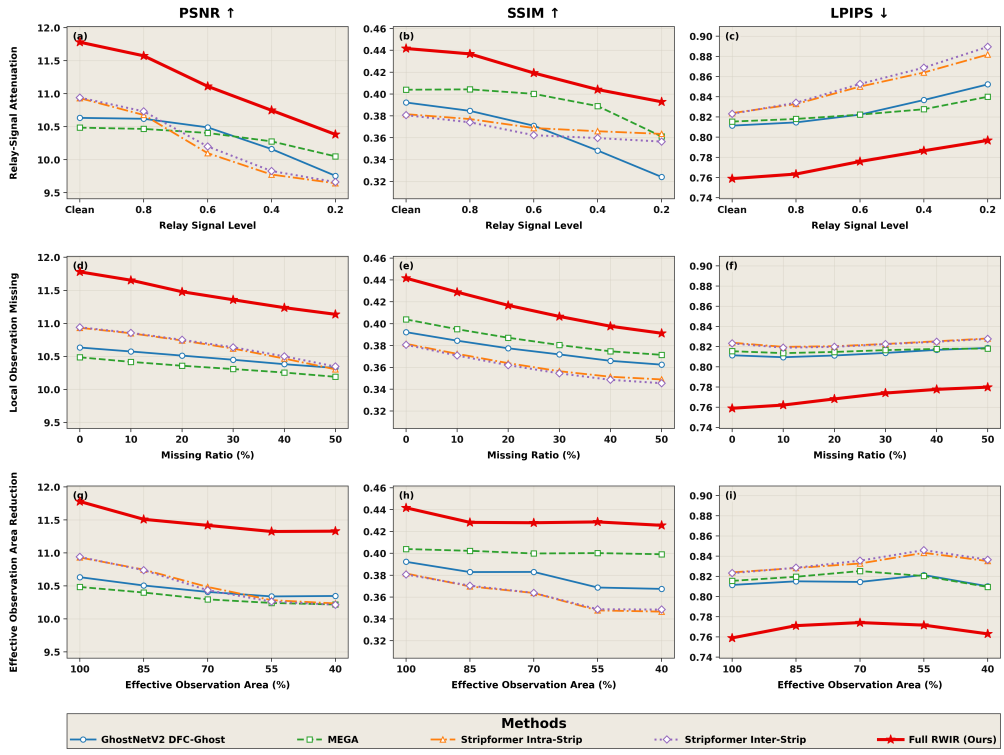

Figure A8: Representation comparison under relay-signal attenuation (top), localized missing observations (middle), and reduced observation area (bottom). Columns report PSNR, SSIM, and LPIPS.

C.4 DSL: Severity-Conditioned Repair

RWIR supplies the preliminary screen-oriented reconstruction Y_{pre}, whose faded strokes, breaks, omissions, and erroneous edges define the repair difficulty. The internal DSL ablations evaluate scalar severity estimation, feature conditioning, and severity-conditioned refinement in Pass 2. The response study connects the reported severity response to the degradation families used in the robustness evaluation.

C.4.1 Internal Ablation of DSL

Table 9 compares six internally constructed Damage-State Learning replacements. GSE (Global Severity Estimator) pools Y_{pre} globally before predicting one severity value, whereas SLSE (Shallow Local Severity Estimator) first extracts shallow local evidence and then aggregates it into the same scalar interface. MODEM introduces the Morton-Order 2D-Selective-Scan Module (MOS2D) and Dual Degradation Estimation Module (DDEM) for degradation estimation Wang et al. (2025); the present study complements this direction by evaluating global and shallow-evidence scalar estimation within DSL. GCSA (Generic Channel-Spatial Attention) and UFR (Uniform Feature Reweighting) replace the internal DACA–DASA conditioning: GCSA applies conventional channel-spatial attention under the scalar condition, whereas UFR converts severity directly into a sample-wise uniform feature scale. PRR (Plain Residual Refinement) and DCR (Direct Convolutional Refinement) replace SDR through, respectively, a bounded 1×1 residual update and a bounded two-layer 3×3 convolutional update. Together, the six configurations evaluate scalar severity estimation, feature conditioning, and residual refinement within DSL.

Table 9: Internal Damage-State Learning ablation on OSHome, OSLock, Screen, and Game. Best values are shown in red and second-best values are shown in blue. Ties at the displayed precision receive the same ranking mark.

Method	Source	OSHome			OSLock			Screen			Game		
		PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Damage-State Estimator Variants													
w/o DSE	—	13.18	0.5214	**0.7699**	15.57	0.5983	0.6642	15.97	0.6186	0.5771	16.09	0.6162	0.6770
DSE → GSE	Internal	13.29	0.5225	0.7815	15.17	0.5871	**0.6578**	16.27	**0.6298**	0.5696	**16.27**	0.6153	**0.6624**
DSE → SLSE	Internal	13.30	0.5218	0.7752	15.53	0.5959	0.6691	16.13	0.6212	0.5816	16.07	0.6188	0.6644
Consumer Feature Selection (DACA/DASA) Variants													
w/o DACA/DASA	—	13.28	0.5255	0.7723	**15.57**	0.5984	0.6641	15.65	0.6113	0.5819	16.09	0.6177	0.6729
DACA/DASA → GCSA	Internal	13.39	**0.5296**	0.7759	15.50	**0.6000**	0.6649	15.87	0.6113	0.5849	16.14	0.6172	0.6664
DACA/DASA → UFR	Internal	13.18	0.5209	0.7706	15.57	0.5984	0.6642	16.19	0.6255	0.5725	16.16	0.6187	0.6688
State-Driven Restoration (SDR) Variants													
w/o SDR	—	13.25	0.5244	0.7715	15.57	0.5983	0.6642	15.89	0.6198	0.5734	15.89	0.6135	0.6755
SDR → PRR	Internal	**13.44**	0.5217	0.7789	15.38	0.5950	0.6582	16.25	0.6233	0.5740	16.16	0.6192	0.6683
SDR → DCR	Internal	13.24	0.5223	0.7761	15.56	0.5979	0.6650	**16.43**	0.6288	**0.5673**	15.99	**0.6199**	0.6686
RSHRL (Ours)	—	**14.52**	**0.5489**	**0.7239**	**16.74**	**0.6240**	**0.6286**	**25.95**	**0.8540**	**0.2048**	**18.45**	**0.6769**	**0.5339**

Across the nine modified rows on Screen, PSNR spans 15.6505 to 16.4315, SSIM spans 0.6113 to 0.6298, and LPIPS spans 0.5673 to 0.5849. Relative to the strongest modified result for each metric, the full row is higher by 9.5212 PSNR and 0.2242 SSIM and lower by 0.3625 LPIPS. The GSE replacement produces the strongest modified SSIM value of 0.6298, while the complete DSL configuration achieves the strongest joint reconstruction performance. The combined pattern shows that scalar severity estimation and conditioned refinement work together to preserve recovered structure and repair the residual damage in the preliminary reconstruction.

Having established the contributions of these components, we next examine how the hierarchical representation and scalar severity respond when the wall observation changes.

C.4.2 Representation and Damage-Severity Responses

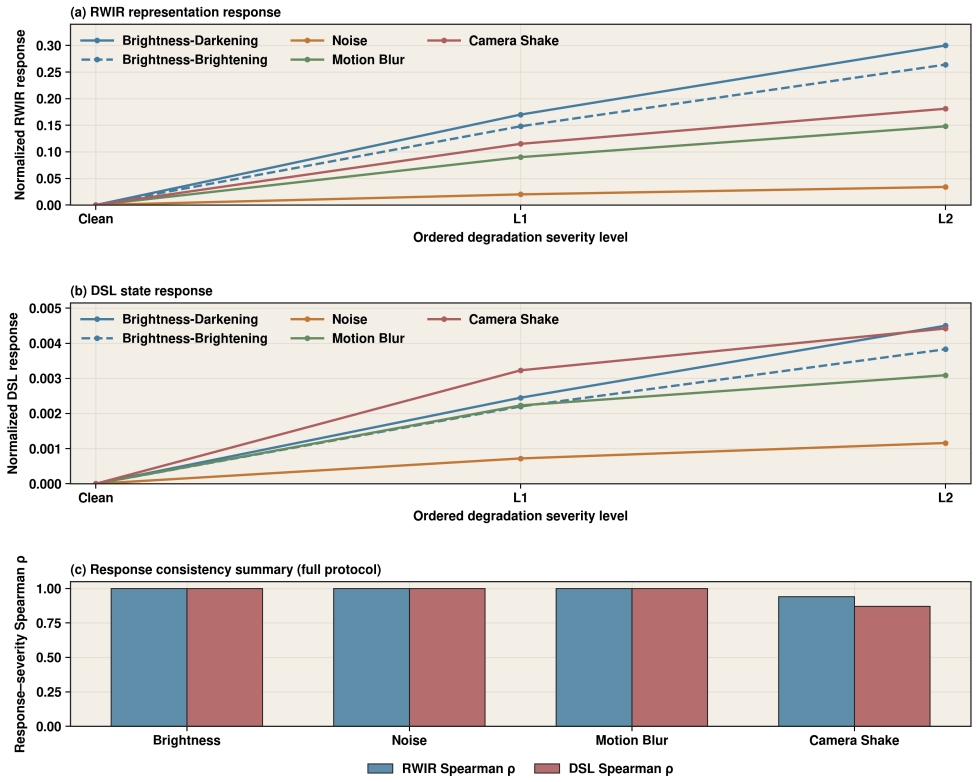

Figure A9: Clean-normalized response of RWIR and DSL across brightness changes, noise, motion blur, and camera shake.

Table 10: RWIR and DSL degradation-response consistency across the four formal datasets. Spearman correlation and monotonicity measure whether each reported response follows increasing degradation, while endpoint response records the change at the most severe tested condition. The DSL response is derived from the scalar severity used by the Pass-2 consumer.

Notation: ρ = Spearman correlation; Mono. = Monotonicity (%); End. = Endpoint response.

Degradation	OSHome						OSLock					
	RWIR			DSL			RWIR			DSL		
	ρ	Mono.	End.	ρ	Mono.	End.	ρ	Mono.	End.	ρ	Mono.	End.
Brightness–Darkening	1.000	100.0	0.1840	1.000	100.0	0.0030	1.000	100.0	0.1270	1.000	100.0	0.0026
Brightness–Brightening	1.000	100.0	0.1699	1.000	100.0	0.0025	1.000	100.0	0.1097	1.000	100.0	0.0024
Noise	1.000	100.0	0.1943	1.000	100.0	0.0043	1.000	100.0	0.1212	1.000	100.0	0.0035
Motion Blur	1.000	99.70	0.3374	1.000	90.25	0.0061	1.000	100.0	0.2705	1.000	98.14	0.0056
Camera Shake	0.9429	81.10	0.3909	0.8286	74.60	0.0094	0.9429	82.44	0.2678	0.9429	80.00	0.0065

Degradation	Screen						Game					
	RWIR			DSL			RWIR			DSL		
	ρ	Mono.	End.	ρ	Mono.	End.	ρ	Mono.	End.	ρ	Mono.	End.
Brightness–Darkening	1.000	100.0	0.4868	1.000	100.0	0.0055	1.000	100.0	0.4078	1.000	98.00	0.0069
Brightness–Brightening	1.000	100.0	0.4215	1.000	100.0	0.0048	1.000	100.0	0.3540	1.000	98.00	0.0056
Noise	1.000	100.0	0.2190	1.000	100.0	0.0047	1.000	100.0	0.1550	1.000	100.0	0.0051
Motion Blur	1.000	100.0	0.3998	1.000	99.04	0.0041	1.000	98.80	0.3195	1.000	95.44	0.0061
Camera Shake	0.9429	78.57	0.2347	0.8286	74.50	0.0028	0.9429	80.48	0.3451	0.8857	76.96	0.0074

The horizontal axis orders the evaluated degradation conditions. Each response is measured relative to the same sample under the transform-free Clean condition. RWIR averages the responses of Z_2, Z_3, Z_4, while DSL reports the change in its scalar severity output.

The hierarchical representation and estimated severity change systematically with the observation. RWIR responds to changes in available relay evidence, while the scalar DSL output provides an observation-dependent condition for refinement.

For any recorded tensor T, the response of sample i, dataset d, degradation family a, and level ℓ is

$$\mathrm{RMS}(T) = \sqrt{|T|^{-1} \sum_u T_u^2}, \qquad R_T(i,d,a,\ell) = \frac{\mathrm{RMS}(T_{i,d,a,\ell} - T_{i,d,a,0})}{\mathrm{RMS}(T_{i,d,a,0}) + 10^{-8}}. \tag{A12}$$

This is a per-sample change relative to Clean, not min–max normalization, and the response is not constrained to $[0, 1]$. The two reported module responses are

$$R_{\mathrm{RWIR}} = \frac{R_{Z_2} + R_{Z_3} + R_{Z_4}}{3}, \qquad R_{\mathrm{DSL}} = \frac{|s_{i,d,a,\ell} - s_{i,d,a,0}|}{|s_{i,d,a,0}| + 10^{-8}}. \tag{A13}$$

Responses are first averaged over samples within each dataset and then macro-averaged with equal weight over OSHome, OSLock, Screen, and Game. The plotted curves contain no error band; the 2,000-sample bootstrap intervals are used only for the tabulated Spearman, monotonicity, and endpoint statistics.

The hierarchical representation and estimated severity change systematically as the observation is altered. RWIR responds to changes in the available relay evidence, while the scalar severity provides an observation-dependent condition for refinement. Their paired trends connect the two-stage design to the changing input that it must reconstruct.

C.5 Complementarity of RWIR and DSL

C.5.1 Extended Component Ablation

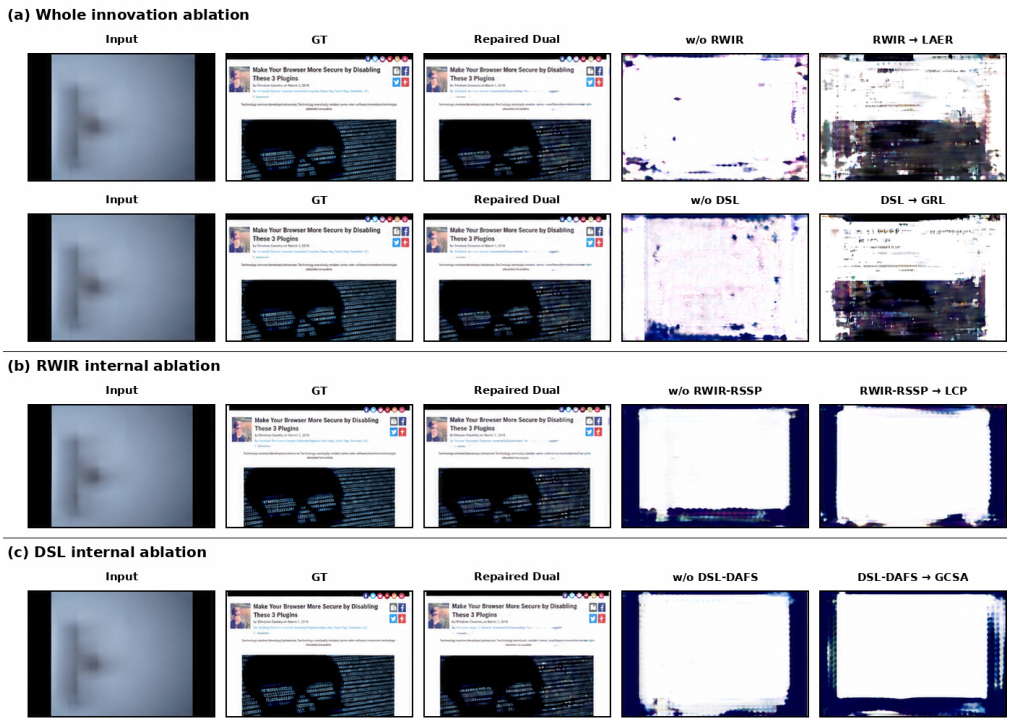

Figure A10: Qualitative component ablation for the proposed two-stage reconstruction.

On the same Screen example, the full configuration preserves the large content regions and window organization, while component changes alter their completeness and separation. The numerical ablations provide the corresponding four-dataset comparison.

On the same Screen example, the full model preserves the large content regions and window organization, while component changes alter their completeness and separation. Together with the numerical ablations, the visual comparison shows the benefit of combining structural correspondence with damage-guided refinement.

C.6 Robustness across Acquisition and Degradation Conditions

We evaluate hidden-screen reconstruction in two acquisition setups and use controlled observation changes to examine the effects of geometry, illumination, and image degradation. Brightness, Gaussian noise, motion blur, and camera shake use the transform-free CLEAN-00 observation as their common reference. Brightness factor 1.0, a no-noise input, blur kernel 1, and shake radius 0 denote that reference condition. The geometry, scattering, added-light, and lens-response studies use the reference value defined on their own parameter axes and are interpreted within their respective sweeps.

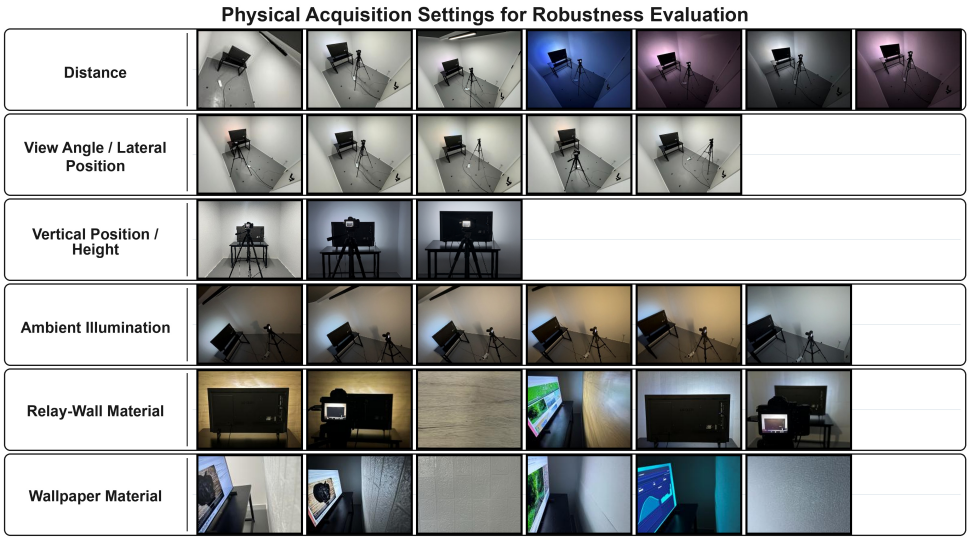

Figure A11: Acquisition scenes and observation factors.

The following studies ask whether the recovered screen organization persists as the wall evidence changes. We first vary its spatial arrangement and then its illumination and local image quality. This sequence follows the same practical question across increasingly altered inputs: what screen information can still be recovered from the wall?

C.6.1 Camera Distance and Viewpoint Conditions

The following experiments use controlled observation changes to examine how the indirect wall response varies with camera-to-wall relation and viewing geometry. The goal is to test whether the learned screen-space organization survives changes in spreading, attenuation, and the relative placement of wall evidence.

Each distance and viewpoint study evaluates eleven conditions and reports both the condition-wise curves and their four-dataset averages.

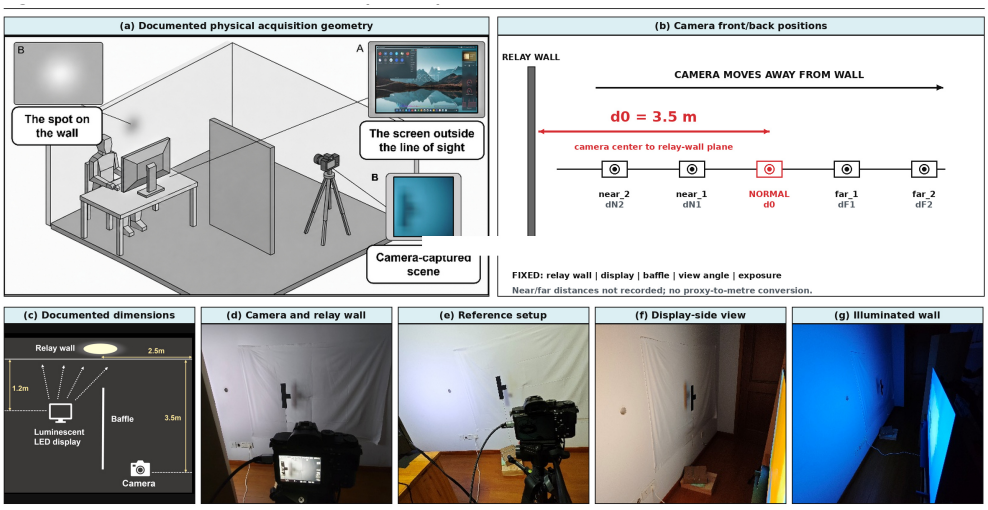

Figure A12: Controlled camera-distance setup.

The diagram defines the factor used to vary the effective camera-distance condition while the paired screen target remains fixed. This experiment isolates the effect of a changing relay observation and

tests whether the proposed representation follows screen structure across changes in wall-response appearance.

The setup varies the effective distance factor while keeping the hidden target fixed. The resulting changes in wall-response spread and strength isolate the observation-side effect examined by the following curves and examples.

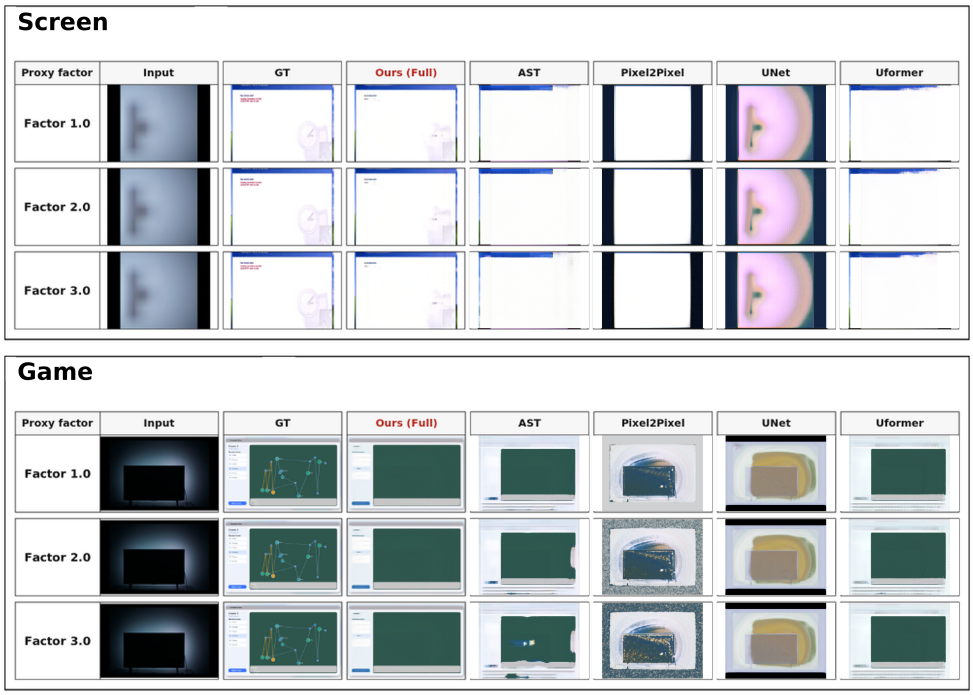

Figure A13: Qualitative camera-distance comparison.

The displayed examples connect the distance conditions to visible screen content: the principal Screen window and the large Game panel remain recognizable across the shown factors.

The curves compare reconstruction quality across all four content datasets as the effective distance factor changes the spread and strength of the observed wall response.

RSHRL preserves a favorable reconstruction trend as the distance factor changes. OSHome and Screen decrease gradually at larger factors, whereas OSLock and Game remain comparatively stable, revealing content-dependent sensitivity to the changing wall response.

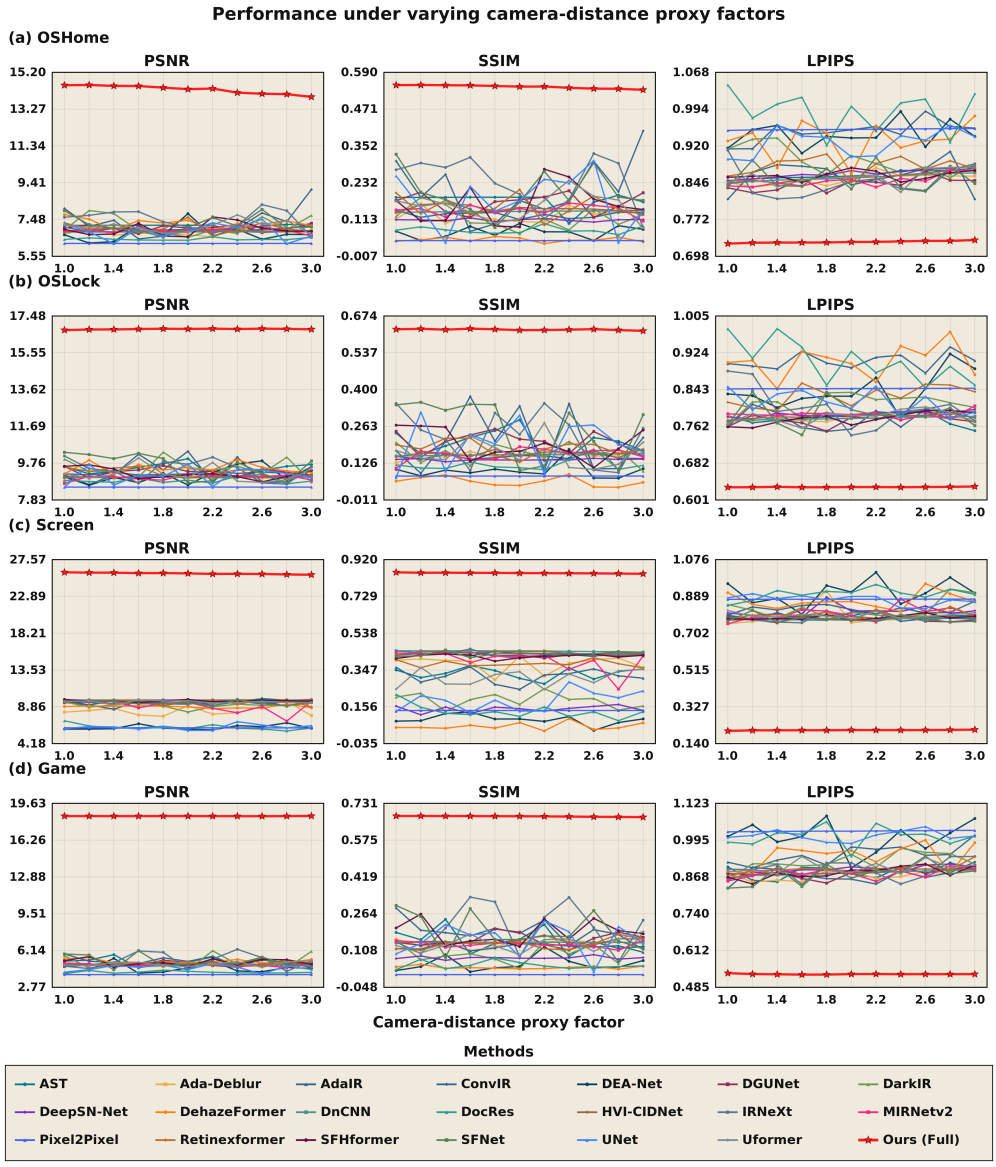

Figure A14: Robustness curves under eleven camera-distance factors from 1.0 to 3.0.

Table 11: Robustness results averaged over eleven camera-distance factors. Best values are shown in red and second-best values in blue.

Method	Source	OSHome			OSLock			Screen			Game		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
General Purpose Image Restoration Frameworks													
AST	CVPR'24	7.146	0.1749	0.8624	9.456	0.2024	0.7800	9.552	0.3300	0.7874	5.084	0.1500	0.8860
AdaIR	ICLR'25	6.977	0.1285	0.9494	9.213	0.1499	0.9056	9.325	0.3157	0.8246	4.865	0.1225	0.9457
ConvIR	TPAMI'24	7.348	0.2104	0.8609	9.330	0.2061	0.7786	9.479	0.4296	0.7828	5.164	0.1877	0.8744
DGUNet	CVPR'22	7.040	0.1679	0.8441	9.318	0.2057	0.7769	9.549	0.4246	0.7714	4.878	0.1538	0.8819
DeepSN-Net	TPAMI'25	6.846	0.1112	0.8653	9.037	0.1392	0.7890	9.557	0.1454	0.8151	4.816	0.07710	0.8926
IRNeXt	ICML'23	7.471	0.2356	0.8476	9.432	0.2292	0.8042	9.493	0.4300	0.7738	5.198	0.2047	0.8949
MIRNetv2	TPAMI'23	6.918	0.1431	0.8502	9.008	0.1586	0.7888	8.951	0.3965	0.8001	4.802	0.1335	0.8802
SFHformer	ECCV'24	6.969	0.1632	0.8613	9.214	0.1973	0.7795	9.616	0.4183	0.7819	5.070	0.1849	0.8829
SFNet	ICLR'23	7.232	0.2005	0.8550	9.638	0.2590	0.7890	9.358	0.4311	0.7743	5.145	0.1908	0.8776
Uformer	CVPR'22	7.118	0.1556	0.8576	9.202	0.1657	0.7842	9.495	0.3061	0.7937	4.912	0.1374	0.9003
Degradation-Specific / Domain-Specific Restoration Models													
Ada-Deblur	TIP'23	6.795	0.1262	0.8536	9.010	0.1538	0.7805	8.365	0.3798	0.7957	4.853	0.1388	0.8734
DEA-Net	TIP'24	6.748	0.07210	0.9429	9.141	0.09820	0.8413	6.246	0.09030	0.9092	4.480	0.05170	1.006
DarkIR	CVPR'25	7.169	0.1185	0.8915	9.503	0.1614	0.8129	9.072	0.1904	0.8252	5.107	0.1009	0.9208
DehazeFormer	TIP'23	7.283	0.04720	0.9325	9.493	0.05780	0.9054	8.870	0.05540	0.8666	5.159	0.03620	0.9410
DnCNN	TIP'17	6.829	0.1390	0.8566	9.020	0.1579	0.7799	9.727	0.4452	0.7738	4.817	0.1423	0.8808
DocRes	CVPR'24	6.452	0.07840	0.9900	8.862	0.1206	0.9058	6.274	0.1327	0.9014	4.152	0.04560	1.008
HVI-CIDNet	CVPR'25	6.888	0.1490	0.8581	9.035	0.1621	0.7861	9.690	0.4364	0.7815	4.843	0.1470	0.8831
Retinexformer	ICCV'23	7.055	0.1537	0.8802	9.334	0.1826	0.8298	9.489	0.3798	0.7940	4.878	0.1347	0.8971
Generic Encoder–Decoder Models													
pix2pix	CVPR'17	6.223	0.04430	0.9537	8.504	0.07830	0.8455	6.190	0.1373	0.8748	3.930	0.005900	1.027
U-Net	MICCAI'15	6.866	0.1762	0.9271	9.118	0.2002	0.8161	6.270	0.1889	0.8722	4.522	0.1523	1.011
RSHRL (Ours)	This work	**14.30**	**0.5434**	**0.7274**	**16.79**	**0.6229**	**0.6291**	**25.81**	**0.8501**	**0.2087**	**18.45**	**0.6751**	**0.5305**

Analysis. RSHRL ranks first for all 12 dataset–metric averages in the distance study. The four-content averages extend the visual examples beyond Screen and Game: screen organization remains recoverable as the spatial spread and strength of the wall evidence change.

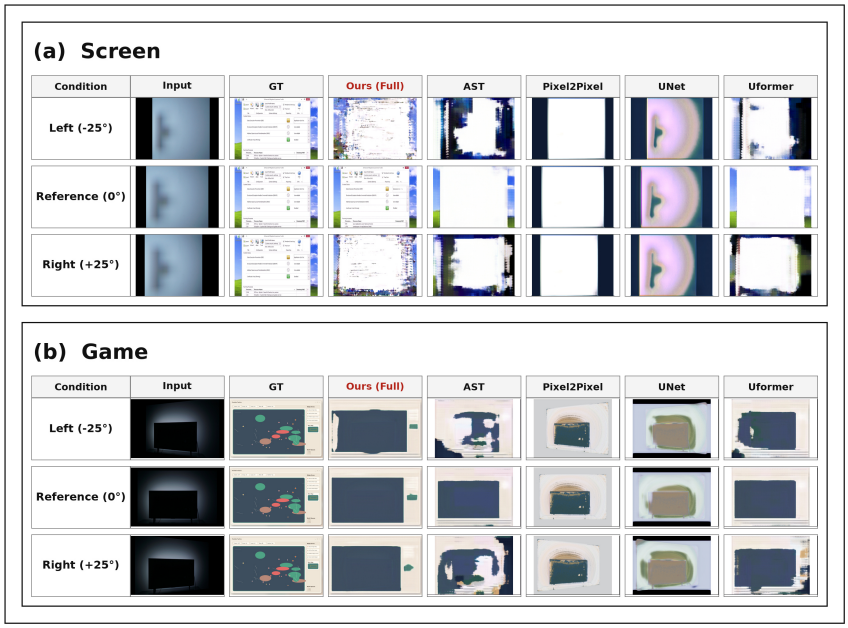

Figure A15: Qualitative reconstruction under horizontal view-angle shifts.

The displayed conditions show how lateral viewpoint change alters the wall pattern and the corresponding reconstruction of the main screen regions.

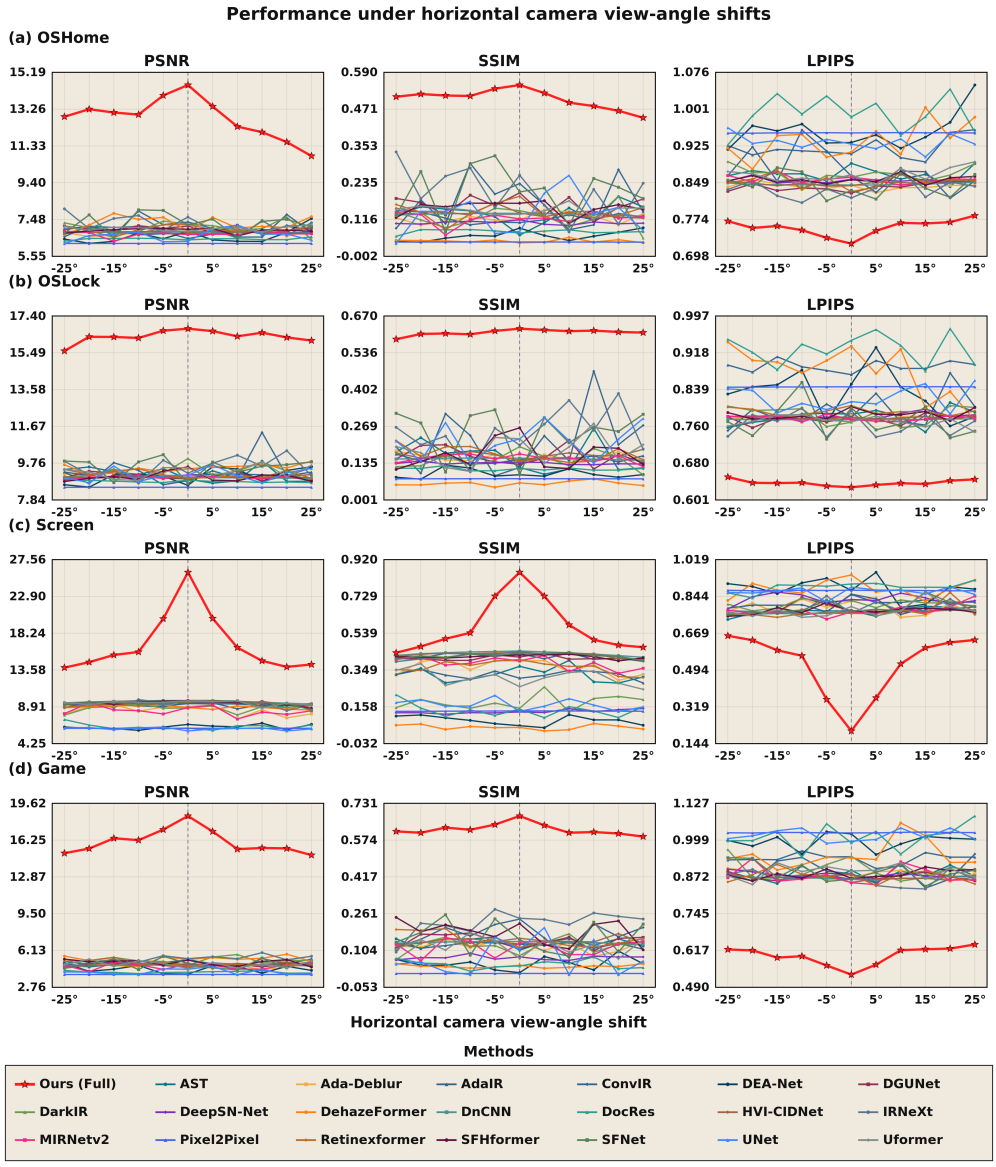

Figure A16: Robustness curves under horizontal view-angle shifts from $-25°$ to $+25°$.

Changing the horizontal viewing angle redistributes the wall response laterally. The curves compare how this redistribution affects reconstruction of the four content types.

Horizontal viewpoint changes alter the spatial arrangement of the wall response. Screen varies more strongly with angle than the broad Game layout, making the content dependence of lateral viewpoint change visible.

Table 12: Robustness results averaged over eleven horizontal view-angle conditions between $-25°$ and $+25°$. Best values are shown in red and second-best values in blue.

Method	Source	OSHome			OSLock			Screen			Game		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
General Purpose Image Restoration Frameworks													
AST	CVPR'24	6.936	0.1356	0.8672	9.173	0.1592	0.7803	9.309	0.3252	0.7723	4.939	0.1395	0.8789
AdaIR	ICLR'25	6.905	0.1221	0.9206	9.160	0.1501	0.8890	9.177	0.3247	0.8106	4.867	0.1273	0.9352
ConvIR	TPAMI'24	7.048	0.1693	0.8743	9.357	0.2156	0.7928	9.228	0.4178	0.7847	4.953	0.1726	0.8907
DGUNet	CVPR'22	6.993	0.1644	0.8407	9.070	0.1606	0.7749	9.422	0.4099	0.7810	4.812	0.1454	0.8742
DeepSN-Net	TPAMI'25	6.819	0.1056	0.8532	9.017	0.1345	0.7822	9.373	0.1333	0.8311	4.792	0.07300	0.8780
IRNeXt	ICML'23	7.236	0.2050	0.8367	9.571	0.2485	0.7638	9.257	0.4198	0.7884	5.335	0.2211	0.8673
MIRNetv2	TPAMI'23	6.720	0.1201	0.8561	9.016	0.1521	0.7771	8.395	0.3847	0.7785	4.692	0.1161	0.8770
SFHformer	ECCV'24	6.898	0.1539	0.8557	9.039	0.1638	0.7849	9.342	0.4176	0.7773	5.030	0.1819	0.8817
SFNet	ICLR'23	7.285	0.2060	0.8453	9.574	0.2515	0.7719	9.312	0.4223	0.7922	4.874	0.1605	0.8905
Uformer	CVPR'22	7.010	0.1475	0.8658	9.309	0.1818	0.7705	9.393	0.3097	0.7777	4.870	0.1229	0.8894
Degradation-Specific / Domain-Specific Restoration Models													
Ada-Deblur	TIP'23	6.769	0.1216	0.8417	9.043	0.1556	0.7784	8.452	0.3640	0.7917	4.700	0.1227	0.8801
DEA-Net	TIP'24	6.591	0.06410	0.9551	9.086	0.09610	0.8320	6.372	0.08740	0.8813	4.431	0.04840	0.9946
DarkIR	CVPR'25	7.065	0.1222	0.8670	9.292	0.1454	0.7846	8.762	0.1820	0.8087	5.216	0.1232	0.8956
DehazeFormer	TIP'23	7.277	0.04750	0.9362	9.504	0.06070	0.8776	8.850	0.05280	0.8819	5.293	0.03650	0.9474
DnCNN	TIP'17	6.802	0.1333	0.8462	8.994	0.1507	0.7745	9.550	0.4345	0.7685	4.782	0.1382	0.8693
DocRes	CVPR'24	6.458	0.07790	0.9898	8.803	0.1136	0.9260	6.407	0.1404	0.8895	4.149	0.04480	1.010
HVI-CIDNet	CVPR'25	6.810	0.1389	0.8483	9.054	0.1648	0.7789	9.501	0.4279	0.7704	4.826	0.1458	0.8677
Retinexformer	ICCV'23	6.963	0.1412	0.8522	9.181	0.1674	0.7916	9.183	0.3637	0.7707	4.993	0.1502	0.8749
Generic Encoder–Decoder Models													
pix2pix	CVPR'17	6.223	0.04410	0.9517	8.503	0.07800	0.8446	6.194	0.1361	0.8725	3.930	0.005800	1.025
U-Net	MICCAI'15	6.638	0.1346	0.9324	9.151	0.2155	0.8204	6.138	0.1703	0.8669	4.187	0.07230	1.015
RSHRL (Ours)	This work	**12.80**	**0.5047**	**0.7572**	**16.34**	**0.6102**	**0.6382**	**16.85**	**0.5733**	**0.5229**	**16.11**	**0.6202**	**0.5998**

Analysis. Horizontal viewpoint changes alter the spatial arrangement of the wall response, while the examples continue to recover the principal window and panel organization. Screen responds more strongly to angle than the broad Game layout, making the content dependence visible; RSHRL ranks first across all 12 averaged comparisons.

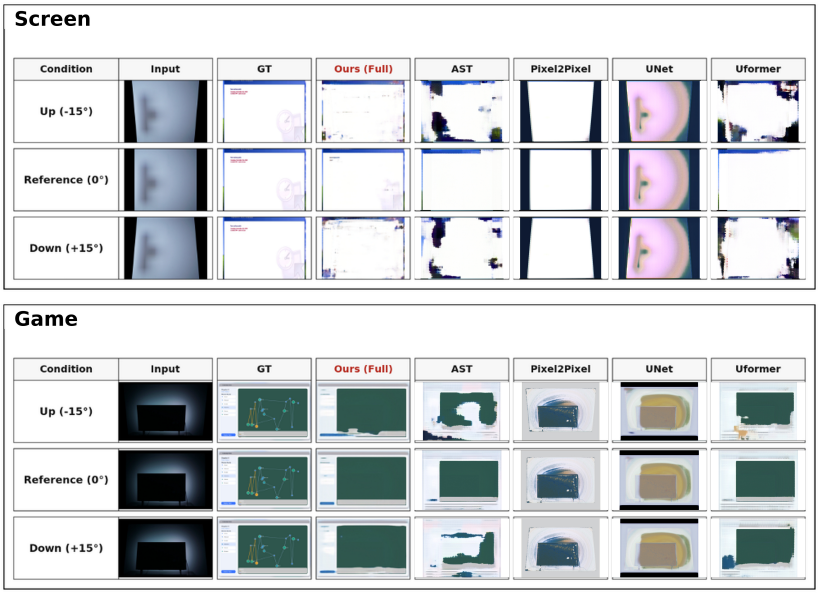

Figure A17: Qualitative recovery under vertical-view shifts.

The displayed pitch settings connect the metric changes to preservation of the principal display regions and their vertical ordering.

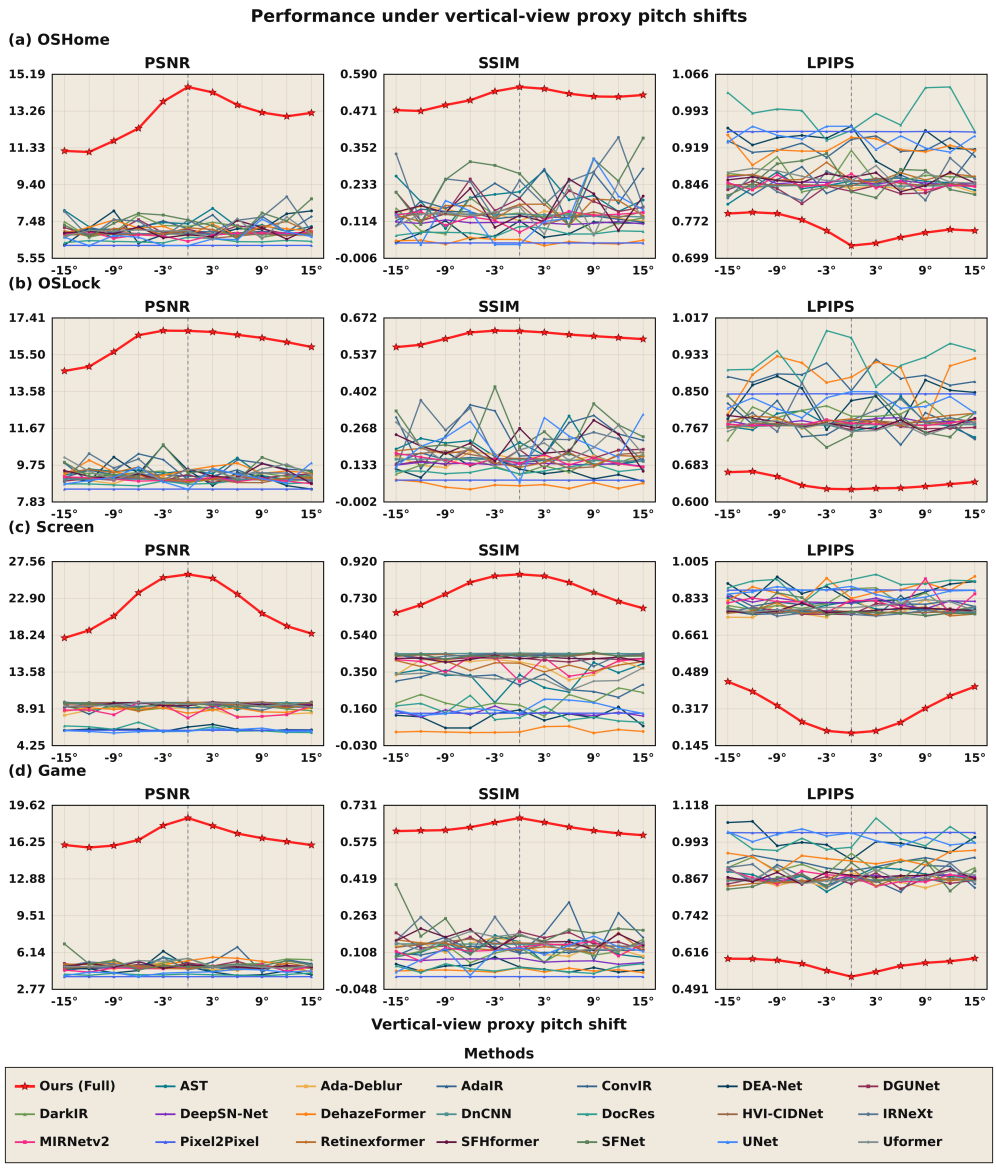

Figure A18: Robustness curves under vertical-view pitch shifts from $-15°$ to $+15°$.

The curves extend the viewpoint study to top-to-bottom changes in the wall response and compare reconstruction quality across the four datasets.

The vertical-view experiment extends the geometry study to top-to-bottom changes. Quality is highest near the central view and decreases toward larger pitch offsets, with the principal screen regions remaining identifiable across the sweep.

Table 13: Robustness results averaged over eleven vertical-view pitch conditions. Best values are shown in red and second-best values in blue.

Method	Source	OSHome			OSLock			Screen			Game		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
General Purpose Image Restoration Frameworks													
AST	CVPR'24	7.333	0.1927	0.8457	9.409	0.1979	0.7802	9.460	0.3227	0.7912	4.888	0.1224	0.8754
AdaIR	ICLR'25	6.951	0.1301	0.9212	9.230	0.1534	0.8842	9.256	0.2987	0.8140	4.786	0.1148	0.9242
ConvIR	TPAMI'24	7.207	0.1959	0.8552	9.583	0.2513	0.7773	9.323	0.4330	0.7726	5.021	0.1655	0.8740
DGUNet	CVPR'22	7.014	0.1695	0.8368	9.086	0.1624	0.7719	9.512	0.4188	0.7785	4.910	0.1561	0.8602
DeepSN-Net	TPAMI'25	6.834	0.1099	0.8520	9.028	0.1368	0.7810	9.537	0.1380	0.8188	4.792	0.07390	0.8710
IRNeXt	ICML'23	7.438	0.2305	0.8489	9.371	0.2243	0.7967	9.451	0.4317	0.7689	4.903	0.1584	0.8768
MIRNetv2	TPAMI'23	6.782	0.1258	0.8470	8.985	0.1498	0.7774	8.670	0.3859	0.8200	4.706	0.1226	0.8706
SFHformer	ECCV'24	6.939	0.1588	0.8557	9.265	0.2061	0.7773	9.444	0.4182	0.7797	4.902	0.1637	0.8758
SFNet	ICLR'23	7.502	0.2418	0.8566	9.578	0.2594	0.7811	9.388	0.4333	0.7805	5.182	0.2029	0.8694
Uformer	CVPR'22	6.991	0.1410	0.8617	9.418	0.1968	0.7699	9.433	0.3197	0.7931	5.011	0.1482	0.8771
Degradation-Specific / Domain-Specific Restoration Models													
Ada-Deblur	TIP'23	6.875	0.1372	0.8403	9.034	0.1558	0.7771	8.547	0.3796	0.7757	4.758	0.1272	0.8592
DEA-Net	TIP'24	7.118	0.08550	0.9285	9.288	0.1150	0.8344	6.342	0.1130	0.8694	4.570	0.04450	0.9940
DarkIR	CVPR'25	7.128	0.1232	0.8644	9.203	0.1345	0.7974	9.061	0.2006	0.8164	5.136	0.1100	0.9111
DehazeFormer	TIP'23	7.096	0.04870	0.9192	9.473	0.06000	0.8835	8.796	0.04590	0.8722	5.144	0.03160	0.9353
DnCNN	TIP'17	6.813	0.1350	0.8433	9.002	0.1522	0.7735	9.707	0.4453	0.7681	4.794	0.1405	0.8619
DocRes	CVPR'24	6.457	0.07890	0.9892	8.825	0.1148	0.9248	6.344	0.1389	0.9026	4.126	0.03980	1.001
HVI-CIDNet	CVPR'25	6.790	0.1353	0.8464	9.032	0.1612	0.7783	9.646	0.4365	0.7717	4.806	0.1424	0.8626
Retinexformer	ICCV'23	7.011	0.1490	0.8626	9.198	0.1634	0.7860	9.321	0.3838	0.7731	4.877	0.1352	0.8639
Generic Encoder–Decoder Models													
pix2pix	CVPR'17	6.223	0.04420	0.9516	8.504	0.07820	0.8447	6.195	0.1370	0.8727	3.930	0.005900	1.025
U-Net	MICCAI'15	6.637	0.1350	0.9391	9.132	0.2121	0.8231	6.137	0.1606	0.8636	4.317	0.1036	1.007
RSHRL (Ours)	This work	**12.89**	**0.5139**	**0.7589**	**16.08**	**0.6027**	**0.6429**	**21.84**	**0.7661**	**0.3125**	**16.73**	**0.6346**	**0.5754**

Analysis. The vertical-view experiment complements the horizontal-angle study: quality is strongest near the central view and changes as pitch moves toward either endpoint. RSHRL ranks first in all 12 dataset–metric averages, and the examples retain the principal regions across the shown settings.

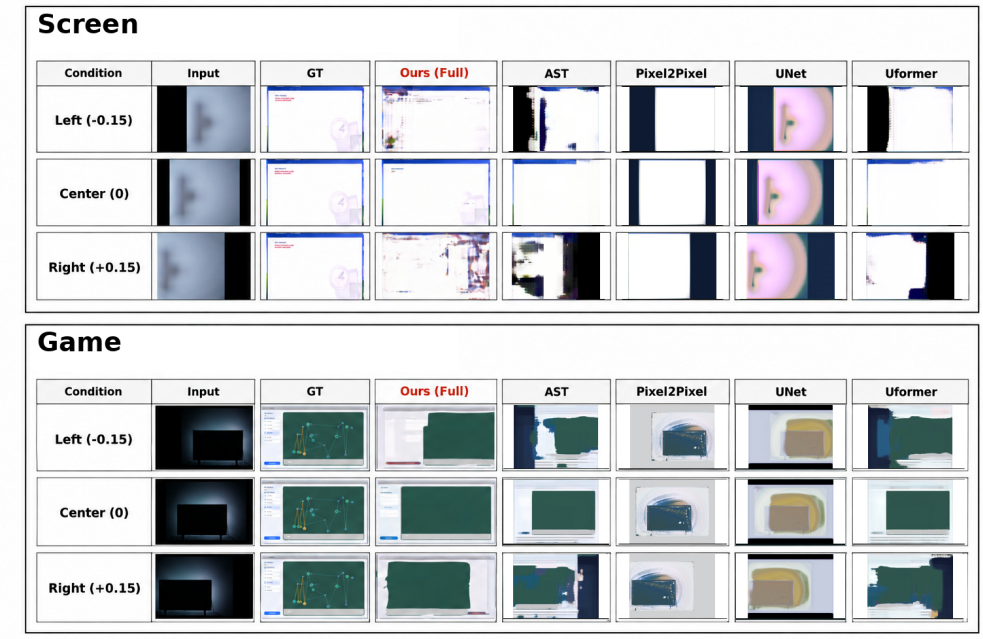

Figure A19: Qualitative reconstruction under horizontal camera-position shifts.

The panel shows how lateral displacement changes the wall pattern and the recovered target window and panel arrangement.

The examples connect this positional trend to the recovered target window and panel arrangement across the displayed shifts.

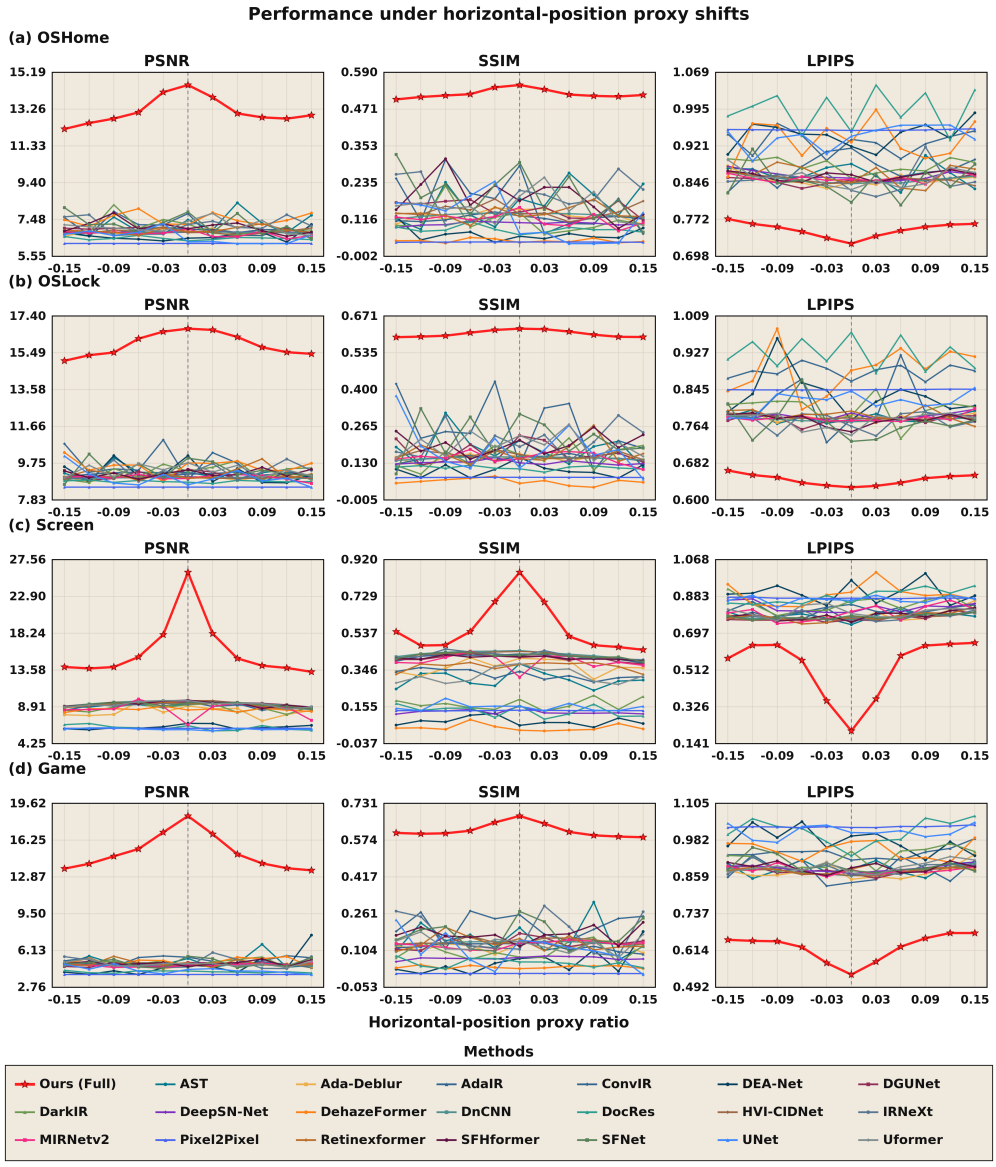

Figure A20: Robustness curves under eleven horizontal camera-position shifts from -0.15 to $+0.15$.

The sweep moves the observation position laterally while holding the hidden target fixed, separating camera displacement from content change.

Horizontal position changes shift the balance and placement of the observed wall pattern. Reconstruction quality is strongest near the reference position and decreases toward larger displacements, especially for Screen and Game.

Table 14: Robustness results averaged over eleven horizontal camera-position conditions. Best values are shown in red and second-best values in blue.

Method	Source	OSHome			OSLock			Screen			Game		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
General Purpose Image Restoration Frameworks													
AST	CVPR'24	7.187	0.1661	0.8648	9.216	0.1825	0.7831	9.057	0.2974	0.7841	5.108	0.1523	0.8884
AdaIR	ICLR'25	6.896	0.1210	0.9286	9.173	0.1484	0.8857	9.081	0.3230	0.8181	4.886	0.1200	0.9397
ConvIR	TPAMI'24	7.059	0.1704	0.8742	9.626	0.2602	0.8037	9.155	0.4123	0.7885	5.269	0.2121	0.8805
DGUNet	CVPR'22	6.897	0.1506	0.8508	9.098	0.1696	0.7791	9.263	0.4075	0.7745	4.810	0.1348	0.8810
DeepSN-Net	TPAMI'25	6.803	0.1030	0.8607	8.999	0.1325	0.7876	9.228	0.1278	0.8283	4.795	0.07070	0.8882
IRNeXt	ICML'23	7.176	0.1918	0.8522	9.309	0.2081	0.7754	9.133	0.4222	0.7823	5.020	0.1771	0.8873
MIRNetv2	TPAMI'23	6.731	0.1165	0.8561	9.008	0.1510	0.7820	8.584	0.3912	0.8008	4.783	0.1287	0.8818
SFHformer	ECCV'24	7.088	0.1883	0.8609	9.218	0.2008	0.7784	9.270	0.4096	0.7867	4.972	0.1720	0.8915
SFNet	ICLR'23	7.221	0.1982	0.8488	9.382	0.2194	0.7742	9.187	0.4200	0.7809	4.996	0.1711	0.8971
Uformer	CVPR'22	7.103	0.1584	0.8566	9.363	0.1964	0.7780	9.231	0.3161	0.8060	4.901	0.1307	0.8934
Degradation-Specific / Domain-Specific Restoration Models													
Ada-Deblur	TIP'23	6.824	0.1285	0.8524	8.990	0.1491	0.7836	8.308	0.3642	0.7869	4.772	0.1219	0.8775
DEA-Net	TIP'24	6.619	0.06930	0.9426	9.244	0.1046	0.8346	6.355	0.07880	0.9004	4.718	0.05370	0.9798
DarkIR	CVPR'25	7.071	0.1117	0.8816	9.308	0.1558	0.8029	8.739	0.1690	0.8457	5.020	0.09950	0.9229
DehazeFormer	TIP'23	7.499	0.04810	0.9318	9.600	0.06230	0.8894	8.549	0.04630	0.8916	5.115	0.03320	0.9500
DnCNN	TIP'17	6.793	0.1309	0.8531	8.983	0.1488	0.7788	9.373	0.4280	0.7781	4.794	0.1358	0.8808
DocRes	CVPR'24	6.487	0.08420	0.9933	8.798	0.1134	0.9246	6.275	0.1324	0.8837	4.172	0.05060	1.012
HVI-CIDNet	CVPR'25	6.849	0.1434	0.8532	9.008	0.1592	0.7838	9.313	0.4229	0.7754	4.846	0.1418	0.8775
Retinexformer	ICCV'23	6.908	0.1367	0.8634	9.204	0.1686	0.7889	9.136	0.3654	0.7818	4.940	0.1353	0.8837
Generic Encoder–Decoder Models													
pix2pix	CVPR'17	6.223	0.04390	0.9528	8.503	0.07790	0.8452	6.193	0.1352	0.8746	3.930	0.005700	1.026
U-Net	MICCAI'15	6.567	0.1195	0.9398	9.002	0.1830	0.8180	6.087	0.1535	0.8676	4.349	0.1058	1.011
RSHRL (Ours)	This work	**13.15**	**0.5212**	**0.7523**	**15.93**	**0.6057**	**0.6454**	**15.95**	**0.5644**	**0.5314**	**15.12**	**0.6157**	**0.6251**

Analysis. Horizontal position changes shift the balance and placement of the observed wall pattern. RSHRL ranks first in all 12 averaged comparisons and continues to recover the main display regions across the sweep, extending the geometry study from changed viewing angle to displaced observation position.

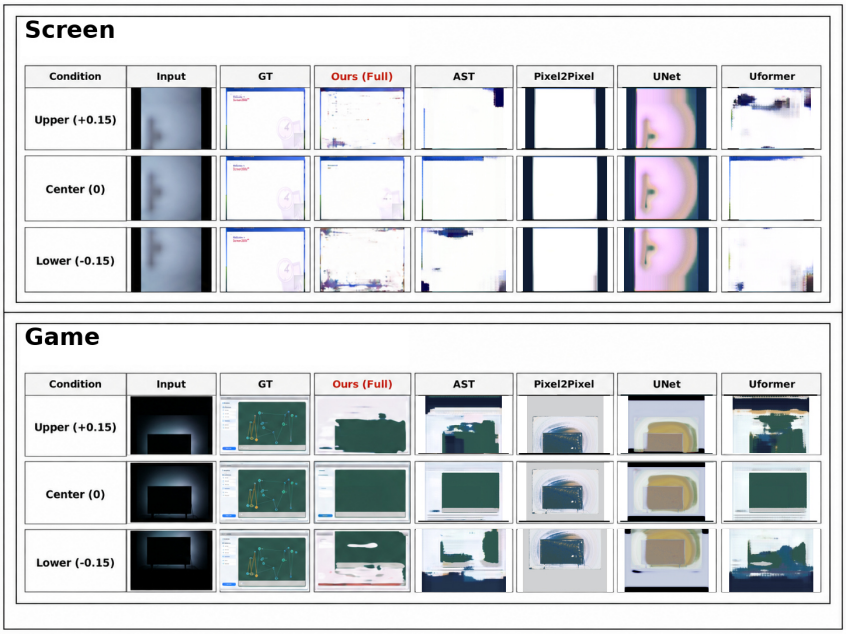

Figure A21: Qualitative recovery under vertical camera-position shifts.

The displayed positions show how the recovered images preserve the main screen partition as the wall evidence moves vertically.

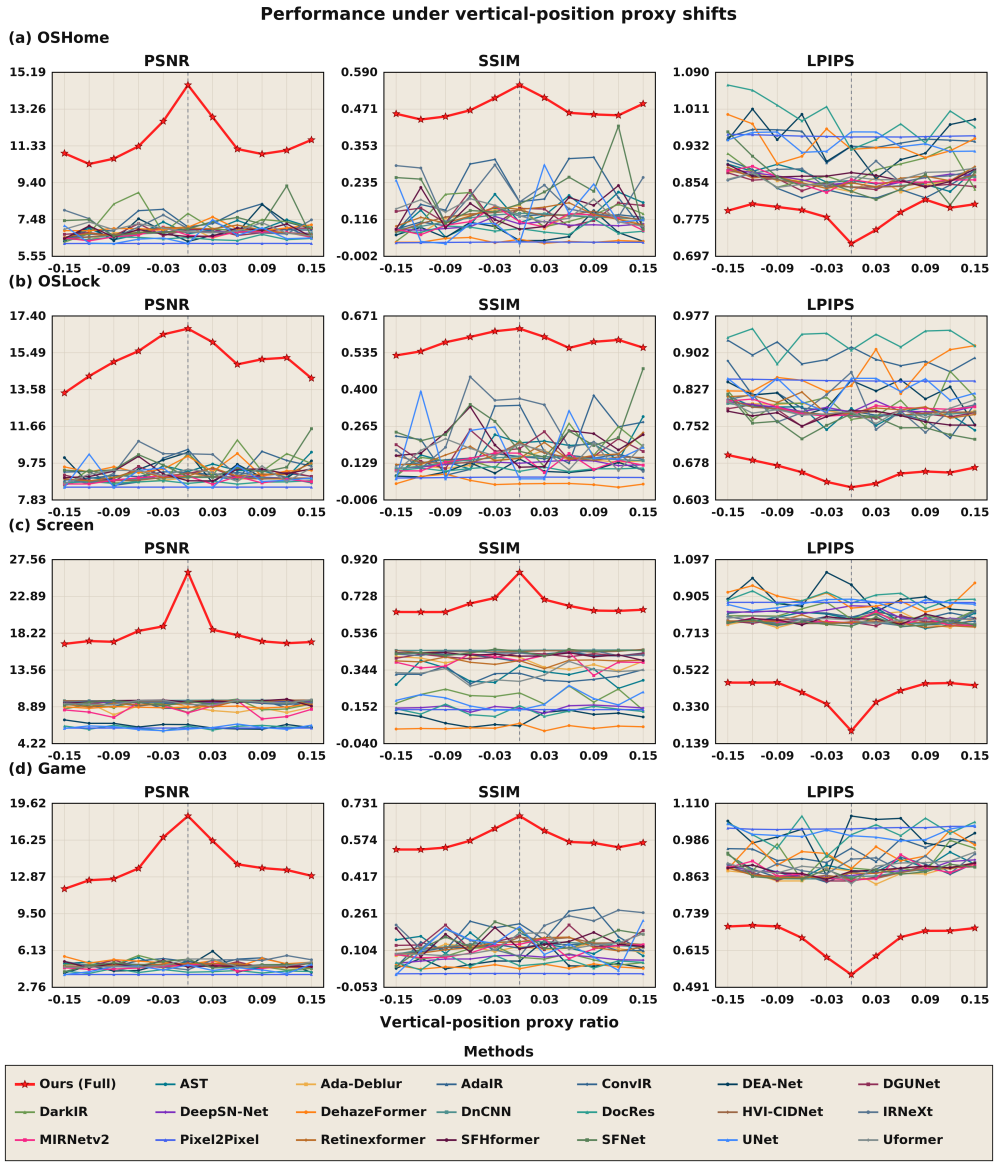

Figure A22: Robustness curves under eleven vertical camera-position shifts from -0.15 to $+0.15$.

The sweep moves the wall response upward and downward relative to the camera while retaining the same hidden target.

Moving the observation upward and downward produces a sharper departure from the reference condition than the corresponding distance sweep. The four datasets nevertheless retain their main screen partitions across the evaluated positions.

Table 15: Robustness results averaged over eleven vertical camera-position conditions. Best values are shown in red and second-best values in blue.

Method	Source	OSHome			OSLock			Screen			Game		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
General Purpose Image Restoration Frameworks													
AST	CVPR'24	6.955	0.1378	0.8667	9.272	0.1790	0.7792	9.499	0.3171	0.7903	4.824	0.1296	0.8968
AdaIR	ICLR'25	6.861	0.1217	0.9436	9.124	0.1470	0.8938	9.346	0.3146	0.8126	4.753	0.1152	0.9512
ConvIR	TPAMI'24	7.303	0.2044	0.8476	9.475	0.2310	0.7930	9.221	0.4220	0.7748	4.936	0.1723	0.8893
DGUNet	CVPR'22	6.854	0.1411	0.8522	9.142	0.1802	0.7780	9.515	0.4190	0.7741	4.815	0.1483	0.8746
DeepSN-Net	TPAMI'25	6.739	0.09840	0.8632	8.946	0.1275	0.7899	9.534	0.1461	0.8171	4.696	0.07000	0.8850
IRNeXt	ICML'23	7.181	0.2019	0.8697	9.514	0.2407	0.7904	9.463	0.4306	0.7789	5.028	0.1893	0.9015
MIRNetv2	TPAMI'23	6.672	0.1082	0.8602	8.863	0.1311	0.7893	8.457	0.3889	0.7754	4.581	0.1066	0.8843
SFHformer	ECCV'24	6.800	0.1387	0.8682	9.260	0.2026	0.7704	9.545	0.4225	0.7834	4.781	0.1419	0.8873
SFNet	ICLR'23	7.282	0.2019	0.8628	9.625	0.2531	0.7573	9.423	0.4375	0.7664	4.803	0.1439	0.8819
Uformer	CVPR'22	6.973	0.1416	0.8642	9.191	0.1647	0.7831	9.437	0.3254	0.7870	4.824	0.1235	0.8979
Degradation-Specific / Domain-Specific Restoration Models													
Ada-Deblur	TIP'23	6.739	0.1160	0.8571	8.890	0.1340	0.7867	8.776	0.3868	0.7750	4.638	0.1099	0.8739
DEA-Net	TIP'24	6.967	0.07660	0.9439	9.419	0.1054	0.8142	6.512	0.09360	0.8998	4.620	0.05320	1.004
DarkIR	CVPR'25	7.321	0.1298	0.8788	9.514	0.1673	0.7976	9.036	0.2049	0.8113	4.817	0.07800	0.9159
DehazeFormer	TIP'23	7.047	0.04720	0.9376	9.505	0.05870	0.8575	8.886	0.04330	0.8938	5.094	0.03430	0.9412
DnCNN	TIP'17	6.719	0.1187	0.8569	8.924	0.1394	0.7807	9.704	0.4474	0.7682	4.689	0.1254	0.8758
DocRes	CVPR'24	6.488	0.08380	0.9913	8.763	0.1074	0.9284	6.271	0.1314	0.8928	4.216	0.05250	1.015
HVI-CIDNet	CVPR'25	6.734	0.1245	0.8568	8.965	0.1519	0.7838	9.658	0.4416	0.7682	4.703	0.1319	0.8762
Retinexformer	ICCV'23	6.843	0.1300	0.8628	9.171	0.1632	0.7928	9.270	0.3883	0.7699	4.767	0.1217	0.8804
Generic Encoder–Decoder Models													
pix2pix	CVPR'17	6.222	0.04330	0.9535	8.503	0.07700	0.8455	6.194	0.1372	0.8740	3.928	0.005300	1.027
U-Net	MICCAI'15	6.629	0.1325	0.9403	9.030	0.1815	0.8263	6.248	0.1854	0.8652	4.426	0.1213	1.009
RSHRL (Ours)	This work	**11.65**	**0.4752**	**0.7887**	**15.17**	**0.5756**	**0.6598**	**18.43**	**0.6875**	**0.4038**	**14.18**	**0.5755**	**0.6530**

Analysis. RSHRL ranks first across all 12 vertical-position averages. Combined with the horizontal-position study, this result shows that the principal screen partition remains recoverable as the observation moves in either spatial direction.

C.6.2 Wall, Illumination and Camera Settings

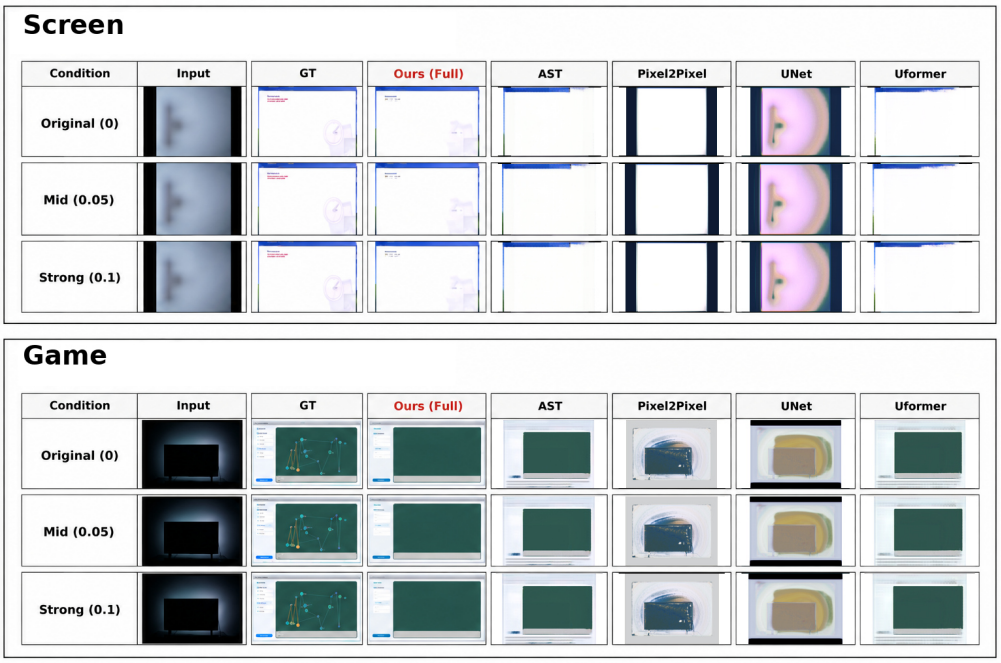

Figure A23: Qualitative reconstruction under wall-scattering changes.

The displayed conditions show how broad screen regions remain recognizable as the simulated relay response becomes more spatially spread.

The examples show that broad screen regions and their arrangement remain recognizable as the simulated wall response becomes more spread out.

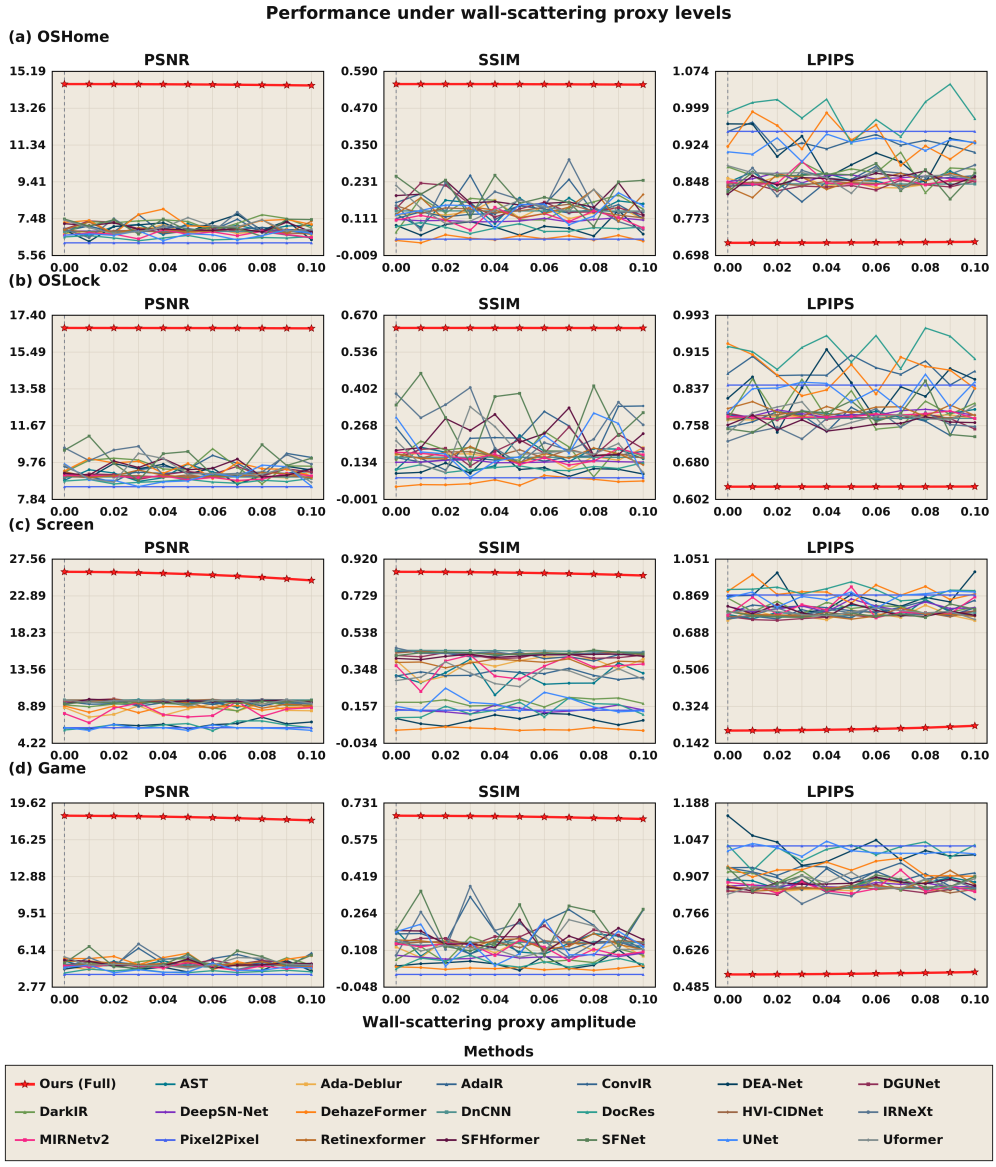

Figure A24: Robustness curves under eleven wall-scattering levels from 0 to 0.1.

The sweep increases the spread and mixing of screen light on the relay surface while retaining the same hidden target.

The curves change only modestly across the wall-scattering range. This stable metric response is consistent with the similar large-region arrangements visible in the corresponding examples.

Table 16: Robustness results for eleven wall-scattering levels. Best values are shown in red and second-best values in blue.

Method	Source	OSHome			OSLock			Screen			Game		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
General Purpose Image Restoration Frameworks													
AST	CVPR'24	6.977	0.1434	0.8535	9.166	0.1591	0.7811	9.403	0.3105	0.7921	4.905	0.1326	0.8829
AdaIR	ICLR'25	6.966	0.1312	0.9313	9.197	0.1514	0.8768	9.411	0.3239	0.8128	4.914	0.1282	0.9268
ConvIR	TPAMI'24	6.974	0.1634	0.8478	9.525	0.2359	0.7752	9.334	0.4227	0.7800	5.192	0.2022	0.8969
DGUNet	CVPR'22	6.963	0.1548	0.8407	9.184	0.1852	0.7749	9.616	0.4188	0.7709	4.938	0.1597	0.8588
DeepSN-Net	TPAMI'25	6.833	0.1057	0.8566	9.027	0.1371	0.7840	9.545	0.1398	0.8139	4.806	0.08210	0.8736
IRNeXt	ICML'23	6.931	0.1573	0.8592	9.713	0.2707	0.7597	9.509	0.4335	0.7827	5.038	0.1715	0.8702
MIRNetv2	TPAMI'23	6.675	0.1116	0.8509	9.010	0.1539	0.7788	8.216	0.3561	0.8122	4.657	0.1140	0.8678
SFHformer	ECCV'24	7.010	0.1745	0.8532	9.446	0.2328	0.7631	9.543	0.4178	0.7927	4.899	0.1600	0.8803
SFNet	ICLR'23	7.138	0.1908	0.8560	9.911	0.2998	0.7760	9.408	0.4325	0.7819	5.244	0.2074	0.8874
Uformer	CVPR'22	7.078	0.1483	0.8591	9.305	0.1830	0.7777	9.509	0.3174	0.7871	4.959	0.1338	0.8804
Degradation-Specific / Domain-Specific Restoration Models													
Ada-Deblur	TIP'23	6.821	0.1312	0.8475	9.040	0.1542	0.7764	8.459	0.3745	0.7754	4.704	0.1208	0.8632
DEA-Net	TIP'24	7.009	0.08170	0.9112	9.299	0.1110	0.8365	6.582	0.08860	0.8677	4.554	0.05360	1.014
DarkIR	CVPR'25	7.248	0.1434	0.8683	9.599	0.1706	0.7980	9.052	0.1808	0.8130	4.977	0.09660	0.8917
DehazeFormer	TIP'23	7.303	0.04650	0.9372	9.486	0.06340	0.8709	8.694	0.03950	0.8910	5.237	0.03210	0.9347
DnCNN	TIP'17	6.819	0.1360	0.8442	9.007	0.1527	0.7739	9.715	0.4439	0.7741	4.809	0.1416	0.8623
DocRes	CVPR'24	6.486	0.08060	0.9905	8.790	0.1137	0.9213	6.429	0.1409	0.8941	4.223	0.05580	1.004
HVI-CIDNet	CVPR'25	6.818	0.1374	0.8458	9.044	0.1640	0.7805	9.622	0.4332	0.7761	4.780	0.1398	0.8620
Retinexformer	ICCV'23	6.999	0.1442	0.8499	9.115	0.1577	0.7918	9.292	0.3842	0.7773	4.918	0.1383	0.8728
Generic Encoder–Decoder Models													
pix2pix	CVPR'17	6.223	0.04430	0.9515	8.504	0.07820	0.8447	6.194	0.1372	0.8741	3.930	0.005900	1.024
U-Net	MICCAI'15	6.642	0.1385	0.9240	9.034	0.1896	0.8288	6.155	0.1703	0.8651	4.437	0.1322	1.007
RSHRL (Ours)	This work	**14.50**	**0.5483**	**0.7247**	**16.73**	**0.6239**	**0.6286**	**25.57**	**0.8473**	**0.2129**	**18.29**	**0.6717**	**0.5372**

Analysis. The recovered screen organization remains stable across the wall-scattering range. The modest metric changes agree with the similar large-region arrangements in the examples, and RSHRL ranks first in all 12 averaged comparisons.

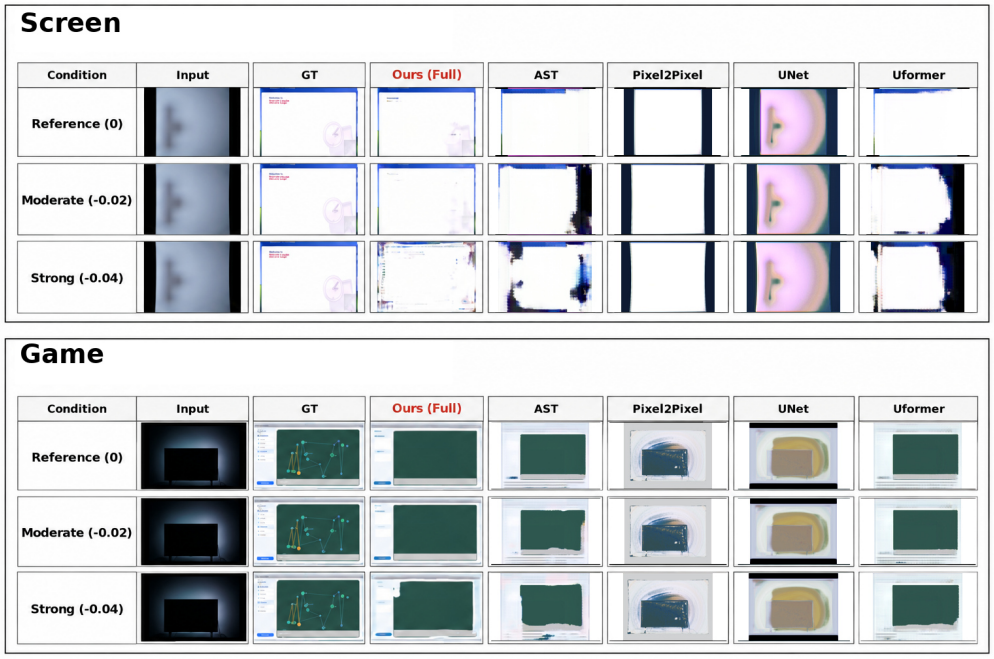

Figure A25: Qualitative reconstruction under lens-distortion changes.

The displayed distortions show how the principal screen regions retain their overall arrangement as local geometry bends.

The displayed reconstructions preserve the overall arrangement of the principal screen regions under the evaluated distortions.

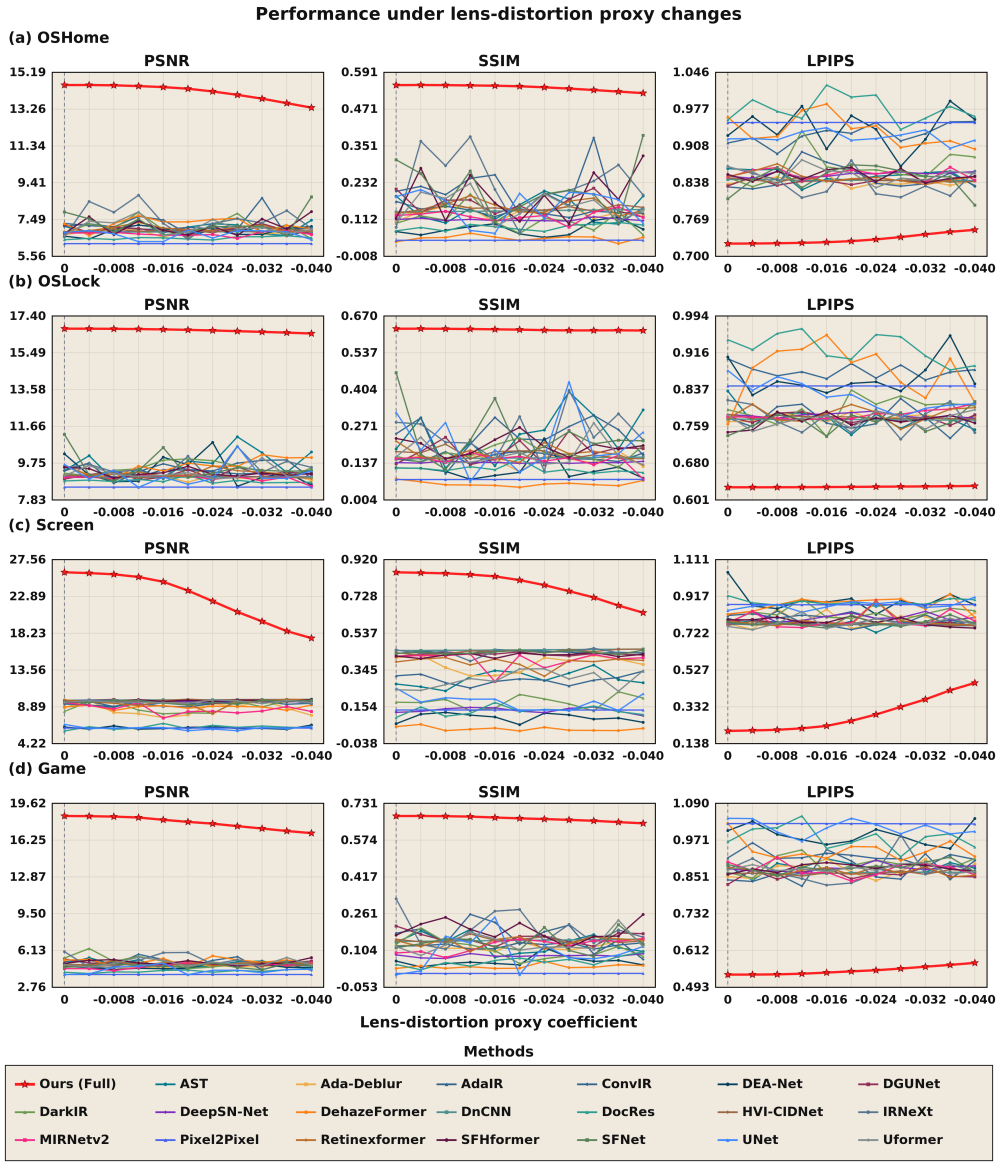

Figure A26: Robustness curves under eleven lens-distortion levels from 0 to -0.04.

The sweep changes local observation geometry and compares the resulting screen reconstruction across all four datasets.

The lens-distortion study changes local observation geometry rather than moving the complete wall pattern. RSHRL retains the main screen regions across the sweep, extending the geometry study from displaced evidence to locally deformed evidence.

Table 17: Robustness results averaged over eleven lens-distortion levels. Best values are shown in red and second-best values in blue.

Method	Source	OSHome			OSLock			Screen			Game		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
General Purpose Image Restoration Frameworks													
AST	CVPR'24	7.027	0.1455	0.8508	9.760	0.2432	0.7762	9.520	0.3015	0.7742	4.904	0.1323	0.8811
AdaIR	ICLR'25	7.050	0.1399	0.9300	9.237	0.1594	0.8764	9.220	0.2954	0.8025	4.860	0.1252	0.9168
ConvIR	TPAMI'24	7.240	0.2026	0.8477	9.465	0.2369	0.7762	9.503	0.4385	0.7787	5.022	0.1652	0.8701
DGUNet	CVPR'22	6.972	0.1597	0.8412	9.163	0.1810	0.7760	9.622	0.4280	0.7757	4.954	0.1581	0.8584
DeepSN-Net	TPAMI'25	6.854	0.1113	0.8543	9.048	0.1399	0.7819	9.562	0.1354	0.8099	4.858	0.08130	0.8793
IRNeXt	ICML'23	7.494	0.2287	0.8379	9.543	0.2476	0.7755	9.532	0.4377	0.7735	5.216	0.1953	0.8747
MIRNetv2	TPAMI'23	6.734	0.1235	0.8549	8.966	0.1476	0.7838	8.570	0.3983	0.7942	4.713	0.1246	0.8758
SFHformer	ECCV'24	7.115	0.1890	0.8518	9.252	0.2001	0.7703	9.635	0.4225	0.7834	5.041	0.1885	0.8805
SFNet	ICLR'23	7.269	0.2117	0.8459	9.540	0.2370	0.7687	9.498	0.4377	0.7782	4.717	0.1369	0.8821
Uformer	CVPR'22	7.086	0.1550	0.8544	9.218	0.1689	0.7731	9.517	0.2954	0.7756	4.974	0.1338	0.8780
Degradation-Specific / Domain-Specific Restoration Models													
Ada-Deblur	TIP'23	6.831	0.1324	0.8401	9.008	0.1498	0.7772	8.594	0.3778	0.7916	4.712	0.1237	0.8729
DEA-Net	TIP'24	6.854	0.08070	0.9405	9.358	0.1139	0.8618	6.234	0.09720	0.8939	4.621	0.05790	0.9859
DarkIR	CVPR'25	7.081	0.1088	0.8671	9.483	0.1727	0.7955	8.875	0.1796	0.8258	5.149	0.1077	0.8906
DehazeFormer	TIP'23	7.237	0.05090	0.9355	9.622	0.06300	0.8757	8.858	0.03940	0.8724	5.128	0.03520	0.9387
DnCNN	TIP'17	6.839	0.1403	0.8433	9.023	0.1562	0.7730	9.778	0.4505	0.7680	4.849	0.1471	0.8641
DocRes	CVPR'24	6.513	0.08720	0.9775	8.828	0.1159	0.9255	6.283	0.1327	0.8863	4.237	0.05710	0.9787
HVI-CIDNet	CVPR'25	6.880	0.1482	0.8434	9.082	0.1717	0.7779	9.704	0.4420	0.7692	4.849	0.1499	0.8638
Retinexformer	ICCV'23	7.007	0.1514	0.8491	9.246	0.1704	0.7840	9.407	0.3858	0.7854	4.931	0.1400	0.8690
Generic Encoder–Decoder Models													
pix2pix	CVPR'17	6.223	0.04440	0.9517	8.504	0.07840	0.8447	6.195	0.1372	0.8731	3.930	0.006000	1.024
U-Net	MICCAI'15	6.695	0.1543	0.9241	9.184	0.2118	0.8237	6.188	0.1787	0.8699	4.328	0.09440	1.011
RSHRL (Ours)	This work	**14.15**	**0.5409**	**0.7326**	**16.65**	**0.6207**	**0.6296**	**22.76**	**0.7850**	**0.2915**	**17.84**	**0.6655**	**0.5475**

Analysis. RSHRL ranks first in all 12 lens-distortion averages and retains the principal screen regions across the sweep. This extends the earlier geometry studies from displaced wall patterns to locally deformed observations.

Table 18: Robustness results averaged over eleven added-light direction conditions, each evaluated with added-light strength 0.22. Best values are shown in red and second-best values in blue.

Method	Source	OSHome			OSLock			Screen			Game		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
General Purpose Image Restoration Frameworks													
AST	CVPR'24	7.041	0.1485	0.8449	9.275	0.1798	0.7792	9.368	0.3049	0.7871	4.830	0.1240	0.8915
AdaIR	ICLR'25	7.017	0.1365	0.9250	9.182	0.1514	0.8812	9.320	0.3292	0.8199	4.875	0.1296	0.9225
ConvIR	TPAMI'24	7.227	0.1970	0.8652	9.605	0.2484	0.7666	9.336	0.4327	0.7766	5.014	0.1747	0.8729
DGUNet	CVPR'22	6.983	0.1618	0.8417	9.092	0.1628	0.7747	9.489	0.4234	0.7701	4.851	0.1480	0.8651
DeepSN-Net	TPAMI'25	6.833	0.1048	0.8557	9.035	0.1369	0.7822	9.522	0.1373	0.8287	4.815	0.07470	0.8837
IRNeXt	ICML'23	7.224	0.2055	0.8547	9.209	0.1859	0.7815	9.395	0.4346	0.7878	5.268	0.2207	0.8922
MIRNetv2	TPAMI'23	6.802	0.1300	0.8523	8.916	0.1371	0.7812	8.751	0.3934	0.7908	4.677	0.1163	0.8773
SFHformer	ECCV'24	7.017	0.1719	0.8524	9.258	0.1959	0.7739	9.454	0.4196	0.7799	4.911	0.1649	0.8821
SFNet	ICLR'23	7.489	0.2475	0.8385	9.600	0.2637	0.7820	9.398	0.4372	0.7729	5.072	0.2007	0.8992
Uformer	CVPR'22	7.142	0.1524	0.8569	9.421	0.2009	0.7755	9.385	0.2979	0.7826	4.961	0.1339	0.8816
Degradation-Specific / Domain-Specific Restoration Models													
Ada-Deblur	TIP'23	6.758	0.1235	0.8476	8.991	0.1479	0.7776	8.622	0.3710	0.7706	4.761	0.1275	0.8692
DEA-Net	TIP'24	6.991	0.08460	0.9257	9.296	0.1131	0.8505	6.210	0.09620	0.8781	4.279	0.03270	1.012
DarkIR	CVPR'25	7.159	0.1265	0.8637	9.818	0.1935	0.7864	9.098	0.1711	0.8085	4.992	0.09560	0.9062
DehazeFormer	TIP'23	7.296	0.04700	0.9242	9.554	0.06200	0.8815	8.708	0.04540	0.8608	5.090	0.03230	0.9542
DnCNN	TIP'17	6.824	0.1366	0.8445	9.013	0.1534	0.7739	9.699	0.4467	0.7696	4.811	0.1425	0.8638
DocRes	CVPR'24	6.490	0.08300	0.9870	8.833	0.1158	0.9245	6.275	0.1331	0.9000	4.191	0.05390	1.010
HVI-CIDNet	CVPR'25	6.848	0.1443	0.8446	9.074	0.1700	0.7793	9.610	0.4371	0.7733	4.802	0.1435	0.8622
Retinexformer	ICCV'23	7.032	0.1535	0.8555	9.150	0.1621	0.7970	9.373	0.3749	0.7764	4.901	0.1469	0.8726
Generic Encoder–Decoder Models													
pix2pix	CVPR'17	6.223	0.04420	0.9515	8.504	0.07830	0.8446	6.194	0.1372	0.8740	3.930	0.005900	1.024
U-Net	MICCAI'15	6.780	0.1728	0.9415	9.132	0.2051	0.8270	6.210	0.1788	0.8703	4.370	0.1122	1.008
RSHRL (Ours)	This work	**14.47**	**0.5485**	**0.7291**	**16.62**	**0.6217**	**0.6313**	**25.69**	**0.8499**	**0.2092**	**17.32**	**0.6440**	**0.5639**

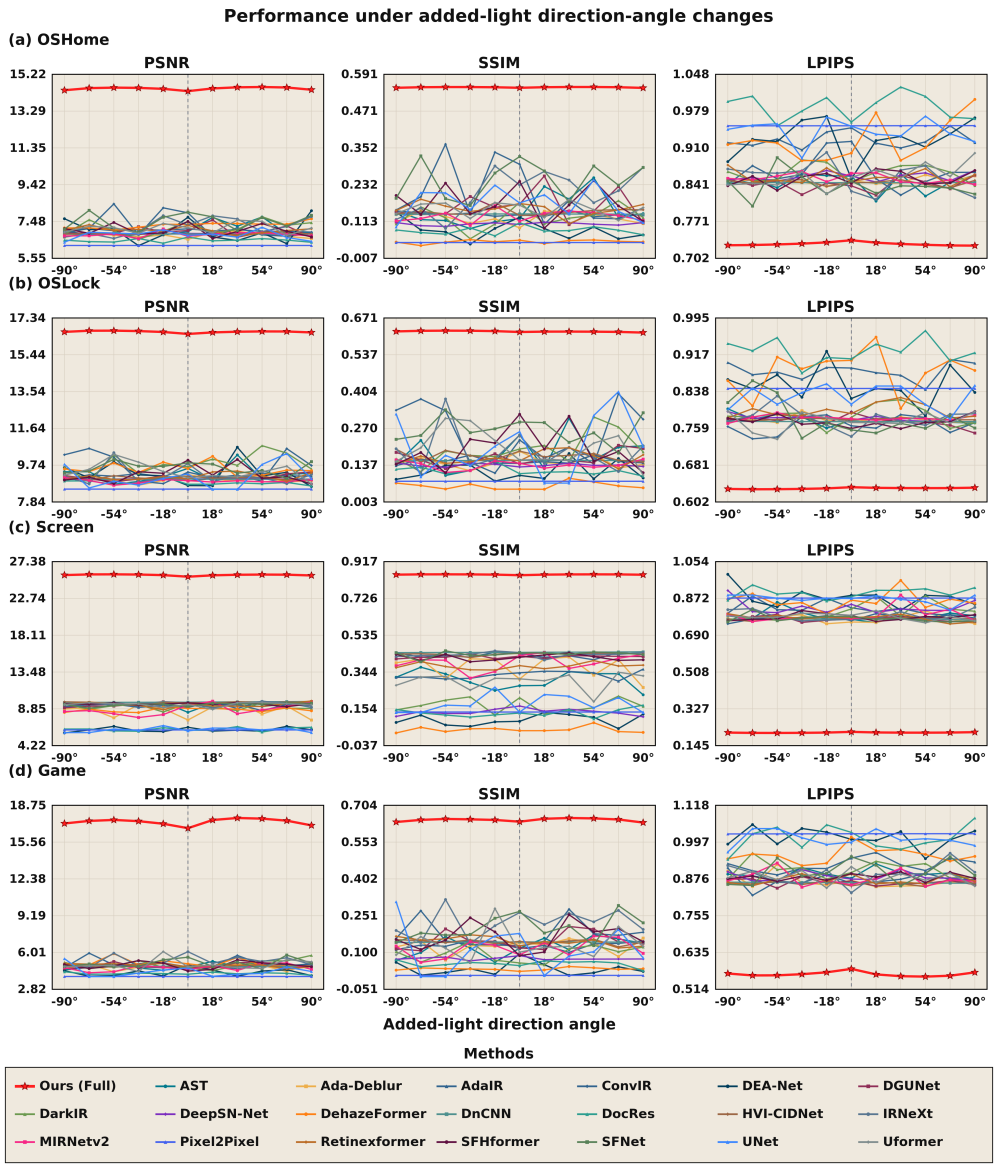

Figure A27: Robustness under changes in added-light direction.

The direction sweep holds added-light strength at 0.22. The 0° condition denotes frontal added light, providing the central direction within this illumination study.

Figure A27 changes the direction of added illumination while preserving the hidden screen target. Every direction, including 0°, uses strength 0.22, so the 0° point is frontal added light rather than a clean or no-light reference.

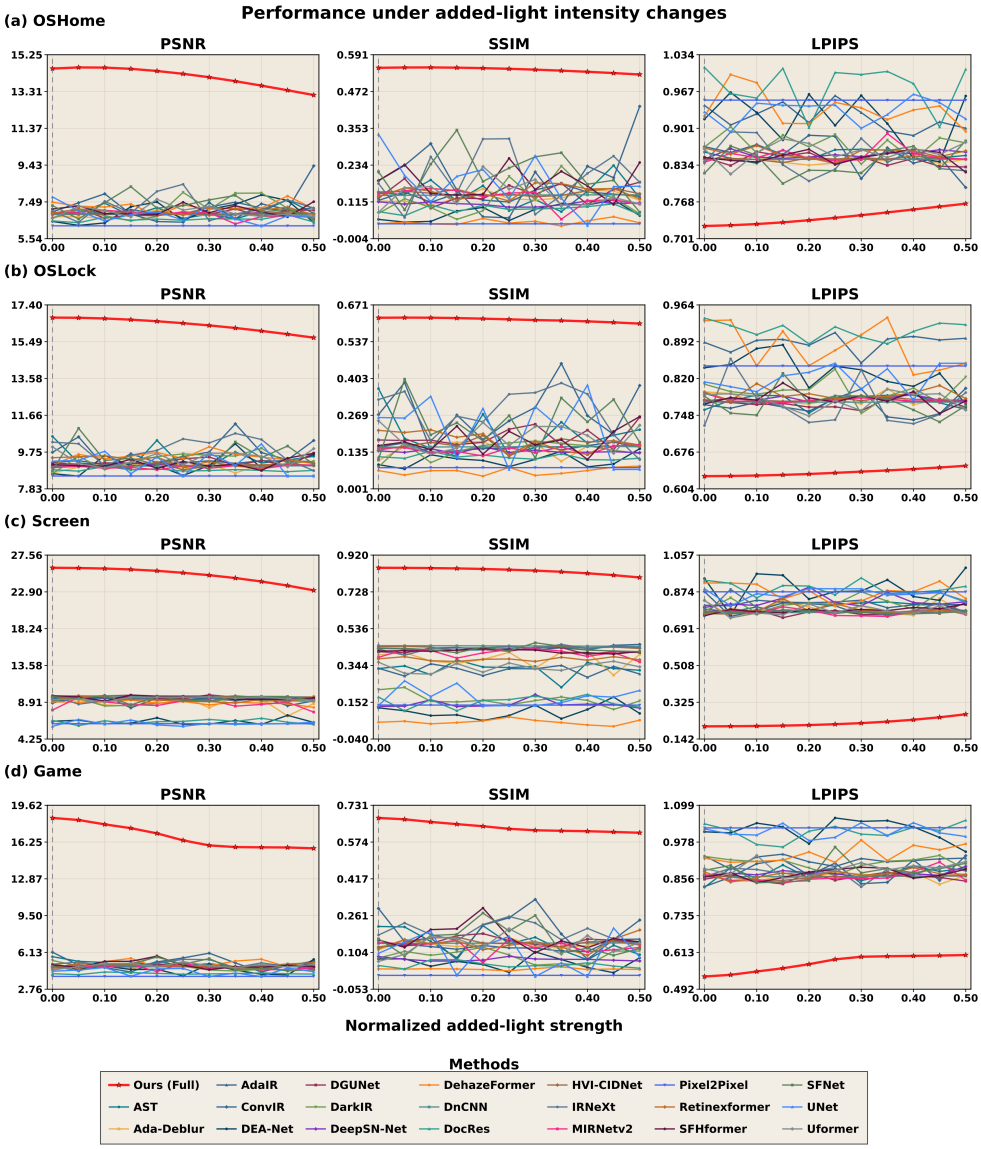

Figure A28: Robustness under changes in added-light strength.

The intensity sweep varies normalized added-light strength and reports reconstruction quality across the four datasets.

The direction curves remain closely grouped, while increasing added-light strength produces a clearer modulation of reconstruction quality. Together, the two sweeps separate where environmental light enters from how strongly it changes the observation.

Analysis. Table 18 examines where added light enters the observation; Table 19 then examines how reconstruction changes with its strength. Every direction in the first study uses strength 0.22, and RSHRL ranks first in all 12 direction-averaged comparisons.

Table 19: Robustness results averaged over eleven added-light intensity conditions. Best values are shown in red and second-best values in blue.

Method	Source	OSHome			OSLock			Screen			Game		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
General Purpose Image Restoration Frameworks													
AST	CVPR'24	6.970	0.1418	0.8523	9.436	0.1994	0.7759	9.441	0.3257	0.7970	5.005	0.1454	0.8736
AdaIR	ICLR'25	7.056	0.1371	0.9199	9.218	0.1527	0.8903	9.341	0.3191	0.8151	4.924	0.1294	0.9177
ConvIR	TPAMI'24	7.265	0.1971	0.8524	9.819	0.2816	0.7672	9.232	0.4325	0.7928	5.170	0.1943	0.8800
DGUNet	CVPR'22	6.897	0.1447	0.8431	9.253	0.1912	0.7701	9.534	0.4226	0.7728	4.890	0.1508	0.8600
DeepSN-Net	TPAMI'25	6.856	0.1089	0.8532	9.049	0.1380	0.7804	9.503	0.1415	0.8281	4.817	0.07400	0.8738
IRNeXt	ICML'23	7.224	0.1949	0.8478	9.890	0.2822	0.7684	9.321	0.4337	0.7880	5.059	0.1931	0.8776
MIRNetv2	TPAMI'23	6.814	0.1304	0.8521	8.967	0.1431	0.7779	8.684	0.4063	0.7718	4.766	0.1285	0.8627
SFHformer	ECCV'24	7.068	0.1800	0.8476	9.156	0.1786	0.7801	9.500	0.4188	0.7831	5.003	0.1672	0.8733
SFNet	ICLR'23	7.280	0.2062	0.8402	9.535	0.2344	0.7727	9.389	0.4301	0.7726	4.905	0.1611	0.8840
Uformer	CVPR'22	6.969	0.1399	0.8585	9.444	0.1913	0.7768	9.389	0.3343	0.7822	4.921	0.1356	0.8913
Degradation-Specific / Domain-Specific Restoration Models													
Ada-Deblur	TIP'23	6.814	0.1293	0.8428	8.996	0.1488	0.7792	8.822	0.3795	0.7889	4.835	0.1374	0.8628
DEA-Net	TIP'24	6.961	0.08150	0.9158	9.111	0.1046	0.8322	6.463	0.1024	0.8926	4.625	0.05910	1.016
DarkIR	CVPR'25	7.128	0.1228	0.8714	9.353	0.1438	0.7963	8.993	0.1664	0.8194	4.953	0.08870	0.9084
DehazeFormer	TIP'23	7.248	0.04970	0.9357	9.493	0.06540	0.8822	8.827	0.04870	0.8745	5.109	0.03280	0.9375
DnCNN	TIP'17	6.831	0.1373	0.8445	9.021	0.1543	0.7738	9.663	0.4459	0.7719	4.814	0.1432	0.8642
DocRes	CVPR'24	6.559	0.09170	0.9758	8.818	0.1171	0.9142	6.442	0.1541	0.8927	4.192	0.05090	1.011
HVI-CIDNet	CVPR'25	6.859	0.1447	0.8454	9.043	0.1634	0.7778	9.563	0.4373	0.7744	4.837	0.1456	0.8622
Retinexformer	ICCV'23	6.994	0.1469	0.8558	9.279	0.1749	0.7894	9.235	0.3766	0.7855	4.913	0.1405	0.8710
Generic Encoder–Decoder Models													
pix2pix	CVPR'17	6.223	0.04430	0.9514	8.504	0.07830	0.8446	6.194	0.1371	0.8740	3.930	0.005900	1.025
U-Net	MICCAI'15	6.827	0.1777	0.9317	9.172	0.2163	0.8213	6.216	0.1783	0.8626	4.367	0.1004	1.012
RSHRL (Ours)	This work	**14.08**	**0.5426**	**0.7409**	**16.36**	**0.6168**	**0.6364**	**25.00**	**0.8380**	**0.2230**	**16.77**	**0.6389**	**0.5781**

Analysis. The illumination studies separate where added light enters from how strongly it changes the observation. RSHRL ranks first across all 12 averages in both studies; the direction curves remain closely grouped, while the intensity sweep produces a clearer change in reconstruction quality.

C.6.3 Brightness Variation

The brightness-variation table reports arithmetic means over five conditions for each dataset and metric: the transform-free Clean-00 reference together with brightness factors 0.8, 0.9, 1.1, and 1.2. Factor 1.0 denotes the reference brightness on the plotted axis.

Table 20: Quantitative comparison on OSHome, OSLock, Screen, and Game. Best values are shown in red and second-best values are shown in blue.

Method	Source	OSHome			OSLock			Screen			Game		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
AST	CVPR'24	10.57	0.3538	0.8007	12.76	0.3977	0.7174	13.21	0.4855	0.6525	11.97	0.4507	0.7256
AdaIR	ICLR'25	6.940	0.1272	0.9324	9.200	0.1519	0.8812	9.180	0.3034	0.8391	4.940	0.1327	0.9129
ConvIR	TPAMI'24	7.790	0.2667	0.8382	9.560	0.2377	0.7765	9.400	0.4339	0.7718	5.210	0.2138	0.8733
DGUNet	CVPR'22	6.870	0.1450	0.8461	9.040	0.1621	0.7800	9.500	0.4197	0.7658	4.840	0.1448	0.8608
DeepSN-Net	TPAMI'25	6.830	0.1062	0.8518	9.030	0.1383	0.7822	9.450	0.1369	0.7989	4.790	0.07520	0.8783
IRNeXt	ICML'23	7.580	0.2540	0.8424	9.880	0.2838	0.7676	9.280	0.4287	0.7755	5.050	0.1831	0.8753
MIRNetv2	TPAMI'23	6.750	0.1232	0.8566	8.960	0.1423	0.7759	8.830	0.4021	0.8002	4.720	0.1205	0.8712
SFHformer	ECCV'24	8.800	0.3359	0.8515	11.27	0.4005	0.7483	9.260	0.4217	0.7900	8.040	0.3496	0.7994
SFNet	ICLR'23	6.990	0.1588	0.8539	9.140	0.1676	0.7677	9.200	0.4369	0.7770	4.800	0.1588	0.9197
Uformer	CVPR'22	11.42	0.3875	0.7844	12.89	0.4281	0.7177	13.08	0.4798	0.6474	11.85	0.4256	0.7286
Degradation-Specific / Domain-Specific Restoration Models													
Ada-Deblur	TIP'23	6.790	0.1271	0.8431	9.040	0.1480	0.7740	8.540	0.3756	0.7874	4.650	0.1124	0.8534
DEA-Net	TIP'24	6.610	0.07570	0.9430	9.130	0.09660	0.8485	6.440	0.09680	0.8800	4.380	0.04800	1.012
DarkIR	CVPR'25	6.580	0.06640	0.8902	8.790	0.09480	0.8044	9.160	0.1748	0.8649	4.460	0.04070	0.9367
DehazeFormer	TIP'23	7.190	0.04650	0.9313	9.560	0.05490	0.9007	8.710	0.04510	0.8944	5.310	0.03330	0.9609
DnCNN	TIP'17	8.170	0.2976	0.8109	10.70	0.3516	0.7420	9.190	0.3885	0.7415	7.210	0.3149	0.7792
DocRes	CVPR'24	6.530	0.08830	0.9720	8.800	0.1129	0.9149	6.410	0.1580	0.8829	4.150	0.04430	1.011
HVI-CIDNet	CVPR'25	6.860	0.1446	0.8449	9.080	0.1710	0.7773	9.550	0.4338	0.7676	4.770	0.1370	0.8667
Retinexformer	ICCV'23	6.870	0.1306	0.8588	9.200	0.1718	0.7961	9.470	0.3688	0.7845	5.100	0.1631	0.8644
Generic Encoder-Decoder Models													
pix2pix	CVPR'17	8.720	0.2476	0.8403	10.83	0.3432	0.7751	8.680	0.3398	0.7513	8.080	0.2764	0.8247
U-Net	MICCAI'15	7.850	0.2682	0.8678	10.76	0.3409	0.7891	7.960	0.2399	0.8281	7.870	0.3203	0.8612
RSHRL (Ours)	This work	**14.41**	**0.5501**	**0.7285**	**16.46**	**0.6181**	**0.6312**	**20.48**	**0.7082**	**0.3824**	**17.43**	**0.6477**	**0.5648**

RSHRL leads all 12 dataset–metric averages across the five brightness conditions. On Screen, its margins over the strongest competing averages are 7.2664 dB PSNR, 0.2226 SSIM, and 0.2651 LPIPS; the result summarizes reconstruction under both darker and brighter wall observations without repeating the per-condition curves below.

C.6.4 Qualitative Brightness Comparison

Figure A29 follows Screen and Game reconstructions from darker observations through the reference condition to brighter observations.

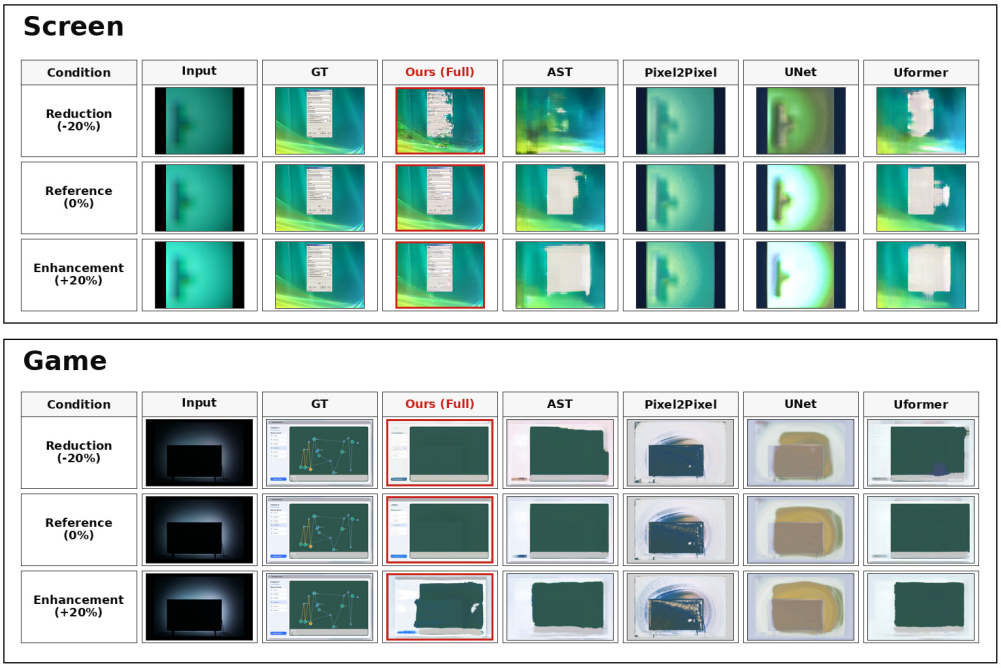

Figure A29: Qualitative brightness comparison on Screen and Game at factors of 0.80, 1.00, and 1.20. Each group shows the wall observation, ground truth, and reconstruction outputs.

The Screen example retains its main window organization across the three brightness settings. The Game example retains the large central panel and its surrounding interface arrangement. Together, they show that changes in wall-light intensity still permit recovery of the main display organization.

The brightness-variation table summarizes the five brightness settings, and the Gaussian-noise table summarizes the clean observation and five Gaussian-noise levels. Together, they examine screen recovery as the wall signal changes in intensity and becomes contaminated by noise.

Appendix Figure A30 shows the Screen reconstruction metrics across the brightness sweep.

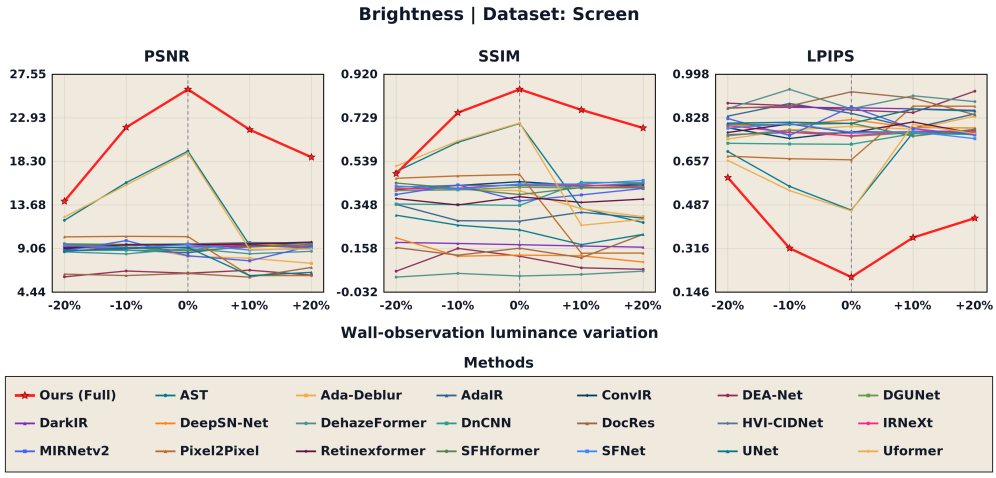

Figure A30: Screen PSNR, SSIM, and LPIPS curves across the evaluated brightness settings. The highlighted RSHRL curve shows its reconstruction trend relative to the compared methods.

Brightness variation tests whether the hidden screen can still be reconstructed as its wall reflection becomes darker or brighter. RSHRL retains a clear comparative advantage across the displayed sweep, with the strongest Screen values near the reference brightness. The examples show the corresponding visual outcome: the main window and panel arrangement remain recoverable in both directions of intensity change.

C.6.5 Gaussian-Noise Variation

The complete Gaussian-noise comparison is reported in Table 3 in the main paper. Each entry is the arithmetic mean over the clean condition and the five Gaussian-noise conditions at 40, 35, 30, 25, and 20 dB.

Figure A31 presents the corresponding Screen Gaussian-noise curves.

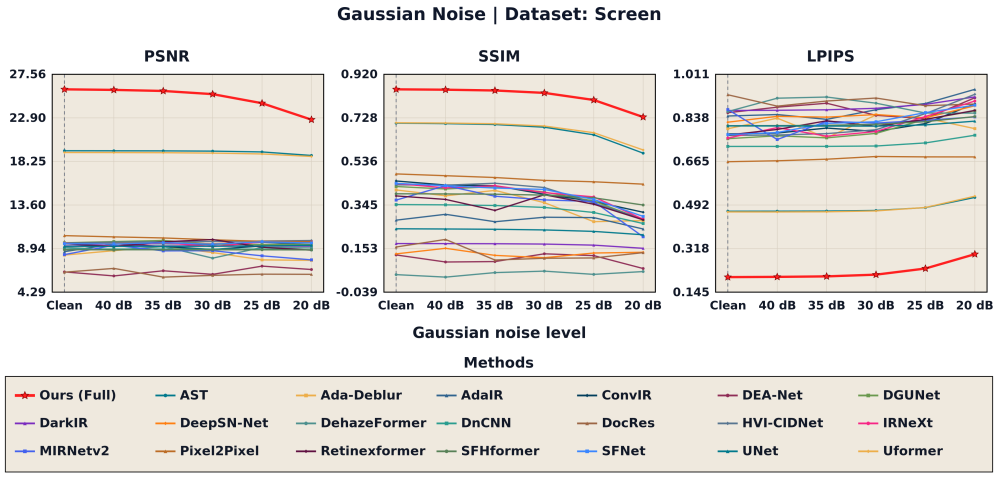

Figure A31: Screen PSNR, SSIM, and LPIPS curves across the evaluated Gaussian-noise conditions. The highlighted curve shows RSHRL relative to the compared methods as the noise level increases.

Added noise competes with the weak spatial variations that carry screen information. RSHRL retains its comparative advantage as noise increases, extending the brightness study from changed signal strength to wall observations containing additional random variation.

The corresponding qualitative Gaussian-noise reconstructions are presented in the main paper in Figure 4.

C.6.6 Motion Blur and Camera Shake

Figure A32 shows the Screen motion-blur curves.

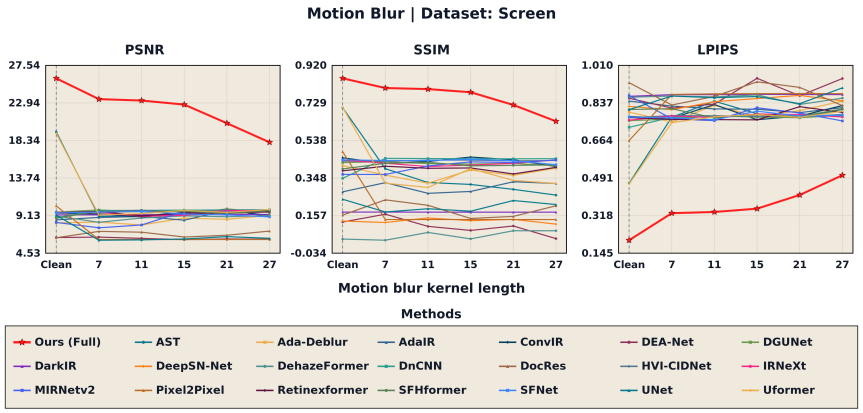

Figure A32: Screen PSNR, SSIM, and LPIPS across the evaluated motion-blur settings.

As motion blur increases, Screen PSNR and SSIM decrease and LPIPS rises, while RSHRL remains favorably positioned across the displayed settings. The trend connects progressive loss of thin strokes and boundaries to the survival of broader screen regions.

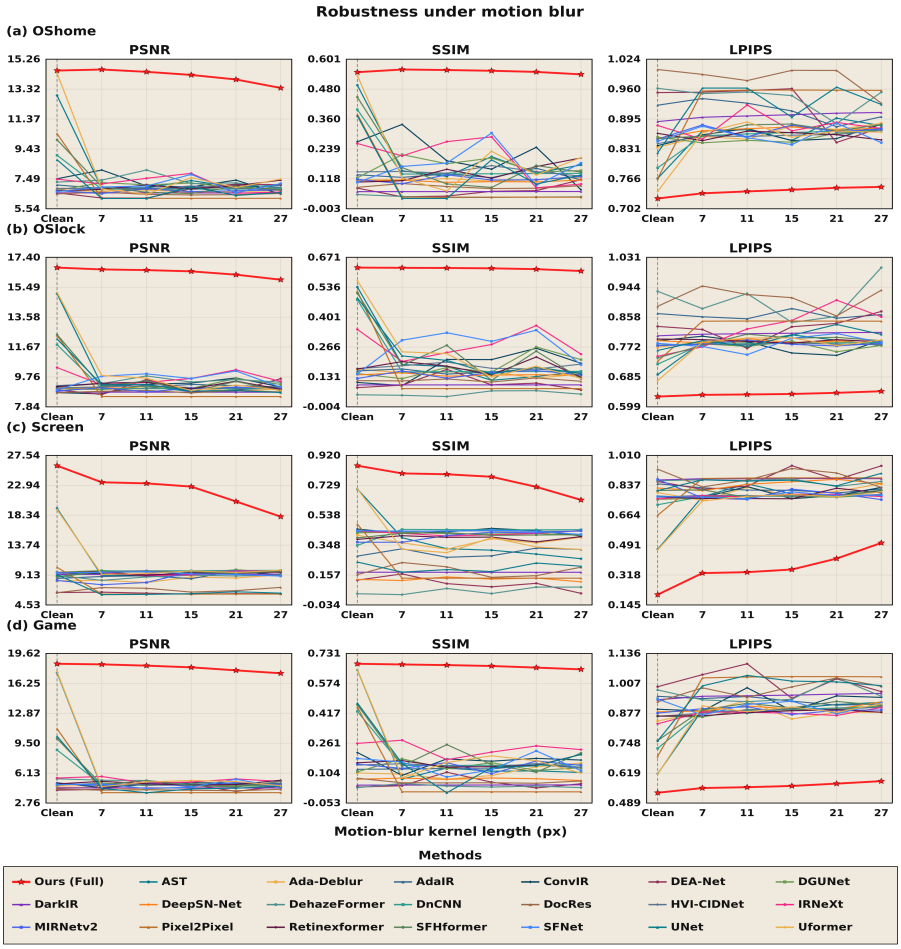

Figure A33: Robustness curves under the transform-free clean observation and five motion-blur conditions with kernel sizes 7, 11, 15, 21, and 27 at fixed angle 0°.

Figure A33 extends the Screen-only view to all four content datasets. Thin structures weaken with larger kernels, while the principal screen regions remain visible across the sweep.

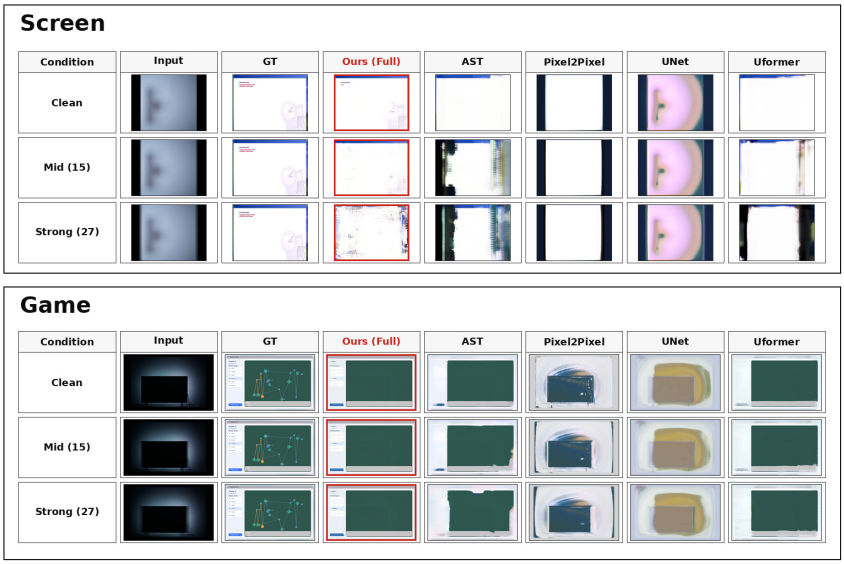

Figure A34: Qualitative motion-blur comparison.

As the blur kernel grows, small strokes and thin boundaries merge. Figure A34 shows that RSHRL preserves the main screen regions and recognizable interface elements across the displayed settings.

Table 21: Robustness results averaged over the clean observation and five motion-blur conditions. Best values are shown in red and second-best values in blue.

Method	**Source**	OSHome			OSLock			Screen			Game		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
General Purpose Image Restoration Frameworks													
AST	CVPR'24	8.016	0.2165	0.8561	10.05	0.2178	0.7729	11.07	0.3821	0.7408	6.872	0.1975	0.8542
AdaIR	ICLR'25	6.966	0.1325	0.9144	9.207	0.1569	0.8638	9.011	0.3003	0.8154	4.889	0.1303	0.9338
ConvIR	TPAMI'24	7.241	0.2123	0.8600	9.149	0.1782	0.7788	9.354	0.4344	0.7884	4.863	0.1670	0.9262
DGUNet	CVPR'22	7.009	0.1696	0.8553	9.182	0.1783	0.7757	9.641	0.4215	0.7691	4.799	0.1374	0.8869
DeepSN-Net	TPAMI'25	6.829	0.1089	0.8706	9.041	0.1403	0.7886	9.601	0.1301	0.8406	4.805	0.07310	0.8999
IRNeXt	ICML'23	7.196	0.2011	0.8818	9.725	0.2767	0.8258	9.469	0.4241	0.7742	5.354	0.2330	0.8743
MIRNetv2	TPAMI'23	6.708	0.1175	0.8693	8.976	0.1429	0.7805	8.819	0.4062	0.7915	4.869	0.1418	0.8892
SFHformer	ECCV'24	7.320	0.1850	0.8720	9.712	0.2419	0.7854	9.481	0.4135	0.7920	5.771	0.2205	0.8857
SFNet	ICLR'23	7.030	0.1718	0.8589	9.597	0.2651	0.7828	9.367	0.4347	0.7757	4.781	0.1474	0.9086
Uformer	CVPR'22	8.272	0.2156	0.8477	10.28	0.2466	0.7706	11.16	0.3968	0.7226	7.077	0.2256	0.8523
Degradation-Specific / Domain-Specific Restoration Models													
Ada-Deblur	TIP'23	6.701	0.1159	0.8638	8.999	0.1487	0.7825	8.543	0.3741	0.7983	4.607	0.1101	0.8792
DEA-Net	TIP'24	6.694	0.07140	0.9241	9.081	0.1021	0.8275	6.335	0.1031	0.8787	4.543	0.05140	1.011
DarkIR	CVPR'25	6.573	0.06670	0.9022	8.785	0.09500	0.8101	9.249	0.1748	0.8741	4.462	0.04150	0.9534
DehazeFormer	TIP'23	7.331	0.04630	0.9416	9.455	0.05520	0.9073	8.914	0.05690	0.8598	5.084	0.03470	0.9604
DnCNN	TIP'17	7.188	0.1820	0.8532	9.477	0.2089	0.7723	9.654	0.4298	0.7659	5.470	0.1923	0.8664
DocRes	CVPR'24	6.511	0.08670	0.9837	8.820	0.1157	0.9119	6.870	0.1850	0.8818	4.244	0.05830	0.9796
HVI-CIDNet	CVPR'25	6.868	0.1472	0.8588	9.110	0.1788	0.7793	9.672	0.4361	0.7763	4.819	0.1442	0.8881
Retinexformer	ICCV'23	6.978	0.1438	0.8600	9.227	0.1728	0.7896	9.447	0.3933	0.7770	5.001	0.1577	0.8794
Generic Encoder–Decoder Models													
pix2pix	CVPR'17	6.923	0.1000	0.9255	9.154	0.1515	0.8270	6.896	0.1947	0.8429	5.131	0.07990	0.9774
U-Net	MICCAI'15	6.856	0.1461	0.9229	9.537	0.2307	0.7959	6.776	0.2057	0.8569	5.347	0.1840	0.9704
RSHRL (Ours)	This work	**14.19**	**0.5518**	**0.7394**	**16.45**	**0.6193**	**0.6360**	**22.32**	**0.7661**	**0.3560**	**18.04**	**0.6650**	**0.5611**

Analysis. Motion blur merges thin strokes and boundaries into broader responses, while the examples retain the principal screen regions and their arrangement. Averaged over the clean observation and five blur settings, RSHRL achieves the strongest reconstruction quality across all 12 dataset–metric comparisons.

Figure A35 presents the Screen camera-shake curves.

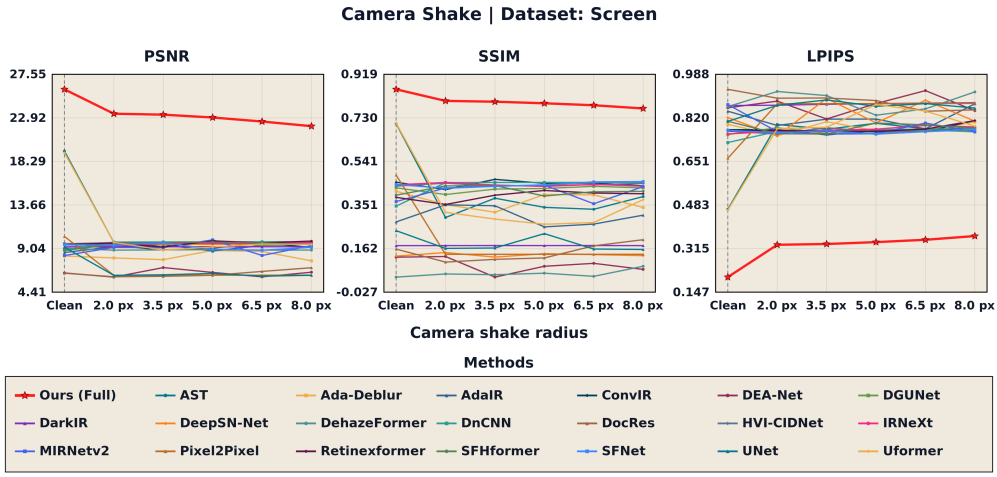

Figure A35: Screen PSNR, SSIM, and LPIPS curves across the evaluated camera-shake settings. The highlighted curve shows RSHRL relative to the compared methods as the camera-shake intensity increases.

As camera-shake intensity increases, Screen PSNR and SSIM decrease and LPIPS rises while RSHRL maintains a strong relative position. The change reflects increasing spatial inconsistency and displacement of weak strokes and edges.

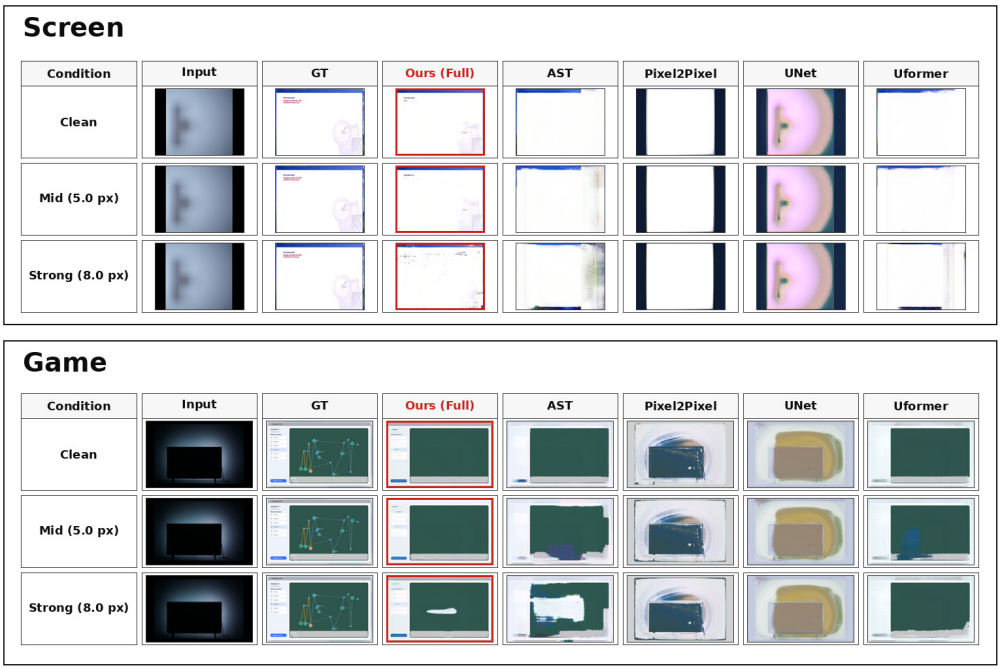

Figure A36: Qualitative camera-shake comparison.

The examples show how motion shifts blur and misplace thin screen structures while the main arrangement and salient regions remain recognizable.

The examples connect the metric trend to visible displacement of fine structures and continued recovery of the principal screen regions and their arrangement.

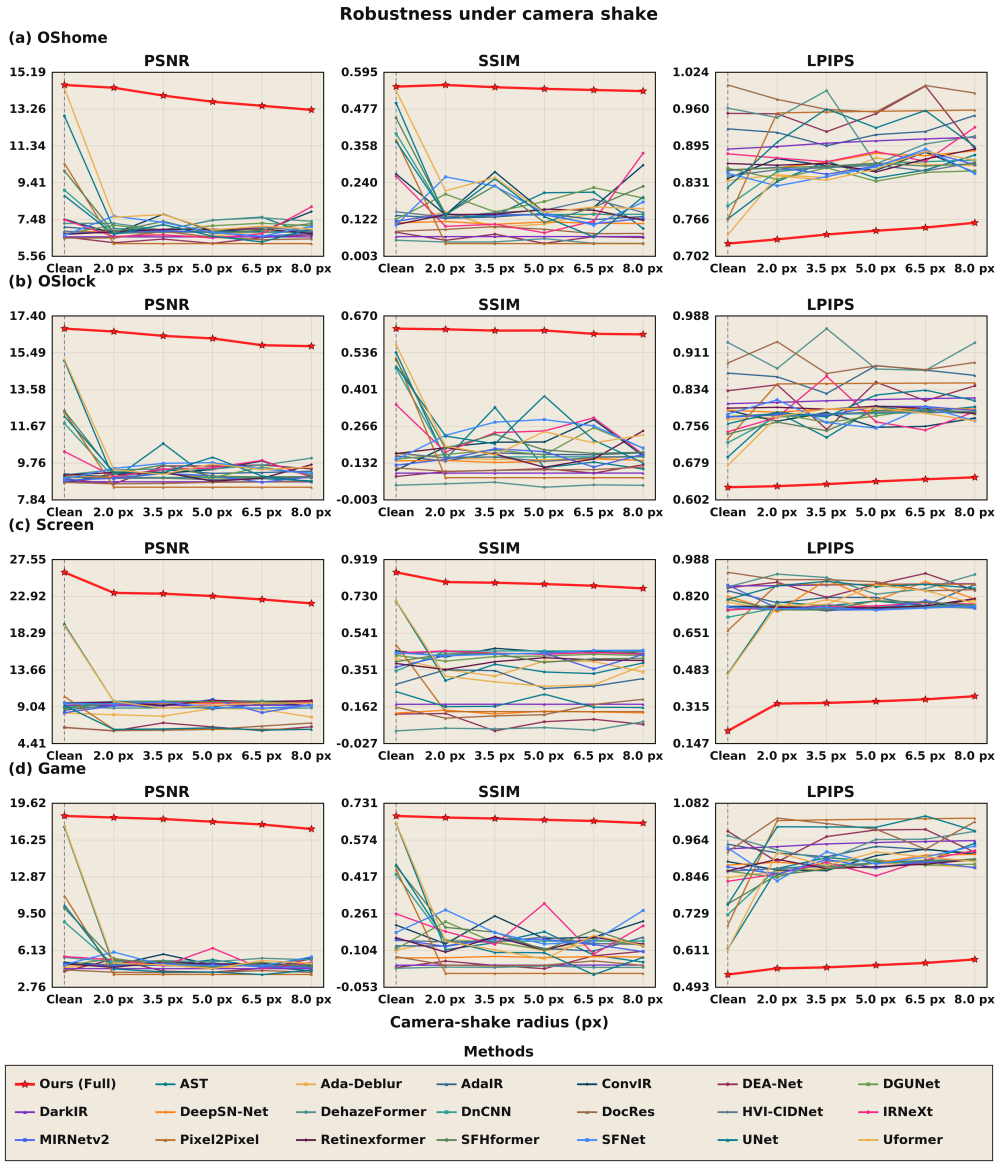

Figure A37: Robustness curves under the transform-free clean observation and five camera-shake conditions with radii 2.0, 3.5, 5.0, 6.5, and 8.0 pixels using 11 motion steps.

Figure A37 extends the Screen curve to all four content datasets and shows how increasing spatial inconsistency affects the three reconstruction metrics.

Camera shake introduces spatial inconsistency and weak edge displacement. Across the evaluated radii, the comparative trend remains favorable while the degradation increasingly affects thin screen structures.

Table 22: Robustness results averaged over the clean observation and five camera-shake conditions. Best values are shown in red and second-best values in blue.

Method	Source	OSHome			OSLock			Screen			Game		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
General Purpose Image Restoration Frameworks													
AST	CVPR'24	8.090	0.2114	0.8399	10.23	0.2297	0.7653	11.24	0.4075	0.7303	6.855	0.2017	0.8571
AdaIR	ICLR'25	6.941	0.1306	0.9207	9.246	0.1603	0.8625	9.161	0.3017	0.8195	4.897	0.1299	0.9307
ConvIR	TPAMI'24	7.275	0.2135	0.8595	9.256	0.1937	0.7711	9.262	0.4432	0.7749	5.106	0.1895	0.8996
DGUNet	CVPR'22	7.065	0.1776	0.8472	9.155	0.1830	0.7796	9.567	0.4209	0.7661	4.860	0.1429	0.8796
DeepSN-Net	TPAMI'25	6.830	0.1107	0.8707	9.028	0.1403	0.7910	9.531	0.1344	0.8268	4.818	0.07490	0.8976
IRNeXt	ICML'23	7.025	0.1662	0.8841	9.549	0.2452	0.7805	9.551	0.4389	0.7748	5.234	0.1957	0.8771
MIRNetv2	TPAMI'23	6.773	0.1268	0.8606	8.994	0.1514	0.7870	9.035	0.4086	0.7864	4.778	0.1339	0.8804
SFHformer	ECCV'24	7.531	0.2197	0.8597	9.751	0.2417	0.7700	9.472	0.4124	0.7825	5.771	0.2105	0.8627
SFNet	ICLR'23	7.032	0.1693	0.8515	9.438	0.2340	0.7794	9.251	0.4402	0.7668	5.107	0.1975	0.9068
Uformer	CVPR'22	8.457	0.2448	0.8380	10.40	0.2679	0.7651	10.99	0.3724	0.7551	6.989	0.2093	0.8587
Degradation-Specific / Domain-Specific Restoration Models													
Ada-Deblur	TIP'23	6.762	0.1215	0.8527	9.013	0.1516	0.7809	8.241	0.3696	0.7825	4.740	0.1268	0.8723
DEA-Net	TIP'24	6.460	0.06420	0.9444	9.324	0.1034	0.8209	6.421	0.09090	0.8676	4.626	0.05610	0.9637
DarkIR	CVPR'25	6.574	0.06670	0.9010	8.785	0.09500	0.8106	9.227	0.1751	0.8728	4.463	0.04120	0.9516
DehazeFormer	TIP'23	7.317	0.05060	0.9289	9.521	0.05220	0.9090	8.919	0.05340	0.8825	5.259	0.03210	0.9562
DnCNN	TIP'17	7.188	0.1811	0.8501	9.478	0.2083	0.7726	9.601	0.4321	0.7653	5.468	0.1908	0.8611
DocRes	CVPR'24	6.557	0.08600	0.9804	8.748	0.1088	0.8899	6.393	0.1458	0.8841	4.187	0.04950	0.9892
HVI-CIDNet	CVPR'25	6.881	0.1503	0.8553	8.975	0.1534	0.7824	9.621	0.4393	0.7710	4.796	0.1419	0.8814
Retinexformer	ICCV'23	6.869	0.1344	0.8670	9.191	0.1721	0.7932	9.602	0.3918	0.7769	4.844	0.1355	0.8849
Generic Encoder–Decoder Models													
pix2pix	CVPR'17	6.923	0.1000	0.9243	9.154	0.1515	0.8271	6.893	0.1946	0.8396	5.131	0.08000	0.9741
U-Net	MICCAI'15	7.095	0.1920	0.9105	9.779	0.2743	0.7981	6.721	0.1855	0.8599	5.215	0.1473	0.9686
RSHRL (Ours)	This work	**13.86**	**0.5441**	**0.7424**	**16.27**	**0.6145**	**0.6383**	**23.35**	**0.8012**	**0.3199**	**17.97**	**0.6625**	**0.5603**

Analysis. Camera shake makes the wall observation spatially inconsistent and displaces weak edges. Across the clean condition and five shake radii, RSHRL leads all 12 averaged comparisons; the examples connect this advantage to preservation of the principal screen regions and their arrangement.

C.7 Data-Limited Representation Evidence

The final study reduces the available paired training data to 50% and evaluates whether RWIR continues to organize dispersed wall evidence when fewer paired layouts are available.

C.7.1 RWIR with 50% Paired Data

Table 23: Representation comparison with 50% of the available training data. Best values are shown in red and second-best values in blue. Ties at the displayed precision receive the same ranking mark.

Representation	OSHome			OSLock			Screen			Game		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
GhostNetV2 DFC-Ghost	10.46	0.3872	0.8427	10.49	0.3214	0.7953	9.228	0.3648	0.8091	9.899	0.3315	0.8216
MEGA	10.23	0.3506	0.8375	10.24	0.3653	0.8063	9.274	0.3488	0.8276	9.898	0.3297	0.8300
Stripformer Intra-Strip	10.76	0.3521	0.8561	10.95	0.3681	0.8248	9.207	0.3134	0.8545	9.995	0.3124	0.8271
Stripformer Inter-Strip	10.77	0.3582	0.8594	10.94	0.3704	0.8223	9.212	0.3144	0.8600	10.01	0.3091	0.8253
Full RWIR (Ours)	**11.27**	**0.4180**	**0.8116**	**11.21**	**0.4521**	**0.7594**	**10.35**	**0.4153**	**0.7508**	**10.92**	**0.3533**	**0.7646**

Analysis. With half of the paired training data, RWIR leads all four content datasets in PSNR, SSIM, and LPIPS. Its comparative advantage therefore persists when fewer wall–screen examples are available, extending the representation result to reduced-data learning.