# RECOVERABLE EVIDENCE IN WALL REFLECTIONS: STRUCTURE-PROTECTED DEGRADATION REPRESENTATION LEARNING FOR SCREEN CONTENT RECOVERY

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

## ABSTRACT

Can an ordinary camera recover the principal content and structure of a hidden screen when it sees only the screen's blurry reflection on a diffuse wall? The core challenge is that spatial propagation and surface reflection weaken text, boundaries, and color while mixing genuine screen structure with wall texture, illumination, reflections, and sensor noise. Moreover, each local region carries only weak evidence, making both the need for correction and its appropriate magnitude difficult to determine. We propose Structure-Protected Degradation Representation Learning (SPDRL) to distinguish reliable screen structure from wall-induced interference. SPDRL combines the wall observation, a frozen base reconstruction, their discrepancy, and local cues into a shared representation with structure and degradation readouts. Representation-Calibrated Correction Admission (RCCA) then uses this representation and correction cues to decide where and how strongly to refine the reconstruction. This structure-protected, spatially calibrated recovery preserves reliable content while strengthening weakly supported regions. Across four screen-content types, the complete method leads all 12 main benchmark comparisons against 21 restoration baselines. Experiments spanning illumination, distance, viewpoint, blur, distortion, compression, and occlusion further demonstrate recovery across varied imaging conditions. Together, these results establish the initial feasibility of recovering the principal content and structure of a hidden screen from wall reflections captured by an ordinary camera.

## 1 INTRODUCTION

A hidden screen can leave visible light on a wall even when its displayed content cannot be seen directly. We ask whether an ordinary camera can turn that diffuse reflection into the principal content and structure of the screen. Backes et al. (2009) reconstructed isolated letter shapes from diffuse-wall reflections using a privacy filter, a black-screen reference, and a measured point-spread function. We investigate this possibility for major regions, boundaries, coarse color, and layout across desktop interfaces, game scenes, charts, and lock-screen content, using ordinary-camera wall observations. Answering this question requires preserving the screen information carried by this weak reflected light while separating it from the wall's own appearance.

Recovery requires two linked decisions: separating faint screen structure from stronger wall-induced interference, then determining which regions need refinement and by how much. Diffuse propagation and wall reflection attenuate screen structure while mixing its residual luminance and chromatic cues with wall texture, ambient light, localized reflections, and sensor noise. Restoration models for directly degraded inputs or shifted noise distributions improve feature organization and adaptation (Wang et al., 2022; Cui et al., 2024; Zhang et al., 2024a; Liao et al., 2025), while wall-reflection recovery additionally requires the model to identify which faint patterns carry genuine text, icon, line, and layout relationships. Once a base reconstruction establishes the main layout and appearance, each local refinement raises a second decision: whether the local evidence supports a correction and how strongly it should be applied. Effective recovery therefore demands both structure-aware evidence organization and spatially calibrated correction.

We address these two challenges with Structure-Protected Degradation Representation Learning (SPDRL), followed by Representation-Calibrated Correction Admission (RCCA). SPDRL combines the wall observation, a stable base reconstruction, their discrepancy, and local cues to organize reliable screen structure and degradation interference within a shared representation. Its structure



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.

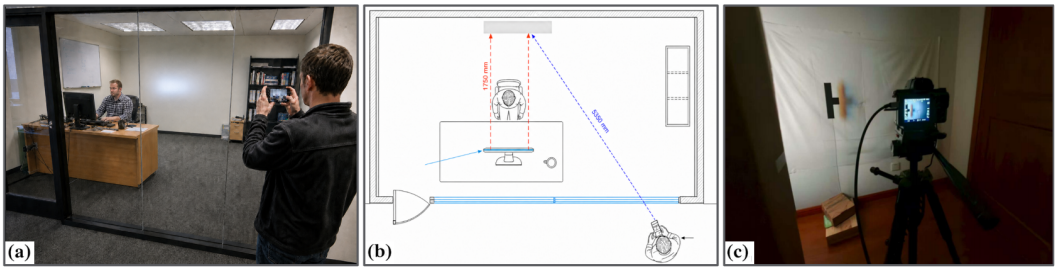



Figure 1: Capture scenario and experimental setup for passive screen-content recovery. (a) An illustrative observation scenario. (b) The acquisition geometry between the display, diffuse wall, and camera, where the camera observes the wall reflection while the emitting screen surface remains outside its direct view. (c) The actual indoor capture setup used to record the wall-reflection observation.

and degradation readouts provide complementary evidence for local refinement. RCCA then generates candidate corrections and determines where and how strongly they should enter the reconstruction, preserving reliable content while strengthening weak regions. Across four screen-content types, the complete method leads all 12 main benchmark comparisons against 21 restoration baselines. Qualitative recovery and imaging-condition variations further show that ordinary-camera wall reflections support the initial recovery of principal screen content and structure.

The main contributions are summarized as follows.

(1) We establish that wall-reflection evidence captured by an ordinary camera while the emitting screen surface remains outside the camera's direct view supports recovery of the principal screen content and structure, demonstrating the initial feasibility of passive screen-content recovery.

(2) We introduce SPDRL to build a structure-protected degradation representation that distinguishes reliable screen structure from wall-induced interference, followed by RCCA to determine position-wise correction admission. Together, they connect structure protection directly to controlled local recovery.

(3) We provide converging evidence across four screen-content types and varied illumination, distance, viewpoint, blur, distortion, compression, and occlusion conditions. The complete method ranks first in all 12 main benchmark columns against 21 restoration baselines, while targeted ablations validate the SPDRL–RCCA progression.

## 2 Related Work

**Recovery from indirect and degraded observations.** Measurement-constrained restoration combines observation models with learned image priors. DDNM uses range–null-space decomposition and a diffusion prior for linear inverse restoration (Wang et al., 2023), while Denoising as Adaptation addresses changing observation statistics through noise-space adaptation (Liao et al., 2025). For screen recovery, Backes et al. (2009) demonstrated diffuse-wall reconstruction through calibrated deconvolution, and IR4Net combines physically regularized inversion with semantic reprojection to reconstruct isolated-screen content (Zheng et al., 2026). Aittala et al. (2019) reconstruct hidden video by jointly estimating scene content and light transport from temporal changes in indirect illumination. Our trained SPDRL–RCCA mapping recovers static screen content, including window layouts and chart-panel structure, from ordinary-camera wall observations through structure protection and position-wise correction (Figure 3).

**Restoration representations and conditioning.** DnCNN learns residual image corrections for denoising (Zhang et al., 2017), Uformer organizes restoration through a U-shaped Transformer (Wang et al., 2022), and Restormer captures long-range interactions with efficient attention and feed-forward designs (Zamir et al., 2022). NAFNet simplifies the restoration architecture (Chen et al., 2022), while FFTformer uses frequency-domain attention and selective frequency processing for deblurring (Kong et al., 2023). Xformer further develops hybrid X-shaped attention for denoising (Zhang et al., 2024a). Conditioning provides another route to adaptation: Retinexformer uses illumination representations (Cai et al., 2023), InstructIR follows human instructions (Conde et al., 2024), and DocRes uses task-specific visual prompts across document restoration tasks (Zhang et al., 2024b). Our comparisons examine these complementary restoration capabilities through reconstruc-

tion results and representation readouts. SPDRL jointly organizes screen structure and localized degradation evidence, and RCCA turns that distinction into spatially calibrated refinement.

# 3 METHOD

## 3.1 OVERALL ARCHITECTURE

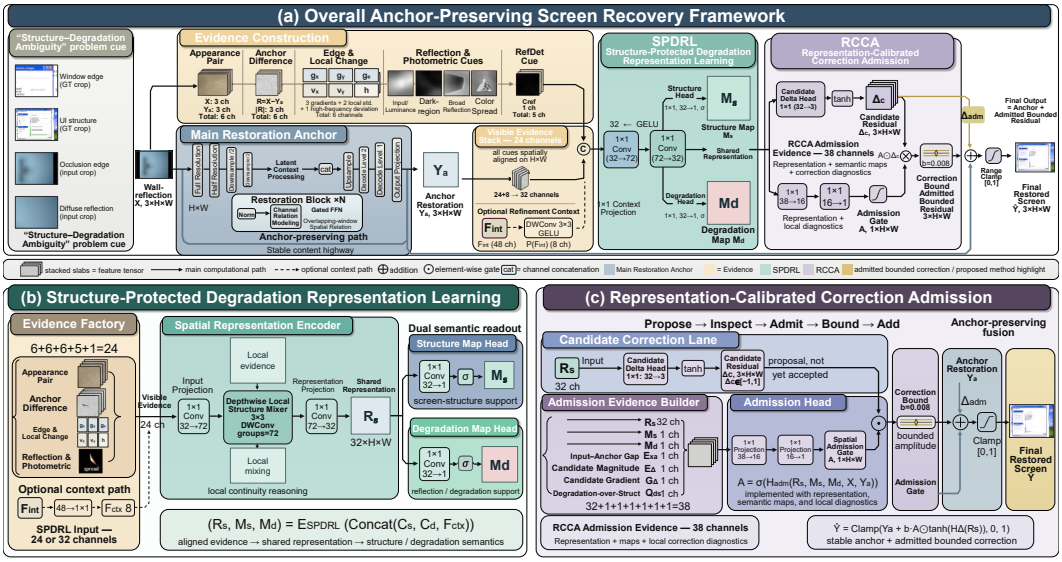



Figure 2: Overview of structure-protected wall-reflection recovery. (a) The complete inference pipeline builds observable evidence around a frozen restoration anchor, then applies an amplitude-bounded correction through SPDRL and RCCA. (b) SPDRL organizes aligned cues into a shared degradation representation with structure and degradation readouts. (c) RCCA proposes a candidate correction and decides its spatial admission before anchor-preserving reconstruction.

Saunders et al. (2019) recover hidden color images with an ordinary camera by modelling an occluder's penumbra using its known shape and size and its estimated position. Our trained SPDRL–RCCA mapping recovers major screen regions and layout from wall-image evidence by protecting reliable structure and calibrating position-wise corrections around a frozen base reconstruction (Figure 3). Wall reflection spatially mixes and attenuates display evidence while adding surface and imaging interference. We first obtain a base reconstruction from a frozen restoration model, then learn an evidence-conditioned, amplitude-bounded correction around this fixed anchor. SPDRL organizes observation-side evidence into a shared degradation representation with structure and degradation readouts; RCCA generates a candidate correction and determines its position-wise admission. The complete inference path is

$$x \to y \to c \to (r, s, d) \to v \to a \to \delta \to o,$$

where $x$ is the wall observation, $o$ is the final reconstruction, and the intermediate quantities form the inference-time computation.

A two-stage radiative-transfer abstraction exposes the spatial mixing that makes wall-reflection recovery difficult:

$$\begin{aligned} x_\kappa(\tau) &= \phi_\kappa\left(\tau; \int_{\Omega_s} K_\kappa(\tau,\xi)\chi_\kappa(\xi)\,d\xi + \beta_\kappa(\tau)\right) + n_\kappa(\tau), \\ K_\kappa(\tau,\xi) &= \int_{\Omega_w} A_\kappa(\tau,\omega)\varrho_\kappa(\omega)G_\kappa(\omega,\xi)\,d\omega, \\ \beta_\kappa(\tau) &= \int_{\Omega_w} A_\kappa(\tau,\omega)\varrho_\kappa(\omega)\iota_\kappa(\omega)\,d\omega, \\ y &= \Omega(\mathcal{F}_{\bar\theta}(x,m))\,. \end{aligned} \tag{1}$$

Here, $\Omega_s$ and $\Omega_w$ denote the screen and wall domains; $\xi$, $\omega$, $\tau$, and $\kappa \in \mathcal{C}$ index a screen point, wall point, camera sample, and color channel. The grouped kernel $K_\kappa$ absorbs screen-to-wall and

wall-to-camera transport, while $\beta_\kappa$ represents the corresponding nuisance illumination term. Here, $\phi_\kappa$ represents the camera response in color channel $\kappa$, and $n_\kappa$ denotes additive observation noise. This abstraction characterizes the spatial mixing that SPDRL must organize directly from observable evidence.

The measured observation is represented by $x \in \mathbb{R}^{B\times C\times H\times W}$, and the optional spatial cue is $m \in [0,1]^{B\times 1\times H\times W}$. The frozen restoration model $\mathcal{F}_{\bar\theta}$ has fixed parameters $\bar\theta$, and $y \in \mathbb{R}^{B\times C\times H\times W}$ is its anchor output. The operator $\Omega$ preserves the forward value while blocking gradients. Thus, the fixed base supplies a stable structural reference while optimization concentrates on the added SPDRL and RCCA functionals.

### 3.2 Structure-Protected Degradation Representation Learning

SPDRL organizes observable discrepancies between the wall observation and the base reconstruction into a structure-protected degradation representation. We first place the available signals on the common image lattice $\mathcal{P}$:

$$\begin{aligned} \Pi_0(t) &= \min\{1, \max\{0, t\}\}, \\ [\Pi_n \eta](p) &= \frac{\eta(p) - \min_{p'\in\mathcal{P}} \eta(p')}{\max_{p'\in\mathcal{P}} \eta(p') - \min_{p'\in\mathcal{P}} \eta(p') + \epsilon_n}, \\ \bar{x} = \Pi_a(\Pi_0(x); y), &\qquad \bar{y} = \Pi_0(y), \qquad e = \bar{x} - \bar{y}, \\ \bar{m} &= \begin{cases} \Pi_a(\Pi_0(m); y), & m \neq \varnothing, \\ 0, & m = \varnothing. \end{cases} \end{aligned} \tag{2}$$

Here, $\Pi_0$ bounds intensities, $\Pi_n$ performs per-sample spatial normalization, and $\Pi_a$ aligns the spatial support with that of the anchor. The signed discrepancy $e$ compares the aligned wall observation with the base reconstruction. SPDRL interprets this difference together with local intensity and structural cues to organize evidence for refinement.

From $\bar{x}$, $\bar{y}$, $e$, and $\bar{m}$, we construct deterministic observation-side statistics describing local intensity, spatial variation, high-frequency deviation, dark-region morphology, and chromatic spread. Their complete definitions, boundary conventions, and numerical stabilizers are given in Appendix A.3. Together with the optional detached base context $u$, these quantities form

$$\begin{aligned} c = \mathcal{X}(x, y, m, f) = \bar{x} \oplus_a \bar{y} \oplus_a e \oplus_a |e| \oplus_a g_x \oplus_a g_y \oplus_a g_e \\ \oplus_a \nu_x \oplus_a \nu_y \oplus_a \ell_x \oplus_a \ell_y \oplus_a h \\ \oplus_a \Pi_n(b) \oplus_a q \oplus_a k \oplus_a \bar{m} \oplus_a u. \end{aligned} \tag{3}$$

The resulting evidence tensor $c \in \mathbb{R}^{B\times Q\times H\times W}$ contains quantities from the current wall observation, frozen base reconstruction, and enabled observation-side cues.

SPDRL then maps the evidence field to one shared representation and two bounded readout maps:

$$\begin{aligned} \gamma(p) &= \Upsilon(c(p); \theta_\gamma), \\ r(p) &= \Phi_r\Big(\widetilde{\gamma}\big|_{\Lambda_{\lambda_4}(p)}; \theta_r\Big), \\ s(p) &= \sigma(\Psi_s(r(p); \theta_s)), \\ d(p) &= \sigma(\Psi_d(r(p); \theta_d)). \end{aligned} \tag{4}$$

The pointwise mapping $\Upsilon$ combines the available cue channels at each location. The spatial mapping $\Phi_r$ then aggregates neighboring evidence over the fixed local support $\Lambda_{\lambda_4}(p)$ into the shared representation $r$, allowing weak local patterns to be interpreted with their surrounding structure. Here, $\sigma$ denotes the sigmoid function. This produces $r \in \mathbb{R}^{B\times R\times H\times W}$ and two bounded maps $s, d \in (0,1)^{B\times 1\times H\times W}$. The structure and degradation readouts $s$ and $d$ provide complementary evidence for RCCA's correction decisions. Their detailed finite-support construction is provided in Appendix A.3.

### 3.3 Representation-Calibrated Correction Admission

RCCA turns the SPDRL representation into an explicit position-wise decision: how much correction should enter the base reconstruction at each location. The candidate correction is generated from

the shared representation $r$, while the shared representation $r$, the readouts $s, d$, and deterministic diagnostics calibrate admission.

The bounded candidate $v$ and four admission diagnostics are

$$\begin{aligned} v_\kappa(p) &= \tanh([\Psi_v(r(p);\theta_v)]_\kappa)\,, \\ \zeta_1 &= \mathcal{M}_c(|e|), \qquad \zeta_2 = \mathcal{M}_c(|v|), \\ \zeta_3 &= \Pi_n(\Delta_\epsilon \mathcal{M}_l(|v|))\,, \qquad \zeta_4 = [d-s]_+. \end{aligned} \tag{5}$$

Thus, $v \in (-1,1)^{B\times C\times H\times W}$ contains a candidate change generated solely from $r$. The fields $\zeta_1, \ldots, \zeta_4$ summarize observation–anchor discrepancy, candidate magnitude, spatial variation in candidate magnitude, and the positive excess of the degradation readout $d$ over the structure readout $s$. The exact definitions of $\mathcal{M}_c$, $\mathcal{M}_l$, and $\Delta_\epsilon$ are given in Appendix A.3.

The admission functional jointly evaluates the representation, the two SPDRL readouts, and the candidate diagnostics:

$$\begin{aligned} z &= r \oplus_a s \oplus_a d \oplus_a \zeta_1 \oplus_a \zeta_2 \oplus_a \zeta_3 \oplus_a \zeta_4, \\ a(p) &= \sigma(\Gamma(z(p);\theta_a))\,. \end{aligned} \tag{6}$$

This defines $a \in (0,1)^{B\times 1\times H\times W}$. The bounded admission map $a$ controls how strongly each candidate change enters the final reconstruction.

The admitted correction is

$$\delta = \rho(a \otimes v), \qquad \|\delta\|_\infty \le \rho, \tag{7}$$

where $\rho \ge 0$ is a fixed configuration coefficient and $\otimes$ denotes elementwise multiplication with $a$ broadcast over the color dimension. Since $a \in (0,1)$ and $v \in (-1,1)$, every element of $\delta$ is bounded by $\rho$ in absolute amplitude. In all main experiments, the correction range is fixed to $\rho = 0.008$.

Finally,

$$\begin{aligned} o &= \Pi_0(y+\delta), \\ |o_\kappa(p) - [\Pi_0(y)]_\kappa(p)| &\le \rho a(p) \le \rho, \\ \|o - \Pi_0(y)\|_\infty &\le \rho. \end{aligned} \tag{8}$$

The candidate $v$ specifies the proposed change, the admission map $a$ determines its local strength, and $\rho$ controls the overall correction range. This division of roles lets RCCA refine weak regions while retaining the base reconstruction as a stable structural reference.

We train SPDRL and RCCA around a frozen dataset-specific XRestormer-Stage6-MicroGate-P1 reconstruction model. The shared representation, structure and degradation readouts, candidate head, and admission gate are jointly optimized through the final reconstruction objective, while the base remains fixed. The structure and degradation readouts $s$ and $d$ are learned entirely through the final reconstruction objective. Appendix A.3 gives the optimization and inference details.

# 4 Experiments and Results

## 4.1 Experimental Setup

We test whether the principal content and structure of a hidden screen remain recoverable across four distinct screen-content types. Screen is the primary realistic benchmark, with approximately 3,800 paired wall-reflection and clean-screen images dominated by desktop interfaces, windows, text, and complex layouts. Game contributes approximately 1,250 paired game-content samples with dark, low-contrast observations; Figure contributes approximately 1,250 paired charts with axes, curves, grids, and other fine geometry; and Lock contributes approximately 1,400 paired lock-screen and wallpaper samples with broad color regions and localized interface elements. Their fixed test partitions contain 251, 250, 250, and 279 samples, respectively. These project-specific datasets jointly probe recovery of text, boundaries, thin lines, color regions, and layout under the captured indoor conditions.

We evaluate recovery using Learned Perceptual Image Patch Similarity (LPIPS) (Zhang et al., 2018), Structural Similarity (SSIM), and peak signal-to-noise ratio (PSNR) (Hore & Ziou, 2010). Lower LPIPS and higher SSIM and PSNR indicate better agreement with the paired screen image.

The baseline pool contains 21 comparison methods spanning general-purpose restoration, visibility and illumination restoration, document and mixed-degradation restoration, and models for related degradations. Table 1 reports the compact main quantitative comparison, and the baseline categories and source references are summarized in Appendix A.5.

### 4.2 Comparison with State-of-the-Art

Our reconstructions reveal the hidden screen's principal content and structure from diffuse wall observations (Figure 3). From diffuse wall observations, our method reconstructs the placement of major windows, foreground regions, and chart panels, together with their coarse color arrangement. These regions align more closely with the targets than the diffuse or distorted outputs of the comparison methods. The examples show how reflected light is converted into the principal content and layout of the hidden screen.

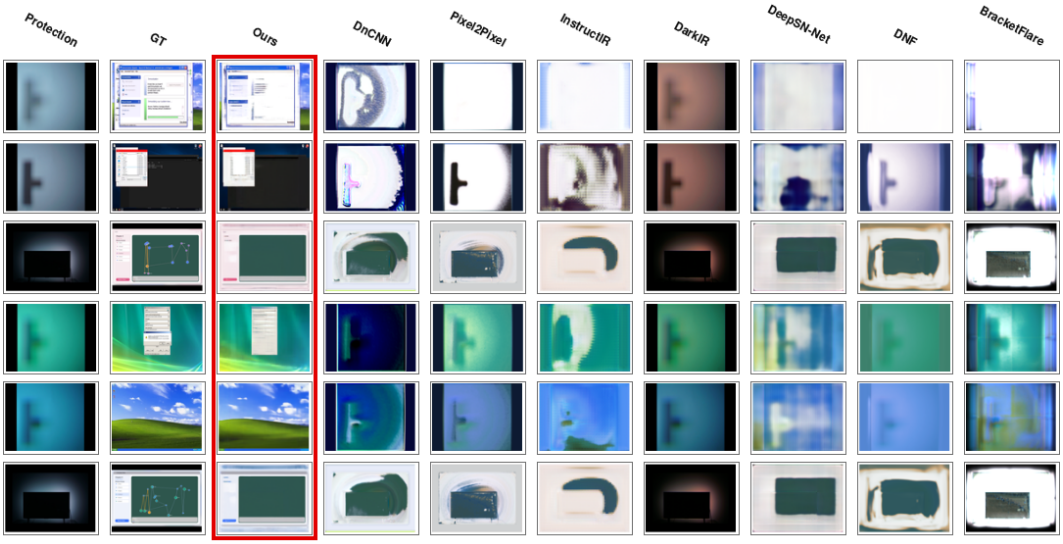



Figure 3: Qualitative comparison of the wall-reflection observation, the corresponding screen target, our reconstruction, and selected restoration baselines. Our output recovers major screen regions, boundaries, and layout from the diffuse observations; Table 1 provides the complete quantitative comparison.

The quantitative comparison supports the recovery pattern visible in Figure 3. Our method leads all 12 dataset–metric comparisons against 21 restoration baselines, with advantages in perceptual similarity, structural similarity, and pixel-level reconstruction. Their agreement across the four content types connects the visible recovery of major screen regions to consistent improvement against the paired targets. Uformer builds a hierarchical Transformer for image restoration (Wang et al., 2022), while ConvIR develops a convolutional restoration architecture (Cui et al., 2024). In our wall-reflection comparison on Screen, the complete SPDRL–RCCA method reaches a PSNR of 21.162 dB and an LPIPS of 0.39788, compared with 19.20 dB and 0.4639 for Uformer and 20.27 dB and 0.4495 for ConvIR (Table 1).

DocRes unifies document dewarping, deshadowing, appearance enhancement, deblurring, and binarization through task-specific visual prompting (Zhang et al., 2024b). For recovery from wall reflections, our complete method surpasses DocRes in all 12 dataset–metric comparisons; on Game, LPIPS improves from 0.57948 to 0.56085, SSIM from 0.64722 to 0.66699, and PSNR from 17.528 to 18.231 dB. The advantage over a model designed for text-bearing document images extends to screen-layout and content recovery from wall reflections.

### 4.3 Ablation Study

#### 4.3.1 Overall Component Ablation

The overall ablation examines how representation construction and correction admission work together to recover screen organization. Our complete model outperforms all six component removals or replacements in all 12 dataset–metric comparisons. Table 2 compares removal of SPDRL or

Table 1: Main comparison on four fixed test sets. Our method ranks first in all 12 dataset–metric columns against 21 comparison methods. Lower LPIPS and higher SSIM and PSNR indicate stronger recovery (Zhang et al., 2018; Hore & Ziou, 2010). The complete method's LPIPS, SSIM, and PSNR values are 0.39788, 0.75015, and 21.162 on Screen; 0.56085, 0.66699, and 18.231 on Game; 0.59595, 0.71863, and 19.285 on Figure; and 0.62275, 0.61520, and 16.930 on Lock. PSNR is measured in dB. Red and blue values indicate the best and second-best results in each dataset–metric column, respectively. The shaded row highlights our method.

| Method | Source | Screen | | | Game | | | Figure | | | Lock | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ |
| **General-Purpose End-to-End Restoration** | | | | | | | | | | | | | |
| AST | CVPR'24 | 19.39 | 0.7050 | 0.4668 | 17.46 | 0.6453 | 0.6158 | 17.89 | 0.6386 | 0.6324 | 15.07 | 0.5366 | 0.6920 |
| Uformer | CVPR'22 | 19.20 | 0.7068 | 0.4639 | 17.55 | 0.6470 | 0.6142 | 17.59 | 0.6187 | 0.6260 | 15.15 | 0.5652 | 0.6743 |
| ConvIR | TPAMI'24 | 20.27 | 0.7283 | 0.4495 | 17.64 | 0.6515 | 0.6013 | 18.54 | 0.7151 | 0.6293 | 14.91 | 0.5925 | 0.6867 |
| U-Net | MICCAI'15 | 9.18 | 0.2401 | 0.8061 | 10.29 | 0.4715 | 0.7581 | 11.81 | 0.5658 | 0.6675 | 12.18 | 0.4874 | 0.7618 |
| DnCNN | TIP'17 | 8.88 | 0.3466 | 0.7245 | 8.74 | 0.4275 | 0.7258 | 14.20 | 0.5738 | 0.6639 | 11.82 | 0.4787 | 0.7221 |
| MIRNet-v2 | TPAMI'23 | 18.24 | 0.6917 | 0.4923 | 17.68 | 0.6559 | 0.6120 | 18.34 | 0.7133 | 0.6536 | 15.21 | 0.6042 | 0.6575 |
| SFHformer | ECCV'24 | 8.82 | 0.3940 | 0.8066 | 10.01 | 0.4620 | 0.7598 | 13.20 | 0.5970 | 0.7134 | 12.46 | 0.5095 | 0.7404 |
| Pix2Pix | CVPR'17 | 10.32 | 0.4816 | 0.6637 | 11.14 | 0.4502 | 0.6917 | 15.48 | 0.6145 | 0.6379 | 12.41 | 0.5178 | 0.7303 |
| InstructIR | ECCV'24 | 13.29 | 0.5227 | 0.6475 | 8.56 | 0.4301 | 0.7585 | 9.69 | 0.5410 | 0.6881 | 13.48 | 0.4524 | 0.7093 |
| **Visibility and Illumination Restoration** | | | | | | | | | | | | | |
| DEA-Net | TIP'24 | 20.65 | **0.7500** | 0.4240 | 17.73 | **0.6579** | 0.5879 | 18.18 | 0.7145 | 0.6425 | 15.96 | 0.6069 | 0.6742 |
| DarkIR | CVPR'25 | 9.22 | 0.1749 | 0.8673 | 4.46 | 0.0409 | 0.9361 | 4.19 | 0.0384 | 0.9290 | 8.79 | 0.0949 | 0.8040 |
| DeepSN-Net | TPAMI'25 | 10.49 | 0.2954 | 0.7787 | 10.87 | 0.4440 | 0.7390 | 14.40 | 0.4993 | 0.6920 | 12.84 | 0.4228 | 0.7711 |
| DehazeFormer | TIP'23 | 19.37 | 0.6931 | 0.4794 | 17.77 | 0.6494 | 0.6074 | **18.92** | 0.7086 | 0.6157 | 14.87 | 0.5375 | 0.6687 |
| Retinexformer | ICCV'23 | 19.57 | 0.6732 | 0.4554 | **17.93** | 0.6556 | 0.5924 | 17.89 | 0.6905 | 0.6186 | 14.60 | 0.5586 | 0.6506 |
| C2PNet | CVPR'23 | 13.41 | 0.5542 | 0.6638 | 9.25 | 0.4532 | 0.7362 | 10.03 | 0.5591 | 0.6870 | 13.50 | 0.4874 | 0.6998 |
| DNF | CVPR'23 | 13.16 | 0.5760 | 0.6646 | 9.96 | 0.4755 | 0.7570 | 9.54 | 0.5877 | 0.7427 | 12.38 | 0.4876 | 0.7527 |
| **Document and Mixed-Degradation Restoration** | | | | | | | | | | | | | |
| DocRes | CVPR'24 | **21.15** | 0.7490 | **0.3999** | 17.53 | 0.6472 | **0.5795** | 18.77 | **0.7184** | **0.6108** | **16.36** | **0.6094** | **0.6467** |
| OneRestore | ECCV'24 | 10.41 | 0.4642 | 0.7142 | 10.31 | 0.5062 | 0.7474 | 12.53 | 0.5863 | 0.7044 | 12.24 | 0.5221 | 0.7212 |
| **Models for Related Degradations** | | | | | | | | | | | | | |
| BracketFlare | CVPR'23 | 12.29 | 0.5432 | 0.6893 | 9.66 | 0.4182 | 0.6990 | 9.72 | 0.5825 | 0.7247 | 11.55 | 0.3733 | 0.6930 |
| DRSformer | CVPR'23 | 14.62 | 0.5930 | 0.6123 | 9.65 | 0.4860 | 0.7028 | 10.15 | 0.5320 | 0.6653 | 13.34 | 0.4822 | 0.6751 |
| WGWS-Net | CVPR'23 | 13.37 | 0.5524 | 0.6588 | 9.70 | 0.4711 | 0.7157 | 9.94 | 0.5511 | 0.6833 | 13.40 | 0.4511 | 0.7173 |
| **Ours** | – | **21.16** | **0.7501** | **0.3979** | **18.23** | **0.6670** | **0.5608** | **19.28** | **0.7186** | **0.5959** | **16.93** | **0.6152** | **0.6227** |

RCCA, replacement of SPDRL by Plain CNN or Restormer (Zamir et al., 2022), and replacement of RCCA by residual or global gating. The w/o RCCA variant retains the learned candidate correction but removes representation-calibrated spatial admission, applying the candidate without the RCCA gate.

Relative to the best tested alternative in each column, the observed gains range from 0.00075300 to 0.090354 in LPIPS, from 0.0024568 to 0.10837 in SSIM, and from 0.10093 to 2.5398 dB in PSNR (Zhang et al., 2018; Hore & Ziou, 2010). Figure S3 connects these gains to the recovered image: the complete design forms more coherent screen regions and refines local content while preserving their overall arrangement. This result motivates the following analyses of what SPDRL must preserve and how RCCA should apply the resulting corrections.

### 4.3.2 SPDRL Internal Ablation

SPDRL's complete structure path and dual readouts lead all 12 dataset–metric comparisons against every tested modification. Table 3 removes the structure path, substitutes generic-edge or wavelet-detail paths, removes the two map heads, and tests shared or complementary-head alternatives.

Relative to the strongest listed alternative in each column, the gains range from 0.0069099 to 0.098870 in LPIPS, from 0.057586 to 0.15809 in SSIM, and from 0.28648 to 2.4616 dB in PSNR (Zhang et al., 2018; Hore & Ziou, 2010). The edge and wavelet variants recover broad screen regions, while the complete SPDRL design produces more coherent boundaries and interface layout in Figure S4. The gain lies in organizing local details into coherent screen structure. The structure path and dual readouts provide complementary information for retaining that organization within the mixed reflection.

### 4.3.3 RCCA Internal Ablation

RCCA addresses the next decision: how a proposed correction should change the reconstructed screen. Table 4 reports mechanism-level RCCA variants: candidate generation determines what change is proposed, and admission determines where and how strongly it enters. The complete model outperforms every admission-gate or candidate-correction variant in all 12 dataset–metric comparisons. In the w/o CCH variant, candidate generation is removed, so the correction vanishes and the output reduces to the frozen anchor, $o = \Pi_0(y)$.

Relative to the strongest listed alternative per column, the gains range from 0.045623 to 0.29181 in LPIPS, from 0.024636 to 0.22299 in SSIM, and from 1.8065 to 8.8816 dB in PSNR (Zhang et al., 2018; Hore & Ziou, 2010). The complete candidate-and-admission design produces more coherent screen regions and boundaries in Figure S5. Candidate generation supplies a possible change, while position-wise admission controls its local application. Their combination links the learned representation to effective refinement of the recovered content. Supplementary Figures S3, S4, and S5 provide the corresponding fixed-sample qualitative comparisons.

Table 2: Overall component ablation on the four fixed test sets. The complete SPDRL–RCCA model leads every tested variant in all 12 dataset–metric columns. The w/o RCCA variant retains the learned candidate correction, removes representation-calibrated spatial admission, and applies the candidate without the RCCA gate. Metric definitions follow (Zhang et al., 2018; Hore & Ziou, 2010). PSNR is measured in dB. Red and blue values indicate the best and second-best results in each dataset–metric column, respectively. The shaded row highlights our method.

| Variant | Screen | | | Game | | | Figure | | | Lock | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ |
| w/o SPDRL | 10.12 | 0.3448 | 0.7629 | 9.06 | 0.2146 | 0.8004 | 13.38 | 0.3502 | 0.6802 | 11.57 | 0.1929 | 0.8311 |
| SPDRL → Plain CNN | 10.03 | 0.3552 | 0.7502 | 9.05 | 0.2118 | 0.8038 | 13.25 | 0.3551 | 0.6795 | 11.67 | 0.2117 | 0.8247 |
| SPDRL → Restormer | 10.06 | 0.3332 | 0.7562 | 9.25 | 0.1967 | 0.8184 | 10.92 | 0.1979 | 0.7888 | 10.16 | 0.1077 | 0.8775 |
| w/o RCCA | 17.99 | 0.5793 | **0.4882** | 18.02 | 0.6246 | **0.5790** | 18.91 | 0.6755 | 0.5991 | 15.97 | **0.5730** | **0.6586** |
| RCCA → Residual Gating | **18.62** | **0.6418** | 0.4927 | **18.13** | **0.6267** | 0.5810 | **19.07** | 0.6855 | **0.5967** | **16.02** | 0.5601 | 0.6722 |
| RCCA → Global Gating | 17.79 | 0.6282 | 0.5001 | 17.87 | 0.6189 | 0.5931 | 17.68 | **0.7162** | 0.6394 | 15.67 | 0.5637 | 0.6656 |
| **Ours** | **21.16** | **0.7501** | **0.3979** | **18.23** | **0.6670** | **0.5608** | **19.28** | **0.7186** | **0.5959** | **16.93** | **0.6152** | **0.6227** |

Table 3: Internal SPDRL ablation on the four fixed test sets. The complete structure path and dual readout heads lead every tested variant in all 12 dataset–metric columns. SP denotes the structure path, and DRH denotes the dual readout heads. Metric definitions follow (Zhang et al., 2018; Hore & Ziou, 2010). PSNR is measured in dB. Red and blue values indicate the best and second-best results in each dataset–metric column, respectively. The shaded row highlights our method.

| Variant | Screen | | | Game | | | Figure | | | Lock | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ |
| w/o SP | 10.13 | 0.3449 | 0.7628 | 9.06 | 0.2146 | 0.8004 | 13.38 | 0.3504 | 0.6802 | 11.57 | 0.1929 | 0.8311 |
| SP → Generic Edge | **18.70** | 0.5907 | **0.4967** | 17.38 | 0.5904 | 0.5897 | 18.89 | 0.6212 | **0.6029** | **15.98** | **0.5571** | **0.6599** |
| SP → Wavelet Detail | 18.68 | **0.5921** | 0.4973 | **17.53** | **0.6094** | **0.5881** | **19.00** | **0.6295** | 0.6057 | 15.96 | 0.5569 | 0.6640 |
| w/o DRH | 10.12 | 0.3423 | 0.7583 | 8.99 | 0.1930 | 0.8225 | 10.47 | 0.2042 | 0.8004 | 10.51 | 0.1164 | 0.8846 |
| DRH → Shared Head | 10.12 | 0.3426 | 0.7583 | 8.99 | 0.1930 | 0.8225 | 10.47 | 0.2042 | 0.8004 | 10.30 | 0.1101 | 0.8658 |
| DRH → Complementary Head | 10.05 | 0.3350 | 0.7566 | 9.25 | 0.1967 | 0.8184 | 10.92 | 0.1979 | 0.7888 | 10.16 | 0.1077 | 0.8775 |
| **Ours** | **21.16** | **0.7501** | **0.3979** | **18.23** | **0.6670** | **0.5608** | **19.28** | **0.7186** | **0.5959** | **16.93** | **0.6152** | **0.6227** |

### 4.4 Recovery Across Imaging Conditions

We next test whether screen recovery persists under changes in color temperature, evaluating Game and Lock at 8000K, 7000K, 6500K, 5000K, and 4000K (Figure 4). Color-temperature variation shifts the chromatic relationship between screen traces and environmental reflections, testing the stability of methods that rely strongly on fixed color distributions or local appearance. Across the evaluated settings, Ours maintains the highest PSNR on both datasets, while its LPIPS remains first and its SSIM stays among the leading results.

Retinexformer uses estimated illumination to guide long-range feature interactions for low-light image enhancement (Cai et al., 2023). Across all five color-temperature settings, our complete method retains the highest PSNR and lowest LPIPS on Game and Lock among the compared methods, including Retinexformer (Figure 4). Screen content remains recoverable across these illumination settings; the appendix extends the tests to geometry and image formation.

Table 4: Mechanism-level RCCA variants on the four fixed test sets. Candidate generation determines what change is proposed, while position-wise admission determines where and how strongly it enters. The w/o CCH variant removes candidate generation, so the correction vanishes and the output reduces to the frozen anchor. The complete mechanism leads every tested variant in all 12 dataset–metric columns. AG denotes the admission gate, and CCH denotes the candidate-correction head. Metric definitions follow (Zhang et al., 2018; Hore & Ziou, 2010). PSNR is measured in dB. Red and blue values indicate the best and second-best results in each dataset–metric column, respectively. The shaded row highlights our method.

| Variant | Screen | | | Game | | | Figure | | | Lock | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ |
| w/o AG | **12.28** | **0.5272** | **0.6897** | **11.07** | 0.5082 | **0.7599** | **17.08** | **0.6497** | **0.6456** | 15.07 | **0.5906** | **0.6684** |
| AG → Plain Confidence | 12.10 | 0.3001 | 0.7967 | 8.38 | 0.2805 | 0.8451 | 10.14 | 0.2962 | 0.7664 | 13.43 | 0.3359 | 0.8232 |
| AG → Global Gate | 12.28 | 0.3499 | 0.7785 | 8.38 | 0.2476 | 0.8277 | 11.91 | 0.3486 | 0.7378 | 13.40 | 0.3611 | 0.8296 |
| w/o CCH | 11.73 | 0.4189 | 0.7389 | 9.15 | **0.5100** | 0.8319 | 15.47 | 0.6291 | 0.6622 | **15.12** | 0.5860 | 0.6731 |
| CCH → RGB-only Head | 11.25 | 0.3746 | 0.7563 | 9.49 | 0.4972 | 0.8205 | 16.72 | 0.6265 | 0.6552 | 13.97 | 0.5243 | 0.7289 |
| CCH → Global Head | 11.41 | 0.4583 | 0.7148 | 10.68 | 0.4418 | 0.7663 | 16.68 | 0.6417 | 0.6823 | 14.07 | 0.5404 | 0.7196 |
| **Ours** | **21.16** | **0.7501** | **0.3979** | **18.23** | **0.6670** | **0.5608** | **19.28** | **0.7186** | **0.5959** | **16.93** | **0.6152** | **0.6227** |

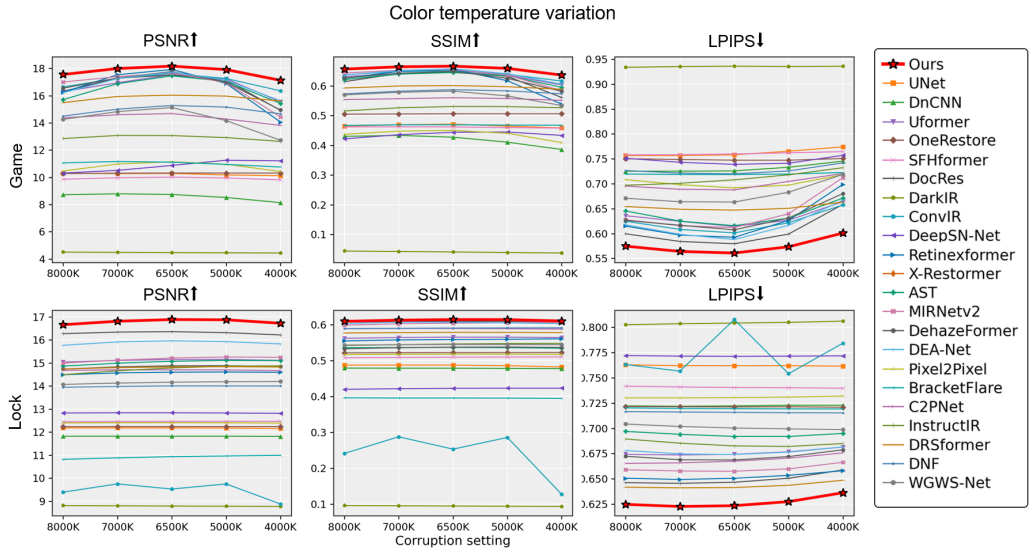



Figure 4: Color-temperature robustness results under five evaluated settings: 8000K, 7000K, 6500K, 5000K, and 4000K. The top row reports results on Game, and the bottom row reports results on Lock. From left to right, the columns show PSNR, SSIM, and LPIPS, respectively. Higher PSNR and SSIM and lower LPIPS indicate more favorable performance (Zhang et al., 2018; Hore & Ziou, 2010). The red curves with black star markers denote the complete proposed model, while the other methods are distinguished by the colors and markers shown in the legend.

## 5 Conclusion

An ordinary camera can recover the principal content and structure of a hidden screen from the light it casts onto a diffuse wall. SPDRL organizes the reflected evidence into a representation that distinguishes reliable screen structure from interference, and RCCA uses this representation to control where and how strongly the reconstruction is refined. The complete method leads all 12 main benchmark comparisons against 21 restoration baselines, while the ablations, imaging-condition variations, and representation analyses explain how this recovery is supported. Alongside earlier diffuse-reflection reconstruction and learned optical inversion (Backes et al., 2009; Zheng et al., 2026), our results establish a structure-protected route to screen-content recovery. Together, these findings demonstrate the initial feasibility of recovering a hidden screen's principal content and structure from wall reflections captured by an ordinary camera.

## REFERENCES


Miika Aittala, Prafull Sharma, Lukas Murmann, Adam 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. URL https://proceedings.neurips.cc/paper/2019/hash/5a2afca61e35f45a7dd44ca46e0225f4-Abstract.html.

Michael Backes, Tongbo Chen, Markus Dürmuth, Hendrik P. A. Lensch, and Martin Welk. Tempest in a teapot: Compromising reflections revisited. In *2009 30th IEEE Symposium on Security and Privacy*, pp. 315–327, 2009. doi: 10.1109/SP.2009.20.

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. 12470–12479, 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, 2022. doi: 10.1007/978-3-031-20071-7_2.

Xiang Chen, Hao Li, Mingqiang Li, and Jinshan Pan. Learning a sparse transformer network for effective image deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5896–5905, 2023. doi: 10.1109/CVPR52729.2023.00571.

Xiangyu Chen, Zheyuan Li, Yuandong Pu, Yihao Liu, Jiantao Zhou, Yu Qiao, and Chao Dong. A comparative study of image restoration networks for general backbone network design. In *Computer Vision – ECCV 2024*, pp. 74–91, 2024a. doi: 10.1007/978-3-031-73209-6_5.

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, 2024b. doi: 10.1109/TIP.2024.3354108.

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

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.

Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16 (8):2080–2095, 2007. doi: 10.1109/TIP.2007.901238.

Yuekun Dai, Yihang Luo, Shangchen Zhou, Chongyi Li, and Chen Change Loy. Nighttime smartphone reflective flare removal using optical center symmetry prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20783–20791, 2023. doi: 10.1109/CVPR52729.2023.01991.

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.

Yu Guo, Yuan Gao, Yuxu Lu, Huilin Zhu, Ryan Wen Liu, and Shengfeng He. OneRestore: A universal restoration framework for composite degradation. In *Computer Vision – ECCV 2024*, pp. 255–272, 2024. doi: 10.1007/978-3-031-72655-2_15.

Alain Hore and Djemel Ziou. Image quality metrics: PSNR vs. SSIM. In *20th International Conference on Pattern Recognition*, pp. 2366–2369, 2010. doi: 10.1109/ICPR.2010.579.

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, 2024. doi: 10.1007/978-3-031-72995-9_22.

Xin Jin, Ling-Hao Han, Zhen Li, Chun-Le Guo, Zhi Chai, and Chongyi Li. DNF: Decouple and feedback network for seeing in the dark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18135–18144, 2023. doi: 10.1109/CVPR52729.2023.01739.

Daniel J. Jobson, Zia-ur Rahman, and Glenn A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing*, 6(7):965–976, 1997. doi: 10.1109/83.597272.

Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5886–5895, 2023. doi: 10.1109/CVPR52729.2023.00570.

Kang Liao, Zongsheng Yue, Zhouxia Wang, and Chen Change Loy. Denoising as adaptation: Noise-space domain adaptation for image restoration. In *The Thirteenth International Conference on Learning Representations, ICLR 2025*. OpenReview.net, 2025. URL https://openreview.net/forum?id=jsBhmOCKYs.

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, 2015. doi: 10.1007/978-3-319-24574-4_28.

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

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.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *The Eleventh International Conference on Learning Representations, ICLR 2023*. OpenReview.net, 2023. URL https://openreview.net/forum?id=mRieQgMtNTQ.

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. 17662–17672, 2022. doi: 10.1109/CVPR52688.2022.01716.

Yang Yang, Wenye Ma, Yin Zheng, Jian-Feng Cai, and Weiyu Xu. Fast single image reflection suppression via convex optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8133–8141, 2019. doi: 10.1109/CVPR.2019.00833.

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

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.

Jiale Zhang, Yulun Zhang, Jinjin Gu, Jiahua Dong, Linghe Kong, and Xiaokang Yang. Xformer: Hybrid X-shaped transformer for image denoising. In *The Twelfth International Conference on Learning Representations, ICLR 2024*. OpenReview.net, 2024a. URL https://openreview.net/forum?id=vXrIQLzIKY.

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, 2024b. 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/CVF Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018. doi: 10.1109/CVPR.2018.00068.

Yu Zheng, Jiahui Zhan, Shengfeng He, Junyu Dong, and Yong Du. Curricular contrastive regularization for physics-aware single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5785–5794, 2023. doi: 10.1109/CVPR52729.2023.00560.

Zhiwen Zheng, Yuheng Qiao, Xiaoshuai Zhang, Zhao Huang, Tao Zhang, Huiyu Zhou, Shaowei Jiang, Jin Liu, Wenwen Tang, and Xingru Huang. Physically-guided optical inversion enable non-contact side-channel attack on isolated screens. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/93e6cab8a85d1c40c65ccd8c49d5343a-Abstract-Conference.html.

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.

Yurui Zhu, Tianyu Wang, Xueyang Fu, Xuanyu Yang, Xin Guo, Jifeng Dai, Yu Qiao, and Xiaowei Hu. Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21747–21758, 2023. doi: 10.1109/CVPR52729.2023.02083.

# APPENDIX FOR RECOVERABLE EVIDENCE IN WALL REFLECTIONS: STRUCTURE-PROTECTED DEGRADATION REPRESEN-TATION LEARNING FOR SCREEN CONTENT RECOVERY

## A STRUCTURE-PROTECTED WALL-REFLECTION RECOVERY: MECHANISM AND EXTENDED EVIDENCE

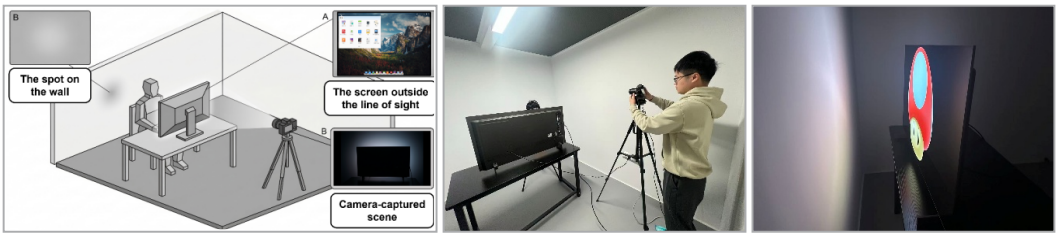
### A.1 EVIDENCE ROADMAP

The appendix first details how SPDRL organizes reflected evidence and how RCCA converts it into bounded local refinement. Additional comparisons and ablations then examine the recovered screen content across examples and imaging variations. The regional analyses locate useful input evidence, the readouts identify the information retained in the representation, and the perturbation study tests how that information supports recovery under internal disturbance. The baseline sources are retained in Appendix A.5.

### A.2 DATASET DESCRIPTION

The dataset is acquired in a real indoor wall-reflection scene, as illustrated in Figure S1. A 60-inch self-emissive LED display faces a white diffuse wall at a separation of approximately 0.5 m. Light emitted by the display illuminates the wall and forms a reflected light pattern. A Sony A7S II camera records the scene from behind the display, so that the observation contains the silhouette of the display back and the surrounding wall reflection, while the emitting screen surface remains outside the camera's direct view.

Each sample pairs the camera observation with the corresponding screen image recorded at capture time. We recover the displayed content from the observation of the display back and surrounding wall reflection, using the paired screen image as the reconstruction target.




(a) Scene schematic (b) Capture setup (c) Side view

Figure S1: Indoor capture setup for wall-reflection screen recovery. (a) Scene schematic. (b) Camera placement behind the display. (c) Side view of the screen facing the diffuse wall. The observation contains the display-back silhouette and the surrounding wall reflection, paired with the corresponding screen content.

The four project-specific test sets examine recovery across different screen structures. Screen contains desktop layouts and window controls; Game emphasizes color regions and game boundaries; Figure emphasizes axes, grids, curves, and other thin geometry; Lock emphasizes large backgrounds with sparse text and login elements. Together, they test whether weak luminance and chromatic traces can support the principal structure of multiple screen-content types under the evaluated indoor conditions.

Spatial propagation, wall material, ambient illumination, exposure, and sensor noise attenuate text strokes, icon contours, window boundaries, and thin chart lines. At the same time, wall texture and localized reflection create edge-like patterns. Recovery therefore hinges on separating genuine screen relationships from stronger relay-surface interference.

### A.3 DETAILED OPERATIONAL DERIVATION

This section completes the operational definitions summarized in the main paper and uses exactly the same variables and forward relations as Section 3. The derivation specifies deterministic cue construction, spatial boundary conventions, the optional base-context path, and the exact bounded-correction interpretation for the recovery process summarized in Section 3.

### A.3.1 Frozen Anchor and Gradient Boundary

The observation model in the main paper is a phenomenological description of spatial mixing. Before grouping the transport terms, the observation can be written as

$$\begin{aligned} x_\kappa(\tau) &= \phi_\kappa\left(\tau; \int_{\Omega_w} A_\kappa(\tau,\omega)\varrho_\kappa(\omega)\left[\int_{\Omega_s} G_\kappa(\omega,\xi)\chi_\kappa(\xi)\,d\xi + \iota_\kappa(\omega)\right] d\omega\right) + n_\kappa(\tau) \\ &= \phi_\kappa\left(\tau; \int_{\Omega_s} K_\kappa(\tau,\xi)\chi_\kappa(\xi)\,d\xi + \beta_\kappa(\tau)\right) + n_\kappa(\tau), \\ K_\kappa(\tau,\xi) &= \int_{\Omega_w} A_\kappa(\tau,\omega)\varrho_\kappa(\omega)G_\kappa(\omega,\xi)\,d\omega, \\ \beta_\kappa(\tau) &= \int_{\Omega_w} A_\kappa(\tau,\omega)\varrho_\kappa(\omega)\iota_\kappa(\omega)\,d\omega. \end{aligned} \tag{S1}$$

Here, $\chi_\kappa : \Omega_s \to \mathbb{R}_{\geq 0}$ denotes the latent screen-emitted color field. The nonnegative functions $G_\kappa$, $A_\kappa$, $\varrho_\kappa$, and $\iota_\kappa$ represent screen-to-wall transport, wall-to-camera coupling, an effective local relay response, and ambient illumination, respectively. Consequently, $K_\kappa$ groups the spatial transport factors and $\beta_\kappa$ groups the corresponding nuisance illumination contribution. Here, $\phi_\kappa$ represents the camera response in color channel $\kappa$, and $n_\kappa$ denotes additive observation noise.

These physical quantities describe the latent transport process, while SPDRL and RCCA operate directly on observable evidence. Their role is to expose the source of severe spatial mixing: the measured value at a camera location may aggregate contributions from multiple screen locations after propagation through the relay surface.

The frozen anchor used by the actual reconstruction pipeline remains

$$y = \Omega(\mathcal{F}_{\bar\theta}(x, m))\,. \tag{S2}$$

The parameters $\bar\theta$ remain fixed. The detachment operator $\Omega$ leaves the forward value unchanged but blocks gradients into the frozen restoration model. The same convention is used for the optional base feature map $f$ described below.

### A.3.2 Alignment, Range Restriction, and Observation–Anchor Discrepancy

All correction-related computations operate on the common lattice

$$\mathcal{P} = \{1,\dots,H\} \times \{1,\dots,W\}.$$

The fixed value restriction and spatial normalization operators are

$$\begin{aligned} \Pi_0(t) &= \min\{1, \max\{0, t\}\}, \\ [\Pi_n\eta](p) &= \frac{\eta(p) - \min_{p'\in\mathcal{P}}\eta(p')}{\max_{p'\in\mathcal{P}}\eta(p') - \min_{p'\in\mathcal{P}}\eta(p') + \epsilon_n}. \end{aligned} \tag{S3}$$

The scalar $\epsilon_n > 0$ prevents division by zero when the field is spatially constant.

The aligned observation, anchor, signed discrepancy, and optional spatial cue are

$$\begin{aligned} \bar x &= \Pi_a(\Pi_0(x); y), \\ \bar y &= \Pi_0(y), \\ e &= \bar x - \bar y, \\ \bar m &= \begin{cases} \Pi_a(\Pi_0(m); y), & m \neq \varnothing, \\ 0, & m = \varnothing. \end{cases} \end{aligned} \tag{S4}$$

Accordingly,

$$\bar x, \bar y, e \in \mathbb{R}^{B\times C\times H\times W}, \qquad \bar m \in [0,1]^{B\times 1\times H\times W}.$$

The operator $\Pi_a$ aligns supported fields to the spatial support of $y$. In the observation–anchor path, $x$ already follows the common lattice, so this notation records the same alignment convention used by optional auxiliary fields.

The signed discrepancy $e$ compares the aligned wall observation with the base reconstruction. SPDRL interprets this difference together with local intensity and structural cues to organize evidence for refinement.

### A.3.3 Luminance and First-Order Spatial Evidence

For a generic RGB field $\alpha : \mathcal{P} \to \mathbb{R}^C$, we use the fixed luminance functional

$$\begin{aligned} \mathcal{M}_l(\alpha)(p) &= w^\mathsf{T}\alpha(p), \\ w \in \mathbb{R}^C_{\geq 0}, &\qquad \mathbf{1}^\mathsf{T} w = 1. \end{aligned} \tag{S5}$$

Fixed weights make the luminance functional consistent across observations, and the unit-sum constraint preserves a constant intensity shared by all RGB channels.

Applying this functional gives

$$\ell_x = \mathcal{M}_l(\bar{x}), \qquad \ell_y = \mathcal{M}_l(\bar{y}), \qquad \ell_e = \mathcal{M}_l(|e|). \tag{S6}$$

For a scalar spatial field $\eta$, define forward horizontal and vertical differences by

$$[\Delta_1\eta](i,j) = \begin{cases} \eta(i,j+1) - \eta(i,j), & j < W, \\ 0, & j = W, \end{cases} \tag{S7}$$

and

$$[\Delta_2\eta](i,j) = \begin{cases} \eta(i+1,j) - \eta(i,j), & i < H, \\ 0, & i = H. \end{cases} \tag{S8}$$

Their regularized magnitude is

$$[\Delta_\epsilon\eta](i,j) = \sqrt{[\Delta_1\eta](i,j)^2 + [\Delta_2\eta](i,j)^2 + \epsilon_g}, \tag{S9}$$

where $\epsilon_g > 0$ is a numerical stabilizer.

The first-order evidence maps used in Equation 3 are then

$$\begin{aligned} g_x &= \Pi_n(\Delta_\epsilon \ell_x)\,, \\ g_y &= \Pi_n(\Delta_\epsilon \ell_y)\,, \\ g_e &= \Pi_n(\Delta_\epsilon \ell_e)\,. \end{aligned} \tag{S10}$$

These maps lie in $[0,1]^{B\times 1\times H\times W}$ after spatial normalization. They measure local change energy visible in the current inference pass and supply observation-derived structural evidence.

### A.3.4 Local Variation and High-Frequency Evidence

To describe local second-order variation, the scalar field $\eta$ is extended outside the valid image lattice by

$$\widetilde{\eta}(p') = \begin{cases} \eta(p'), & p' \in \mathcal{P}, \\ 0, & p' \notin \mathcal{P}. \end{cases} \tag{S11}$$

For the fixed neighborhood $\Lambda_{\lambda_1}(p)$, the local first moment is

$$\mathcal{M}_{\lambda_1}[\eta](p) = \frac{1}{|\Lambda_{\lambda_1}|} \sum_{p' \in \Lambda_{\lambda_1}(p)} \widetilde{\eta}(p'), \tag{S12}$$

and the local variation functional is

$$\mathcal{S}_{\lambda_1}[\eta](p) = \sqrt{\left[\mathcal{M}_{\lambda_1}[\eta^2](p) - \left(\mathcal{M}_{\lambda_1}[\eta](p)\right)^2\right]_+ + \epsilon_s}. \tag{S13}$$

Here, $\epsilon_s > 0$ is a stabilizer and $[\cdot]_+$ retains the nonnegative part. The neighborhood cardinality remains fixed even at image boundaries because the same zero-extension convention is used.

The resulting evidence maps are

$$\begin{aligned} \nu_x &= \Pi_n(\mathcal{S}_{\lambda_1}[\ell_x])\,, \\ \nu_y &= \Pi_n(\mathcal{S}_{\lambda_1}[\ell_y])\,, \\ h &= \Pi_n(|\ell_x - \mathcal{M}_{\lambda_1}[\ell_x]|)\,. \end{aligned} \tag{S14}$$

Thus, $\nu_x$ and $\nu_y$ measure normalized local variation in the observation and anchor, whereas $h$ records deviation of the observed luminance from its local first moment.

The support symbol $\lambda_1$ identifies a fixed operational neighborhood, separating computational support from the latent physical scene parameters.

### A.3.5 Low-Intensity Morphology and Relative Chromatic Spread

Two additional cues describe appearance behavior that is complementary to first- and second-order spatial variation.

The local lower envelope is

$$b(p) = \min_{\substack{p' \in \Lambda_{\lambda_2}(p) \cap \mathcal{P} \\ \kappa \in \mathcal{C}}} \bar{x}_\kappa(p'), \tag{S15}$$

and its locally smoothed normalized field is

$$q = \Pi_n(\mathcal{M}_{\lambda_3}[b]) \,. \tag{S16}$$

Relative channel spread is

$$k(p) = \Pi_0\left( \frac{\max_{\kappa \in \mathcal{C}} \bar{x}_\kappa(p) - \min_{\kappa \in \mathcal{C}} \bar{x}_\kappa(p)}{\max\left\{\max_{\kappa \in \mathcal{C}} \bar{x}_\kappa(p), \epsilon_c\right\}} \right), \tag{S17}$$

where $\epsilon_c > 0$ prevents numerical instability in dark regions.

The quantities $b$, $q$, and $k$ are deterministic observation-side measurements that provide complementary local appearance coordinates for the evidence field. Together they expose low-intensity and chromatic manifestations of the mixed wall reflection.

### A.3.6 Optional Base Context

When the optional base-context path is enabled, a base feature map is mapped to the evidence lattice:

$$u = \begin{cases} \Phi_u(\Pi_a(\Omega(f); y); \theta_u) \,, & f \neq \varnothing, \\ 0, & f = \varnothing. \end{cases} \tag{S18}$$

Here,

$$f \in \mathbb{R}^{B \times F \times H_f \times W_f}$$

is a base feature map of the frozen restoration model, and

$$u \in \mathbb{R}^{B \times U \times H \times W}$$

is its mapped representation. The detachment operator $\Omega(f)$ blocks gradients into the frozen restoration model while preserving its forward value. The learned mapping $\Phi_u$ may receive gradients when this path is enabled.

This context path is optional. An enabled path contributes $u$ to the concatenation; a disabled path omits that term. Therefore, the feature dimension $Q$ of $c$ follows the presence of $u$, while the downstream definitions remain unchanged.

### A.3.7 Complete Observable Evidence Inventory

Combining the quantities above reproduces the evidence field used by SPDRL:

$$\begin{aligned} c = \mathcal{X}(x, y, m, f) = \bar{x} \oplus_a \bar{y} \oplus_a e \oplus_a |e| \oplus_a g_x \oplus_a g_y \oplus_a g_e \\ \oplus_a \nu_x \oplus_a \nu_y \oplus_a \ell_x \oplus_a \ell_y \oplus_a h \\ \oplus_a \Pi_n(b) \oplus_a q \oplus_a k \oplus_a \bar{m} \oplus_a u. \end{aligned} \tag{S19}$$

The operator $\oplus_a$ concatenates feature coordinates after spatial alignment. Thus,

$$c \in \mathbb{R}^{B \times Q \times H \times W} .$$

Every coordinate in $c$ is determined by the current observation $x$, its detached base reconstruction $y$, the optional observation-side cue $m$, and the optional detached base feature map $f$.

This inventory exposes several complementary manifestations of weak reflected evidence, allowing the learned SPDRL representation to combine them into a coherent structure-protected degradation representation.

### A.3.8 FINITE-SUPPORT SPDRL REPRESENTATION

The pointwise evidence lifting used in Equation 4 is

$$\gamma(p) = \Upsilon(c(p); \theta_\gamma), \qquad p \in \mathcal{P}. \tag{S20}$$

This produces

$$\gamma \in \mathbb{R}^{B\times J\times H\times W}.$$

To state the local boundary rule explicitly, define

$$\widetilde{\gamma}(p') = \begin{cases} \gamma(p'), & p' \in \mathcal{P}, \\ 0_J, & p' \notin \mathcal{P}, \end{cases} \tag{S21}$$

where $0_J$ is the zero vector in the lifted feature space.

The shared representation is then

$$r(p) = \Phi_r\Big(\widetilde{\gamma}\big|_{\Lambda_{\lambda_4}(p)}; \theta_r\Big), \tag{S22}$$

with

$$r \in \mathbb{R}^{B\times R\times H\times W}.$$

The two bounded SPDRL readout maps satisfy

$$s(p) = \sigma(\Psi_s(r(p); \theta_s)), \tag{S23}$$

and

$$d(p) = \sigma(\Psi_d(r(p); \theta_d)). \tag{S24}$$

Consequently,

$$s, d \in (0,1)^{B\times 1\times H\times W}.$$

Both are read from the same shared representation $r$. Their bounded domains provide finite internal structure and degradation readouts for downstream RCCA.

The support $\Lambda_{\lambda_4}(p)$ makes the SPDRL mapping explicitly local after the pointwise lifting step. Its purpose is to allow neighboring evidence coordinates to be interpreted jointly while retaining a finite computational neighborhood.

### A.3.9 CANDIDATE GENERATION IN RCCA

RCCA first proposes a candidate from the shared representation alone. For each $\kappa \in \mathcal{C}$,

$$v_\kappa(p) = \tanh([\Psi_v(r(p); \theta_v)]_\kappa). \tag{S25}$$

This is equivalent to placing the learned candidate in

$$v \in (-1,1)^{B\times C\times H\times W}.$$

The computational graph deliberately separates proposal from admission: $v$ is obtained from $r$, and $r$, $s$, $d$, and the candidate diagnostics then determine how much of that candidate enters the anchor.

### A.3.10 RCCA ADMISSION DIAGNOSTICS

For a generic color field $\alpha$, the fixed channel moment is

$$\mathcal{M}_c(\alpha) = \frac{1}{C}\sum_{\kappa\in\mathcal{C}} \alpha_\kappa. \tag{S26}$$

The four RCCA diagnostics are

$$\begin{aligned} \zeta_1 &= \mathcal{M}_c(|e|), \\ \zeta_2 &= \mathcal{M}_c(|v|), \\ \zeta_3 &= \Pi_n(\Delta_\epsilon \mathcal{M}_l(|v|)), \\ \zeta_4 &= [d - s]_+. \end{aligned} \tag{S27}$$

The first field summarizes the absolute observation–anchor discrepancy. The second measures the RGB-averaged candidate magnitude. The third describes spatial variation in the magnitude of the proposed candidate using exactly the luminance and regularized differential operators defined above. The fourth retains only the nonnegative excess of $d$ over $s$.

All four diagnostics are deterministic after $e$, $v$, $s$, and $d$ have been computed. The admission gate combines the shared representation $r$, the structure and degradation scores $s$ and $d$, and the four correction diagnostics to decide how strongly each candidate is applied.

#### A.3.11 Admission Readout

The complete RCCA admission field is

$$z = r \oplus_a s \oplus_a d \oplus_a \zeta_1 \oplus_a \zeta_2 \oplus_a \zeta_3 \oplus_a \zeta_4. \tag{S28}$$

Its bounded admission map satisfies

$$a(p) = \sigma(\Gamma(z(p); \theta_a)), \tag{S29}$$

yielding

$$a \in (0,1)^{B\times 1\times H\times W}.$$

The gate therefore has access to the complete shared representation, both bounded SPDRL readout maps, and all four candidate-related diagnostics produced during the same inference pass. The bounded value $a$ directly controls the position-wise admission of each candidate correction.

#### A.3.12 Exact Bounded-Correction Interpretation

The main paper writes the implemented admitted correction as

$$\delta = \rho(a \otimes v).$$

Its relationship to an explicit feasible correction set can also be stated exactly.

Define

$$\mathbb{B}_\rho = \left\{\delta' \in \mathbb{R}^{B\times C\times H\times W} : \|\delta'\|_\infty \le \rho\right\}. \tag{S30}$$

The corresponding Euclidean projection characterization is

$$\begin{aligned} \delta &= \Pi_{\mathbb{B}_\rho}(\rho(a \otimes v)) \\ &= \arg\min_{\delta' \in \mathbb{B}_\rho} \frac{1}{2} \|\delta' - \rho(a \otimes v)\|_2^2 \\ &= \rho(a \otimes v). \end{aligned} \tag{S31}$$

The final equality follows directly from

$$a \in (0,1), \qquad v \in (-1,1),$$

which implies

$$\|\rho(a \otimes v)\|_\infty \le \rho.$$

Thus, the proposed correction already lies inside $\mathbb{B}_\rho$, and the projection leaves it unchanged.

This projection interpretation is an equivalent mathematical description of the same single-pass forward computation.

#### A.3.13 Final Reconstruction and Anchor-Centered Bounds

The sole reconstructed output remains

$$o = \Pi_0(y + \delta). \tag{S32}$$

Because

$$\delta = \rho(a \otimes v),$$

the elementwise displacement satisfies

$$|o_\kappa(p) - [\Pi_0(y)]_\kappa(p)| \le \rho\, a(p)\, |v_\kappa(p)| \le \rho a(p) \le \rho. \tag{S33}$$

Consequently,

$$\|o - \Pi_0(y)\|_\infty \le \rho. \tag{S34}$$

At a spatial location $p$, the color-vector displacement additionally satisfies

$$\|o(p) - [\Pi_0(y)](p)\|_2 \le \sqrt{C}\,\rho a(p). \tag{S35}$$

For any second admissible pair $a'$ and $v'$ with the same domains,

$$\begin{aligned}&\|\rho(a \otimes v) - \rho(a' \otimes v')\|_\infty \\ &\quad\le \rho\left(\|a - a'\|_\infty + \|v - v'\|_\infty\right).\end{aligned} \tag{S36}$$

This relation bounds the sensitivity of the admitted correction with respect to changes in its two bounded internal coordinates.

Under the stated initialization,

$$v^{(0)} = 0 \quad\Longrightarrow\quad \delta^{(0)} = 0 \quad\Longrightarrow\quad o^{(0)} = \Pi_0(y). \tag{S37}$$

Thus, the initial output coincides with the range-restricted base reconstruction before a nonzero candidate correction is produced.

These properties establish anchor-centered amplitude and perturbation control. SPDRL organizes the observable evidence, and RCCA uses it to control local refinement; both modules learn through the final reconstruction objective.

#### A.3.14 Training and Inference Boundary

SPDRL and RCCA are trained around the frozen dataset-specific XRestormer-Stage6-MicroGate-P1 reconstruction model. Its parameters are omitted from the optimizer, all base parameters use `requires_grad=False`, and the model remains in evaluation mode. The base reconstruction and optional internal feature are detached before entering the trainable extension.

The complete trainable portion of the proposed extension consists of the learned SPDRL and RCCA functionals. Gradients may pass through

$$\Upsilon, \quad \Phi_r, \quad \Psi_s, \quad \Psi_d, \quad \Psi_v, \quad \Gamma,$$

and through $\Phi_u$ when the optional base-context path is enabled.

Throughout optimization,

$$y = \Omega(\mathcal{F}_{\bar\theta}(x, m))$$

remains detached, and the optional base feature map is used through $\Omega(f)$. Therefore, the frozen restoration model supplies a fixed base reconstruction throughout optimization.

**Optimization.** The final reconstruction objective is

$$\mathcal{L} = \mathcal{L}_{\text{char}} + \lambda_s \mathcal{L}_{\text{SSIM}} + \lambda_p \mathcal{L}_{\text{PSNR}} + \lambda_l \mathcal{L}_{\text{LPIPS}} + \lambda_g \mathcal{L}_{\text{grad}}. \tag{S38}$$

It combines masked Charbonnier reconstruction with SSIM, PSNR, LPIPS, and gradient-preservation terms. The base coefficients are Charbonnier 1, $\lambda_s = 0.055$, $\lambda_p = 0.005$, $\lambda_l = 0.004$, and $\lambda_g = 0.006$. Some configured dataset-specific runs increase the designated primary-metric coefficient to SSIM $0.085$, PSNR $0.010$, or LPIPS $0.010$. Optimization uses AdamW with learning rate $3 \times 10^{-5}$, weight decay $0$, $128 \times 128$ crops, batch size $16$ with an $8$-sample memory fallback, up to $16$ epochs, early-stopping patience $5$, gradient clipping at $0.5$, and seed $20260803$.

The representation, structure and degradation readouts, candidate head, and admission gate are learned end-to-end through these final reconstruction losses. The training configuration for $s$ and $d$ uses reconstruction-only supervision, with separate structure/degradation ground truth and BCE, Dice, segmentation, edge, and RefDet-mask supervision disabled.

During inference, the complete dependency chain is

$$(x, m) \to y \to (\bar x, \bar y, e, \bar m) \to c \to (\gamma, r, s, d) \to v \to (\zeta_1, \zeta_2, \zeta_3, \zeta_4) \to z \to a \to \delta \to o,$$

with $f$ and $u$ participating only when the optional base-context path is enabled.

At inference, SPDRL forms the shared representation from the wall observation, the frozen base reconstruction, and the enabled observation-side cues. RCCA generates a candidate correction and determines its spatial admission from this representation and the correction diagnostics. The final composition applies the admitted update around the base reconstruction.

### A.4 Baseline Coverage

The 21 baselines cover four complementary restoration capabilities: general-purpose reconstruction, visibility and illumination restoration, document and composite-degradation recovery, and restoration under related degradations. The groups contain nine, seven, two, and three methods, respectively, providing a broad comparison for recovering screen content from wall reflections.

### A.5 Baseline Pool

The baseline pool contains 21 comparison methods. The general-purpose restoration group includes AST (Zhou et al., 2024), Uformer (Wang et al., 2022), ConvIR (Cui et al., 2024), U-Net (Ronneberger et al., 2015), DnCNN (Zhang et al., 2017), MIRNet-v2 (Zamir et al., 2023), SFHformer (Jiang et al., 2024), Pix2Pix (Isola et al., 2017), and InstructIR (Conde et al., 2024). The visibility and illumination restoration group includes DEA-Net (Chen et al., 2024b), DarkIR (Feijoo et al., 2025), DeepSN-Net (Deng et al., 2025), DehazeFormer (Song et al., 2023), Retinexformer (Cai et al., 2023), C2PNet (Zheng et al., 2023), and DNF (Jin et al., 2023). The document and mixed-degradation group includes DocRes (Zhang et al., 2024b) and OneRestore (Guo et al., 2024). The related-degradation group includes BracketFlare (Dai et al., 2023), DRSformer (Chen et al., 2023), and WGWS-Net (Zhu et al., 2023). Table 1 reports the compact main quantitative comparison.

### A.6 Supplementary Main Comparison Visualization

Supplementary Figure S2 extends the qualitative evidence with five paired examples and additional baseline outputs. Each row aligns the wall-reflection observation, the corresponding screen target, our reconstruction, and the comparison outputs.

The recovered window regions, foreground organization, and coarse color arrangement provide further examples of screen content emerging from diffuse wall observations.

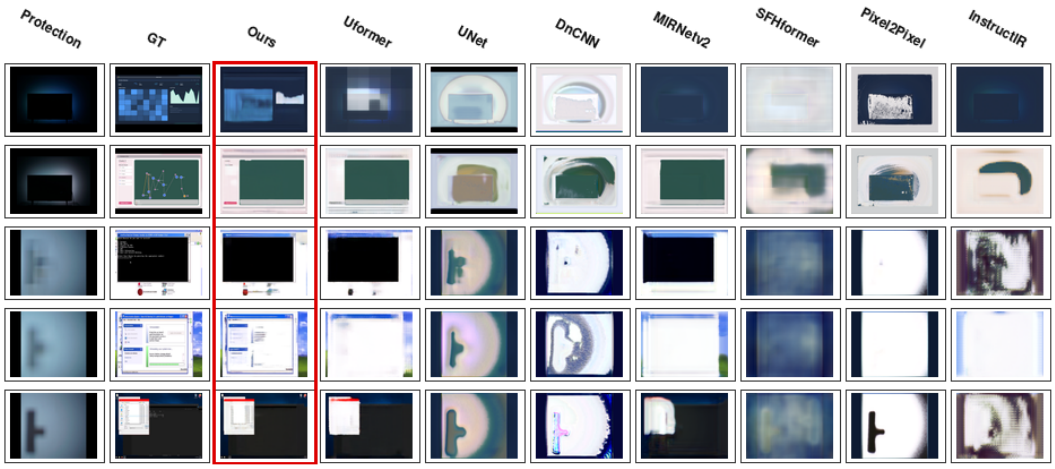



Figure S2: Supplementary qualitative comparison on five paired examples. Each row shows the wall-reflection observation, corresponding screen target, our reconstruction, and representative baseline outputs.

### A.7 Qualitative Ablation Results

The following fixed-sample visualizations expose how each stage changes the recovered screen structure alongside the quantitative results in Tables 2, 3, and 4. All three comparisons use the same four evaluated cases. Blue borders mark the ground-truth images, while red borders mark the complete model. Together, they show a clear progression: SPDRL organizes evidence about reliable boundaries and layout within the mixed reflection, and RCCA converts that representation into controlled local corrections toward the target structure. For the SPDRL readout ablation, Complementary Head is defined by $s_{\mathrm{cmp}} = \sigma(\Psi_{\mathrm{cmp}}(r))$, $s = s_{\mathrm{cmp}}$, and $d = 1 - s_{\mathrm{cmp}}$, where $\Psi_{\mathrm{cmp}}$ is the shared readout head.

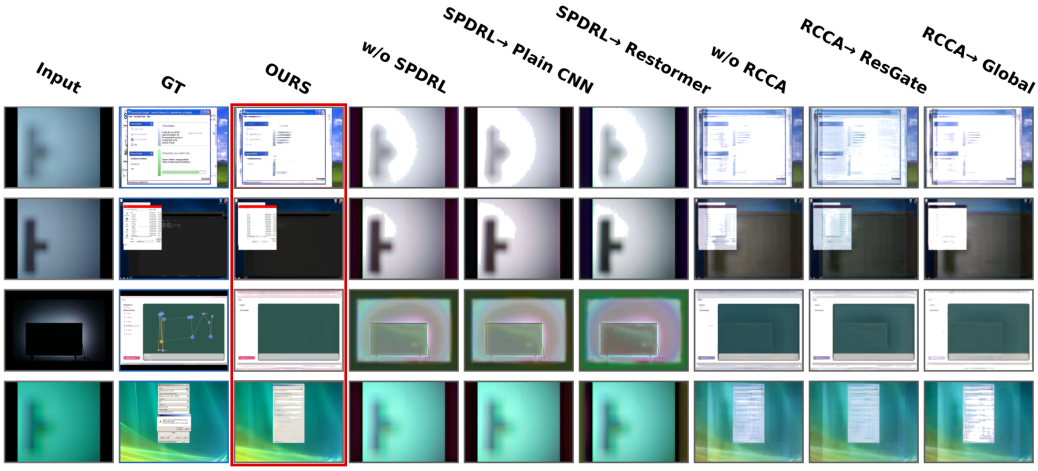



Figure S3: Overall component ablation on fixed samples. SPDRL organizes evidence for screen structure from the diffuse input, and RCCA further converts that representation into controlled corrections. Table 2 provides the complete-test-set metrics.

The overall comparison distinguishes changes to the representation stage from changes to the refinement stage. In Figure S3, the variants without the original SPDRL retain diffuse bright regions or fail to recover the target interface arrangement, whereas the RCCA variants recover more of the coarse layout but differ in local structure and color. These examples motivate examining the internal design of both stages. Each output is compared with the paired target in its row, so that a visually sharper region is not by itself evidence of more faithful recovery. Table 2 provides the corresponding evaluation over the full test sets.

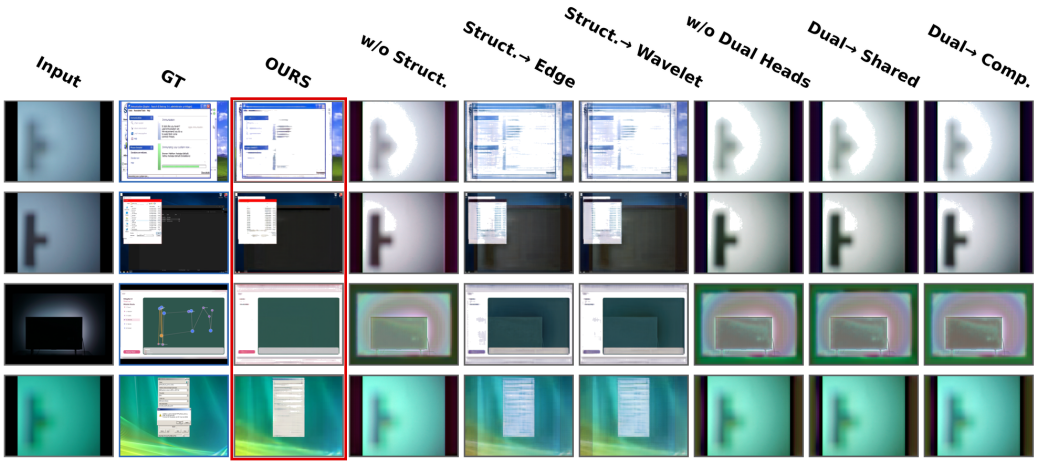



Figure S4: Internal SPDRL ablation on fixed samples. The complete structure path and dual readouts recover screen boundaries and layout more coherently than the tested variants. Table 3 provides the complete-test-set metrics.

The SPDRL comparison separates the choice of structure features from the way the two readouts are produced. In Figure S4, the generic edge and wavelet detail paths recover recognizable screen regions in some examples, while differences remain in the boundaries and internal arrangement. The shared and complementary head variants examine a separate design choice. In particular, the complementary head forces the two maps to sum to one, so increasing one necessarily decreases the other. This constraint differs from estimating the two readouts separately. The comparison therefore examines both the structure path and the readout design, with quantitative results reported in Table 3.

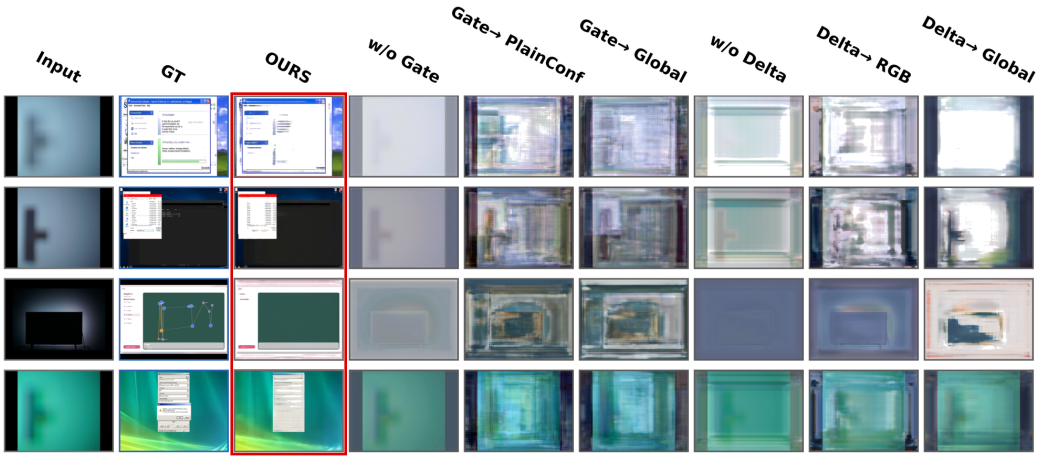



Figure S5: Internal RCCA ablation on fixed samples. Position-wise admission produces cleaner screen regions and more stable boundaries than the tested gate and candidate-correction variants. Table 4 provides the corresponding complete-test-set metrics.

The RCCA variants distinguish the proposed correction from the gate that controls its use. Figure S5 compares both choices against the same targets, with quantitative results in Table 4. These ablations change the reconstruction model itself; the following robustness experiments instead examine variations in the capture conditions and observations.

## A.8 Robustness Experiments

### A.8.1 Physical Scene Settings

Figures S6–S8 show variations in the relay surface, camera position, and illumination. These settings change the surface response, observation geometry, and lighting through which the screen content reaches the camera, providing the physical context for the following recovery evaluations. The photographs show the same screen-back reflection configuration as the dataset: the camera records the display back and the light reflected by the relay surface.




Figure S6: Wall-surface settings. Columns show wood-finish, tile-finish, and matte white surfaces. The top row shows the capture setups; the bottom row shows the corresponding surface appearance.

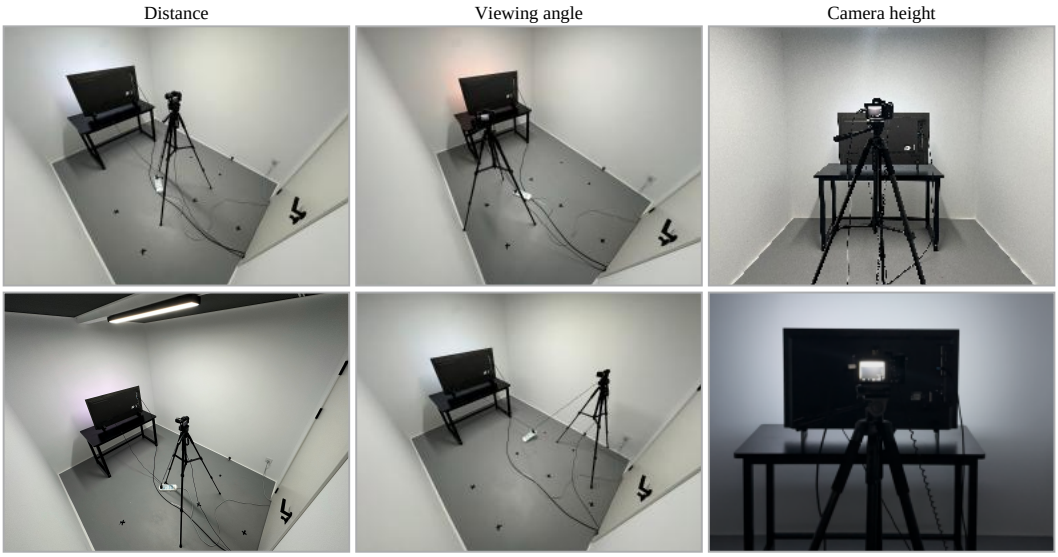



Figure S7: Camera-position settings. Columns illustrate changes in distance, viewing angle, and height, respectively. Each column shows two documented configurations.

The position examples describe acquisition geometry rather than reconstruction results. Distance affects the apparent scale of the observed setup, while viewing angle and height affect its framing. The quantitative comparisons below focus on viewpoint and relative distance. The height photographs document an additional capture configuration and are not presented as a separate quantitative sweep.

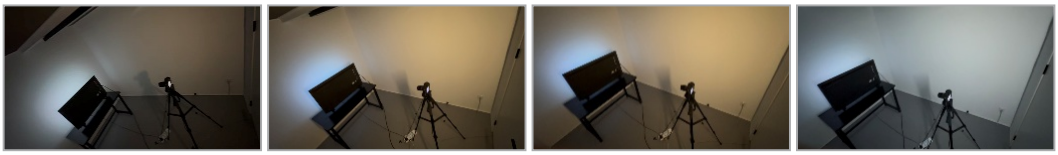



Figure S8: Illumination settings. The left pair shows ambient-light intensity settings; the right pair shows ambient-light color-temperature settings.

The illumination examples separate ambient-light intensity from color temperature. The former concerns the strength of ambient illumination, whereas the latter concerns its chromatic balance. This distinction connects the scene photographs to the color temperature evaluation in the main text and to the brightness and spatial illumination evaluations below. The photographs document capture conditions; performance comparisons are provided by the corresponding metric curves.

### A.8.2 Wall-Surface Variation

Figure S6 shows wood-finish, tile-finish, and matte white relay surfaces. Their different textures and reflection appearances alter the spatial and chromatic distribution of the screen-emitted light, connecting surface variation to the structure–interference mixture addressed by SPDRL.

### A.8.3 Camera-Position Robustness

Figure S7 illustrates changes in camera distance, viewing angle, and height. Supplementary Figures S9 and S10 evaluate the corresponding changes in viewpoint and observation scale. The robustness comparisons additionally include X-Restormer, which combines spatial and channel attention for general-purpose image restoration (Chen et al., 2024a).

**Viewpoint robustness.** Viewpoint variation changes the geometry through which the reflected light is observed. Supplementary Figure S9 shows that our method retains its recovery advantage on Figure and Lock across settings 3, 5, 8, 10, and 15. The result extends the recovery evidence from changes in illumination to changes in observation geometry.

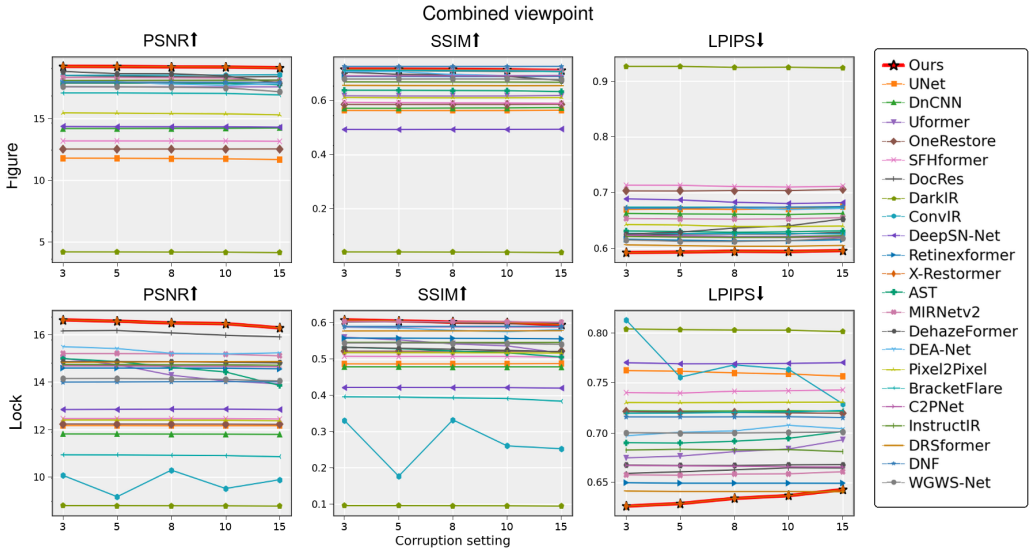



Figure S9: Viewpoint results on Figure and Lock at settings 3, 5, 8, 10, and 15. Higher PSNR and SSIM and lower LPIPS indicate more favorable values (Zhang et al., 2018; Hore & Ziou, 2010).

**Camera-distance robustness.** Supplementary Figure S10 evaluates camera-distance variation on Game and Lock at relative distance multipliers of 1.0×, 1.125×, 1.25×, 1.375×, and 1.5×. The method retains its recovery advantage as the apparent scale of the wall observation changes. Unlike the preceding viewpoint experiment, this evaluation focuses on observation scale rather than viewing direction. Together, the two comparisons examine complementary aspects of camera placement, assessing whether reflected screen cues continue to support reconstruction across changes in both viewing geometry and apparent scale.

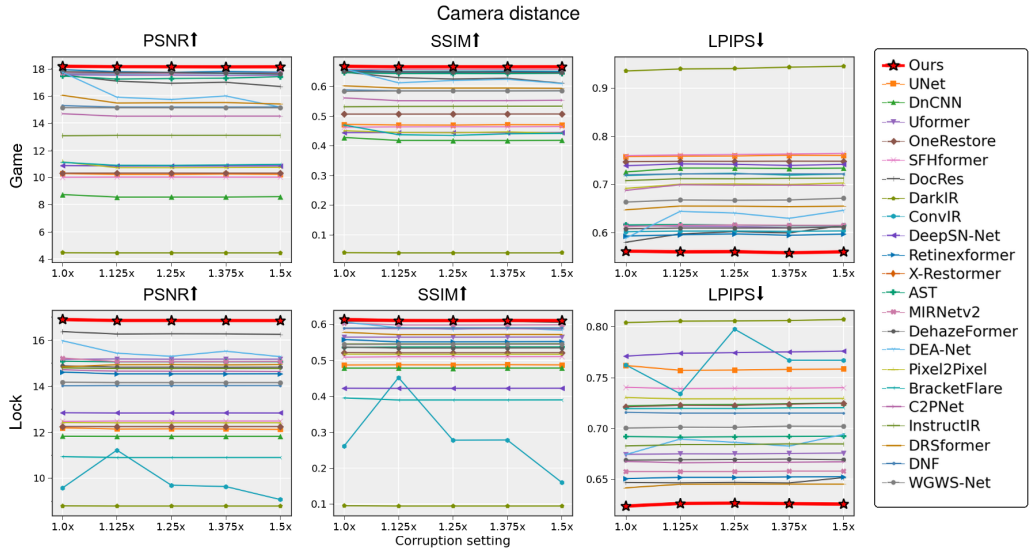



Figure S10: Camera-distance results at relative distance multipliers 1.0×, 1.125×, 1.25×, 1.375×, and 1.5×. The top row shows Game, and the bottom row reports Lock. From left to right, the columns show PSNR, SSIM, and LPIPS, respectively. Higher PSNR and SSIM and lower LPIPS indicate more favorable performance (Zhang et al., 2018; Hore & Ziou, 2010). The complete proposed model, denoted Ours, is shown by the red curves with black star markers.

The recovered screen organization remains stable across these distance settings, extending the recovery result to changes in observation scale.

### A.8.4 Brightness Reduction Robustness

The next question is whether recovery remains possible as the reflected observation becomes dimmer. Supplementary Figure S11 evaluates brightness reduction on Figure and Lock at settings of 4, 8, 12, 18, and 24. Figure contains fine chart structures, whereas Lock combines broad color regions with sparse interface elements. Evaluating both allows the response to reduced brightness to be examined across these different forms of screen content.

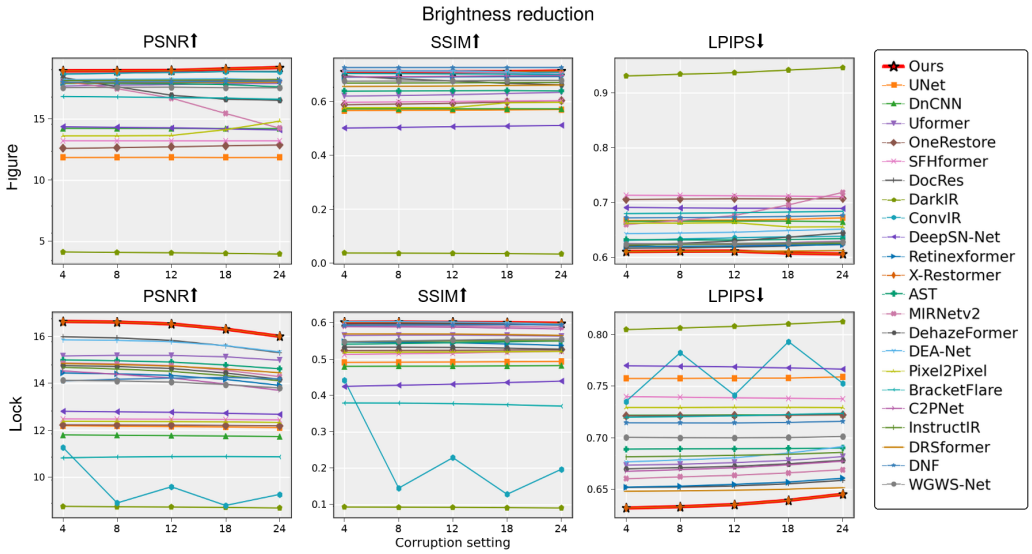



Figure S11: Brightness-reduction results at evaluated settings 4, 8, 12, 18, and 24. The top row reports Figure, and the bottom row reports Lock. From left to right, the columns show PSNR, SSIM, and LPIPS, respectively. Higher PSNR and SSIM and lower LPIPS indicate more favorable performance (Zhang et al., 2018; Hore & Ziou, 2010). The red curves with black star markers denote Ours.

Our method retains its recovery advantage across the brightness settings, showing that the remaining reflection cues continue to support the principal screen content and structure.

In Figure S11, the overall trends for Ours are a reduction in PSNR and SSIM and an increase in LPIPS as the observation becomes dimmer. Thus, retaining a comparative advantage does not imply that reconstruction quality is unchanged. The comparison considers both the metric level at each setting and the change across settings: a nearly flat curve at a low recovery level would not, by itself, demonstrate effective robustness. The Figure and Lock rows provide separate views of this response, since their content differs. The results support recovery over the evaluated brightness range without implying that information lost under stronger attenuation can always be recovered.

### A.8.5 Non-Uniform Illumination Robustness

Non-uniform illumination changes the strength of reflection cues differently across the image. Supplementary Figure S12 reports Game and Screen at perturbation levels 0, 0.02, 0.04, 0.06, and 0.08. Our method maintains leading recovery across the five illumination gradients, showing that useful screen information can be recovered when local contrast varies by position. This spatial variation complements the overall brightness experiment and directly matches the need for location-dependent refinement.

Unlike an overall reduction in brightness, a spatial illumination gradient changes the relative strength of cues at different image locations. This motivates evaluating recovery when different regions offer unequal visual evidence. The Game and Screen curves in Figure S12 show how the final output responds to this spatial variation. Their aggregate scores evaluate reconstruction quality, but do not by themselves identify which locations caused a change. The task-conditioned diagnostics in Appendix A.11 provide complementary evidence about structure and degradation-location alignment, keeping this output-level robustness test distinct from the analysis of the intermediate representation.

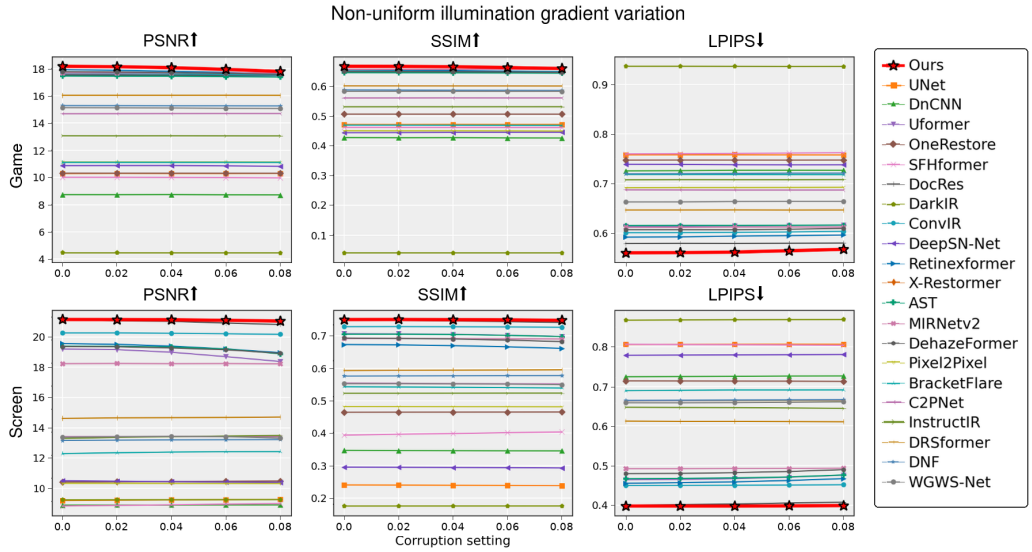



Figure S12: Non-uniform illumination-gradient results at perturbation levels 0, 0.02, 0.04, 0.06, and 0.08. The top row reports Game, and the bottom row reports Screen. From left to right, the columns show PSNR, SSIM, and LPIPS, respectively. Higher PSNR and SSIM and lower LPIPS indicate more favorable performance (Zhang et al., 2018; Hore & Ziou, 2010). The complete proposed model, denoted Ours, is shown by the red curves with black star markers.

### A.8.6 Perspective-Distortion Robustness

Perspective distortion changes the spatial arrangement of the wall observation. Across the five distortion levels 0, 0.005, 0.01, 0.015, and 0.02, our method maintains its recovery advantage on Game and Lock, showing that the reflected cues continue to support screen reconstruction under altered perspective.

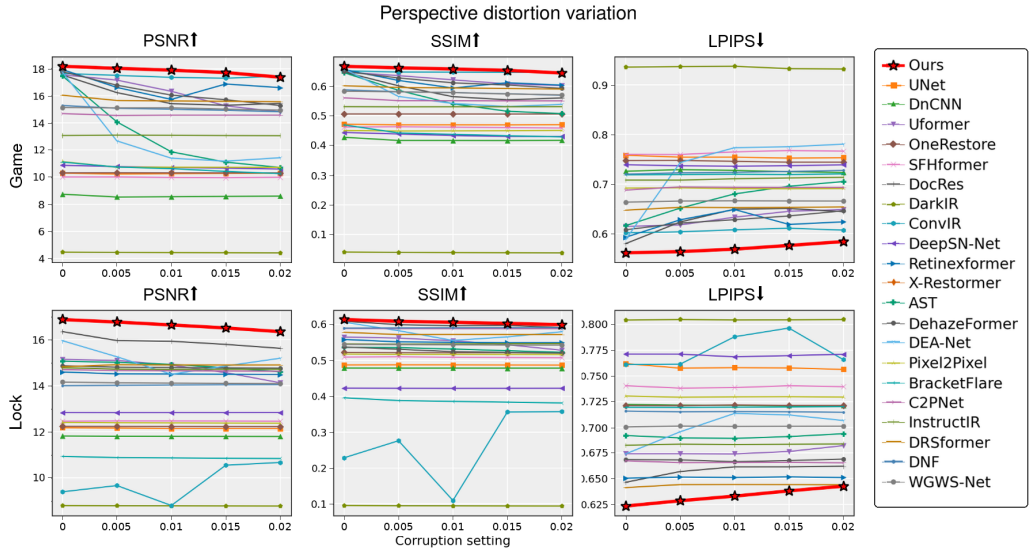



Figure S13: Perspective-distortion results at perturbation levels 0, 0.005, 0.01, 0.015, and 0.02. The top row reports Game, and the bottom row reports Lock. From left to right, the columns show PSNR, SSIM, and LPIPS, respectively. Higher PSNR and SSIM and lower LPIPS indicate more favorable performance (Zhang et al., 2018; Hore & Ziou, 2010). The complete proposed model, denoted Ours, is shown by the red curves with black star markers.

### A.8.7 LENS-DISTORTION ROBUSTNESS

Figure S14 evaluates nonlinear image deformation on Game and Lock. Our method retains leading recovery across all five tested distortion levels.

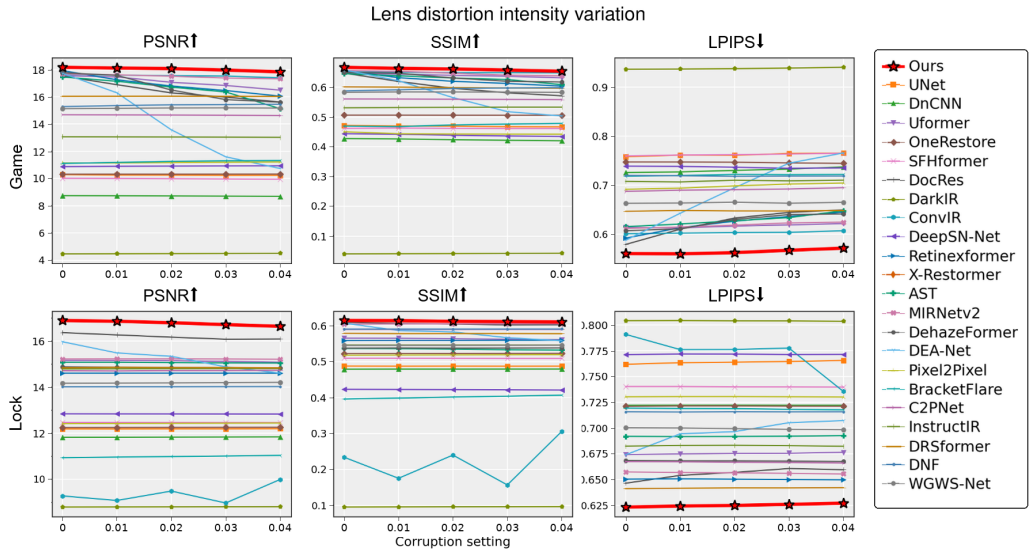



Figure S14: Lens-distortion results at distortion levels 0, 0.01, 0.02, 0.03, and 0.04. The top row reports Game, and the bottom row reports Lock. From left to right, the columns show PSNR, SSIM, and LPIPS, respectively. Higher PSNR and SSIM and lower LPIPS indicate more favorable performance (Zhang et al., 2018; Hore & Ziou, 2010). The complete proposed model, denoted Ours, is shown by the red curves with black star markers.

### A.8.8 MILD CAMERA-SHAKE ROBUSTNESS

Figure S15 evaluates mild camera shake on Figure and Lock. Our method retains its comparative advantage, with little variation across the five tested settings.

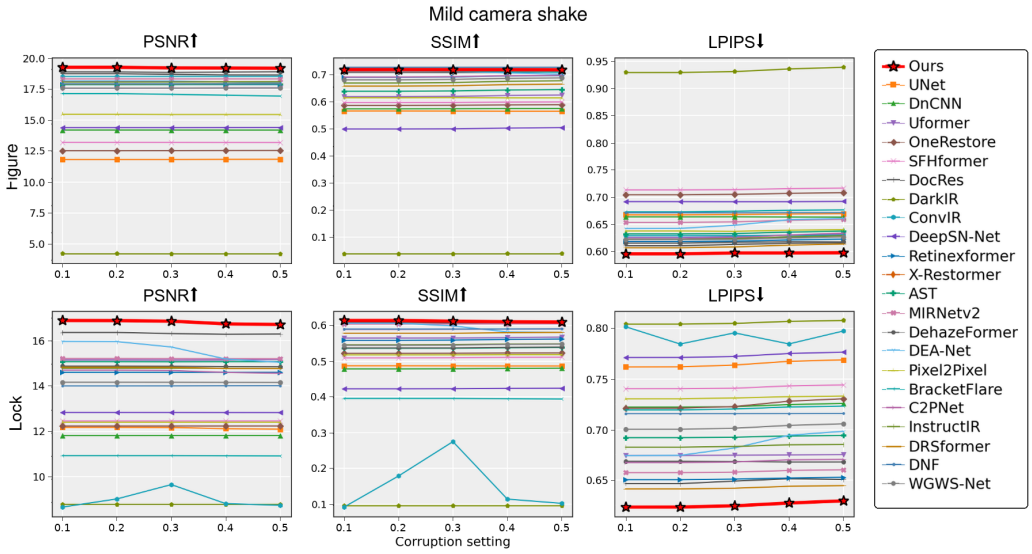



Figure S15: Mild camera-shake results at perturbation levels 0.1, 0.2, 0.3, 0.4, and 0.5. The top row reports Figure, and the bottom row reports Lock. From left to right, the columns show PSNR, SSIM, and LPIPS, respectively. Higher PSNR and SSIM and lower LPIPS indicate more favorable performance (Zhang et al., 2018; Hore & Ziou, 2010). The complete proposed model, denoted Ours, is shown by the red curves with black star markers.

### A.8.9 Motion Blur Robustness

Motion blur further spreads the already diffuse reflection cues. Our method retains its recovery advantage as the blur setting increases, showing that the remaining spatial organization continues to support reconstruction of the principal screen content.

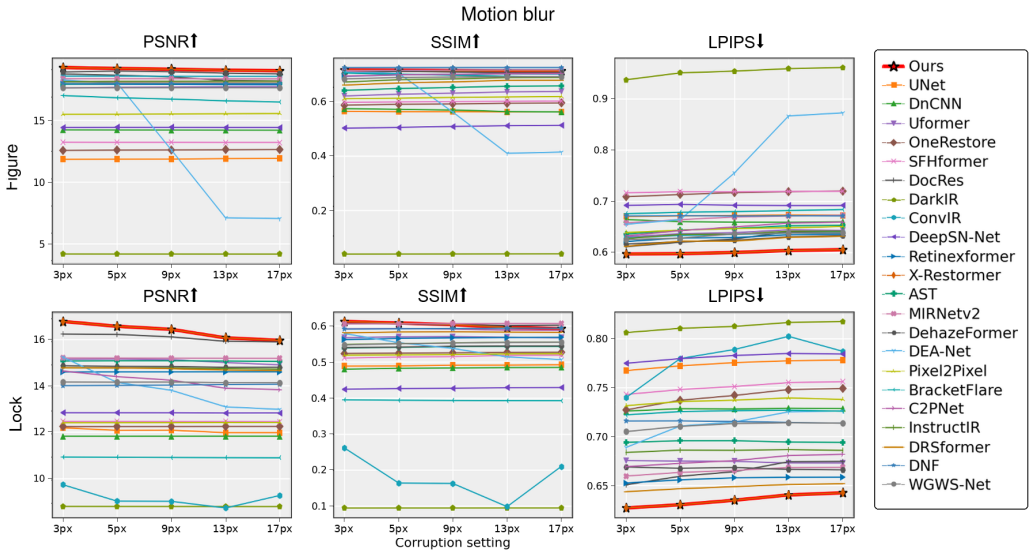



Figure S16: Motion-blur results on Figure and Lock at settings of 3, 5, 9, 13, and 17 pixels. Columns show PSNR, SSIM, and LPIPS. Higher PSNR and SSIM and lower LPIPS indicate more favorable values (Zhang et al., 2018; Hore & Ziou, 2010).

### A.8.10 Defocus Blur Robustness

Defocus weakens local contrast and spreads fine variations in the wall observation. Our method maintains its recovery advantage on Game and Lock across the five blur levels 0, 0.2, 0.4, 0.6, and 0.8, showing that useful screen information remains available as the reflected details become less distinct.

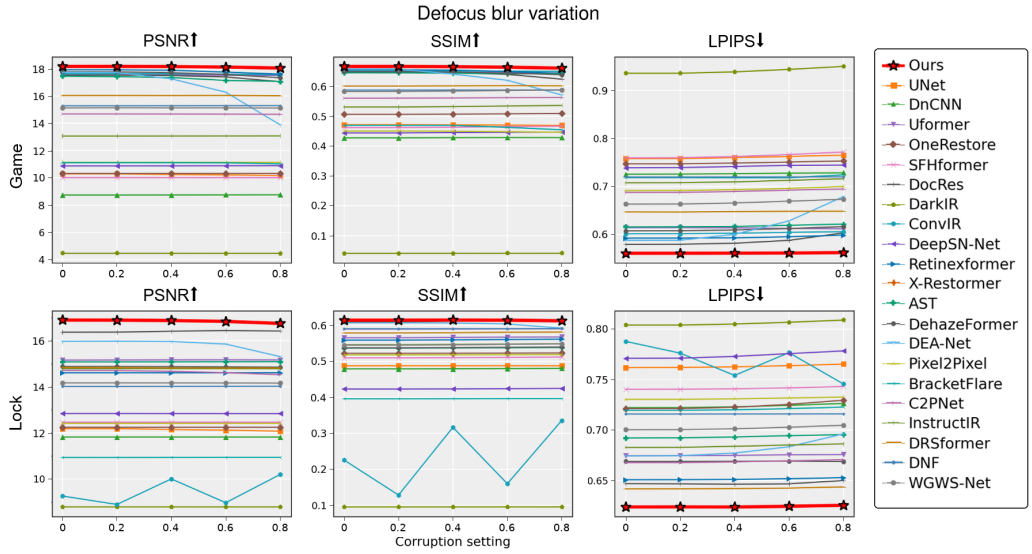



Figure S17: Defocus-blur results at blur levels 0, 0.2, 0.4, 0.6, and 0.8. The top row reports Game, and the bottom row reports Lock. From left to right, the columns show PSNR, SSIM, and LPIPS, respectively. Higher PSNR and SSIM and lower LPIPS indicate more favorable performance (Zhang et al., 2018; Hore & Ziou, 2010). The complete proposed model, denoted Ours, is shown by the red curves with black star markers.

### A.8.11 JPEG Compression Robustness

We test whether useful screen cues persist through JPEG compression, evaluating Figure and Lock at quality settings 100, 98, 96, 94, and 92 (Supplementary Figure S18). Our method maintains leading recovery across the five JPEG quality settings. The small changes in reconstruction quality show that the screen information used by the model remains effective through this compression range.

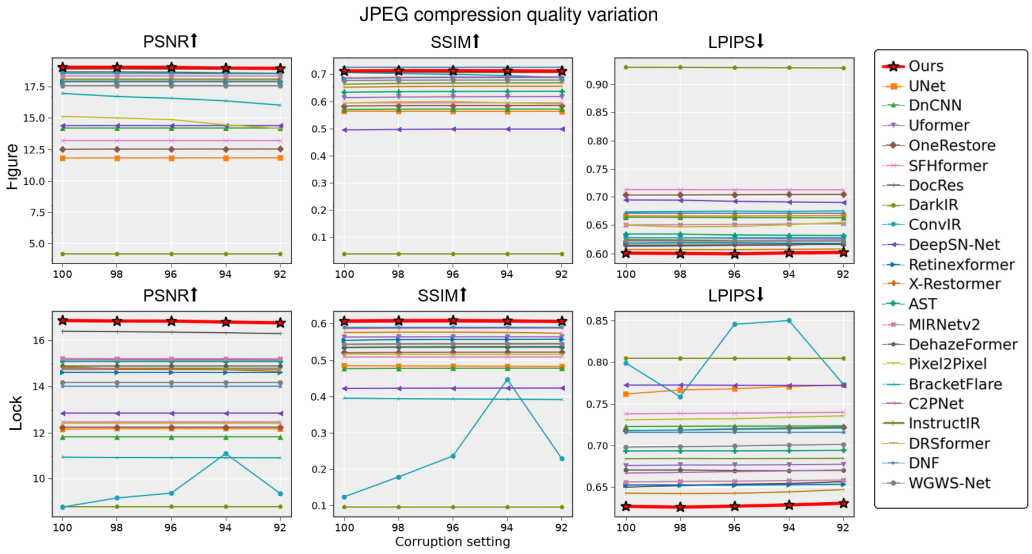



Figure S18: JPEG-compression results at quality settings 100, 98, 96, 94, and 92. The top row reports Figure, and the bottom row reports Lock. From left to right, the columns show PSNR, SSIM, and LPIPS, respectively. Higher PSNR and SSIM and lower LPIPS indicate more favorable performance (Zhang et al., 2018; Hore & Ziou, 2010). The complete proposed model, denoted Ours, is shown by the red curves with black star markers.

JPEG compression tests a different change in the observation from the preceding blur experiments. The relevant question is not whether the compressed wall image appears visually unchanged, but whether its remaining cues still support recovery of the hidden screen. In Figure S18, PSNR, SSIM, and LPIPS provide complementary summaries of this downstream effect. The small variations for Ours are consistent with useful recovery cues surviving within the evaluated quality range. The Figure and Lock rows also keep the two content types separate, allowing compression sensitivity to be assessed without averaging away their different recovery levels. Since the tested quality factors span 100 to 92, the conclusion is limited to this range and does not establish robustness to severe compression.

### A.8.12 Screen-Boundary Occlusion Robustness

We test whether the remaining reflection cues support screen recovery by evaluating Screen and Lock at boundary-occlusion levels of 0%, 1%, 2%, 3%, and 4% (Supplementary Figure S19). Masking near the display boundary removes part of the local observation and its neighboring reflection cues.

Boundary occlusion is more localized: it explicitly masks part of the observation near the screen boundary. This makes the experiment complementary to perturbations that change image appearance without applying such a mask. The same occlusion percentage need not produce the same effect for Screen and Lock, since the distribution of useful reflection cues can depend on content. The two dataset rows are therefore interpreted separately. This experiment evaluates tolerance to the specified boundary masks; the later nine-region analysis asks which input locations contribute to recovery. Together, they relate robustness under missing evidence to the spatial organization of that evidence.

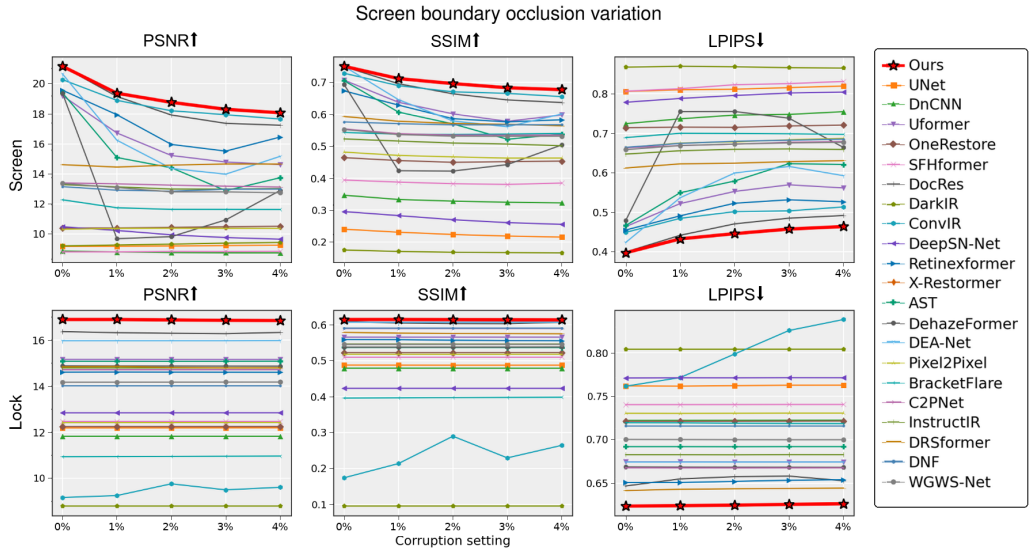



Figure S19: Screen-boundary occlusion results at occlusion levels 0%, 1%, 2%, 3%, and 4%. The top row reports Screen, and the bottom row reports Lock. From left to right, the columns show PSNR, SSIM, and LPIPS, respectively. Higher PSNR and SSIM and lower LPIPS indicate more favorable performance (Zhang et al., 2018; Hore & Ziou, 2010). The complete proposed model, denoted Ours, is shown by the red curves with black star markers.

Our method retains its comparative advantage as boundary-region evidence is removed. The remaining input cues continue to support recovery on Screen and Lock, connecting the occlusion results to the spatially distributed evidence examined in the following regional analysis.

### A.9 Comparison with Conventional Methods

BM3D performs collaborative transform-domain denoising, Multi-Scale Retinex adjusts image appearance through multiscale illumination processing, and Fast SIRS suppresses image reflections through convex optimization (Dabov et al., 2007; Jobson et al., 1997; Yang et al., 2019). Our learned screen-recovery mapping recovers the hidden screen's content and leads all 36 pairwise metric comparisons across the four content types (Table S1), achieving a Screen SSIM of 0.7501 compared with 0.4238–0.4653 for these operators. This advantage connects the treatment of reflected observations to the reconstruction of the hidden screen itself; the component ablations explain how SPDRL and RCCA contribute to that reconstruction.

Table S1: Comparison with three conventional restoration methods on four fixed test sets. Our method wins all 36 pairwise metric comparisons. Metric definitions follow (Zhang et al., 2018; Hore & Ziou, 2010). PSNR is measured in dB. Red and blue values indicate the best and second-best results in each dataset–metric column, respectively. The shaded row highlights our method.

| **Method** | **Screen** | | | **Game** | | | **Figure** | | | **Lock** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ |
| BM3D | 9.73 | 0.4523 | 0.7662 | 4.81 | 0.1428 | 0.8641 | 4.58 | 0.1528 | 0.8613 | 9.00 | 0.1527 | 0.7749 |
| MSR | 6.83 | 0.4238 | 0.7890 | **7.26** | **0.3323** | **0.8099** | **7.04** | **0.3999** | **0.8032** | 8.81 | **0.3605** | 0.8494 |
| Fast SIRS | **9.92** | **0.4653** | **0.7319** | 5.22 | 0.2139 | 0.8382 | 4.86 | 0.2199 | 0.8344 | **9.13** | 0.1887 | **0.7586** |
| **Ours** | **21.16** | **0.7501** | **0.3979** | **18.23** | **0.6670** | **0.5608** | **19.28** | **0.7186** | **0.5959** | **16.93** | **0.6152** | **0.6227** |

Supplementary Table S1 shows a consistent advantage in recovering screen content over the three conventional operators. Supplementary Figure S20 complements this comparison with eight paired examples, while Figure 3 shows how the learned recovery mapping reconstructs major screen regions and layout. Together, these results connect the quantitative advantage to the recovery of screen organization from reflected light.

The paired examples below distinguish enhancement of the wall observation from reconstruction of the hidden screen. Each output is compared with its corresponding target in terms of layout, region boundaries, and coarse color organization. A brighter or sharper wall image does not necessar-

ily reproduce the arrangement of the screen content. The comparison therefore considers whether the recovered regions correspond to the target, rather than whether the observation itself appears cleaner. These eight paired examples complement the aggregate metrics in Table S1 by showing the correspondence between individual reconstructions and their screen targets.

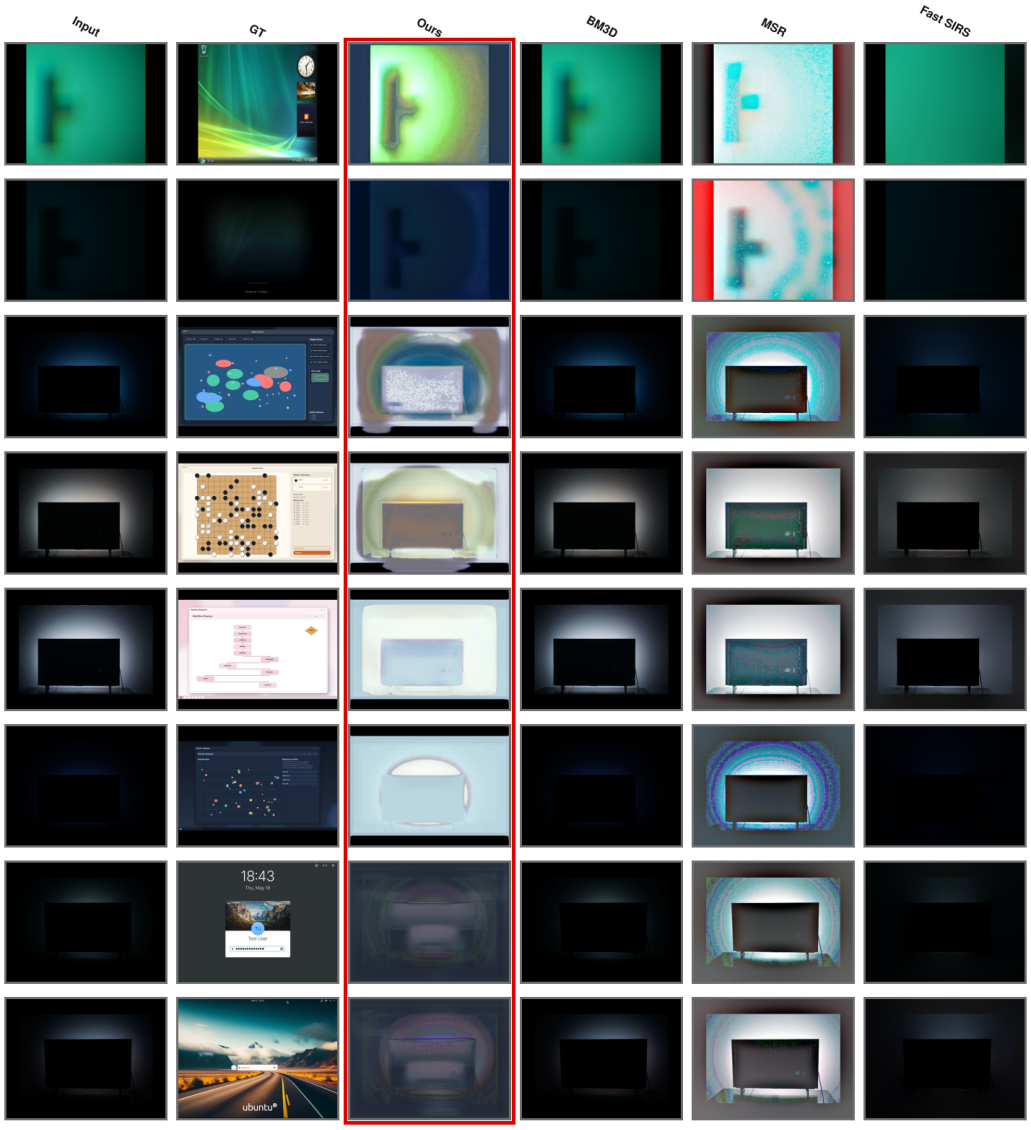



Figure S20: Fixed sample comparison of BM3D, Multi-Scale Retinex, Fast SIRS, and Ours with the shown inputs and targets (Dabov et al., 2007; Jobson et al., 1997; Yang et al., 2019).

These comparisons strengthen the evidence that wall-reflection observations support reconstruction of the hidden screen's principal content and structure.

### A.10 Nine-Region Input Contribution Analysis

Which regions of the wall observation support the recovered screen content? Dividing the input into a fixed $3 \times 3$ grid allows us to relate changes in the available reflection evidence to changes in the reconstructed screen.

Supplementary Figure S21 masks one grid region at a time. The resulting changes in window regions and boundaries reveal the contributions of spatially distributed input cues to the reconstruction.

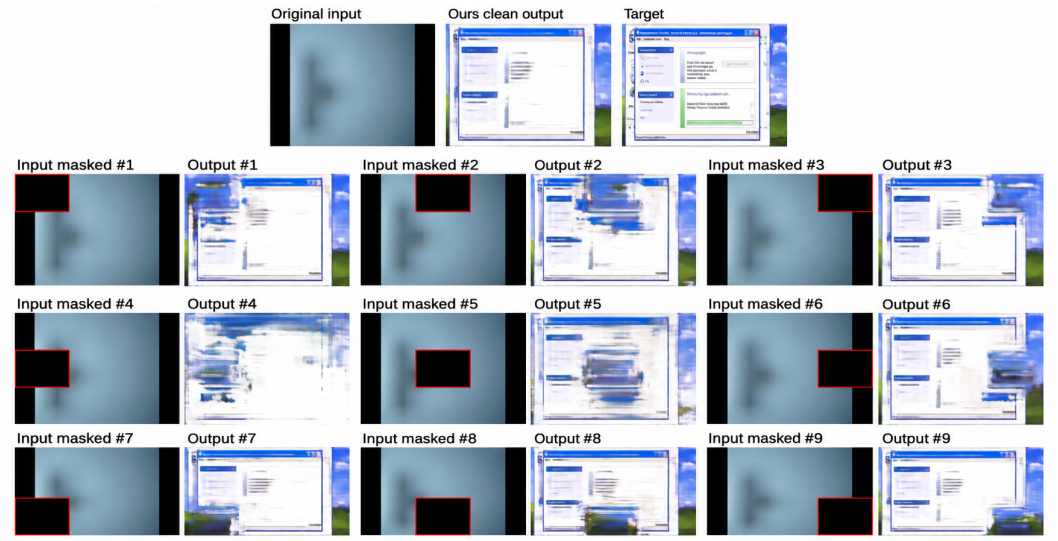



Figure S21: Nine-region input occlusion analysis for one Screen sample. One grid region is masked at a time, and the corresponding restored output is shown together with the unmasked input, original output, and target.

Supplementary Figure S22 complements the masking analysis by retaining one grid region at its original position and masking the remaining input. The resulting reconstructions reveal what screen information each retained region can support.

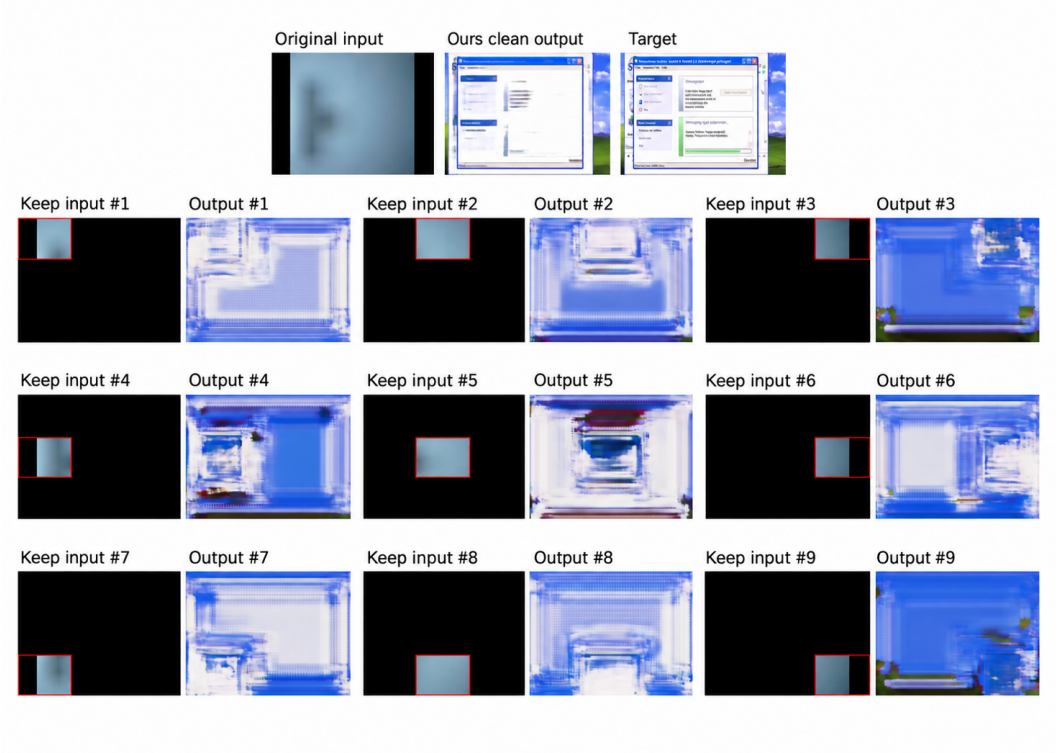



Figure S22: Nine-region keep-only analysis for one Screen sample. Each panel retains one input region at its original position and shows the corresponding reconstruction.

Supplementary Figure S23 visualizes how the contribution of the observed reflection varies across input regions. Supplementary Table S2 extends this regional analysis across the four screen-content types. The heat map provides a compact spatial summary of the diagnostic and should be read alongside the masking and keep-only examples. Masking examines how recovery changes when a

region is removed, whereas the keep-only setting shows what reconstruction remains possible from that region alone. These complementary views help distinguish sensitivity to missing evidence from recovery supported by restricted evidence. The regional scores describe the evaluated inputs and diagnostic settings; they do not imply that the same location is always dominant across different screen layouts. Comparing the four content types therefore places the illustrated Screen example in a broader context while preserving this sample-dependent interpretation.

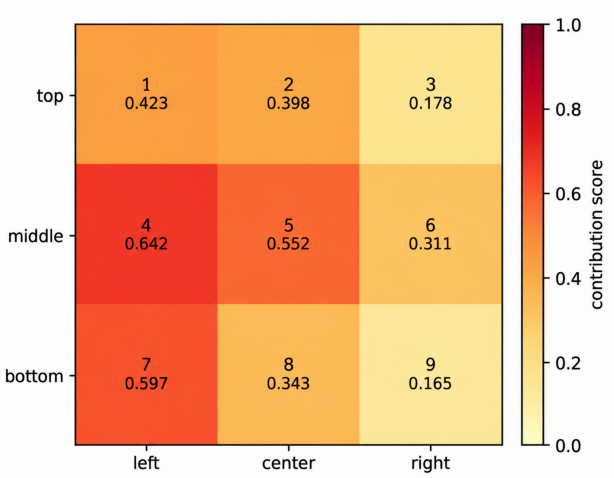



Figure S23: Nine-region contribution heat map for the Screen sample. The heat map summarizes the relative contribution of each grid region under the defined input-region diagnostic.

Table S2: Nine-region contribution scores for the input occlusion and keep-only diagnostics. Higher values indicate stronger regional contribution under the defined diagnostic setting. Red and blue values mark the highest and second-highest contribution scores among the nine regions for each screen-content type, respectively.

| Dataset | Top | | | Middle | | | Bottom | | |
|---|---|---|---|---|---|---|---|---|---|
| | Left | Center | Right | Left | Center | Right | Left | Center | Right |
| Screen | 0.3380 | 0.6034 | 0.0773 | **0.7404** | **0.8888** | 0.6196 | 0.5235 | 0.6500 | 0.0418 |
| Game | 0.2150 | 0.2124 | 0.2021 | 0.6443 | **0.9818** | 0.5777 | 0.4475 | **0.6554** | 0.2988 |
| Figure | 0.4788 | 0.2341 | 0.5067 | **0.7091** | **0.8925** | 0.7041 | 0.0995 | 0.2198 | 0.6053 |
| Lock | 0.2278 | **1.0000** | 0.3438 | 0.3569 | 0.0000 | 0.5282 | 0.0222 | **0.5569** | 0.0615 |

Together, the masking, keep-only, and contribution-map analyses connect the recovered screen to spatially distributed reflection evidence. The strongest regional contribution in Table S2 occurs at the middle center for Screen, Game, and Figure, and at the top center for Lock. These content-dependent contributions provide complementary support for reconstruction and motivate SPDRL's joint organization of neighboring cues.

### A.11 Task-Conditioned Representation Diagnostics

The regional analysis identifies where useful input evidence is located; we now move from those input regions to the representations that organize the evidence for recovery. The diagnostics examine alignment with task-conditioned structure, degradation-location, and coupled structure–degradation targets. These targets combine target-side structural evidence, observation-derived degradation cues, and SPDRL mechanism responses, yielding task-conditioned measures of alignment with the recovery mechanism. We measure alignment using the area under the precision–recall curve (AUPRC), F1 score, and intersection over union (IoU), with higher values indicating stronger performance.

**Implementation.** All variants use a common U-Net restoration framework with NAFNet-style, Restormer-style, FFTformer-style, and SPDRL representation producers at spatial size $192 \times 256$. For each non-SPDRL representation, a trained $1 \times 1$ Conv probe uses BCEWithLogits, AdamW with learning rate $10^{-2}$ and weight decay $10^{-5}$, 12 epochs, batch size 8, and checkpoint selection by validation mean AUPRC. SPDRL uses its explicit mechanism scores. The structure diagnostic is a weighted combination of target-side edge evidence, input edge evidence, the RefDet cue, and the SPDRL structure response. The degradation diagnostic combines the input–target residual, RefDet cue, and SPDRL degradation response; the coupled diagnostic multiplies the two diagnostic scores.

Within each valid image region, the top 10% of locations are positive. The F1/IoU threshold is selected on validation by maximum F1 and then fixed for test.

Supplementary Table S3 evaluates task-conditioned structure alignment. NAFNet provides a simplified restoration architecture for denoising and deblurring (Chen et al., 2022). SPDRL reaches an AUPRC of 0.7585 on Screen, compared with 0.3761 for NAFNet, and leads all 12 dataset–metric comparisons within this diagnostic. These results identify strong alignment with the structural evidence used to preserve screen edges, text contours, and interface boundaries during recovery.

Table S3: Task-conditioned structure-diagnostic alignment results. Higher AUPRC, F1, and IoU indicate stronger diagnostic alignment. Red and blue values indicate the best and second-best readout results in each column, respectively. The shaded row highlights SPDRL.

| Representation | Screen | | | Game | | | Figure | | | Lock | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AUPRC↑ | F1↑ | IoU↑ | AUPRC↑ | F1↑ | IoU↑ | AUPRC↑ | F1↑ | IoU↑ | AUPRC↑ | F1↑ | IoU↑ |
| NAFNet | **0.3761** | **0.4657** | **0.3036** | **0.3844** | **0.4053** | **0.2542** | **0.5805** | **0.5478** | **0.3773** | **0.3541** | **0.3409** | **0.2055** |
| Restormer | 0.3332 | 0.4289 | 0.2730 | 0.2524 | 0.3375 | 0.2030 | 0.4925 | 0.4791 | 0.3150 | 0.3248 | 0.3200 | 0.1905 |
| FFTformer | 0.3211 | 0.3971 | 0.2477 | 0.2120 | 0.3046 | 0.1797 | 0.2155 | 0.3339 | 0.2004 | 0.3079 | 0.3061 | 0.1807 |
| **SPDRL** | **0.7585** | **0.7258** | **0.5696** | **0.7717** | **0.6738** | **0.5081** | **0.7773** | **0.6959** | **0.5336** | **0.3640** | **0.3526** | **0.2141** |

SPDRL leads all 12 task-conditioned degradation-location comparisons in Supplementary Table S4. On Figure, its AUPRC reaches 0.9463, compared with 0.6331 for the strongest alternative in that column. Together with structure alignment, degradation-location alignment provides complementary spatial evidence for evaluating local correction.

Table S4: Task-conditioned degradation-location alignment results. Higher AUPRC, F1, and IoU indicate stronger diagnostic alignment. Red and blue values indicate the best and second-best readout results in each column, respectively. The shaded row highlights SPDRL.

| Representation | Screen | | | Game | | | Figure | | | Lock | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AUPRC↑ | F1↑ | IoU↑ | AUPRC↑ | F1↑ | IoU↑ | AUPRC↑ | F1↑ | IoU↑ | AUPRC↑ | F1↑ | IoU↑ |
| NAFNet | 0.4031 | 0.4761 | 0.3124 | **0.6045** | 0.5935 | 0.4220 | **0.6331** | **0.6648** | **0.4979** | **0.2534** | **0.2697** | **0.1559** |
| Restormer | **0.4632** | **0.4775** | **0.3136** | 0.5952 | **0.5990** | **0.4275** | 0.3458 | 0.4187 | 0.2647 | 0.2331 | 0.2409 | 0.1370 |
| FFTformer | 0.3820 | 0.4770 | 0.3132 | 0.4354 | 0.5573 | 0.3863 | 0.3397 | 0.4096 | 0.2575 | 0.2199 | 0.2520 | 0.1442 |
| **SPDRL** | **0.8336** | **0.7666** | **0.6216** | **0.8610** | **0.7656** | **0.6202** | **0.9463** | **0.8858** | **0.7950** | **0.3692** | **0.3384** | **0.2037** |

The coupled diagnostic examines alignment with screen structure and degradation location together. Restormer models long-range image interactions through efficient attention (Zamir et al., 2022), while FFTformer uses frequency-domain attention and selective frequency processing for deblurring (Kong et al., 2023). SPDRL reaches an AUPRC of 0.8460 on Screen, compared with 0.4037 for Restormer and 0.3449 for FFTformer, and leads all 12 dataset–metric comparisons within the coupled diagnostic (Supplementary Table S5).

Table S5: Task-conditioned coupled structure–degradation alignment results. Higher AUPRC, F1, and IoU indicate stronger diagnostic alignment. Red and blue values indicate the best and second-best readout results in each column, respectively. The shaded row highlights SPDRL.

| Representation | Screen | | | Game | | | Figure | | | Lock | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AUPRC↑ | F1↑ | IoU↑ | AUPRC↑ | F1↑ | IoU↑ | AUPRC↑ | F1↑ | IoU↑ | AUPRC↑ | F1↑ | IoU↑ |
| NAFNet | 0.3857 | **0.4597** | **0.2984** | 0.1961 | 0.2400 | 0.1363 | **0.4708** | **0.4999** | **0.3333** | **0.2982** | **0.3076** | **0.1817** |
| Restormer | **0.4037** | 0.4401 | 0.2822 | 0.2064 | 0.2568 | 0.1473 | 0.2552 | 0.3334 | 0.2000 | 0.2642 | 0.2633 | 0.1516 |
| FFTformer | 0.3449 | 0.4400 | 0.2821 | **0.2523** | **0.3111** | **0.1842** | 0.2736 | 0.3259 | 0.1947 | 0.2219 | 0.2281 | 0.1287 |
| **SPDRL** | **0.8460** | **0.7825** | **0.6427** | **0.6669** | **0.5731** | **0.4017** | **0.7867** | **0.7050** | **0.5444** | **0.4081** | **0.3629** | **0.2217** |

Within this task-conditioned setting, SPDRL achieves the strongest alignment with both structure and degradation location, including their coupled response, providing complementary evidence for RCCA's position-wise correction. The following experiment tests how that organization supports recovery when the representation is disturbed.

### A.12 Representation Stress Test

We next apply a structured internal disturbance to test whether this organization continues to support recovery. All variants use the same U-Net restoration framework with NAFNet-style, Restormer-

style, FFTformer-style, and SPDRL representation producers. The common perturbation stress test uses selected checkpoints from the corresponding exploratory representation runs, retaining their representation-specific training budgets.

For every representation, the shared perturbation is

$$\begin{aligned} r_{\text{shift}} &= \text{roll}(r, (2, 2)), \\ r' &= r(1 - \alpha M_s) + \alpha M_d(r_{\text{shift}} - r), \end{aligned} \tag{S39}$$

where $M_s$ is the structure-region mask, $M_d$ is the degradation-region mask, and $\alpha \in \{0, 0.25, 0.50, 0.75\}$. Structure regions undergo representation attenuation, while degradation regions undergo spatially shifted feature mixing. The same mask construction, shift, formula, and severity are applied to all four representations.

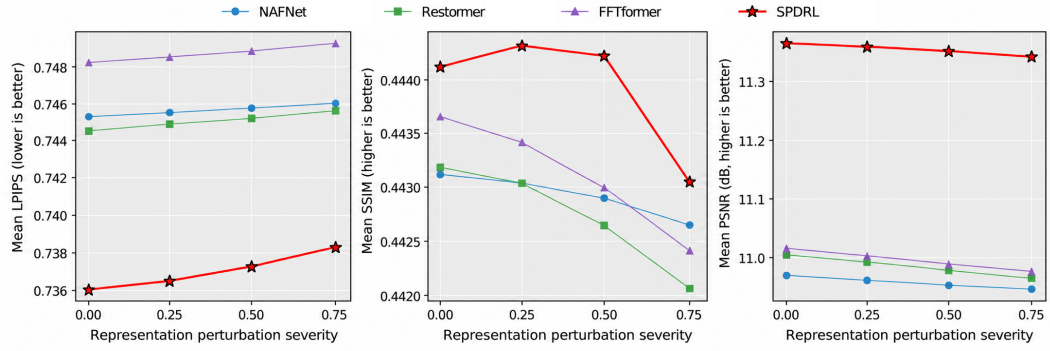



Figure S24: Representation stress test under the shared perturbation for the NAFNet, Restormer, FFTformer, and SPDRL representations. The three columns report average LPIPS, SSIM, and PSNR on the four test datasets. Different curves correspond to different intermediate representations. In this stress test, SPDRL retains the strongest absolute recovery quality across the severity levels.

Table S6 provides the corresponding numerical results averaged over Screen, Game, Figure, and Lock.

Table S6: Average results in the representation stress test under the shared perturbation. The reported values are averaged over Screen, Game, Figure, and Lock. Lower LPIPS and higher SSIM and PSNR indicate better restoration quality. LPIPS and SSIM are shown to four decimal places, and PSNR is shown to three decimal places. Best and second-best highlighting follows the underlying unrounded values. PSNR is measured in dB. Red and blue values indicate the best and second-best results at each perturbation level, respectively. The shaded row highlights SPDRL.

| **Representation** | **0.00** | | | **0.25** | | | **0.50** | | | **0.75** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | LPIPS↓ | SSIM↑ | PSNR↑ | LPIPS↓ | SSIM↑ | PSNR↑ | LPIPS↓ | SSIM↑ | PSNR↑ | LPIPS↓ | SSIM↑ | PSNR↑ |
| NAFNet | 0.7453 | 0.4431 | 10.971 | 0.7454 | 0.4431 | 10.963 | 0.7455 | 0.4429 | 10.956 | 0.7456 | **0.4427** | 10.950 |
| Restormer | **0.7446** | 0.4432 | 11.009 | **0.7447** | 0.4430 | 10.996 | **0.7450** | 0.4427 | 10.982 | **0.7454** | 0.4421 | 10.970 |
| FFTformer | 0.7481 | **0.4436** | **11.015** | 0.7482 | **0.4434** | **11.002** | 0.7485 | **0.4430** | **10.990** | 0.7490 | 0.4424 | **10.978** |
| **SPDRL** | **0.7362** | **0.4441** | **11.363** | **0.7366** | **0.4443** | **11.360** | **0.7373** | **0.4442** | **11.354** | **0.7384** | **0.4431** | **11.342** |

The task-conditioned diagnostics establish how the representations align with the recovery mechanism; this stress test then examines recovery under a common internal disturbance. In this stress test, SPDRL retains the lowest LPIPS and highest SSIM and PSNR at all four severity levels, averaged across Screen, Game, Figure, and Lock (Table S6 and Figure S24). These results extend the evidence from task-conditioned representation alignment to screen reconstruction under internal disturbance.