CAN REFLECTED LIGHT REVEAL SCREEN CONTENT ? LEARNING DEGRADATION-ROUTED EVIDENCE REPRESENTATIONS FOR HIDDEN SCREEN CONTENT RECOVERY

Anonymous authors
Paper under double-blind review

ABSTRACT

Can an ordinary camera recover what a hidden screen displays from the light it casts onto a wall? Propagation and diffuse reflection attenuate and spatially mix the screen signal, while surface texture, ambient illumination, and camera imaging further alter the observation. The resulting faint glow contains screen cues entangled with unrelated appearance. To recover screen content from this mixed glow, we propose Reflection-Evidence Routed Representation Learning (RERL). Consensus-Driven Evidence Representation Learning (CDERL) organizes screen-related cues through encoder-decoder agreement and masked prototype transport. Degradation-Aware Routing (DAR) combines complementary expert transformations with spatial guidance to adapt reconstruction to uneven degradation across the reflection. Across four screen-content groups, RERL outperforms 24 image-restoration baselines in PSNR, SSIM, and LPIPS. Component ablations reveal the complementary contributions of evidence organization and adaptive reconstruction, while perturbation tests and visual analyses demonstrate recovery under varied observation conditions. The reconstructions retain screen layouts, colour regions, and major contours, establishing the feasibility of recovering hidden screen content from wall reflections captured by an ordinary camera.

1 INTRODUCTION

A hidden screen lies outside the camera's direct field of view, while its light reaches the camera through reflection from an ordinary wall (Figure 1). Can the screen content be recovered using only these faint wall reflections? Passive non-line-of-sight imaging and computational periscopy have shown that occluders and diffuse surfaces can encode recoverable spatial information using steady-state intensity measurements Faccio et al. (2020); Yedidia et al. (2019); Saunders et al. (2019). We study recovery of the current front-view RGB screen content from a camera observation of a wall or door illuminated by the hidden display. The task exposes a privacy risk previously illustrated by screen-reflection attacks, and provides a computational-imaging setting for learning from weak reflected screen evidence Backes et al. (2008).

Modern image-restoration networks combine convolution, attention, and multiscale decoding to recover images affected by blur, noise, haze, and illumination changes Zamir et al. (2022); Chen et al. (2022). Their observation models, however, generally assume that the degraded image remains spatially registered with the desired image. A wall-reflected screen observation violates that assumption: unknown propagation and diffuse reflection mix screen emission with surface texture, ambient light, and camera response before the measurement reaches the network. Two coupled difficulties follow: useful screen evidence is compressed and mixed with nuisance appearance, so skip features may carry weak screen traces together with unrelated wall structure. Evidence quality is also spatially variable, yet a common recovery path applies the same transformation family everywhere. Recovery therefore requires selecting screen-related evidence and adapting reconstruction to its uneven quality.

To address the above challenges, we propose Reflection-Evidence Routed Representation Learning (RERL), a recovery path designed to promote evidence alignment and degradation-aware feature transformation. Consensus-Driven Evidence Representation Learning (CDERL) combines alignment of fused decoder and encoder skip features with masked learned-token residuals. Degradation-Aware

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.

Routing (DAR) combines globally weighted expert branches with a spatial self-guidance residual at the high-level decoder stages and uses a lighter global mixture at lower levels. Together, modules select and organize reflected evidence and adapt its reconstruction to spatially varying degradation.

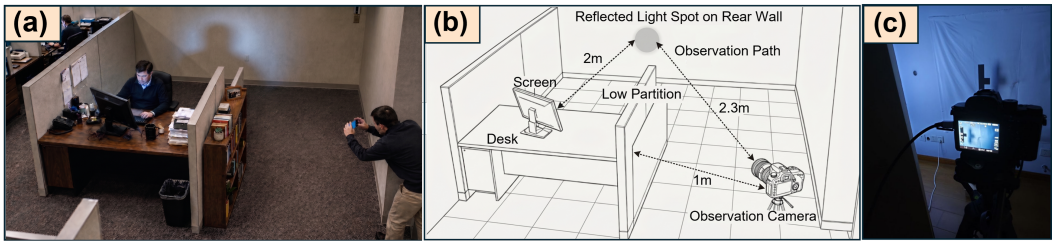

Figure 1: Hidden-screen observation setting. (a) The observation scenario. (b) The screen–surface–camera geometry. (c) A camera recording the reflected screen light in the acquisition environment. The displayed content lies outside the camera's direct view.

The main contributions are summarized as follows.

(1) We introduce RERL for recovering hidden screen content from weak wall reflections captured by an ordinary RGB camera.

(2) We develop CDERL to organize weak screen cues through encoder–decoder feature agreement and masked prototype transport, and DAR to adapt reconstruction through complementary expert transformations and spatial guidance, jointly supporting recovery under spatially variable evidence quality.

(3) Experiments across four content groups demonstrate recovery of screen layouts, colour regions, and major contours from ordinary-camera wall reflections, with advantages over 24 restoration baselines and complementary evidence from ablations, perturbations, and occlusion analysis.

2 Related Work

According to observation regime and output space, existing studies can be broadly divided into passive indirect imaging and learned restoration under mixed degradation.

Passive Indirect Imaging. Passive indirect imaging recovers information outside a direct field of view from steady-state intensity measurements. Yedidia et al. proposed a hidden-scene recovery method, which uses unknown occluders as accidental coding elements to handle hidden-content recovery Yedidia et al. (2019). Saunders et al. proposed a computational periscopy method, which uses corner-induced transport recorded by an ordinary digital camera to handle hidden-content recovery Saunders et al. (2019). These studies establish that steady-state indirect observations can contain structured information, while their sensing configurations and outputs differ from diffuse wall-reflection screen recovery.

Our task recovers the hidden screen's front-view RGB content from its reflected light on a visible surface. Its input is an indoor RGB observation of screen light after propagation, surface reflection, and camera imaging. The target is the current screen image, and the observed evidence is mixed with reflecting-surface appearance and camera noise. We treat the physical transport as unknown and learn the reflection-to-screen mapping from paired observations and screen targets.

Learned Restoration Under Mixed Degradation. Learned restoration under mixed degradation maps corrupted observations to spatially organized image estimates with convolutional or attention-based operators. Zamir et al. proposed a restoration model named Restormer, which uses efficient attention and multiscale processing to handle high-resolution restoration Zamir et al. (2022). Chen et al. proposed a restoration model named NAFNet, which uses activation-free nonlinear blocks to handle general image restoration Chen et al. (2022). Song et al. proposed a dehazing model named DehazeFormer, which uses a Transformer restoration design to handle single-image haze Song et al. (2023). Cai et al. proposed a low-light restoration model named Retinexformer, which uses a one-stage Retinex-based Transformer to handle illumination degradation Cai et al. (2023). These methods target degraded images that retain a more direct relation to their clean targets, whereas this paper studies an indirect wall observation with unknown transport and surface mixing. The comparison pool tests how the restoration designs transfer to hidden-screen recovery. RERL extends restoration families with an evidence-alignment, masked token-transport, and the degradation-aware

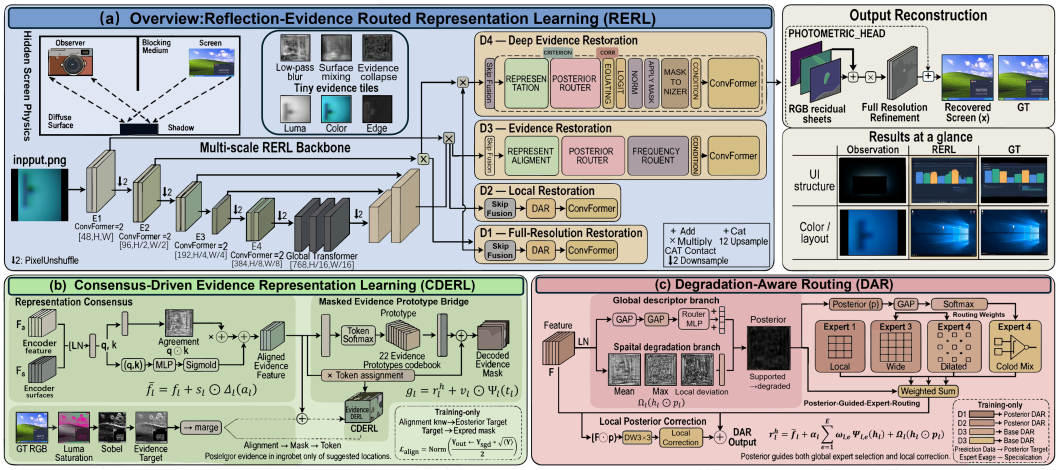

Figure 2: Overview architecture of RERL. RERL consists of CDERL and DAR.

routing path, enabling the decoder to organize weak reflected cues before reconstructing the target image.

3 METHOD

3.1 OVERALL ARCHITECTURE

Recovering a front-view screen image from a wall-reflected RGB observation differs from conventional restoration because screen emission is spatially dispersed by propagation and diffuse reflection before being mixed with surface appearance, ambient illumination, and camera response. We therefore formulate RERL as a multiscale encoder–decoder in which high-level decoding explicitly couples reflected-evidence representation with degradation-dependent transformation. Given a reflected frame $x \in \mathbb{R}^{B\times 3\times H\times W}$, the encoder extracts multiscale observations $\{e_l\}_{l=1}^{4}$ and a bottleneck state z. At the high-level decoder stages $l \in \mathcal{L}_h = \{3, 4\}$, CDERL aligns the evolving recovery representation with encoder-side reflected evidence, DAR transforms the aligned state according to its inferred degradation configuration, and a masked token bridge transports reusable evidence representations before decoding. At the lower stages $l \in \mathcal{L}_\ell = \{1, 2\}$, DAR retains a lighter globally routed form. A residual RGB readout maps the final decoder state to the recovered screen image $\hat{y}$. The overall architecture is illustrated in Figure 2.

We use the following continuous observation and recovery relation ($\xi_0 \equiv x$); ρ_η is a propagation–reflection kernel, χ_c is the channel response, and β and n collect surface bias and interference:

$$x_{b,c}(u) = \chi_c\left(\int_\Omega \sum_{c'=1}^{3} \rho_{\eta,cc'}(u,v)\, y_{b,c'}(v)\,\mathrm{d}v + \beta_c(u)\right) + n_{b,c}(u), \qquad \hat{y} = \mathcal{R}_\theta\big(z, \{e_l\}_{l=1}^{4}, x\big)\,. \tag{1}$$

Here y denotes the hidden screen image, and the nested integral expresses attenuation, spatial diffusion, surface mixing, and camera response with the optical transport treated as unknown. The recovery mapping combines multiscale features with evidence alignment, degradation-aware routing, masked prototype transport, and RGB refinement. It converts the reflected observation into the reconstructed screen image.

3.2 CONSENSUS-DRIVEN EVIDENCE REPRESENTATION LEARNING

Screen emission reaching a wall or door undergoes distance attenuation, spatial diffusion, surface-dependent mixing, and camera compression, so text boundaries, chromatic transitions, and icon contours appear only as weak and partially entangled responses. Direct skip propagation carries these screen-related traces together with wall texture and illumination structure, while the decoder representation already contains an evolving estimate whose local organization may differ from the encoder observation. CDERL therefore constructs evidence representations through agreement between encoder-side reflected measurements and decoder-side recovery states, and complements this alignment with a learned token field that represents reusable feature configurations at high-level decoding stages.

Let d_{l+1} denote the decoder state from the next coarser level, with $d_5 = z$, and ξ_j the state entering encoder level $j+1$. The encoder–bottleneck chain is written recursively, making the dependency between successive scales explicit:

$$\begin{aligned} &\xi_0 \equiv x, \qquad e_l = \mathfrak{E}_l[\xi_{l-1}], \qquad \xi_l = \mathfrak{S}_l[e_l]\ (l = 1, \dots, 4), \qquad z = \mathfrak{M}[\xi_4], \\ &f_l = \mathfrak{F}_l[\mathfrak{U}_l[d_{l+1}], e_l], \qquad l = 1, \dots, 4, \\ &f_{l,b,c}(u) = \sum_{q=1}^{C_{l+1}} \int_{\Omega_{l+1}} \kappa^{\uparrow}_{l,cq}(u, v)\, d_{l+1,b,q}(v)\, \mathrm{d}v + \sum_{s=1}^{C_l} \kappa^{\mathrm{skip}}_{l,cs}(u)\, e_{l,b,s}(u), \qquad l \in \mathcal{L}_h. \end{aligned} \tag{2}$$

where $\mathfrak{E}_l$ and $\mathfrak{S}_l$ collect the level-wise ConvFormer transformations and stride-two resampling, while $\mathfrak{M}$ denotes the bottleneck transformation. The upsampling operator $\mathfrak{U}_l$ and fusion transformation $\mathfrak{F}_l$ link f_l to d_{l+1} and e_l at every level; the last line expands the high-level skip fusion. Encoder and decoder streams are then projected into a common D_l-dimensional representation space. We write $(K \star g)_{b,j}(u) = \sum_q \int K_{jq}(u, v) g_{b,q}(v)\, \mathrm{d}v$ for an integral kernel action; its full channel and neighbourhood expansion is given in Appendix C.2. The normalized projections, agreement field, and confidence-gated transport are

$$\begin{aligned} \begin{bmatrix} q_l \\ k_l \end{bmatrix} &= \mathcal{N}_c^{\mathrm{blk}} \left(\begin{bmatrix} K_l^q \star \mathcal{N}[f_l] \\ K_l^k \star \mathcal{N}[e_l] \end{bmatrix} \right), \\ a_l &= q_l \odot k_l, \qquad s_l = \sigma(\mathcal{C}_l[q_l; k_l])\,, \\ \bar{f}_l &= f_l + s_l \odot \mathcal{V}_l[a_l] = \oplus_a(f_l, e_l), \qquad l \in \mathcal{L}_h. \end{aligned} \tag{3}$$

Here $\mathcal{N}$ normalizes channels, while $\mathcal{N}_c^{\mathrm{blk}}$ independently normalizes query and key blocks. The kernels K_l^q, K_l^k project the two streams. The confidence map $\mathcal{C}_l$ uses two pointwise convolutions separated by GELU; $\mathcal{V}_l$ combines depthwise convolution, GELU, and pointwise projection. These maps add a confidence-gated evidence residual before DAR. At lower stages, $\bar{f}_l = f_l$, and DAR directly adapts the fused features through its lightweight routing path. The masked prototype bridge transports reusable evidence configurations after DAR and the intervening stage transformations.

3.3 Degradation-Aware Routing

Aligned screen evidence still varies across regions because propagation, surface reflectance, ambient interference, and sensor response affect image regions to different degrees. Some locations retain measurable intensity or chromatic organization, whereas others preserve only weak structural variation or are dominated by nuisance appearance. A decoder based on a single transformation path applies the same operator family despite these different feature states. DAR therefore decomposes recovery into multiple transformation branches and conditions their aggregation on the inferred degradation configuration. High-level stages combine image-level expert routing with spatial self-guidance, whereas lower stages retain only the global mixture.

For an aligned state $\bar{f}_l$, let $h_l = \mathcal{N}(\bar{f}_l)$. The global router $\mathcal{G}_l[h_l]$ combines spatially pooled feature means and maxima with a summary of the spatial degradation map at high-level stages. Define $\mathcal{P}_l[h_l]$ as the spatial guidance residual built from channel statistics and neighbourhood variation, and $\mathcal{E}_{l,e}[h_l]$ as the e-th expert transformation (expanded in Appendix C.3). The compact main-text routing relation is

$$\begin{aligned} \omega_{l,b,e} &= \frac{\exp\big([\mathcal{G}_l[h_l]]_{b,e}\big)}{\sum_{r=1}^{E} \exp\big([\mathcal{G}_l[h_l]]_{b,r}\big)}, \\ r_{l,b}(u) &= \bar{f}_{l,b}(u) + \alpha_l \sum_{e=1}^{E} \omega_{l,b,e}\, \mathcal{E}_{l,e}[h_l]_b(u) + \mathbf{1}_{\{l \in \mathcal{L}_h\}}\, \mathcal{P}_l[h_l]_b(u), \qquad l \in \mathcal{L}_h \cup \mathcal{L}_\ell. \end{aligned} \tag{4}$$

Here $\mathcal{G}_l$ returns one logit per expert from the pooled descriptor and high-level degradation summary, so $\omega_{l,b,e}$ is constant over the spatial coordinates of sample b. The operators $\mathcal{E}_{l,e}$ encode the local, wide, dilated, and gated branches, while $\mathcal{P}_l$ is the spatial guidance residual. Thus, DAR adapts recovery through complementary expert transformations and spatial guidance. The expert mixture responds to the overall feature state, while the guidance residual adjusts the correction across image regions. For $l \in \mathcal{L}_\ell$, the global mixture uses pooled feature statistics alone, and the indicator omits the spatial guidance residual.

After the final decoder stage, a residual RGB projection and a full-resolution refinement map produce the network output,

$$\begin{aligned} \tilde{y} &= x + \mathcal{O}[d_1], \\ h_{\mathrm{ref}} &= \mathrm{Concat}(\mathcal{Q}[d_1], \mathcal{P}[\tilde{y}; x])\,, \qquad \hat{y} = \tilde{y} + \mathcal{K}_r[h_{\mathrm{ref}}]. \end{aligned} \tag{5}$$

where $\mathcal{O}$ is the RGB residual head. The feature branch $\mathcal{Q}$ normalizes and projects d_1, while $\mathcal{P}$ projects the six-channel reconstruction–reflection pair; each produces $C_r = \max(\lfloor C_1/2 \rfloor, 16)$ channels. The learned refinement map $\mathcal{K}_r$ processes their concatenation through a 3×3 convolution, GELU, a 3×3 depthwise convolution, GELU, and a pointwise RGB projection. Thus the refinement stage combines decoder features, the preliminary reconstruction, and the reflected input to produce the final RGB correction, with $\hat{y}$ as the returned prediction.

Training combines RGB reconstruction with complementary supervision of alignment, prototypes, and degradation-aware routing:

$$\mathcal{L}_{\text{train}} = \mathcal{L}_1^{\text{mask}}(\hat{y}, y) + \sum_{l \in \mathcal{L}_h} \boldsymbol{\lambda}^\top \boldsymbol{\ell}_l. \tag{5a}$$

Here $\mathcal{L}_1^{\text{mask}}$ is the mean absolute RGB error within the target mask, and $\boldsymbol{\ell}_l$ collects the auxiliary terms defined in Appendix C.3; these terms connect evidence alignment, masked prototype transport, and the spatial degradation map to complementary reconstruction signals.

4 Experiment

4.1 Experimental Setup

Each sample pairs a captured wall reflection with corresponding screen image on four content groups. Screen uses 2,121 training, 124 validation, and 251 test pairs; Figure and Game each use 900, 100, and 250; Lock uses 1,008, 112, and 279. These fixed splits define the dataset sizes used in paper. Each input and target is represented as a 384 by 512 RGB image. The network returns the raw residual-based RGB prediction. All estimates are evaluated on the corresponding fixed test partitions.

The main comparisons use the recorded valid target bounding box after clipping predictions to $[0, 1]$. Peak Signal-to-Noise Ratio (PSNR ↑) and Structural Similarity Index Measure (SSIM ↑) are higher-is-better fidelity measures Horé & Ziou (2010); Wang et al. (2004), while Learned Perceptual Image Patch Similarity (LPIPS ↓) is a lower-is-better learned perceptual distance Zhang et al. (2018).

The comparison includes 20 baselines in Table 1 and four additional baselines in Table 2, for a total of 24. Transformer-oriented entries include AST, AdaIR, Restormer, SFHformer, Uformer, DehazeFormer, and Retinexformer Zhou et al. (2024); Cui et al. (2025); Zamir et al. (2022); Jiang et al. (2024); Wang et al. (2022); Song et al. (2023); Cai et al. (2023).

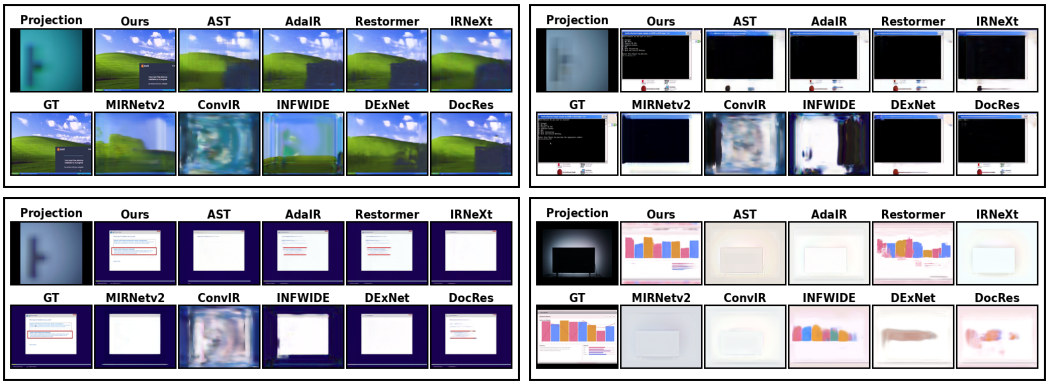

Figure 3: Qualitative comparison of RERL and the representative baseline.

Convolutional and specialized entries include ConvIR, IRNeXt, MIRNetv2, DnCNN, U-Net, pix2pix, DeepSN-Net, DEA-Net, DExNet, DarkIR, DocRes, HVI-CIDNet, and INFWIDE Cui et al. (2024; 2023); Zamir et al. (2023); Zhang et al. (2017); Ronneberger et al. (2015); Isola et al. (2017); Deng et al. (2025); Chen et al. (2024); Huang et al. (2025); Feijoo et al. (2025); Zhang et al. (2024); Yan et al. (2025); Zhang et al. (2023). Each baseline is trained independently for each content group with the same 384 by 512 reflected input, screen target, fixed split, and valid-region evaluation protocol. At inference, each baseline uses the captured reflected RGB observation as its sole input.

Training updates all parameters with AdamW using a learning rate of 5×10^{-5} and weight decay of 10^{-4} Loshchilov & Hutter (2019). Validation selects the model weights, and test data are reserved for final measurement. RERL uses four encoder levels, places CDERL at the two higher decoder levels, applies DAR at all four decoder levels, and uses 32 learned tokens per prototype bridge and four expert branches per router.

4.2 Comparison with State-of-the-Art

We evaluate whether recovery from wall reflections extends across Lock, Game, Screen, and Figure, comparing RERL with the baseline pool using the training and validation splits in Section 4.1. Table 1 and Figure 3 provide complementary quantitative and qualitative views of this comparison. Table 1 reports the 20 main baselines and RERL across the four content groups and three metrics. On Screen, RERL achieves 26.699 dB PSNR, 0.868 SSIM, and 0.189 LPIPS. Consistent trends are observed across the other content groups, with RERL obtaining the strongest reported values. Figure 3 provides a partial qualitative comparison on representative samples. It shows visible differences in recovered layout and colour organization, whereas several baselines recover mainly coarse bright regions or blurred reflection patterns, with fine structures being washed out or replaced by wall-related responses.

Method	Source	Lock			Game			Screen			Figure		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Transformer-Based Restoration													
AST	CVPR'24	15.072	0.537	0.692	17.456	0.645	0.616	19.391	0.705	0.467	17.891	0.639	0.632
AdaIR	ICLR'25	15.168	0.573	0.646	17.436	0.639	0.575	21.714	0.762	0.387	17.930	0.669	0.623
Restormer	CVPR'22	16.754	0.607	0.633	17.368	0.644	0.568	22.601	0.777	0.345	18.921	0.710	0.551
SFHformer	ECCV'24	12.461	0.509	0.740	10.010	0.462	0.760	8.817	0.394	0.807	13.205	0.597	0.713
Uformer	CVPR'22	15.149	0.565	0.674	17.547	0.647	0.614	19.204	0.707	0.464	17.586	0.619	0.626
DehazeFormer	TIP'23	14.873	0.537	0.669	17.774	0.649	0.607	19.366	0.693	0.479	18.917	0.709	0.616
Retinexformer	ICCV'23	14.600	0.559	0.651	17.925	0.656	0.592	19.567	0.673	0.455	17.895	0.690	0.619
Convolution-Based Restoration													
ConvIR	TPAMI'24	14.911	0.593	0.687	17.642	0.651	0.601	20.271	0.728	0.449	18.540	0.715	0.629
IRNeXt	ICML'23	16.134	0.608	0.678	18.136	0.668	0.574	19.956	0.723	0.454	17.479	0.668	0.633
MIRNetv2	TPAMI'23	15.211	0.604	0.658	17.679	0.656	0.612	18.237	0.692	0.492	18.340	0.713	0.654
DnCNN	TIP'17	11.818	0.479	0.722	8.738	0.427	0.726	8.883	0.347	0.724	14.204	0.574	0.664
U-Net	MICCAI'15	12.180	0.487	0.762	10.285	0.471	0.758	9.184	0.240	0.806	11.815	0.566	0.667
pix2pix	CVPR'17	12.405	0.518	0.730	11.140	0.450	0.692	10.324	0.482	0.664	15.476	0.615	0.638
DeepSN-Net	TPAMI'25	12.836	0.423	0.771	10.868	0.444	0.739	10.489	0.295	0.779	14.401	0.499	0.692
Hybrid / Degradation-Specific Restoration													
DEA-Net	TIP'24	15.962	0.607	0.674	17.729	0.658	0.588	20.651	0.750	0.424	18.179	0.715	0.642
DExNet	TPAMI'25	15.518	0.608	0.652	17.629	0.659	0.595	20.005	0.732	0.438	18.602	0.721	0.621
DarkIR	CVPR'25	8.787	0.095	0.804	4.465	0.041	0.936	9.221	0.175	0.867	4.185	0.038	0.929
DocRes	CVPR'24	16.357	0.609	0.647	17.528	0.647	0.579	21.145	0.749	0.400	18.772	0.718	0.611
HVI-CIDNet	CVPR'25	15.606	0.576	0.671	17.774	0.656	0.598	19.692	0.718	0.472	19.415	0.721	0.596
INFWIDE	TIP'23	15.053	0.595	0.695	16.807	0.641	0.633	21.564	0.755	0.380	19.054	0.720	0.617
Reflection-Evidence Routed Representation Learning													
RERL	Ours	17.084	0.620	0.630	18.695	0.681	0.542	26.699	0.868	0.189	19.523	0.723	0.546

Table 1: Complete comparison on four content groups. Red text and shading highlight the best row; blue text and shading highlight the selected rows within their respective method groups.

Method	Source	Lock			Screen			Game			Figure		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
MPRNet	CVPR'21	14.494	0.555	0.734	15.051	0.561	0.637	15.977	0.594	0.703	18.480	0.717	0.671
HINet	CVPRW'21	14.795	0.579	0.653	14.814	0.585	0.606	16.688	0.616	0.653	18.550	0.708	0.628
SwinIR	ICCVW'21	11.567	0.508	0.713	10.065	0.461	0.692	10.220	0.459	0.737	14.793	0.595	0.653
MIMO-UNet	ICCV'21	14.499	0.560	0.700	15.057	0.573	0.617	15.702	0.580	0.652	18.071	0.693	0.620
RERL	Ours	17.084	0.620	0.630	26.699	0.868	0.189	18.695	0.681	0.542	19.523	0.723	0.546

Table 2: Comparison with four additional learned-restoration baselines on Lock, Screen, Game, and Figure. Red text and shading highlight the best row; blue text and shading identify HINet.

Table 2 and Figure 4 extend the comparison to MPRNet Zamir et al. (2021), HINet Chen et al. (2021), SwinIR Liang et al. (2021), and MIMO-UNet Cho et al. (2021). This comparison tests whether the recovery advantage extends across four additional learned-restoration designs. The results show that additional learned-restoration baselines tend to transfer salient blurred edges, bright contours, and input intensity patterns, leaving the recovered screen layout dominated by the reflection's intensity structure. RERL produces outputs closer to the ground truth, with more coherent graphic boxes,

interface boundaries, and chart layouts.

These results highlight the value of organizing screen-related evidence before reconstruction. Generic restoration models assume that the target remains spatially registered with the degraded input, whereas this task requires identifying screen-specific evidence within a spatially mixed reflection. CDERL aligns decoder states with encoder-side evidence and adds confidence-weighted evidence residuals, while DAR adapts the recovery path to the inferred degradation state. This evidence-first formulation is therefore better suited to hidden-screen recovery than direct restoration of the observed reflection.

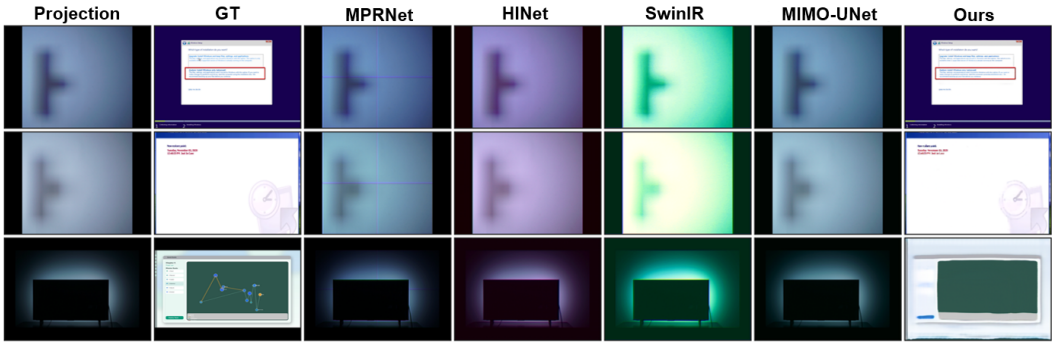

Figure 4: Screen reconstructions from MPRNet, HINet, SwinIR, MIMO-UNet, and RERL, alongside the reflected observation and target screen.

4.3 Ablation Study

Ablation experiments are conducted on RERL under the same strategy as Section 4.1 to examine how RERL addresses the difficulties in reflected-screen observations. Table 3 examines the joint contribution of CDERL and DAR by removing either module or replacing it with generic bridging and routing schemes. The results show that evidence representation learning and degradation-aware routing are both essential to the recovery pipeline. The full model achieves the best results across generic bridge and standard routing replacements.

Method	Model	Source	Lock			Figure			Game			Screen		
			PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Consensus-Driven Evidence Representation Learning (CDERL)														
w/o CDERL	-	-	15.827	0.573	0.656	18.870	0.690	0.624	17.346	0.602	0.619	18.142	0.629	0.528
CDERL → Concat Bridge	U-Net	MICCAI'15	17.034	0.610	0.632	19.459	0.721	0.556	18.678	0.681	0.545	25.264	0.851	0.238
CDERL → Metric-Latent Bridge	ProtoNet	NeurIPS'17	17.051	0.613	0.635	19.479	0.721	0.586	18.582	0.680	0.549	25.270	0.849	0.233
CDERL → Cross-Attn Bridge	Transformer	NeurIPS'17	16.599	0.612	0.637	19.258	0.723	0.611	17.518	0.646	0.620	18.338	0.694	0.499
Degradation-Aware Routing (DAR)														
w/o DAR	-	-	16.021	0.564	0.674	18.788	0.681	0.623	17.331	0.593	0.626	18.106	0.637	0.535
DAR → Global-only MoE	Two-expert mixture	-	17.023	0.616	0.633	19.410	0.720	0.546	18.677	0.681	0.546	25.136	0.847	0.239
DAR → Complex Router	UniRes	ICCV'25	15.878	0.576	0.661	18.733	0.685	0.616	17.242	0.593	0.622	17.745	0.624	0.543
DAR → Spatial-Attn Router	CBAM	ECCV'18	16.624	0.611	0.630	19.118	0.722	0.616	17.348	0.637	0.630	18.204	0.674	0.497
Reflection-Evidence Routed Representation Learning (RERL)														
Ours (Full)	RERL	-	17.084	0.620	0.630	19.523	0.723	0.546	18.695	0.681	0.542	26.699	0.868	0.189

Table 3: Main ablation of CDERL and DAR variants on four content groups. The global-only MoE replacement uses an image-level mixture of two feature-conditioned experts. Model and Source identify reference designs of the replacement modules, including U-Net Ronneberger et al. (2015), ProtoNet Snell et al. (2017), Transformer Vaswani et al. (2017), UniRes Zhou et al. (2025), and CBAM Woo et al. (2018). Red highlights the best row; blue highlights the selected replacement rows.

Tables 4 and 5 further isolate the internal designs of CDERL and DAR. For CDERL, cross-attention, 1×1 fusion, SE gating, token projection, latent bottlenecks, and VQ codebooks improve PSNR over their corresponding removal variants on Screen, Figure, and Lock; the full design achieves the best results across all four groups. This pattern indicates that its advantage arises from coupling encoder-side reflected evidence with the evolving decoder state before reintroducing reusable evidence prototypes. Alignment adds confidence-gated responses derived from encoder–decoder agreement, while the prototype bridge provides structured residual cues that can be reused during decoding.

In Table 5, the learned degradation estimator yields stronger reconstruction than spatial attention, gradient–variance rules, and global scores, supporting its role in characterizing the reflected feature state for routing. The full DAR gives the best or tied-best results over assignment removal, top-1 gating, fixed dual-branch averaging, and static expert weights, with the largest PSNR decrease under static weights. The results indicate that the estimator and adaptive assignment operate as a coupled

mechanism: the former characterizes the degradation condition of the reflected evidence, and the latter converts this into an appropriate expert mixture. Their consistent advantage over removal and replacement variants establishes DAR as an essential component of RERL.

Method	Model	Source	Lock			Figure			Game			Screen		
			PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Reflection-Evidence and Restoration-Representation Alignment														
w/o Rep. Align.	-	-	15.867	0.575	0.658	18.918	0.692	0.623	17.217	0.599	0.618	18.171	0.654	0.519
Align. → Cross-Attn	Transformer	NeurIPS'17	16.549	0.619	0.631	19.184	0.723	0.617	17.387	0.642	0.623	18.495	0.693	0.502
Align. → 1×1 Fusion	U-Net	MICCAI'15	16.618	0.615	0.634	19.208	0.722	0.613	17.280	0.640	0.626	18.555	0.692	0.500
Align. → SE Gating	SENet	CVPR'18	16.496	0.610	0.630	19.195	0.722	0.612	17.276	0.640	0.624	18.658	0.699	0.502
Evidence Prototype Bridge														
w/o Proto. Bridge	-	-	15.983	0.572	0.656	18.833	0.689	0.619	17.334	0.595	0.619	17.992	0.618	0.532
Bridge → Token Proj.	ViT	ICLR'21	16.555	0.615	0.631	19.219	0.721	0.610	17.354	0.642	0.624	18.609	0.701	0.495
Bridge → Latent Bottl.	Perceiver	ICML'21	16.630	0.616	0.635	19.245	0.721	0.610	17.324	0.642	0.624	18.601	0.698	0.500
Bridge → VQ Codebook	VQ-VAE	NeurIPS'17	16.605	0.615	0.635	19.226	0.722	0.612	17.314	0.641	0.624	18.614	0.697	0.497
Reflection-Evidence Routed Representation Learning (RERL)														
Ours (Full)	RERL	-	17.084	0.620	0.630	19.523	0.723	0.546	18.695	0.681	0.542	26.699	0.868	0.189

Table 4: Internal ablation of representation alignment and token bridging on four content groups. Red highlights the best; blue highlights the selected replacement rows. Replacement modules are instantiated within the recovery backbone; Model and Source identify their reference designs: Transformer Vaswani et al. (2017), U-Net Ronneberger et al. (2015), SENet Hu et al. (2018), ViT Dosovitskiy et al. (2021), Perceiver Jaegle et al. (2021), and VQ-VAE van den Oord et al. (2017).

Method	Model	Source	Lock			Figure			Game			Screen		
			PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Spatial Degradation Estimator														
w/o Degradation Est.	-	-	16.625	0.614	0.634	19.109	0.720	0.616	17.388	0.643	0.622	18.424	0.695	0.502
Est. → Spatial Attn.	CBAM	ECCV'18	16.540	0.612	0.635	19.106	0.720	0.616	17.353	0.641	0.623	18.377	0.693	0.502
Est. → Grad-Var Rule	Grad-Var Rule	-	16.585	0.612	0.635	19.084	0.721	0.618	17.399	0.642	0.622	18.453	0.696	0.495
Est. → Global Score	SENet	CVPR'18	16.587	0.613	0.635	19.072	0.720	0.617	17.235	0.640	0.624	18.366	0.695	0.503
Adaptive Expert Assignment														
w/o Local Assign.	-	-	16.592	0.614	0.635	19.103	0.723	0.618	17.425	0.643	0.623	18.291	0.691	0.498
Assign. → Static expert weights	Shared expert mixture	-	14.860	0.518	0.692	18.272	0.660	0.622	15.760	0.559	0.666	15.467	0.554	0.617
Assign. → Top-1 Gate	Switch Transformer	JMLR'22	16.551	0.613	0.634	19.080	0.721	0.617	17.254	0.640	0.625	18.400	0.696	0.497
Assign. → Fixed Dual Avg	Dual-Branch Avg	-	16.595	0.615	0.633	19.113	0.722	0.617	17.319	0.641	0.626	18.401	0.695	0.497
Reflection-Evidence Routed Representation Learning (RERL)														
Ours (Full)	RERL	-	17.084	0.620	0.630	19.523	0.723	0.546	18.695	0.681	0.542	26.699	0.868	0.189

Table 5: Internal ablation of spatial degradation estimation and routing assignment on four content groups. Model and Source identify reference designs of the replacement modules: CBAM Woo et al. (2018), SENet Hu et al. (2018), and Switch Transformer Fedus et al. (2022). Red highlights the best; blue highlights the selected replacement rows.

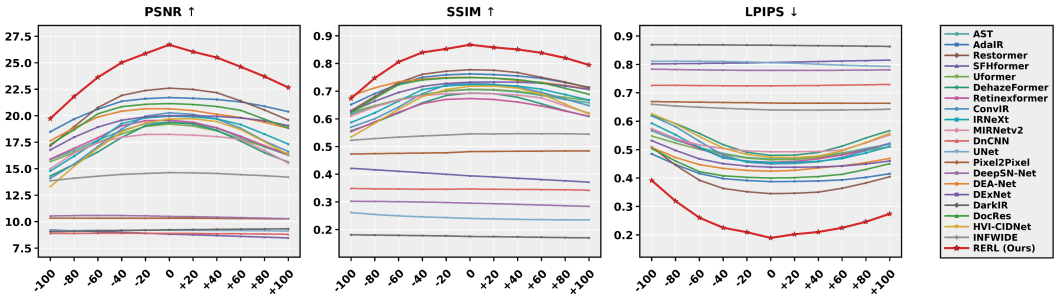

Figure 5: Screen PSNR, SSIM, and LPIPS across global brightness indices.

4.4 Robustness to Brightness Variation

We evaluate global brightness variation on Screen with brightness indices from -100 to $+100$, corresponding to RGB multipliers of 0.90–1.10 (Appendix F). Figure 5 reports PSNR, SSIM, and LPIPS across this sweep, where negative and positive indices correspond to global darkening and brightening of the reflection. RERL maintains the highest PSNR and lowest LPIPS throughout the brightness sweep. Its advantage persists under both darkening and brightening, showing that screen

recovery remains effective as the reflection's overall intensity changes. Brightness shifts alter the contrast between interface regions and change the reflected traces of window boundaries and colour partitions. Several baselines exhibit larger metric changes at the ends of the sweep.

This behavior is consistent with the two-stage design of RERL. CDERL prioritizes local structures jointly supported by encoder-side reflection evidence and the evolving decoder representation, reducing reliance on a fixed brightness pattern. DAR then adjusts the image-level expert mixture according to degraded evidence state and uses spatial guidance to correct regions according to their local evidence strength.

4.5 Qualitative Results

Figure 6 tests which regions of the reflection support the recovered screen content. The first column shows the complete reflection, ground-truth screen and full-model prediction. The remaining panels occlude one reflection region at a time and display the corresponding reconstructions. Masking different grid cells produces distinct changes in page structure, texture, and colour organization, while the PSNR drop identifies regions whose removal has the largest impact on reconstruction. The spatial pattern of PSNR drops connects each masked region to its contribution to the recovered screen layout and colour organization. The recovery therefore makes spatially selective use of the reflected observation. This dependence on where evidence is available motivates combining evidence selection with region-adaptive correction.

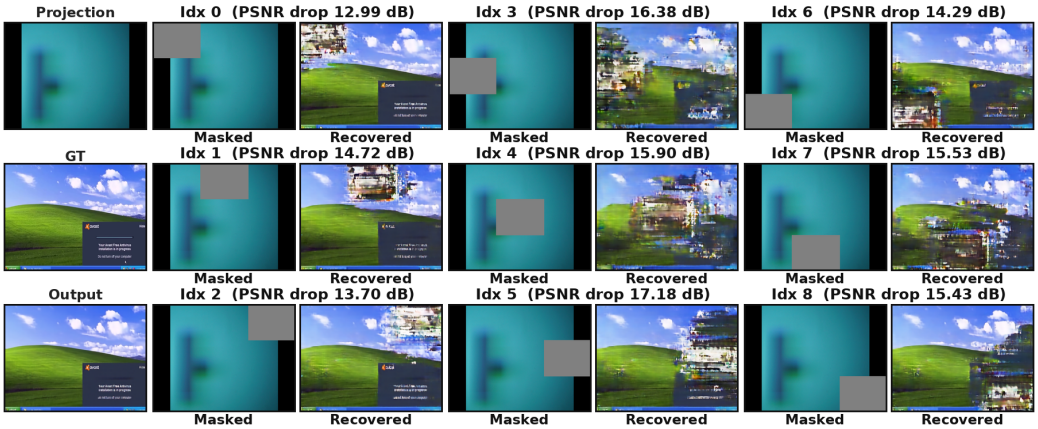

Figure 6: Occlusion analysis on a 3×3 reflection grid. The left column shows the complete observation, target screen, and unmasked reconstruction. Each remaining pair shows a masked observation and its reconstruction, together with the PSNR decrease relative to the unmasked result.

CDERL identifies reflection responses that remain consistent with the evolving recovery state and organizes them into evidence-supported representations. DAR combines image-level expert weighting with spatial guidance to adjust the correction for weak or heavily mixed regions. Thus, RERL couples evidence organization with spatially guided reconstruction to exploit the reflection's spatial structure.

5 Conclusion

We presented RERL to recover hidden screen content from wall reflections captured by an ordinary camera. Propagation and diffuse reflection weaken and mix the screen signal, leaving screen-related cues entangled with surface appearance and imaging noise. CDERL organizes these weak cues through feature agreement and masked prototype transport, while DAR adapts reconstruction through complementary expert transformations and spatial guidance.

Across four content groups, RERL achieves the strongest results in the main comparisons and recovers coherent layouts, colour regions, and major contours. Controlled comparisons and component ablations identify the contribution of these designs, while perturbation and occlusion analyses connect recovery to changes in the reflected observation.

These results establish the feasibility of recovering hidden-screen layouts, colour regions, and major contours from ordinary-camera wall reflections through RERL's evidence organization and adaptive reconstruction. Future work will extend the approach to broader sensing environments, additional reflecting surfaces, and physically grounded models of indirect light transport.

REFERENCES

Michael Backes, Markus Dürmuth, and Dominique Unruh. Compromising reflections, or how to read LCD monitors around the corner. In *Proceedings of the IEEE Symposium on Security and Privacy*, pp. 158–169, 2008. doi: 10.1109/SP.2008.25.

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

Liangyu Chen, Xin Lu, Jie Zhang, Xiaojie Chu, and Chengpeng Chen. HINet: Half instance normalization network for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 182–192, 2021. URL https://openaccess.thecvf.com/content/CVPR2021W/NTIRE/html/Chen_HINet_Half_Instance_Normalization_Network_for_Image_Restoration_CVPRW_2021_paper.html.

Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *Computer Vision – ECCV 2022*, volume 13667 of *Lecture Notes in Computer Science*, pp. 17–33. Springer, 2022. doi: 10.1007/978-3-031-20071-7_2.

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

Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4641–4650, 2021. URL https://openaccess.thecvf.com/content/ICCV2021/html/Cho_Rethinking_Coarse-To-Fine_Approach_in_Single_Image_Deblurring_ICCV_2021_paper.html.

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

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

Yuning Cui, Syed Waqas Zamir, Salman Khan, Alois Knoll, Mubarak Shah, and Fahad Shahbaz Khan. AdaIR: Adaptive all-in-one image restoration via frequency mining and modulation. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=M5t0WvjfCg.

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.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=YicbFdNTTy.

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

William Fedus, Barret Zoph, and Noam Shazeer. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *Journal of Machine Learning Research*, 23(120):1–39, 2022. URL https://www.jmlr.org/papers/v23/21-0998.html.

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.

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

Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, 2018. URL https://openaccess.thecvf.com/content_cvpr_2018/html/Hu_Squeeze-and-Excitation_Networks_CVPR_2018_paper.html.

Jun-Jie Huang, Tianrui Liu, Zihan Chen, Xinwang Liu, Meng Wang, and Pier Luigi Dragotti. A lightweight deep exclusion unfolding network for single image reflection removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(6):4957–4973, 2025. doi: 10.1109/TPAMI.2025.3548148.

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.

Robert A. Jacobs, Michael I. Jordan, Steven J. Nowlan, and Geoffrey E. Hinton. Adaptive mixtures of local experts. *Neural Computation*, 3(1):79–87, 1991. doi: 10.1162/neco.1991.3.1.79. URL https://direct.mit.edu/neco/article/3/1/79/5560/Adaptive-Mixtures-of-Local-Experts.

Andrew Jaegle, Felix Gimeno, Andy Brock, Oriol Vinyals, Andrew Zisserman, and Joao Carreira. Perceiver: General perception with iterative attention. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 4651–4664. PMLR, 2021. URL https://proceedings.mlr.press/v139/jaegle21a.html.

Liming Jiang, Bo Dai, Wayne Wu, and Chen Change Loy. Focal frequency loss for image reconstruction and synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 13919–13929, 2021. URL https://openaccess.thecvf.com/content/ICCV2021/html/Jiang_Focal_Frequency_Loss_for_Image_Reconstruction_and_Synthesis_ICCV_2021_paper.html.

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.

Justin Johnson, Alexandre Alahi, and Fei-Fei Li. Perceptual losses for real-time style transfer and super-resolution. In *Proceedings of the European Conference on Computer Vision*, 2016. URL https://cs.stanford.edu/people/jcjohns/eccv16/.

Dmitry Lepikhin, HyoukJoong Lee, Yuanzhong Xu, Dehao Chen, Orhan Firat, Yanping Huang, Maxim Krikun, Noam Shazeer, and Zhifeng Chen. GShard: Scaling giant models with conditional computation and automatic sharding. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=qrwe7XHTmYb.

Mike Lewis, Shruti Bhosale, Tim Dettmers, Naman Goyal, and Luke Zettlemoyer. BASE layers: Simplifying training of large, sparse models. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 6265–6274. PMLR, 2021. URL https://proceedings.mlr.press/v139/lewis21a.html.

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pp. 1833–1844, 2021. URL https://openaccess.thecvf.com/content/ICCV2021W/AIM/html/Liang_SwinIR_Image_Restoration_Using_Swin_Transformer_ICCVW_2021_paper.html.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10012–10022, 2021. URL https://openaccess.thecvf.com/content/ICCV2021/html/Liu_Swin_Transformer_Hierarchical_Vision_Transformer_Using_Shifted_Windows_ICCV_2021_paper.html.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, volume 9351 of *Lecture Notes in Computer Science*, pp. 234–241. Springer, 2015. doi: 10.1007/978-3-319-24574-4_28.

Charles Saunders, John Murray-Bruce, and Vivek K. Goyal. Computational periscopy with an ordinary digital camera. *Nature*, 565(7740):472–475, 2019. doi: 10.1038/s41586-018-0868-6.

Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=B1ckMDqlg.

Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/cb8da6767461f2812ae4290eac7cbc42-Abstract.html.

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.

Aaron van den Oord, Oriol Vinyals, and Koray Kavukcuoglu. Neural discrete representation learning. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/7a98af17e63a0ac09ce2e96d03992fbc-Abstract.html.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018. URL https://arxiv.org/abs/1711.07971.

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.

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

Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. CBAM: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision*, pp. 3–19, 2018. URL https://openaccess.thecvf.com/content_ECCV_2018/html/Sanghyun_Woo_Convolutional_Block_Attention_ECCV_2018_paper.html.

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

Adam B. Yedidia, Manel Baradad, Christos Thrampoulidis, William T. Freeman, and Gregory W. Wornell. Using unknown occluders to recover hidden scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12231–12239, 2019. doi: 10.1109/CVPR.2019.01251.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14821–14831, 2021. URL https://openaccess.thecvf.com/content/CVPR2021/html/Zamir_Multi-Stage_Progressive_Image_Restoration_CVPR_2021_paper.html.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5728–5739, 2022. doi: 10.1109/CVPR52688.2022.00564.

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

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

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

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

Zhihong Zhang, Yuxiao Cheng, Jinli Suo, Liheng Bian, and Qionghai Dai. INFWIDE: Image and feature space wiener deconvolution network for non-blind image deblurring in low-light conditions. *IEEE Transactions on Image Processing*, 32:1390–1402, 2023. doi: 10.1109/TIP.2023.3244417.

Mo Zhou, Keren Ye, Mauricio Delbracio, Peyman Milanfar, Vishal M. Patel, and Hossein Talebi. UniRes: Universal image restoration for complex degradations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 13237–13247, 2025. URL https://openaccess.thecvf.com/content/ICCV2025/html/Zhou_UniRes_Universal_Image_Restoration_for_Complex_Degradations_ICCV_2025_paper.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.

Yanqi Zhou, Tao Lei, Hanxiao Liu, Nan Du, Yanping Huang, Vincent Zhao, Andrew M. Dai, Zhifeng Chen, Quoc V. Le, and James Laudon. Mixture-of-experts with expert choice routing. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/2f00ecd787b432c1d36f3de9800728eb-Abstract-Conference.html.

SUPPLEMENTARY MATERIAL FOR RERL: RECOVERING HIDDEN SCREEN CONTENT FROM WALL REFLECTIONS

This supplement provides additional details and ablation experiments to support the main RERL paper. It elaborates the reflection-evidence routed representation pipeline and evaluates the recovery performance under varied degradation conditions.

A APPENDIX OVERVIEW

This appendix follows the central evidence chain of Reflection-Evidence Routed Representation Learning (RERL). A camera observes a reflection spot on a wall, door, or another surface while the frontal screen is outside the direct view. The spot retains coarse luminance distribution, colour organization, spatial layout, and major contours because screen emission remains spatially structured after propagation and diffuse reflection. The same processes also spread, mix, and weaken those traces, which makes reflected evidence representation and degradation-aware routing the central recovery questions.

We first compare representation, training, and routing designs within a common recovery backbone. Internal ablations then locate the contribution of alignment, prototype transport, degradation estimation, and expert assignment. Condition sweeps examine recovery under changes in brightness, noise, geometry, blur, surface response, and ambient light. Visual diagnostics follow how the reflected observation develops into the reconstructed screen through intermediate features and final outputs.

B SCENARIO AND EVIDENCE CONDITIONS

B.1 OBSERVATION SETTING

The camera receives an RGB reflection spot produced by screen light on a wall, door, glass, or another surface. The frontal screen is hidden from the camera, and the input records a dispersed surface response. Screen emission leaves structured traces in that response: broad intensity regions, colour partitions, aligned boundaries, chart bars, icon contours, and layout relationships. These traces are the reflected evidence organized by RERL.

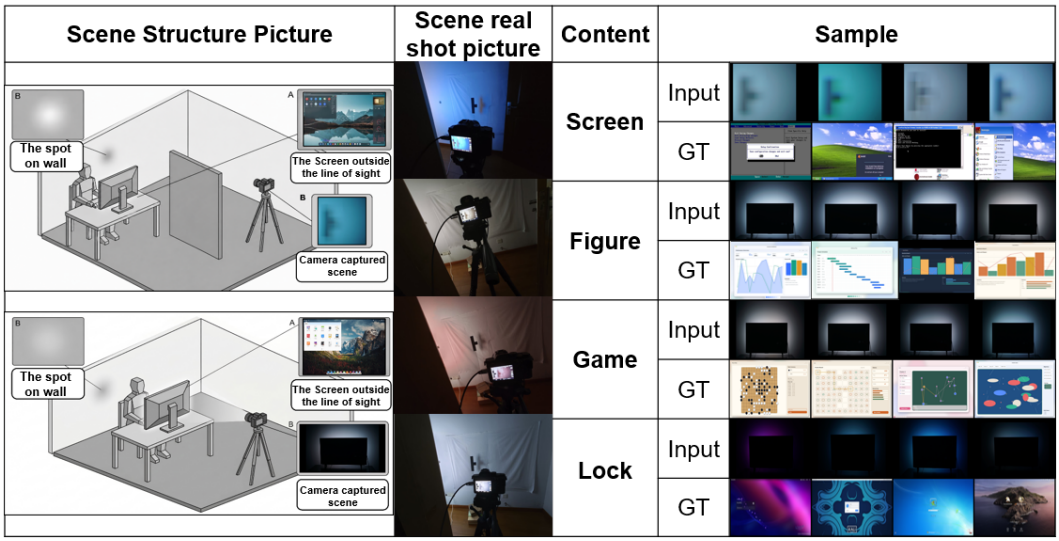

Figure 7: Dataset acquisition configurations and representative samples. The first column illustrates two non-line-of-sight acquisition geometries: an intervening partition or wall between the screen and the observation region, and a back-facing configuration in which the camera is positioned behind the computer. The second column shows corresponding real-shot scenes. The remaining columns present Input and GT examples from the Screen, Figure, Game, and Lock content groups.

Figure 7 summarizes the acquisition configurations, real-world recording environments, and representative samples in the dataset. The first column presents two schematic acquisition geometries. In the first configuration, an intervening physical partition or wall separates the screen from the observation region, placing the screen outside the camera's direct line of sight while its emitted light produces a visible spot on the observation surface. In the second configuration, the camera is positioned behind the computer and faces away from the screen, placing the screen outside the direct view and making the captured reflected or diffused light the input for content recovery.

The second column shows corresponding real-shot scenes, including the screen, camera, observation surface, and surrounding environment. The remaining columns provide representative examples from the four content groups: Screen, Figure, Game, and Lock. For each group, the Input row contains the captured wall or surface observations with attenuated, diffused, and colour-mixed screen evidence, whereas the GT row contains the corresponding frontal screen images used as reconstruction targets. The examples illustrate that the dataset covers interface layouts, graphical plots, game-like visual structures, and lock-screen content from surface reflections captured outside the screen's direct view. The four content groups in the main paper are Screen, Figure, Game, and Lock. Screen contains interface-like layouts; Lock contains lock-screen content; Figure emphasizes lines, bars, and colour partitions; Game contains stronger texture and abrupt colour transitions. Across these groups, PSNR, SSIM, and LPIPS measure reconstruction quality over the valid target region.

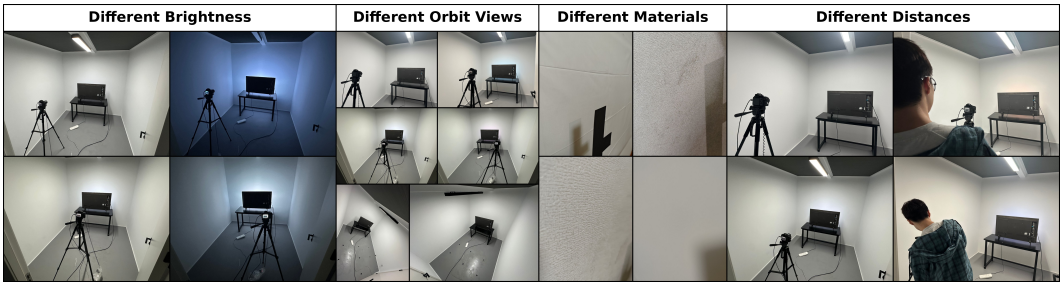

Figure 8: Representative real-world acquisition scenes under different illumination conditions, observation viewpoints, reflecting-surface materials, and observation distances.

Figure 8 summarizes representative real-world acquisition scenes under different illumination conditions, observation viewpoints, reflecting-surface materials, and observation distances. The first column presents two pairs of brighter and darker scenes, with the brighter captures on the left and the darker captures on the right, illustrating the visible lighting differences in the recording environment. The second column shows the experimental setup from different orbit-view positions, including lateral and elevated views that reveal the spatial arrangement of the screen, camera, and surrounding surfaces. The third column provides close-up views of reflecting surfaces with different textures and finishes, documenting the physical appearance of representative material conditions. The fourth column presents distance-related recording scenes, showing the placement of the screen and acquisition equipment within the environment. Together, these photographs complement the acquisition geometries and Input/GT examples in Figure 7 by documenting physical recording conditions complementary to the four screen-content groups. These photographs document the acquisition environment and complement the condition definitions in Appendix F, providing context for the reflection-spot degradation examined in the following subsection.

B.2 How the Spot Degrades

Propagation attenuates screen energy, diffuse reflection spreads local responses, surface appearance mixes with the screen contribution, and camera imaging compresses the resulting signal. Brightness changes shift global energy and contrast; noise introduces responses that can resemble edges; distance changes weaken and spread the spot; viewpoint changes alter its position and spatial correspondence; motion blur smooths contours; material changes alter diffusion and mixing; and ambient light adds intensity gradients and colour bias. Each condition therefore changes the availability of luminance, colour, contour, and layout evidence in a different way.

The reflection spot provides evidence for recovering the screen's coarse structure. Broad luminance distribution and large colour regions survive as thin boundaries weaken, while major contours remain

recoverable when several local responses support the same layout. This ordering motivates CDERL: evidence is organized before the decoder decides how strongly to transform each region.

C Mechanistic Evidence Model

C.1 Approximate Observation Relation

The main paper relates the reflected observation to the hidden screen through unknown optical transport. The following observation relation describes how propagation, surface reflection, and camera response redistribute the screen signal:

$$\begin{aligned} x_{b,c}(u) &= \chi_c\Bigg(\int_\Omega \sum_{r=1}^{3} \rho_{\eta,cr}(u,v) \underbrace{y_{b,r}(v)}_{\text{screen emission}} \,\mathrm{d}v \\ &\qquad + \beta_c(u) \Bigg) + n_{b,c}(u), \\ \rho_{\eta,cr}(u,v) &= e^{-\mu_\eta(u,v)} \int_\Sigma s_{\eta,c}(u,w)\, d_\eta(w,v)\, m_{\eta,r}(w)\,\mathrm{d}w, \\ \beta_c(u) &= \int_\Omega a_{\eta,c}(u,v)\,\mathrm{d}v. \end{aligned} \tag{6}$$

Here Ω is the hidden-screen domain and Σ the reflecting surface. The attenuation coefficient μ_η, diffuse kernel d_η, surface mixing factor m_η, ambient field a_η, and camera response χ_c describe attenuation, diffuse transport, surface mixing, ambient illumination, and camera response. The effective kernel ρ_η includes the attenuation factor along with the spatial transport. The interference term accommodates spatially and chromatically correlated surface and imaging effects.

In Equation 6, the screen contributes an attenuated and spatially mixed response to the observed surface. RERL learns the reverse mapping from paired reflections and screen images, using feature agreement and adaptive reconstruction to recover useful content from this indirect observation. Figures 3 and 4 show the resulting recovery of screen layout and colour organization.

C.2 Consensus-Driven Evidence Representation Learning

Let $v_x(u)$, $v_d(u)$, and $v_t(u)$ denote encoder-side reflected evidence, the evolving decoder state, and a reusable evidence prototype at location u. For a candidate screen structure r, define the normalized local agreement and its Gibbs posterior by

$$\begin{aligned} \langle a, b\rangle_{\Omega_u} &= \frac{\displaystyle\sum_{j=1}^{D} \int_{\Omega_u} a_j(v) b_j(v)\,\mathrm{d}v}{\sqrt{\displaystyle\sum_{j=1}^{D} \int_{\Omega_u} a_j^2(v)\,\mathrm{d}v + \varepsilon}\sqrt{\displaystyle\sum_{j=1}^{D} \int_{\Omega_u} b_j^2(v)\,\mathrm{d}v + \varepsilon}}, \\ \kappa(u,r) &= \frac{1}{3} \sum_{\nu\in\{x,d,t\}} \langle r, v_\nu\rangle_{\Omega_u}, \qquad \pi(r \mid u) = \frac{\exp(\kappa(u,r)/\tau)}{\sum_{r'\in\mathcal{R}} \exp(\kappa(u,r')/\tau)}, \\ \bar{v}(u) &= \sum_{r\in\mathcal{R}} \pi(r \mid u) r(u), \qquad \mathcal{E}_{\mathrm{con}} = \int_\Omega \sum_{r\in\mathcal{R}} \pi(r \mid u)\big(1 - \kappa(u,r)\big)^2\,\mathrm{d}u. \end{aligned} \tag{7}$$

The inner product is an explicitly normalized integral similarity over a neighbourhood Ω_u, τ controls posterior sharpness, and $\mathcal{R}$ is the set of reusable token structures. Candidates with stronger aggregate agreement receive larger weights, and their weighted combination forms the consensus representation $\bar{v}$. The agreement energy $\mathcal{E}_{\mathrm{con}}$ summarizes discrepancies between candidate structures and the three feature streams.

Prototype transport complements evidence alignment at high-level stages. Let t_l be the state entering the bridge after DAR and the intervening stage transformations, and $\mathbf{C}_l \in \mathbb{R}^{T\times D_l^{\mathrm{tok}}}$ the learned

prototype matrix. A pointwise projection $\mathcal{Z}_l$ predicts assignment logits; learned maps $\mathcal{M}_l$ and $\mathcal{D}_l^{\mathrm{tok}}$ produce mask logits and decode the prototype field, respectively. The transported state is

$$\begin{aligned} h_l^{\mathrm{tok}} &= \mathcal{N}(t_l), \qquad \alpha_{l,t}(u) = \mathrm{softmax}_t\big(\mathcal{Z}_l[h_l^{\mathrm{tok}}](u)\big)\,, \\ v_l(u) &= \sum_{t=1}^{T} \alpha_{l,t}(u)\mathbf{C}_{l,t}, \qquad m_l(u) = \sigma\big(\mathcal{M}_l[h_l^{\mathrm{tok}}](u)\big)\,, \\ \widetilde{F}_l(u) &= t_l(u) + m_l(u)\mathcal{D}_l^{\mathrm{tok}}[v_l](u), \qquad l \in \mathcal{L}_h. \end{aligned} \tag{7a}$$

The bridge transports a weighted combination of reusable prototypes into the evolving recovery state. Learned assignments organize prototype contributions, while the spatial mask controls their residual influence before the next decoder transformation reconstructs screen content.

Writing $F_l = f_l$ and $E_l = e_l$, the channel-level agreement operator in the main text is

$$\begin{aligned} q_{l,b,j}(u) &= \mathcal{N}_c\left(\sum_{q=1}^{C_l}\int_{\Omega_l} K^q_{l,jq}(u,v)\,\mathcal{N}[F_l]_{b,q}(v)\,\mathrm{d}v\right), \\ k_{l,b,j}(u) &= \mathcal{N}_c\left(\sum_{s=1}^{C_l}\int_{\Omega_l} K^k_{l,js}(u,v)\,\mathcal{N}[E_l]_{b,s}(v)\,\mathrm{d}v\right), \\ a_{l,b,j}(u) &= q_{l,b,j}(u)k_{l,b,j}(u), \\ s_{l,b}(u) &= \sigma(\mathcal{C}_l[q_l;k_l]_b(u))\,, \\ \bar{F}_{l,b,c}(u) &= F_{l,b,c}(u) + s_{l,b}(u)\mathcal{V}_l[a_l]_{b,c}(u). \end{aligned} \tag{8}$$

The normalized decoder and encoder projections form pointwise agreement. The confidence map $\mathcal{C}_l$ processes their concatenated $2D_l$ channels, including the intermediate GELU nonlinearity. The last line transports the confidence-weighted evidence residual through $\mathcal{V}_l$, completing $\oplus_a$.

C.3 Degradation-Aware Routing

After evidence alignment, reflected features can still vary in strength across the image. Let h be the normalized aligned feature. The spatial degradation map uses its channel moments and a neighbourhood discrepancy:

$$\begin{aligned} \mu_h(u) &= \frac{1}{C}\sum_{c=1}^{C} h_c(u), \qquad \nu_h(u) = \max_{1\le c\le C} h_c(u), \\ \bar{\mu}_h(u) &= \int_{\mathcal{N}(u)} w(u,v)\mu_h(v)\,\mathrm{d}v, \qquad \zeta_h(u) = |\mu_h(u) - \bar{\mu}_h(u)|, \\ \delta(u) &= \sigma(\mathcal{D}_{\mathrm{sp}}[\mu_h;\nu_h;\zeta_h](u))\,. \end{aligned} \tag{9}$$

The map $\delta(u)$ summarizes channel statistics and neighbourhood variation for spatial guidance. Here $\mathcal{D}_{\mathrm{sp}}$ combines a 3×3 convolution, GELU, and pointwise projection. DAR forms global logits from pooled features and the map summary, then normalizes the expert weights:

$$\begin{aligned} g_h &= \mathrm{Concat}(\mathrm{AvgPool}_\Omega\, h, \mathrm{MaxPool}_\Omega\, h)\,, \qquad \bar{\delta} = |\Omega|^{-1}\int_\Omega \delta(v)\,\mathrm{d}v, \\ \ell_e &= \mathcal{A}_e[g_h] + \lambda\mathcal{C}_e[\bar{\delta}], \\ \omega_e &= \frac{\exp(\ell_e)}{\sum_{r=1}^{E}\exp(\ell_r)}, \qquad \sum_{e=1}^{E}\omega_e = 1, \end{aligned} \tag{10}$$

and applies

$$\begin{aligned} R_c(u) = \bar{f}_c(u) + \alpha\sum_{e=1}^{E}\omega_e\,\mathcal{E}_e[h]_c(u) \\ + \mathcal{H}_\delta[\delta\odot h]_c(u). \end{aligned} \tag{11}$$

The operators $\mathcal{E}_e$ implement the local depthwise, wide, dilated, and gated pointwise branches. The

spatial map $\mathcal{H}_\delta$ applies depthwise convolution, GELU, and pointwise projection to the modulated feature. The learned maps $\mathcal{A}_e, \mathcal{C}_e$ transform pooled descriptors into expert logits. The weights select one mixture per image, while δ guides spatial correction. Lower stages use pooled features alone, providing a lightweight continuation of adaptive recovery.

The routing equations give the perturbation results a common interpretation. CDERL determines which responses deserve transmission, and DAR determines how the transmitted representation should be transformed under its local evidence condition. The two mechanisms address different stages of the same reflected-screen recovery problem.

Training objective. In Equation 5a, the masked RGB loss averages absolute prediction error over target-mask pixels and colour channels. At each high-level stage, $\boldsymbol{\ell}_l$ contains nine terms: cosine alignment; target-evidence and confidence supervision; batch contrastive alignment; decoded-prototype reconstruction and mask supervision; token usage regularization; spatial degradation supervision; and expert specialization. Their weights, in this order, are

$$\boldsymbol{\lambda} = (0.003, 0.003, 0.002, 0.002, 0.002, 0.002, 0.001, 0.004, 0.001)^\top.$$

Target evidence combines screen luminance, Sobel variation, and saturation after resizing to the feature resolution. Smooth ℓ_1 terms match the standardized query–key agreement to this evidence and supervise confidence and prototype masks with its min–max-normalized form. Decoded prototypes match the detached normalized bridge input using Smooth ℓ_1; the spatial degradation map matches the resized, min–max-normalized detached mean absolute RGB error. Batch contrastive alignment compares pooled query and key descriptors. Token and expert regularizers combine assignment entropy with a penalty for low aggregate usage entropy, supporting selective assignments and usage across the prototype bank and expert branches.

Representation	Model	Source	Screen			Figure			Game		
			PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Direct skip	U-Net	MICCAI'15	10.688	0.491	0.676	12.849	0.497	0.667	15.642	0.474	0.603
CBAM-style bridge	CBAM	ECCV'18	10.658	0.490	0.675	12.846	0.491	0.689	16.002	0.489	0.598
Non-local bridge	Non-local	CVPR'18	10.633	0.488	0.682	14.171	0.517	0.609	16.289	0.502	0.574
Window-attention bridge	Swin Transformer	ICCV'21	10.615	0.489	0.679	13.513	0.518	0.643	16.252	0.496	0.588
MDTA bridge	MDTA	CVPR'22	10.097	0.467	0.698	15.239	0.513	0.603	16.184	0.493	0.572
NAFBlock bridge	NAFBlock	ECCV'22	10.687	0.485	0.688	13.945	0.528	0.632	16.236	0.493	0.583
CDERL	Ours Representation	-	13.922	0.542	0.626	17.047	0.571	0.562	16.516	0.517	0.551

Table 6: Representation comparison under the same U-Net-style recovery framework. Replacement modules are instantiated within the recovery backbone; Model and Source identify their reference designs, including U-Net Ronneberger et al. (2015), CBAM Woo et al. (2018), Non-local Wang et al. (2018), Swin Transformer Liu et al. (2021), MDTA Zamir et al. (2022), and NAFBlock Chen et al. (2022). Red text and shading highlight the best row; blue text and shading highlight the selected bridge row across the three content groups.

D Controlled Mechanism Comparisons

D.1 Representation, Learning, and Routing

We compare representation modules, training settings, and routing mechanisms within a common U-Net-style recovery framework (Tables 6–8). These comparisons examine how each design contributes at its corresponding stage of screen recovery, relating the choice of representation, supervision, and expert processing to reconstruction quality.

CDERL gives the clearest representation gain on Figure, where its PSNR reaches 17.047, compared with 15.239 for the MDTA bridge and 13.945 for the NAFBlock bridge; on Game, its LPIPS is also the lowest listed value at 0.551. These results show that evidence organization preserves line structure, colour partitions, and major contours in content whose reflected traces are spatially mixed.
This representation gain is consistent with CDERL's feature agreement and prototype transport. CDERL compares encoder-side reflected evidence with the evolving decoder state before transporting reusable evidence prototypes, so feature aggregation favors mutually supported screen structure and gives later decoding a stronger basis for spatial layout.

The training comparison shows that supervision contributes alongside the representation module. The proposed setting gives the strongest Screen result in every metric: 14.314 dB PSNR, 0.527 SSIM,

Learning objective	Model	Source	Screen			Figure			Game		
			PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Coarse-scale ℓ_1	Pixel L1	-	13.755	0.508	0.663	13.941	0.504	0.630	16.198	0.449	0.586
Charbonnier	Robust Pixel Loss	CVPR'17	13.901	0.512	0.623	13.926	0.509	0.631	16.207	0.505	0.574
ℓ_1 + SSIM	SSIM Loss	TIP'04	14.043	0.517	0.651	14.170	0.522	0.615	16.237	0.486	0.571
ℓ_1 + Perceptual	Perceptual Loss	ECCV'16	14.035	0.514	0.652	13.746	0.503	0.635	16.013	0.447	0.583
ℓ_1 + FFL	Focal Frequency Loss	ICCV'21	13.813	0.507	0.656	13.506	0.501	0.642	16.283	0.446	0.578
ℓ_1 + cGAN	Conditional GAN	CVPR'17	13.034	0.403	0.684	12.967	0.252	0.657	14.866	0.262	0.667
Consensus-driven learning	Ours Learning	-	14.314	0.527	0.618	16.752	0.564	0.565	16.376	0.505	0.568

Table 7: Training-setting comparison under the same U-Net-style recovery framework. The compared training settings use a common recovery backbone; Model and Source identify the source objectives, including SSIM Wang et al. (2004), Perceptual Loss Johnson et al. (2016), Focal Frequency Loss Jiang et al. (2021), and Conditional GAN Isola et al. (2017). Consensus-driven learning denotes ℓ_1 with auxiliary supervision. Red highlights the best learning rows; blue highlights the selected row.

Routing mechanism	Model	Source	Screen			Figure			Game		
			PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Prior-smoothed soft routing	Adaptive MoE	NeurComp'91	13.153	0.504	0.645	16.995	0.579	0.557	16.310	0.512	0.562
Top-1 expert routing	Switch Transformer	JMLR'22	12.181	0.492	0.676	16.989	0.578	0.560	16.370	0.515	0.560
Top-2 expert routing	GShard	ICLR'21	10.884	0.451	0.710	16.628	0.570	0.560	16.380	0.517	0.558
Noisy top-2 routing	Sparsely-Gated MoE	ICLR'17	12.315	0.486	0.665	16.787	0.575	0.557	16.390	0.517	0.559
Balanced soft routing	BASE-inspired	ICML'21	11.103	0.428	0.688	16.865	0.577	0.550	16.306	0.510	0.557
Expert-choice routing	Expert Choice MoE	NeurIPS'22	11.114	0.454	0.686	16.906	0.575	0.555	16.457	0.516	0.558
Degradation-aware routing	RERL (Ours)	-	13.982	0.511	0.627	17.171	0.586	0.550	16.728	0.521	0.554

Table 8: Routing comparison under the same representation and recovery framework. Replacement modules are instantiated within the recovery backbone; Model and Source identify their reference designs, including Adaptive MoE Jacobs et al. (1991), Switch Transformer Fedus et al. (2022), GShard Lepikhin et al. (2021), Sparsely-Gated MoE Shazeer et al. (2017), BASE Lewis et al. (2021), and Expert Choice MoE Zhou et al. (2022). Balanced soft routing is the BASE-inspired soft-routing variant. Red highlights the best routing row; blue highlights the selected routing row.

and 0.618 LPIPS. On Figure, it raises PSNR from the strongest alternative value of 14.170 dB to 16.752 dB, linking the training design to the recovery of organized screen content.

The proposed setting combines full-resolution ℓ_1 reconstruction with auxiliary evidence supervision. The alternatives span pixel, structural, perceptual, frequency, and adversarial objectives, including coarse-scale output comparisons. The results connect complementary supervision of reconstructed content and internal representations to improved screen recovery.

DAR produces the strongest routing result on Game, reaching 16.728 PSNR and 0.521 SSIM, and it reaches 17.171 PSNR on Figure compared with 16.995 for prior-smoothed soft routing. These gains support conditioning the expert mixture on reflected feature quality and refining local recovery with spatial guidance.

The alternatives in Table 8 compare input-dependent expert selection and balanced allocation Fedus et al. (2022); Lewis et al. (2021); Zhou et al. (2022). DAR combines an observation-conditioned expert mixture with spatial guidance, reaching 16.728 dB PSNR on Game versus 16.457 dB for the strongest alternative in Table 8. This connects adaptive expert processing with spatial correction for reflected-screen recovery.

D.2 Same-Backbone Interference

We next examine whether the training and representation advantages persist as the reflected observation becomes noisy. Tables 9 and 10 compare the corresponding designs across five SNR settings on Screen within the common U-Net-style backbone. The sweep tracks how the proposed training setting and CDERL preserve their reconstruction advantage as random fluctuations alter the observed reflection. The proposed setting leads at every SNR level, with 14.872 dB PSNR at 40 dB SNR and 0.606 LPIPS, compared with the strongest alternative LPIPS of 0.612.

Learned perceptual losses compare reconstructed and target images through feature representations, complementing pixel fidelity.

Noise introduces local responses that compete with screen boundaries and colour transitions during reconstruction. The proposed training setting combines full-resolution image agreement with auxiliary

Learning objective	Source	20			25			30			35			40		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Coarse-scale ℓ_1	-	11.549	0.467	0.736	11.557	0.477	0.729	11.559	0.480	0.727	11.560	0.481	0.727	11.560	0.482	0.728
Charbonnier	CVPR'17	12.607	0.495	0.696	12.660	0.504	0.692	12.678	0.507	0.691	12.690	0.508	0.691	12.693	0.509	0.691
ℓ_1 + SSIM	TIP'04	13.955	0.527	0.650	13.978	0.535	0.649	13.979	0.537	0.651	13.980	0.538	0.652	13.978	0.539	0.653
ℓ_1 + Perceptual	ECCV'16	12.536	0.404	0.618	12.564	0.410	0.614	12.576	0.411	0.612	12.582	0.412	0.612	12.585	0.412	0.612
ℓ_1 + FFL	ICCV'21	11.944	0.489	0.704	11.998	0.497	0.700	12.021	0.501	0.699	12.030	0.502	0.699	12.034	0.502	0.699
ℓ_1 + cGAN	CVPR'17	13.208	0.374	0.695	13.214	0.380	0.689	13.216	0.383	0.687	13.218	0.383	0.686	13.221	0.384	0.685
Consensus-driven learning	Ours	14.889	0.536	0.610	14.879	0.543	0.606	14.876	0.546	0.606	14.876	0.547	0.606	14.872	0.547	0.606

Table 9: Training-setting comparison on Screen across SNR (dB) settings in the common U-Net-style backbone. Consensus-driven learning denotes ℓ_1 with auxiliary supervision. Red text and shading highlight the best learning row; blue text and shading highlight the selected row.

supervision of the evidence representation. Its improvement across PSNR, SSIM, and LPIPS therefore extends across image fidelity, structural agreement, and perceptual similarity across the tested noise levels.

Representation	Source	20			25			30			35			40		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Direct skip	MICCAI'15	10.758	0.467	0.745	10.770	0.479	0.735	10.775	0.483	0.735	10.778	0.484	0.736	10.779	0.485	0.736
CBAM-style bridge	ECCV'18	10.776	0.460	0.742	10.794	0.473	0.729	10.803	0.477	0.727	10.807	0.479	0.728	10.810	0.479	0.729
Non-local bridge	CVPR'18	10.716	0.459	0.728	10.725	0.473	0.720	10.729	0.477	0.721	10.732	0.479	0.722	10.733	0.479	0.723
Window-attention bridge	ICCV'21	10.748	0.458	0.741	10.763	0.473	0.725	10.771	0.479	0.721	10.775	0.481	0.721	10.777	0.481	0.721
MDTA bridge	CVPR'22	10.179	0.447	0.696	10.170	0.456	0.694	10.165	0.460	0.694	10.162	0.461	0.694	10.161	0.461	0.694
NAFBlock bridge	ECCV'22	10.778	0.457	0.691	10.791	0.471	0.684	10.797	0.477	0.684	10.799	0.479	0.685	10.800	0.479	0.685
CDERL	Ours	15.829	0.586	0.576	15.854	0.592	0.572	15.864	0.594	0.571	15.864	0.595	0.571	15.865	0.595	0.571

Table 10: Representation comparison on Screen across SNR (dB) settings in the common U-Net-style backbone. Red text and shading highlight the best row; blue text and shading highlight the selected bridge row across the five SNR settings.

CDERL maintains a clear representation advantage across the SNR sweep. At 40 dB SNR, it achieves 15.865 dB PSNR and 0.571 LPIPS, compared with the strongest alternative values of 10.810 dB and 0.685. Organizing weak reflected cues therefore provides a stronger input for reconstructing screen content.

Interference first weakens fine boundaries and colour transitions, and generic aggregation averages these cues with surface responses. CDERL retains mutually supported screen traces, so the decoder receives a screen-oriented evidence representation before DAR applies degradation-dependent transformations.

E Internal Ablations of CDERL and DAR

E.1 Joint Module Contribution

Table 3 quantifies the joint contribution of CDERL and DAR across four content groups, and Figure 9 shows how these changes appear in three reconstructions. The full model preserves clearer interface regions and chart organization, while module removal changes the arrangement and separation of recovered regions. This connects the joint-ablation gains to visible screen content.

The visual variants show complementary changes in recovered structure. CDERL organizes evidence before decoding, and DAR increases selectivity across weak and strong regions. The full RERL output retains the clearest interface regions, chart layout, and colour organization across the shown previews.

The image-level evidence follows the recovery chain in the quantitative tables. CDERL weights the evidence residuals added to decoder features, and DAR adapts the expert mixture and spatial correction to the resulting feature state. Their roles form a continuous evidence-to-recovery sequence in the visualization.

E.2 Internal CDERL Design

Table 4 separates alignment from prototype transport. Alignment links the decoder state to the reflection, while prototype transport adds reusable feature configurations. Figures 25 and 26 complement this component analysis through final reconstructions and intermediate features, respectively.

The feature visualizations show distinct organization from encoder states through deeper recovery layers. RERL retains clearer screen-related structure across the previews, while generic representations preserve more diffuse spot or surface responses.

The alignment and prototype pathways contribute complementary feature operations. CDERL aligns encoder-side reflected evidence with the evolving decoder state and transports reusable evidence prototypes, giving later decoding a stronger basis for screen layout and colour organization.

E.3 Internal DAR Design

Table 5 isolates degradation estimation and adaptive assignment within DAR. Table 8 and Figure 10 complement this analysis by comparing alternative routing mechanisms in the common recovery backbone, including prior-smoothed soft, top-1, top-2, noisy top-2, balanced soft, and expert-choice routing.

The routing visualization shows more coherent screen regions for DAR across the three previews, while common allocations produce broader bright fields, smoothed regions, or transferred spot contours. The spatial difference follows the condition of reflected evidence: expert selection changes with the observed feature state.

The routing table supports the same mechanism. DAR delivers the strongest or tied-strongest metrics across Screen, Figure, and Game, because its degradation estimate converts uneven evidence quality into a suitable transformation path.

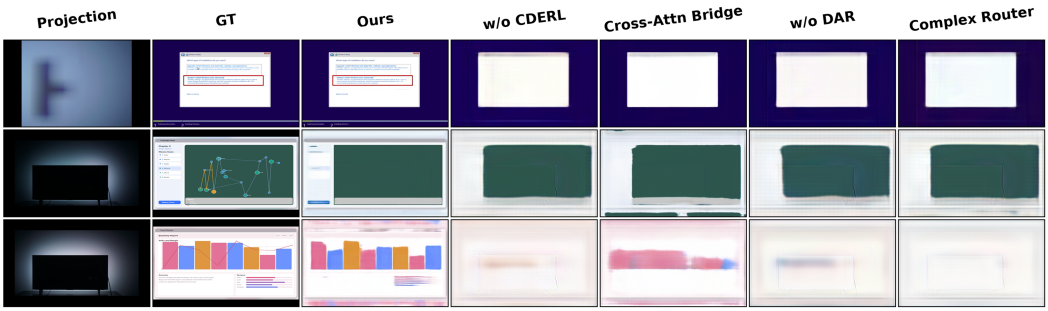

Figure 9: Qualitative comparison of the full RERL model and the CDERL and DAR ablation variants.

The two routing components act at different scales. The estimator describes local feature conditions; its spatial summary and pooled features condition the image-level expert mixture. Static assignment fixes the expert weights across observations. DAR adapts the mixture to the observed features and uses spatial guidance to refine the correction across regions.
The main-ablation visualization (Figure 9) shows how the two modules change the screen structure recovered from the same reflection spot. The full output preserves the clearest interface regions, chart layout, and colour organization across the shown samples, while the ablations retain broader spot appearance with weaker screen-specific structure.

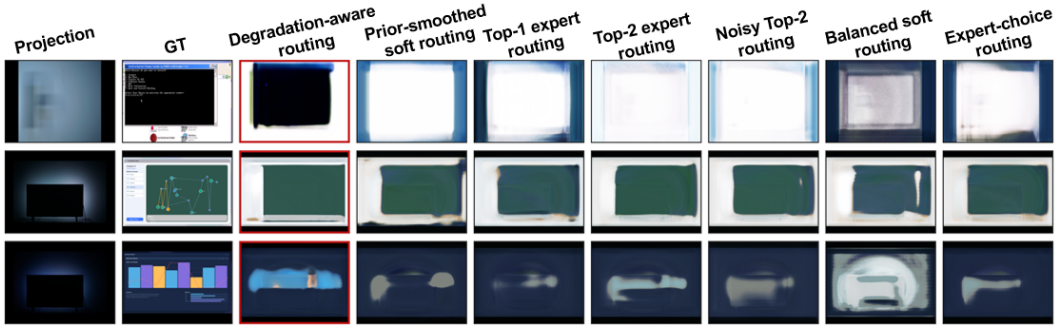

Figure 10: Qualitative comparison of DAR and expert-assignment strategies under the same reflected observations.

The visual patterns match the representation and routing comparisons. CDERL organizes evidence before decoding, and DAR increases selectivity in transforming weak and strong regions. The full RERL chain therefore links the qualitative comparison to a visible evidence-to-recovery difference.

The routing visualization (Figure 10) shows how expert allocation preserves the screen layout when the reflection spot contains uneven evidence. DAR keeps the large screen regions and their internal organization more coherently across the three previews, while alternative allocations produce broad bright fields, smoothed regions, or transferred spot contours.
The visual pattern explains the routing rows in Table 8. DAR combines a feature-conditioned expert mixture with spatially varying guidance to adapt the correction to the available reflected evidence.

Method	Source	-100			-80			-60			-40			-20			0		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Transformer-Based Restoration																			
AST	CVPR'24	16.928	0.635	0.627	17.116	0.638	0.624	17.249	0.641	0.621	17.330	0.642	0.619	17.397	0.644	0.617	17.456	0.645	0.616
AdaIR	ICLR'25	16.938	0.618	0.591	16.997	0.620	0.590	17.028	0.621	0.589	17.051	0.621	0.588	17.055	0.622	0.587	17.436	0.639	0.575
Restormer	CVPR'22	14.436	0.594	0.612	14.542	0.596	0.611	14.680	0.599	0.608	14.679	0.600	0.608	14.780	0.601	0.606	17.368	0.644	0.568
SFHformer	ECCV'24	10.025	0.468	0.761	10.028	0.466	0.761	10.031	0.465	0.761	10.033	0.464	0.760	10.033	0.463	0.760	10.010	0.462	0.760
Uformer	CVPR'22	16.858	0.634	0.628	17.044	0.638	0.625	17.203	0.641	0.621	17.344	0.643	0.618	17.448	0.645	0.616	17.547	0.647	0.614
DehazeFormer	TIP'23	17.078	0.630	0.617	17.337	0.636	0.614	17.522	0.641	0.612	17.642	0.645	0.610	17.711	0.647	0.609	17.774	0.649	0.607
Retinexformer	ICCV'23	15.758	0.600	0.621	16.061	0.609	0.617	16.263	0.617	0.612	16.387	0.623	0.610	16.485	0.628	0.608	17.925	0.656	0.592
Convolution-Based Restoration																			
ConvIR	TPAMI'24	17.299	0.645	0.609	17.430	0.648	0.606	17.537	0.650	0.604	17.606	0.652	0.602	17.633	0.652	0.601	17.642	0.651	0.601
IRNeXt	ICML'23	17.622	0.651	0.583	17.860	0.658	0.579	18.028	0.663	0.575	18.126	0.666	0.573	18.159	0.668	0.573	18.136	0.668	0.574
MIRNetv2	TPAMI'23	17.533	0.650	0.619	17.577	0.652	0.617	17.608	0.653	0.616	17.621	0.654	0.615	17.619	0.655	0.614	17.679	0.656	0.612
DnCNN	TIP'17	8.363	0.417	0.738	8.346	0.416	0.738	8.342	0.415	0.738	8.334	0.416	0.738	8.325	0.416	0.737	8.738	0.427	0.726
U-Net	MICCAI'15	10.184	0.472	0.758	10.186	0.471	0.757	10.193	0.471	0.758	10.194	0.470	0.757	10.190	0.470	0.757	10.285	0.471	0.758
pix2pix	CVPR'17	10.244	0.451	0.701	10.256	0.451	0.701	10.269	0.450	0.701	10.279	0.451	0.701	10.291	0.450	0.700	11.140	0.450	0.692
DeepSN-Net	TPAMI'25	10.470	0.444	0.748	10.555	0.444	0.745	10.633	0.444	0.743	10.705	0.444	0.741	10.769	0.444	0.740	10.868	0.444	0.739
Hybrid / Degradation-Specific Restoration																			
DEA-Net	TIP'24	17.158	0.641	0.608	17.330	0.645	0.603	17.469	0.650	0.597	17.563	0.653	0.593	17.614	0.655	0.590	17.729	0.658	0.588
DExNet	TPAMI'25	17.139	0.646	0.605	17.282	0.649	0.603	17.367	0.652	0.602	17.430	0.654	0.601	17.474	0.655	0.600	17.629	0.659	0.595
DarkIR	CVPR'25	4.398	0.039	0.942	4.410	0.039	0.941	4.421	0.039	0.940	4.434	0.039	0.938	4.445	0.040	0.937	4.465	0.041	0.936
DocRes	CVPR'24	16.938	0.625	0.598	17.023	0.629	0.596	17.074	0.631	0.595	17.154	0.633	0.593	17.179	0.634	0.591	17.528	0.647	0.579
HVI-CIDNet	CVPR'25	16.809	0.623	0.634	16.816	0.624	0.633	16.832	0.625	0.632	16.865	0.626	0.632	16.853	0.626	0.631	17.774	0.656	0.598
INFWIDE	TIP'23	13.045	0.594	0.677	13.084	0.594	0.676	13.061	0.594	0.676	13.033	0.593	0.675	13.083	0.594	0.674	16.807	0.641	0.633
Reflection-Evidence Routed Representation Learning																			
RERL	-	17.902	0.652	0.562	18.143	0.660	0.556	18.338	0.666	0.551	18.503	0.672	0.547	18.610	0.676	0.544	18.695	0.681	0.542

Table 11: Game brightness comparison at representative brightness indices. Red text and shading highlight the best row; blue text and shading highlight the DehazeFormer, IRNeXt, and DEA-Net rows within their respective method groups. (Part 1: indices -100 to 0)

Method	Source	0			+20			+40			+60			+80			+100		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Transformer-Based Restoration																			
AST	CVPR'24	17.456	0.645	0.616	17.442	0.645	0.616	17.397	0.645	0.616	17.298	0.643	0.618	17.175	0.641	0.620	16.994	0.639	0.623
AdaIR	ICLR'25	17.436	0.639	0.575	17.439	0.640	0.575	17.450	0.641	0.575	17.451	0.640	0.575	17.456	0.641	0.575	17.463	0.641	0.575
Restormer	CVPR'22	17.368	0.644	0.568	17.360	0.644	0.568	17.351	0.643	0.568	17.363	0.644	0.567	17.361	0.644	0.567	17.364	0.644	0.567
SFHformer	ECCV'24	10.010	0.462	0.760	10.006	0.461	0.760	9.998	0.460	0.760	9.987	0.459	0.760	9.972	0.458	0.760	9.953	0.457	0.761
Uformer	CVPR'22	17.547	0.647	0.614	17.540	0.647	0.614	17.512	0.646	0.614	17.445	0.645	0.616	17.283	0.642	0.619	17.043	0.639	0.624
DehazeFormer	TIP'23	17.774	0.649	0.607	17.756	0.649	0.608	17.697	0.648	0.610	17.587	0.645	0.613	17.398	0.641	0.617	17.127	0.636	0.622
Retinexformer	ICCV'23	17.925	0.656	0.592	17.931	0.656	0.593	17.889	0.656	0.594	17.827	0.655	0.596	17.741	0.653	0.598	17.654	0.650	0.601
Convolution-Based Restoration																			
ConvIR	TPAMI'24	17.642	0.651	0.601	17.619	0.651	0.602	17.536	0.649	0.604	17.435	0.647	0.607	17.286	0.644	0.611	17.125	0.640	0.616
IRNeXt	ICML'23	18.136	0.668	0.574	18.071	0.666	0.575	17.989	0.665	0.578	17.877	0.662	0.581	17.722	0.658	0.585	17.535	0.653	0.590
MIRNetv2	TPAMI'23	17.679	0.656	0.612	17.669	0.656	0.612	17.652	0.656	0.613	17.625	0.655	0.613	17.587	0.654	0.614	17.532	0.653	0.615
DnCNN	TIP'17	8.738	0.427	0.726	8.737	0.427	0.726	8.738	0.425	0.726	8.729	0.426	0.726	8.726	0.425	0.726	8.719	0.423	0.726
U-Net	MICCAI'15	10.285	0.471	0.758	10.283	0.470	0.758	10.278	0.470	0.758	10.272	0.470	0.757	10.283	0.470	0.758	10.277	0.469	0.757
pix2pix	CVPR'17	11.140	0.450	0.692	11.138	0.449	0.692	11.148	0.447	0.692	11.152	0.447	0.692	11.156	0.446	0.692	11.160	0.446	0.692
DeepSN-Net	TPAMI'25	10.868	0.444	0.739	10.891	0.443	0.739	10.937	0.443	0.739	10.973	0.442	0.738	11.005	0.441	0.738	11.031	0.439	0.738
Hybrid / Degradation-Specific Restoration																			
DEA-Net	TIP'24	17.729	0.658	0.588	17.755	0.659	0.588	17.738	0.659	0.588	17.682	0.658	0.590	17.588	0.656	0.593	17.469	0.654	0.598
DExNet	TPAMI'25	17.629	0.659	0.595	17.628	0.659	0.596	17.615	0.658	0.597	17.583	0.657	0.597	17.540	0.656	0.599	17.470	0.654	0.601
DarkIR	CVPR'25	4.465	0.041	0.936	4.470	0.041	0.936	4.481	0.041	0.934	4.492	0.041	0.932	4.504	0.041	0.931	4.516	0.042	0.930
DocRes	CVPR'24	17.528	0.647	0.579	17.532	0.647	0.579	17.539	0.648	0.579	17.533	0.647	0.579	17.534	0.647	0.580	17.516	0.646	0.580
HVI-CIDNet	CVPR'25	17.774	0.656	0.598	17.766	0.655	0.598	17.760	0.656	0.598	17.760	0.656	0.598	17.750	0.656	0.598	17.742	0.656	0.598
INFWIDE	TIP'23	16.807	0.641	0.633	16.809	0.641	0.633	16.804	0.641	0.633	16.805	0.641	0.632	16.774	0.640	0.632	16.782	0.640	0.632
Reflection-Evidence Routed Representation Learning																			
RERL	-	18.695	0.681	0.542	18.654	0.680	0.543	18.585	0.678	0.544	18.465	0.675	0.547	18.288	0.671	0.550	18.100	0.666	0.554

Table 11: Game brightness comparison at representative brightness indices. Red text and shading highlight the best row; blue text and shading highlight the Retinexformer, IRNeXt, and HVI-CIDNet rows within their respective method groups. (Part 2: indices 0 to $+100$)

F PERTURBATION RESULTS

We apply controlled input perturbations to captured reflected observations, examining brightness, noise, geometry, blur, surface response, and ambient illumination while retaining the corresponding screen targets. Brightness indices $b \in \{-100, -80, \ldots, 100\}$ are dimensionless and specify an RGB multiplier of $1 + b/1000$. Noise is parameterized by SNR in dB, using the mean-square power of normalized RGB observations. Surface and geometric conditions describe transformations of these observations; acquisition photographs document the recording environment.

F.1 BRIGHTNESS

The brightness experiment tracks screen luminance distribution and colour organization as the reflection spot is globally darkened or brightened. Table 11 reports the full Game brightness sweep; Table 12 does the same for chart-like Figure content.

RERL leads the Game brightness comparison at every selected brightness index in PSNR, SSIM, and LPIPS. Its PSNR rises from 17.902 at -100 to 18.695 at 0, then falls to 18.100 at $+100$, while LPIPS remains between 0.542 and 0.562. Global brightness changes shift the energy and contrast of reflected screen-light evidence, so generic restoration paths tend to follow the altered intensity pattern and lose screen layout. The advantage spans both directions of intensity change: Game recovery remains ahead in pixel fidelity, structural similarity, and perceptual distance as the observation is darkened or brightened. This comparison connects intensity variation in the reflection to reconstruction quality across the full sweep.

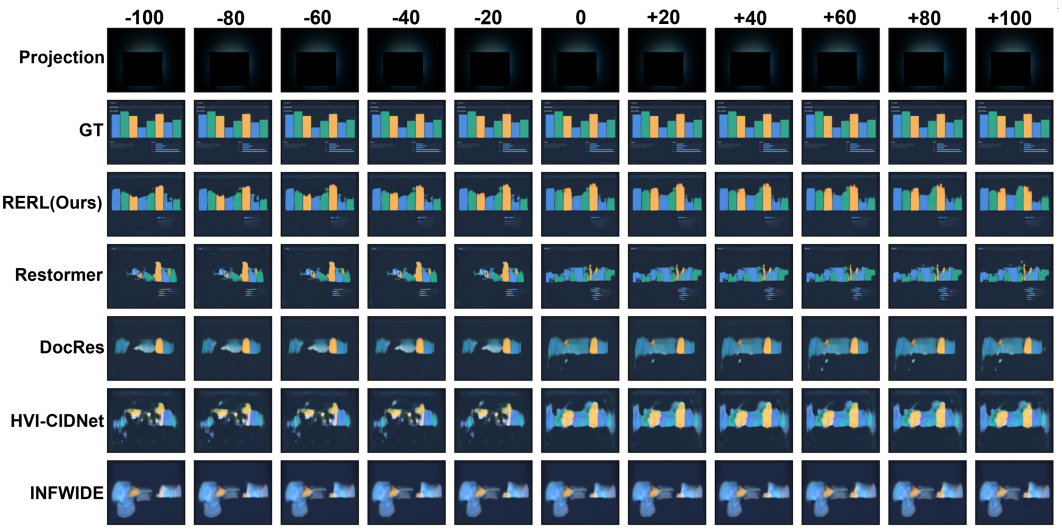

Figure 11: Qualitative hidden-screen recovery across the global brightness indices.

CDERL preserves local luminance, colour, and boundary traces that remain jointly supported, and DAR adjusts the image-level expert mixture and spatial correction as brightness shifts alter feature strength. The qualitative trend is visible in Figure 11.

RERL leads in PSNR and LPIPS at every selected Figure brightness index and reaches the top SSIM level from the reference through the bright settings; its PSNR varies by less than 0.1 dB across the selected points. Figure content contains many lines, borders, and colour partitions, and global brightness shifts can merge these traces into broad bright or dark responses that generic restoration paths process uniformly. CDERL preserves compatible line and colour responses, and DAR processes regions according to their altered degradation state. This pairing keeps chart organization available as the reflected evidence changes in global energy.

The brightness visualization shows the recovered screen as the reflection spot moves from -100 to $+100$. RERL preserves the main bar layout and colour organization across the sweep, while the projection changes substantially in visible energy and the comparison outputs follow more diffuse spot structure.

The visual pattern complements both brightness tables. Global intensity changes alter the appearance of reflected evidence, while CDERL keeps screen-supported relationships available to the decoder and DAR adapts the recovery path to the shifted local condition.

Method	Source	-100			-80			-60			-40			-20			0		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Transformer-Based Restoration																			
AST	CVPR'24	17.872	0.641	0.631	17.879	0.640	0.632	17.885	0.640	0.632	17.886	0.639	0.632	17.885	0.639	0.632	17.892	0.639	0.632
AdaIR	ICLR'25	17.967	0.669	0.626	17.954	0.669	0.625	17.941	0.669	0.625	17.929	0.669	0.624	17.915	0.669	0.624	17.930	0.669	0.623
Restormer	CVPR'22	18.717	0.704	0.564	18.722	0.704	0.562	18.719	0.704	0.562	18.708	0.703	0.562	18.715	0.704	0.561	18.921	0.710	0.551
SFHformer	ECCV'24	13.199	0.600	0.712	13.201	0.599	0.713	13.202	0.599	0.713	13.204	0.598	0.713	13.206	0.598	0.713	13.205	0.597	0.713
Uformer	CVPR'22	17.766	0.625	0.624	17.735	0.623	0.624	17.705	0.622	0.625	17.670	0.621	0.625	17.639	0.620	0.625	17.586	0.619	0.626
DehazeFormer	TIP'23	17.237	0.681	0.627	17.611	0.685	0.625	17.994	0.690	0.623	18.370	0.696	0.621	18.659	0.701	0.618	18.917	0.709	0.616
Retinexformer	ICCV'23	17.970	0.692	0.620	17.954	0.692	0.620	17.939	0.692	0.620	17.924	0.692	0.620	17.909	0.691	0.620	17.895	0.690	0.619
Convolution-Based Restoration																			
ConvIR	TPAMI'24	18.753	0.715	0.634	18.714	0.715	0.633	18.673	0.715	0.632	18.631	0.715	0.631	18.592	0.715	0.630	18.540	0.715	0.629
IRNeXt	ICML'23	17.620	0.667	0.635	17.595	0.667	0.634	17.569	0.667	0.634	17.542	0.667	0.633	17.517	0.668	0.633	17.479	0.668	0.633
MIRNetv2	TPAMI'23	17.093	0.712	0.671	17.456	0.712	0.667	17.778	0.713	0.663	18.043	0.713	0.660	18.228	0.713	0.657	18.340	0.713	0.654
DnCNN	TIP'17	14.214	0.574	0.666	14.215	0.574	0.666	14.214	0.573	0.666	14.214	0.573	0.665	14.211	0.573	0.665	14.204	0.574	0.664
U-Net	MICCAI'15	11.854	0.568	0.667	11.852	0.568	0.667	11.852	0.567	0.667	11.847	0.567	0.667	11.841	0.567	0.667	11.815	0.566	0.667
pix2pix	CVPR'17	13.615	0.577	0.663	13.613	0.578	0.662	13.610	0.577	0.663	13.608	0.577	0.662	13.606	0.577	0.663	15.476	0.615	0.638
DeepSN-Net	TPAMI'25	14.286	0.506	0.690	14.305	0.504	0.690	14.324	0.503	0.690	14.343	0.502	0.691	14.361	0.500	0.691	14.401	0.499	0.692
Hybrid / Degradation-Specific Restoration																			
DEA-Net	TIP'24	18.233	0.715	0.645	18.219	0.715	0.644	18.208	0.715	0.644	18.192	0.715	0.643	18.187	0.715	0.643	18.179	0.715	0.642
DExNet	TPAMI'25	17.870	0.702	0.626	17.962	0.704	0.625	18.047	0.706	0.624	18.124	0.708	0.624	18.176	0.709	0.623	18.602	0.721	0.621
DarkIR	CVPR'25	4.095	0.036	0.935	4.112	0.036	0.934	4.128	0.036	0.933	4.144	0.037	0.931	4.160	0.037	0.929	4.185	0.038	0.929
DocRes	CVPR'24	18.751	0.713	0.617	18.727	0.714	0.617	18.701	0.714	0.617	18.676	0.714	0.617	18.647	0.713	0.616	18.771	0.718	0.611
HVI-CIDNet	CVPR'25	19.119	0.717	0.607	19.125	0.717	0.607	19.128	0.717	0.607	19.135	0.717	0.607	19.138	0.718	0.606	19.415	0.721	0.597
INFWIDE	TIP'23	19.028	0.719	0.622	19.038	0.719	0.622	19.041	0.719	0.621	19.054	0.719	0.621	19.051	0.719	0.621	19.054	0.720	0.617
Reflection-Evidence Routed Representation Learning																			
RERL	-	19.456	0.715	0.544	19.455	0.715	0.545	19.482	0.716	0.544	19.439	0.714	0.547	19.460	0.715	0.544	19.523	0.723	0.546

Table 12: Figure brightness comparison at representative brightness indices. Red text and shading identify RERL; blue text and shading identify the highlighted comparison methods. (Part 1: indices -100 to 0)

Method	Source	+20			+40			+60			+80			+100		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Transformer-Based Restoration																
AST	CVPR'24	17.890	0.638	0.634	17.887	0.638	0.633	17.883	0.638	0.633	17.877	0.637	0.633	17.871	0.636	0.633
AdaIR	ICLR'25	17.922	0.669	0.623	17.910	0.669	0.623	17.899	0.669	0.622	17.889	0.669	0.622	17.879	0.669	0.622
Restormer	CVPR'22	18.916	0.709	0.552	18.906	0.709	0.551	18.888	0.709	0.552	18.888	0.710	0.552	18.895	0.709	0.551
SFHformer	ECCV'24	13.210	0.596	0.714	13.212	0.596	0.714	13.215	0.595	0.714	13.217	0.594	0.714	13.220	0.594	0.714
Uformer	CVPR'22	17.573	0.618	0.627	17.542	0.618	0.627	17.510	0.617	0.627	17.481	0.616	0.628	17.454	0.615	0.628
DehazeFormer	TIP'23	18.943	0.710	0.616	18.942	0.712	0.616	18.913	0.713	0.617	18.877	0.713	0.618	18.835	0.713	0.619
Retinexformer	ICCV'23	17.886	0.690	0.619	17.872	0.690	0.619	17.861	0.690	0.619	17.850	0.689	0.619	17.840	0.689	0.619
Convolution-Based Restoration																
ConvIR	TPAMI'24	18.517	0.715	0.630	18.486	0.715	0.630	18.457	0.715	0.630	18.429	0.714	0.631	18.401	0.714	0.631
IRNeXt	ICML'23	17.462	0.669	0.633	17.437	0.669	0.632	17.413	0.670	0.632	17.389	0.671	0.631	17.365	0.671	0.631
MIRNetv2	TPAMI'23	18.319	0.713	0.653	18.213	0.713	0.651	18.044	0.714	0.651	17.852	0.715	0.651	17.688	0.717	0.652
DnCNN	TIP'17	14.202	0.574	0.664	14.199	0.573	0.663	14.196	0.572	0.663	14.193	0.571	0.663	14.193	0.571	0.664
U-Net	MICCAI'15	11.811	0.565	0.668	11.802	0.565	0.669	11.792	0.564	0.669	11.779	0.563	0.670	11.771	0.563	0.671
pix2pix	CVPR'17	15.476	0.614	0.639	15.473	0.614	0.639	15.471	0.614	0.637	15.465	0.613	0.638	15.455	0.613	0.637
DeepSN-Net	TPAMI'25	14.410	0.498	0.693	14.424	0.495	0.694	14.436	0.493	0.695	14.447	0.491	0.696	14.457	0.488	0.696
Hybrid / Degradation-Specific Restoration																
DEA-Net	TIP'24	18.179	0.714	0.642	18.175	0.714	0.642	18.165	0.714	0.642	18.158	0.713	0.642	18.150	0.712	0.642
DExNet	TPAMI'25	18.599	0.721	0.621	18.589	0.721	0.621	18.580	0.721	0.622	18.571	0.721	0.622	18.560	0.721	0.622
DarkIR	CVPR'25	4.195	0.038	0.929	4.211	0.039	0.928	4.227	0.039	0.926	4.243	0.039	0.924	4.259	0.040	0.923
DocRes	CVPR'24	18.753	0.718	0.611	18.725	0.718	0.611	18.701	0.719	0.611	18.673	0.719	0.611	18.648	0.719	0.612
HVI-CIDNet	CVPR'25	19.413	0.721	0.597	19.406	0.721	0.597	19.411	0.721	0.598	19.407	0.721	0.597	19.407	0.721	0.598
INFWIDE	TIP'23	19.053	0.719	0.617	19.049	0.719	0.617	19.046	0.718	0.617	19.049	0.719	0.617	19.051	0.718	0.617
Reflection-Evidence Routed Representation Learning																
RERL	-	19.503	0.721	0.548	19.490	0.721	0.546	19.501	0.722	0.547	19.496	0.722	0.546	19.501	0.722	0.546

Table 12: Figure brightness comparison at representative brightness indices. Red text and shading identify RERL; blue text and shading identify the highlighted comparison methods. (Part 2: indices $+20$ to $+100$)

F.2 NOISE

The noise experiment tracks how random responses displace screen evidence in highly textured scenes. Table 13 reports Figure at five noise levels, and Table 14 reports Screen. Table 15 reports Game at the same selected SNR levels.

Method	Source	20			25			30			35			40		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Transformer-Based Restoration																
AST	CVPR'24	10.024	0.445	0.762	10.036	0.488	0.717	10.042	0.507	0.699	10.046	0.513	0.695	10.048	0.515	0.693
AdaIR	ICLR'25	10.018	0.427	0.781	10.128	0.460	0.734	10.544	0.482	0.713	11.135	0.483	0.707	11.937	0.470	0.703
Restormer	CVPR'22	10.364	0.475	0.718	10.653	0.512	0.695	10.803	0.527	0.679	10.925	0.535	0.670	11.000	0.540	0.662
SFHformer	ECCV'24	7.327	0.454	0.805	7.332	0.472	0.778	7.335	0.478	0.767	7.336	0.480	0.765	7.337	0.481	0.765
Uformer	CVPR'22	9.493	0.433	0.759	9.508	0.474	0.713	9.516	0.491	0.695	9.520	0.497	0.690	9.522	0.499	0.689
DehazeFormer	TIP'23	10.280	0.562	0.694	10.312	0.569	0.690	10.329	0.571	0.688	10.337	0.572	0.687	10.341	0.572	0.687
Retinexformer	ICCV'23	9.787	0.487	0.718	9.706	0.526	0.691	9.685	0.547	0.679	9.703	0.554	0.675	9.734	0.557	0.673
Convolution-Based Restoration																
ConvIR	TPAMI'24	9.587	0.534	0.697	9.604	0.561	0.692	9.614	0.572	0.691	9.619	0.576	0.691	9.623	0.577	0.691
IRNeXt	ICML'23	9.465	0.498	0.723	9.463	0.523	0.695	9.463	0.534	0.688	9.463	0.538	0.686	9.463	0.539	0.686
MIRNetv2	TPAMI'23	9.894	0.497	0.745	9.919	0.545	0.711	9.932	0.565	0.708	9.939	0.572	0.709	9.944	0.575	0.710
DnCNN	TIP'17	6.138	0.413	0.760	6.086	0.440	0.734	6.770	0.435	0.717	6.669	0.455	0.709	6.635	0.461	0.706
U-Net	MICCAI'15	12.249	0.593	0.594	11.975	0.601	0.590	11.839	0.603	0.589	11.796	0.604	0.589	11.782	0.603	0.589
pix2pix	CVPR'17	7.152	0.416	0.771	7.281	0.438	0.769	7.368	0.458	0.766	7.433	0.469	0.763	7.539	0.465	0.761
DeepSN-Net	TPAMI'25	6.971	0.335	0.821	6.985	0.376	0.773	6.993	0.394	0.753	6.997	0.400	0.745	7.000	0.402	0.743
Hybrid / Degradation-Specific Restoration																
DEA-Net	TIP'24	9.885	0.577	0.705	9.875	0.577	0.704	9.869	0.577	0.704	9.866	0.577	0.704	9.865	0.577	0.703
DExNet	TPAMI'25	9.535	0.583	0.694	9.619	0.583	0.690	9.649	0.583	0.689	9.655	0.582	0.689	9.657	0.582	0.689
DarkIR	CVPR'25	5.119	0.179	0.890	5.114	0.188	0.868	5.112	0.190	0.861	5.111	0.191	0.858	5.111	0.191	0.857
DocRes	CVPR'24	10.155	0.460	0.785	10.344	0.506	0.764	10.086	0.540	0.747	10.012	0.560	0.720	10.258	0.567	0.702
HVI-CIDNet	CVPR'25	10.043	0.420	0.801	10.034	0.445	0.793	10.160	0.466	0.789	10.307	0.480	0.784	10.451	0.490	0.777
INFWIDE	TIP'23	8.673	0.493	0.795	8.930	0.527	0.748	9.101	0.549	0.728	9.260	0.574	0.696	9.265	0.580	0.681
Reflection-Evidence Routed Representation Learning																
RERL	-	19.170	0.662	0.569	19.392	0.695	0.550	19.458	0.710	0.544	19.497	0.715	0.543	19.487	0.716	0.543

Table 13: Figure noise comparison at representative SNR levels (dB). Red text and shading highlight the best row; blue text and shading highlight the selected rows within their respective method groups.

Method	Source	20			25			30			35			40		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Transformer-Based Restoration																
AST	CVPR'24	18.892	0.572	0.522	19.251	0.654	0.481	19.354	0.687	0.470	19.380	0.699	0.467	19.388	0.703	0.467
AdaIR	ICLR'25	16.721	0.473	0.586	17.911	0.574	0.526	18.015	0.596	0.518	17.806	0.597	0.527	17.885	0.606	0.523
Restormer	CVPR'22	12.565	0.363	0.732	13.264	0.436	0.683	13.799	0.480	0.649	14.321	0.509	0.624	14.585	0.519	0.620
SFHformer	ECCV'24	8.801	0.344	0.843	8.809	0.376	0.823	8.813	0.388	0.813	8.815	0.392	0.809	8.816	0.393	0.807
Uformer	CVPR'22	18.827	0.586	0.526	19.122	0.663	0.481	19.208	0.692	0.468	19.208	0.702	0.465	19.207	0.705	0.464
DehazeFormer	TIP'23	18.715	0.665	0.497	19.219	0.684	0.484	19.339	0.690	0.480	19.357	0.692	0.480	19.361	0.693	0.480
Retinexformer	ICCV'23	15.530	0.506	0.569	18.390	0.617	0.490	19.009	0.646	0.470	19.316	0.660	0.461	19.473	0.667	0.456
Convolution-Based Restoration																
ConvIR	TPAMI'24	19.125	0.557	0.516	20.092	0.667	0.457	20.233	0.708	0.448	20.266	0.721	0.448	20.272	0.726	0.449
IRNeXt	ICML'23	16.499	0.509	0.583	18.743	0.647	0.483	19.774	0.702	0.456	19.932	0.717	0.454	19.949	0.721	0.454
MIRNetv2	TPAMI'23	18.053	0.675	0.493	18.121	0.684	0.493	18.193	0.688	0.493	18.223	0.690	0.492	18.232	0.691	0.492
DnCNN	TIP'17	8.815	0.263	0.769	8.881	0.312	0.739	8.887	0.334	0.726	8.886	0.342	0.725	8.885	0.345	0.725
U-Net	MICCAI'15	9.293	0.214	0.825	9.218	0.229	0.810	9.190	0.235	0.805	9.182	0.238	0.805	9.180	0.239	0.806
pix2pix	CVPR'17	9.680	0.436	0.683	9.713	0.447	0.683	9.879	0.453	0.685	10.079	0.466	0.674	10.199	0.474	0.668
DeepSN-Net	TPAMI'25	10.340	0.219	0.867	10.416	0.264	0.817	10.451	0.284	0.794	10.469	0.291	0.784	10.479	0.294	0.781
Hybrid / Degradation-Specific Restoration																
DEA-Net	TIP'24	19.112	0.718	0.467	20.200	0.742	0.435	20.547	0.748	0.427	20.616	0.749	0.425	20.636	0.750	0.424
DExNet	TPAMI'25	19.935	0.725	0.441	19.977	0.730	0.439	20.001	0.732	0.438	20.003	0.732	0.438	20.005	0.732	0.438
DarkIR	CVPR'25	9.217	0.154	0.921	9.221	0.168	0.892	9.222	0.173	0.876	9.222	0.174	0.870	9.222	0.175	0.868
DocRes	CVPR'24	15.527	0.526	0.589	16.652	0.589	0.529	17.108	0.611	0.505	17.296	0.619	0.497	17.336	0.619	0.497
HVI-CIDNet	CVPR'25	12.165	0.348	0.760	12.251	0.404	0.719	12.313	0.430	0.704	12.358	0.443	0.699	12.456	0.453	0.693
INFWIDE	TIP'23	8.568	0.338	0.804	8.787	0.356	0.786	9.978	0.415	0.739	10.793	0.453	0.708	11.428	0.477	0.688
Reflection-Evidence Routed Representation Learning																
RERL	-	23.204	0.730	0.272	24.931	0.813	0.224	25.974	0.849	0.202	26.463	0.861	0.193	26.588	0.865	0.190

Table 14: Screen noise comparison at representative SNR levels (dB). Red text and shading highlight the best row; blue text and shading highlight the selected rows within their respective method groups.

RERL's Figure PSNR is 19.170 dB at 20 dB SNR and 19.487 dB at 40 dB SNR, reaching 19.497 dB at 35 dB SNR. Added noise introduces local fluctuations that can resemble chart lines and borders, while the qualitative examples retain the major bar arrangement and colour ordering of the hidden screen.

CDERL retains responses jointly supported by reflected evidence and the decoder state, and DAR changes processing for regions whose features become interference-heavy. The resulting output stays tied to major chart lines and colour partitions. RERL leads the Screen noise comparison at all selected levels, with PSNR increasing from 23.204 to 26.588 and LPIPS decreasing from 0.272 to 0.190 over the settings. Noise perturbs local interface edges while broad luminance regions and colour boundaries remain available, making it important to distinguish noise-induced edge responses from the surviving luminance and colour boundaries.CDERL preserves the compatible layout cues, and DAR applies condition-dependent transformations to noise-affected regions. This evidence-first routing explains the stable recovery of interface content.

RERL's Game reconstructions remain near 18.7 dB PSNR and 0.54 LPIPS across the selected noise conditions. Added noise produces random local responses that can resemble meaningful screen boundaries, placing textured appearance and screen-related cues in competition as the reconstruction draws on the reflected observation. CDERL strengthens responses supported by encoder–decoder agreement, and DAR supplies a condition-dependent transformation for regions where noise dominates the feature state. The Figure, Screen, and Game tables therefore connect noise stability to the same evidence-transmission mechanism.

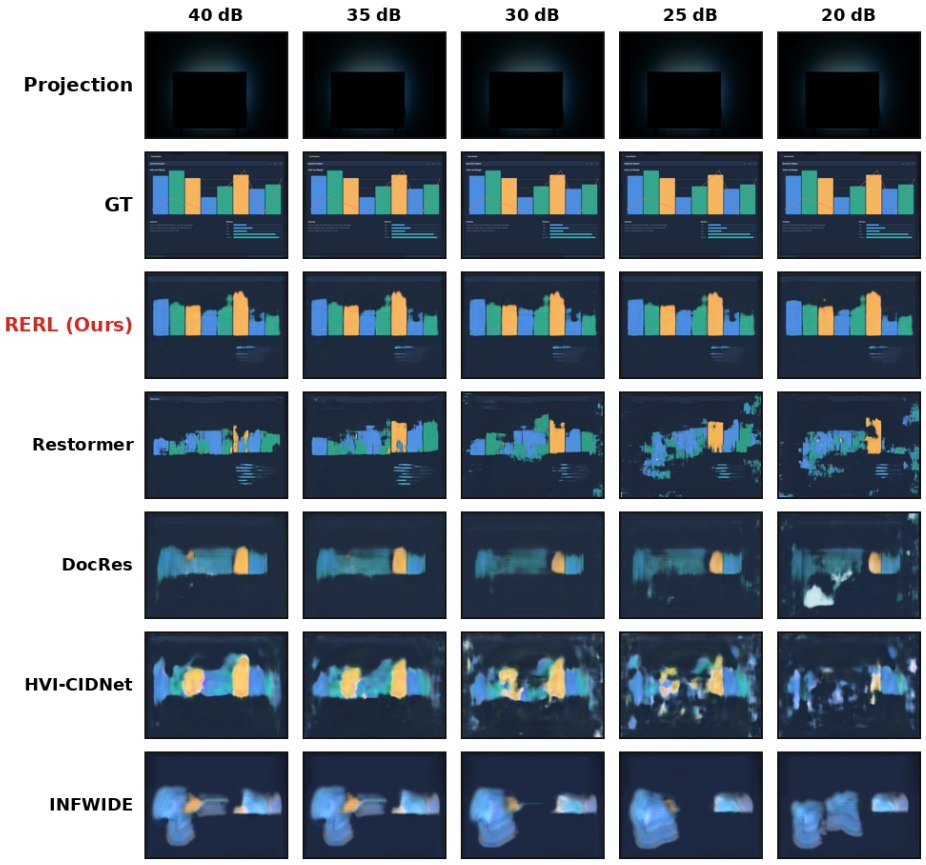

Figure 12: Qualitative recovery as the noise condition changes.

The noise visualization (Figure 12) shows how the recovered chart retains its major bars and colour ordering as the reflection becomes noisier. RERL keeps the dominant bar arrangement across the shown conditions, while comparison outputs contain more diffuse or irregular responses.

The visual result supports the table trends. Noise changes local fluctuations more directly than broad screen layout, and RERL uses the supported layout and colour organization as a stable recovery direction.

Method	Source	20.0			25.0			30.0			35.0			40.0		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Transformer-Based Restoration																
AST	CVPR'24	9.797	0.484	0.689	9.800	0.508	0.677	9.801	0.517	0.675	9.801	0.520	0.675	9.802	0.521	0.675
AdaIR	ICLR'25	8.619	0.330	0.782	8.774	0.344	0.771	8.905	0.351	0.762	9.441	0.359	0.745	10.413	0.375	0.719
Restormer	CVPR'22	8.831	0.341	0.765	8.935	0.368	0.753	8.831	0.388	0.750	8.777	0.412	0.736	8.939	0.450	0.702
SFHformer	ECCV'24	6.941	0.361	0.798	6.962	0.369	0.785	6.974	0.372	0.782	6.981	0.373	0.783	6.985	0.373	0.784
Uformer	CVPR'22	9.335	0.486	0.691	9.336	0.509	0.680	9.332	0.518	0.677	9.333	0.521	0.677	9.333	0.522	0.677
DehazeFormer	TIP'23	9.920	0.522	0.673	9.935	0.524	0.672	9.942	0.525	0.671	9.946	0.526	0.671	9.948	0.526	0.671
Retinexformer	ICCV'23	10.254	0.468	0.695	10.199	0.505	0.676	10.033	0.525	0.660	10.055	0.529	0.654	10.060	0.530	0.651
Convolution-Based Restoration																
ConvIR	TPAMI'24	9.499	0.492	0.659	9.509	0.514	0.655	9.512	0.522	0.656	9.514	0.525	0.656	9.515	0.526	0.656
IRNeXt	ICML'23	9.161	0.522	0.636	9.180	0.534	0.637	9.189	0.538	0.638	9.193	0.539	0.638	9.194	0.540	0.638
MIRNetv2	TPAMI'23	9.872	0.528	0.675	9.891	0.530	0.671	9.905	0.530	0.670	9.912	0.530	0.670	9.915	0.530	0.670
DnCNN	TIP'17	6.754	0.260	0.765	6.240	0.288	0.785	5.542	0.335	0.777	5.510	0.343	0.773	5.487	0.344	0.772
U-Net	MICCAI'15	9.924	0.483	0.696	10.096	0.501	0.688	10.114	0.507	0.685	10.147	0.509	0.685	10.173	0.509	0.685
pix2pix	CVPR'17	8.713	0.256	0.792	8.967	0.269	0.774	8.609	0.288	0.769	8.198	0.297	0.776	7.675	0.316	0.783
DeepSN-Net	TPAMI'25	6.797	0.324	0.825	6.790	0.345	0.796	6.789	0.354	0.786	6.788	0.356	0.783	6.788	0.357	0.782
Hybrid / Degradation-Specific Restoration																
DEA-Net	TIP'24	9.713	0.517	0.691	9.795	0.529	0.665	9.800	0.531	0.657	9.798	0.531	0.655	9.799	0.531	0.655
DExNet	TPAMI'25	10.003	0.511	0.677	10.058	0.527	0.659	10.029	0.531	0.654	10.003	0.532	0.654	9.985	0.532	0.654
DarkIR	CVPR'25	5.393	0.185	0.890	5.389	0.190	0.876	5.387	0.192	0.871	5.387	0.192	0.870	5.387	0.193	0.869
DocRes	CVPR'24	8.114	0.369	0.854	8.200	0.364	0.830	8.519	0.359	0.795	9.094	0.357	0.758	10.243	0.371	0.714
HVI-CIDNet	CVPR'25	10.176	0.347	0.792	10.297	0.378	0.784	10.516	0.397	0.782	10.683	0.407	0.777	11.028	0.425	0.755
INFWIDE	TIP'23	7.488	0.345	0.856	7.582	0.364	0.803	8.148	0.435	0.753	8.844	0.483	0.718	9.022	0.509	0.695
Reflection-Evidence Routed Representation Learning																
RERL	-	18.694	0.634	0.547	18.726	0.663	0.539	18.722	0.674	0.539	18.726	0.678	0.540	18.730	0.679	0.540

Table 15: Game noise comparison at representative SNR levels (dB). Red text and shading highlight the best row; blue text and shading highlight the selected rows within their respective method groups.

F.3 Distance and Viewpoint

Method	Source	1.0			1.25			1.5			2.0			3.0		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Transformer-Based Restoration																
AST	CVPR'24	19.391	0.705	0.467	19.207	0.697	0.476	18.807	0.689	0.486	18.957	0.691	0.482	16.874	0.647	0.531
AdaIR	ICLR'25	21.714	0.762	0.387	20.601	0.736	0.420	19.586	0.699	0.447	18.217	0.625	0.507	16.416	0.544	0.579
Restormer	CVPR'22	22.601	0.777	0.345	22.241	0.760	0.372	20.504	0.712	0.423	14.562	0.512	0.634	13.373	0.473	0.665
SFHformer	ECCV'24	8.817	0.394	0.807	8.822	0.396	0.809	8.820	0.396	0.810	8.819	0.395	0.812	8.806	0.392	0.815
Uformer	CVPR'22	19.204	0.707	0.464	18.882	0.699	0.473	18.673	0.695	0.478	18.716	0.696	0.476	17.752	0.674	0.498
DehazeFormer	TIP'23	19.366	0.693	0.479	19.209	0.690	0.484	18.909	0.683	0.490	18.529	0.673	0.499	15.003	0.574	0.594
Retinexformer	ICCV'23	19.567	0.673	0.455	17.740	0.618	0.499	17.559	0.611	0.510	17.439	0.608	0.508	10.170	0.371	0.701
Convolution-Based Restoration																
ConvIR	TPAMI'24	20.271	0.728	0.449	20.162	0.717	0.464	19.995	0.711	0.469	20.078	0.713	0.467	19.088	0.680	0.494
IRNeXt	ICML'23	19.956	0.723	0.454	19.881	0.714	0.465	19.784	0.710	0.469	19.817	0.711	0.468	19.406	0.696	0.482
MIRNetv2	TPAMI'23	18.237	0.692	0.492	18.255	0.692	0.492	18.254	0.692	0.492	18.271	0.692	0.492	18.274	0.691	0.494
DnCNN	TIP'17	8.883	0.347	0.724	8.806	0.344	0.728	8.799	0.344	0.728	8.798	0.344	0.729	8.805	0.338	0.731
U-Net	MICCAI'15	9.184	0.240	0.806	9.184	0.241	0.808	9.186	0.241	0.808	9.182	0.241	0.807	9.189	0.239	0.806
pix2pix	CVPR'17	10.324	0.482	0.664	10.300	0.479	0.667	10.288	0.479	0.669	10.282	0.478	0.672	10.161	0.470	0.693
DeepSN-Net	TPAMI'25	10.489	0.295	0.779	10.503	0.294	0.775	10.502	0.293	0.774	10.500	0.293	0.773	10.486	0.288	0.776
Hybrid / Degradation-Specific Restoration																
DEA-Net	TIP'24	20.651	0.750	0.424	20.583	0.748	0.428	20.461	0.746	0.431	20.461	0.746	0.431	16.358	0.653	0.532
DExNet	TPAMI'25	20.005	0.732	0.438	19.962	0.730	0.440	19.926	0.729	0.441	19.918	0.729	0.442	19.577	0.721	0.450
DarkIR	CVPR'25	9.221	0.175	0.867	9.218	0.176	0.870	9.218	0.175	0.867	9.218	0.175	0.871	9.222	0.175	0.877
DocRes	CVPR'24	21.145	0.749	0.400	20.442	0.734	0.416	19.394	0.696	0.447	16.689	0.591	0.523	15.469	0.555	0.561
HVI-CIDNet	CVPR'25	19.692	0.718	0.472	17.406	0.650	0.540	16.603	0.616	0.562	12.293	0.458	0.693	12.033	0.444	0.704
INFWIDE	TIP'23	14.618	0.545	0.640	10.342	0.444	0.713	9.470	0.402	0.739	10.194	0.437	0.716	8.908	0.375	0.763
Reflection-Evidence Routed Representation Learning																
RERL	-	26.699	0.868	0.189	26.599	0.865	0.192	26.561	0.864	0.192	26.447	0.863	0.193	26.130	0.857	0.199

Table 16: Screen recovery at representative relative-distance factors. Red text and shading highlight the best row; blue text and shading highlight the selected rows within their respective method groups.

The distance experiment tests reflected-evidence recovery under input transformations indexed by relative-distance factors. Table 16 reports the relative factors 1.0, 1.25, 1.5, 2.0, and 3.0. The yaw-tilt comparison averages whole-image metrics across Screen, Game, and Figure; Table 17 reports the sign-symmetric -5, -2, 0, $+2$, and $+5$ degree settings.

Method	Source	Tilt-Y -5 deg			Tilt-Y -2 deg			Tilt-Y 0 deg			Tilt-Y +2 deg			Tilt-Y +5 deg		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Transformer-Based Restoration																
AST	CVPR'24	12.878	0.566	0.624	12.584	0.561	0.631	13.081	0.581	0.612	12.584	0.561	0.631	12.878	0.566	0.624
AdaIR	ICLR'25	13.420	0.590	0.584	13.317	0.590	0.586	13.909	0.607	0.567	13.317	0.590	0.586	13.420	0.590	0.584
Restormer	CVPR'22	14.112	0.618	0.544	14.032	0.617	0.546	14.217	0.624	0.532	14.032	0.617	0.546	14.112	0.618	0.544
SFHformer	ECCV'24	7.718	0.416	0.785	7.720	0.416	0.786	7.714	0.416	0.785	7.720	0.416	0.786	7.718	0.416	0.785
Uformer	CVPR'22	12.638	0.572	0.613	12.541	0.571	0.614	12.687	0.577	0.610	12.541	0.571	0.614	12.638	0.572	0.613
DehazeFormer	TIP'23	11.223	0.541	0.668	12.582	0.581	0.633	13.220	0.597	0.613	12.582	0.581	0.633	11.223	0.541	0.668
Retinexformer	ICCV'23	12.878	0.579	0.600	12.747	0.578	0.603	13.134	0.587	0.592	12.747	0.578	0.603	12.878	0.579	0.600
Convolution-Based Restoration																
ConvIR	TPAMI'24	12.820	0.596	0.609	12.868	0.602	0.608	13.137	0.611	0.599	12.868	0.602	0.608	12.820	0.596	0.609
IRNeXt	ICML'23	12.637	0.591	0.602	12.643	0.592	0.602	12.871	0.601	0.593	12.643	0.592	0.602	12.637	0.591	0.602
MIRNetv2	TPAMI'23	12.676	0.598	0.626	12.698	0.599	0.625	12.700	0.599	0.624	12.698	0.599	0.625	12.676	0.598	0.626
DnCNN	TIP'17	6.985	0.383	0.733	6.979	0.384	0.734	6.990	0.385	0.733	6.979	0.384	0.734	6.985	0.383	0.733
U-Net	MICCAI'15	10.378	0.450	0.693	10.384	0.450	0.693	10.381	0.451	0.694	10.384	0.450	0.693	10.378	0.450	0.693
pix2pix	CVPR'17	8.352	0.445	0.701	8.357	0.445	0.701	8.357	0.447	0.698	8.357	0.445	0.701	8.352	0.445	0.701
DeepSN-Net	TPAMI'25	8.082	0.349	0.765	8.086	0.349	0.764	8.093	0.352	0.767	8.086	0.349	0.764	8.082	0.349	0.765
Hybrid / Degradation-Specific Restoration																
DEA-Net	TIP'24	12.733	0.590	0.641	12.450	0.583	0.651	13.438	0.620	0.594	12.450	0.583	0.651	12.733	0.590	0.641
DExNet	TPAMI'25	13.023	0.611	0.598	13.188	0.614	0.595	13.214	0.616	0.593	13.188	0.614	0.595	13.023	0.611	0.598
DarkIR	CVPR'25	6.571	0.186	0.865	6.577	0.186	0.867	6.573	0.186	0.864	6.577	0.186	0.867	6.571	0.186	0.865
DocRes	CVPR'24	13.299	0.594	0.591	13.743	0.604	0.584	13.993	0.617	0.568	13.743	0.604	0.584	13.299	0.594	0.591
HVI-CIDNet	CVPR'25	12.746	0.588	0.623	12.639	0.583	0.629	13.314	0.610	0.598	12.639	0.583	0.629	12.746	0.588	0.623
INFWIDE	TIP'23	9.004	0.472	0.738	9.544	0.490	0.728	10.983	0.548	0.669	9.544	0.490	0.728	9.004	0.472	0.738
Reflection-Evidence Routed Representation Learning																
RERL	-	15.036	0.658	0.484	15.141	0.659	0.482	15.293	0.667	0.479	15.141	0.659	0.482	15.036	0.658	0.484

Table 17: Yaw-tilt recovery, averaged across Screen, Game, and Figure using whole-image metrics, including at 0°. The transform depends on yaw-tilt magnitude, so opposite signs yield the same projection. Red highlights the best row; blue highlights the selected rows within their respective method groups.

RERL retains the strongest Screen result from factor 1.0 through factor 3.0, changing from 26.699 to 26.130 PSNR, while Restormer changes from 22.601 to 13.373. Increasing distance attenuates and spreads the reflection spot, reducing local contrast and weakening screen contours; generic restoration paths lose correspondence as the evidence changes scale and strength.

CDERL preserves traces that remain mutually supported, and DAR adjusts its expert mixture and spatial guidance as attenuation and spreading alter feature reliability. The gradual RERL trend therefore reflects evidence organization followed by suitable transformation paths.

RERL is strongest at each selected yaw-tilt setting, with the three-group mean PSNR ranging from 15.036 to 15.293 dB. The simulated projection changes the spatial distribution of the reflection, altering the positions of luminance and contour responses. The comparison thus examines recovery as screen-related structure is redistributed across the observed image.

CDERL compares compatible structure across the altered observation, and DAR processes the resulting local evidence condition through an adapted transformation mixture. RERL thus preserves layout and colour relationships under the simulated projection changes.

The distance visualization (Figure 13) shows the hidden screen layout as the reflection spot changes scale and concentration. RERL retains the large interface region and its spatial arrangement across the shown factors; comparison outputs develop stronger border artifacts and surface-related responses at larger distance.

The figure makes the quantitative distance trend interpretable. The spot's local contrast changes with relative scaling, and CDERL keeps screen-supported traces available and DAR changes the recovery path where those traces become weaker.

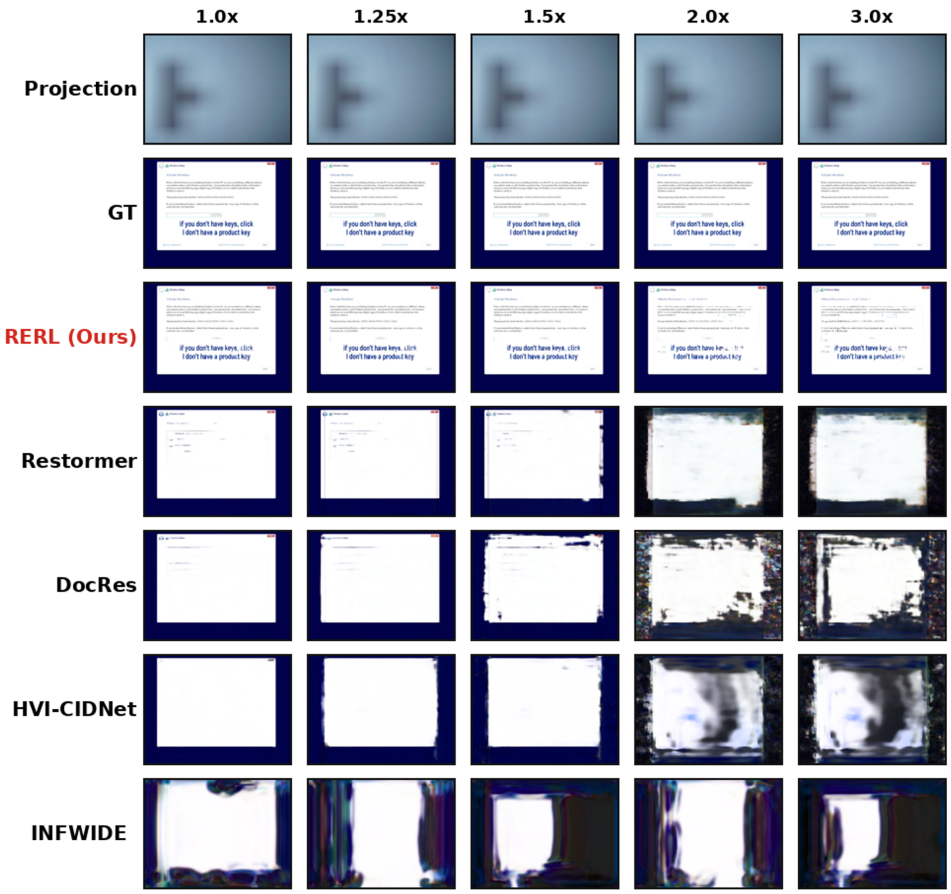

Figure 13: Qualitative recovery under the relative-distance conditions.

The angle visualization (Figure 14) shows recovered screen structure under small viewpoint changes. Across the shown angles, RERL preserves the main interface layout and colour regions; several comparison outputs emphasize border artifacts or surface appearance.

The visual stability agrees with Table 17. Viewpoint changes alter where evidence lands, and RERL maintains the structural relationships that remain jointly compatible with the reflected observation.

F.4 Blur and Viewpoint Curves

The blur experiment tracks contour organization as motion blur removes local detail. The plots (Figure 15, Figure 16) report LPIPS or the full metric curves over defocus and motion settings, while the camera-jitter curves evaluate viewpoint perturbations on Game and Screen.

The blur curve shows the recovery metrics as defocus and motion blur smooth local contours. RERL remains the leading plotted method across the settings, retaining a clear separation in PSNR, SSIM, and LPIPS. Blur removes high-frequency edge evidence and merges neighboring screen regions.

Broad luminance and layout cues remain available longer than fine boundaries. CDERL preserves these mutually supported cues, and DAR assigns suitable transformations where blur reduces local evidence.

The Game LPIPS plot (Figure 16) shows perceptual screen organization under varied blur types and orientations. RERL has the lowest plotted LPIPS values over the conditions, including defocus and motion settings. Blur weakens local contours and makes textured reflection responses compete with screen structure.

CDERL keeps the remaining screen-related organization available, and DAR changes the transformation path for blur-affected regions. The recovered content therefore stays perceptually closer to the target as local contours weaken.

The Screen camera-jitter curve (Figure 17) shows the recovery ranking under a small orbit perturbation. RERL remains above the comparison methods in PSNR and SSIM and below them in LPIPS across the shown settings. Camera jitter shifts the reflected spot and changes spatial correspondence while preserving portions of screen structure.

CDERL emphasizes compatible layout and contour relationships, and DAR adapts to the feature condition produced by the shifted spot. RERL therefore maintains the recovery ordering as viewpoint changes.

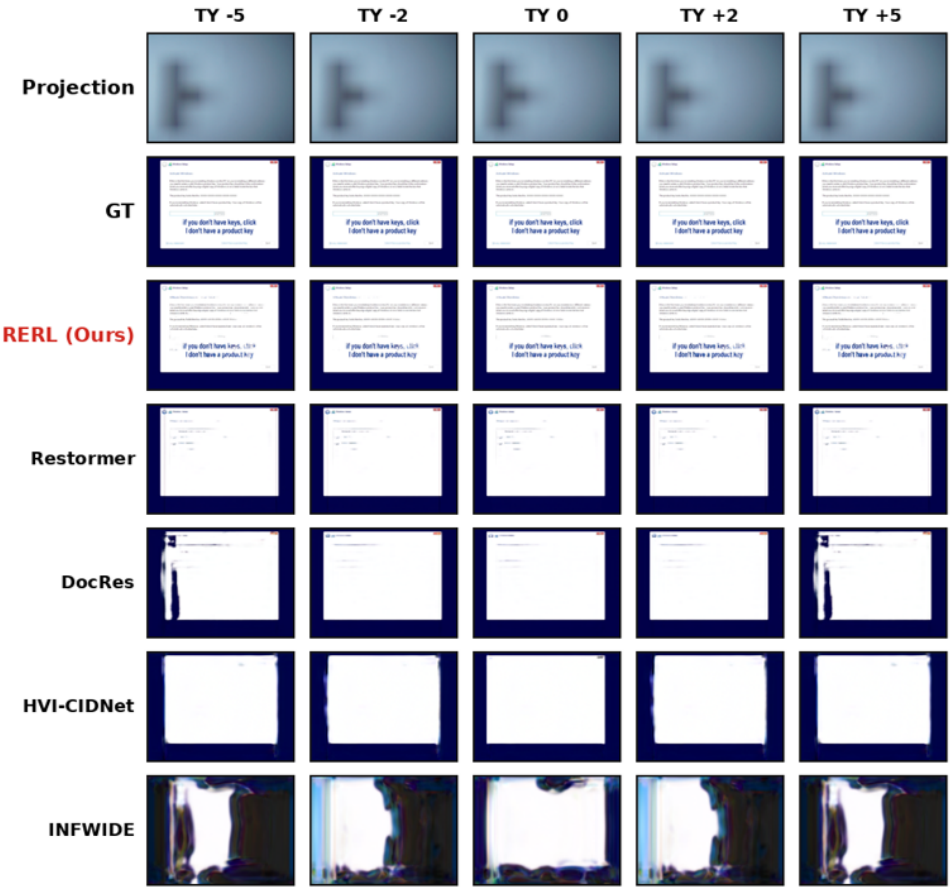

Figure 14: Qualitative recovery under the viewpoint-angle conditions.

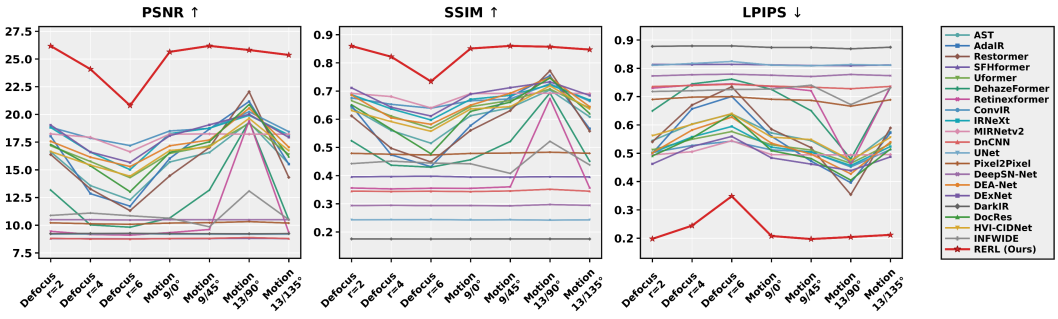

Figure 15: PSNR, SSIM, and LPIPS across the defocus and motion blur conditions.

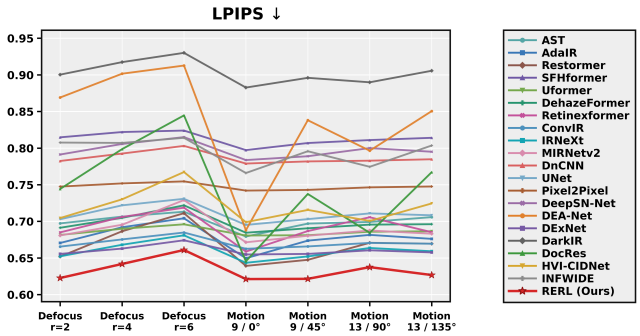

Figure 16: Game LPIPS across the defocus and motion-blur conditions.

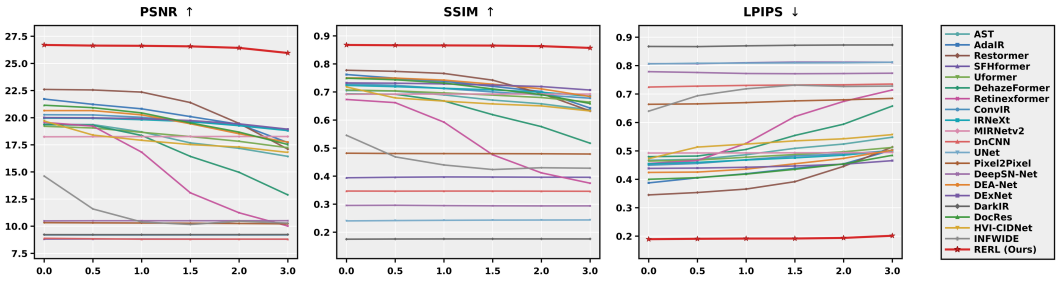

Figure 17: Screen PSNR, SSIM, and LPIPS under the camera jitter settings.

The Game camera-jitter plot (Figure 18) shows the same geometric change in textured content. RERL has the lowest plotted LPIPS across the orbit settings. The shifted reflection position changes local appearance and correspondence. CDERL represents compatible structure across changing local appearances, and DAR handles the altered degradation state. The perceptual result remains tied to screen organization as the reflection position changes.

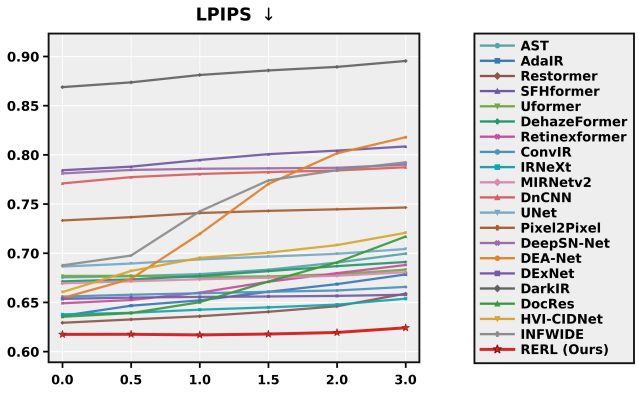

Figure 18: Game LPIPS under the camera-jitter settings.

F.5 Surface Material and Ambient Light

The surface experiment tracks screen-content recovery under surface-response transformations that vary diffusion and mixing. Table 18 uses the Smooth Paint, Matte Paper, White Plaster, Fine Wallpaper, and Glossy Tile columns. The illumination experiment tracks screen evidence under external light. Tables 19 and 20 use the full listed ambient-light and one-sided-intensity conditions. RERL leads across Smooth Paint, Matte Paper, White Plaster, Fine Wallpaper, and Glossy Tile, with the largest gap on the glossy condition. Material changes reshape the diffusion, smoothness, and mixing of reflected screen light, so generic restoration paths follow surface-specific appearance instead of stable screen organization.

Method	Source	Smooth Paint			Matte Paper			White Plaster			Fine Wallpaper			Glossy Tile		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Transformer-Based Restoration																
AST	CVPR'24	17.852	0.673	0.503	14.915	0.582	0.616	16.082	0.626	0.557	16.387	0.634	0.547	14.779	0.630	0.545
AdaIR	ICLR'25	15.066	0.504	0.621	15.073	0.499	0.643	15.360	0.508	0.628	15.253	0.505	0.630	13.807	0.490	0.645
Restormer	CVPR'22	16.308	0.604	0.523	12.804	0.465	0.688	13.481	0.496	0.643	13.954	0.514	0.623	12.055	0.449	0.718
SFHformer	ECCV'24	8.956	0.405	0.803	9.155	0.424	0.801	9.037	0.414	0.802	9.050	0.414	0.802	8.658	0.390	0.820
Uformer	CVPR'22	17.936	0.681	0.492	15.366	0.602	0.579	16.736	0.647	0.530	16.962	0.651	0.523	15.596	0.650	0.527
DehazeFormer	TIP'23	18.050	0.655	0.519	14.281	0.549	0.643	16.046	0.597	0.578	16.299	0.602	0.573	16.278	0.617	0.551
Retinexformer	ICCV'23	18.823	0.647	0.479	9.782	0.362	0.709	14.529	0.502	0.601	16.207	0.553	0.561	9.624	0.358	0.738
Convolution-Based Restoration																
ConvIR	TPAMI'24	18.947	0.689	0.477	14.684	0.565	0.611	16.791	0.627	0.531	16.891	0.630	0.527	15.779	0.612	0.549
IRNeXt	ICML'23	19.105	0.696	0.475	15.159	0.583	0.591	17.787	0.649	0.512	17.672	0.645	0.516	17.154	0.654	0.514
MIRNetv2	TPAMI'23	18.060	0.680	0.505	14.728	0.582	0.611	17.168	0.643	0.529	17.318	0.648	0.526	15.446	0.605	0.565
DnCNN	TIP'17	8.834	0.353	0.725	8.453	0.356	0.739	8.602	0.354	0.734	8.654	0.355	0.732	8.470	0.326	0.759
U-Net	MICCAI'15	9.283	0.249	0.809	9.387	0.272	0.802	9.345	0.261	0.808	9.326	0.258	0.808	9.146	0.232	0.796
pix2pix	CVPR'17	9.773	0.455	0.680	9.502	0.436	0.704	9.573	0.440	0.694	9.602	0.445	0.691	9.957	0.464	0.668
DeepSN-Net	TPAMI'25	10.360	0.295	0.781	9.889	0.287	0.789	10.150	0.293	0.783	10.221	0.294	0.783	10.037	0.284	0.784
Hybrid / Degradation-Specific Restoration																
DEA-Net	TIP'24	20.297	0.742	0.436	16.689	0.663	0.526	18.666	0.711	0.475	19.040	0.717	0.469	18.423	0.703	0.477
DExNet	TPAMI'25	19.156	0.704	0.462	14.337	0.536	0.645	16.984	0.620	0.533	17.405	0.632	0.520	17.263	0.650	0.510
DarkIR	CVPR'25	9.187	0.175	0.873	9.095	0.169	0.890	9.153	0.174	0.880	9.150	0.174	0.877	9.208	0.161	0.867
DocRes	CVPR'24	16.395	0.583	0.535	13.315	0.470	0.669	14.327	0.512	0.611	15.136	0.528	0.592	13.795	0.489	0.631
HVI-CIDNet	CVPR'25	11.647	0.463	0.713	10.356	0.404	0.773	11.043	0.421	0.744	10.982	0.424	0.749	12.622	0.490	0.683
INFWIDE	TIP'23	10.921	0.452	0.712	10.326	0.420	0.754	9.648	0.411	0.743	10.470	0.434	0.730	11.116	0.438	0.728
Reflection-Evidence Routed Representation Learning																
RERL	-	24.973	0.841	0.220	19.541	0.678	0.393	22.264	0.774	0.295	22.414	0.777	0.293	23.382	0.804	0.258

Table 18: Screen recovery under surface-response transformations representing the listed materials. Red text and shading highlight the best row; blue text and shading highlight the selected rows within their respective method groups.

Method	Source	Left Soft			Right Soft			Top Warm			Bottom Cool			Diag Mixed			Spot Glare		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Transformer-Based Restoration																			
AST	CVPR'24	13.329	0.572	0.595	12.725	0.564	0.594	11.281	0.504	0.658	10.861	0.506	0.662	11.428	0.508	0.650	9.439	0.428	0.726
AdaIR	ICLR'25	11.084	0.363	0.749	11.457	0.370	0.751	10.437	0.328	0.785	10.609	0.335	0.780	11.110	0.355	0.762	9.721	0.304	0.815
Restormer	CVPR'22	10.796	0.388	0.771	11.182	0.400	0.773	10.403	0.371	0.799	10.108	0.372	0.789	10.529	0.380	0.781	9.706	0.357	0.837
SFHformer	ECCV'24	8.056	0.353	0.824	8.713	0.392	0.824	8.285	0.366	0.829	8.125	0.362	0.826	8.489	0.383	0.822	8.041	0.348	0.829
Uformer	CVPR'22	14.682	0.608	0.548	13.194	0.592	0.569	11.866	0.540	0.616	12.383	0.556	0.604	12.716	0.569	0.592	10.014	0.442	0.689
DehazeFormer	TIP'23	13.778	0.560	0.617	14.247	0.572	0.601	11.412	0.506	0.683	11.771	0.519	0.669	12.688	0.540	0.642	9.876	0.474	0.721
Retinexformer	ICCV'23	8.823	0.336	0.763	9.281	0.346	0.748	8.308	0.319	0.767	9.017	0.349	0.750	9.183	0.348	0.750	8.701	0.357	0.765
Convolution-Based Restoration																			
ConvIR	TPAMI'24	13.087	0.536	0.617	13.481	0.548	0.606	11.484	0.505	0.682	11.624	0.503	0.668	12.200	0.515	0.646	9.861	0.476	0.726
IRNeXt	ICML'23	13.629	0.547	0.610	14.297	0.567	0.585	11.749	0.504	0.674	12.138	0.512	0.661	12.809	0.530	0.633	9.722	0.476	0.719
MIRNetv2	TPAMI'23	12.576	0.533	0.631	13.022	0.533	0.625	10.981	0.506	0.670	11.221	0.508	0.662	11.793	0.513	0.652	9.139	0.468	0.710
DnCNN	TIP'17	8.456	0.353	0.753	8.477	0.356	0.757	7.905	0.347	0.757	7.785	0.344	0.766	8.069	0.350	0.764	7.591	0.348	0.782
U-Net	MICCAI'15	9.036	0.240	0.813	9.307	0.234	0.807	9.107	0.242	0.805	9.160	0.238	0.798	9.301	0.236	0.801	9.027	0.247	0.800
pix2pix	CVPR'17	9.711	0.447	0.681	9.720	0.447	0.679	9.308	0.438	0.688	9.476	0.444	0.677	9.603	0.446	0.678	9.078	0.425	0.695
DeepSN-Net	TPAMI'25	9.385	0.277	0.792	9.162	0.270	0.798	8.711	0.269	0.807	8.829	0.268	0.796	8.918	0.267	0.797	8.134	0.261	0.802
Hybrid / Degradation-Specific Restoration																			
DEA-Net	TIP'24	16.892	0.651	0.532	15.150	0.619	0.577	13.594	0.578	0.632	13.563	0.582	0.619	14.220	0.601	0.602	11.189	0.516	0.695
DExNet	TPAMI'25	15.796	0.609	0.541	15.766	0.601	0.555	13.360	0.517	0.626	13.400	0.541	0.611	13.945	0.548	0.602	10.862	0.461	0.698
DarkIR	CVPR'25	9.132	0.139	0.889	9.224	0.138	0.892	9.030	0.125	0.906	9.146	0.124	0.891	9.217	0.124	0.896	9.082	0.104	0.922
DocRes	CVPR'24	11.766	0.394	0.711	12.002	0.403	0.712	10.771	0.349	0.773	10.555	0.347	0.774	11.239	0.359	0.756	9.777	0.335	0.809
HVI-CIDNet	CVPR'25	10.468	0.403	0.756	10.594	0.410	0.744	10.165	0.401	0.771	11.137	0.418	0.751	10.980	0.424	0.739	9.384	0.401	0.781
INFWIDE	TIP'23	10.035	0.414	0.738	11.425	0.453	0.717	10.046	0.399	0.762	10.512	0.420	0.746	10.597	0.420	0.744	9.559	0.396	0.763
Reflection-Evidence Routed Representation Learning																			
RERL	-	20.240	0.708	0.345	20.230	0.714	0.337	16.238	0.592	0.463	15.813	0.582	0.473	17.971	0.641	0.407	11.980	0.444	0.636

Table 19: Screen recovery under representative ambient-light conditions. Red text and shading identify RERL; blue text and shading identify the highlighted comparison methods.

CDERL retains luminance, colour, and contour traces that remain compatible with the screen, and DAR processes the changed local condition through an adapted route. The metric gains support screen recovery across the tested surface-response conditions.

Directional illumination, colour bias, and local glare introduce different interference patterns into the wall reflection. RERL achieves the highest PSNR and lowest LPIPS across all six ambient-light conditions, showing a consistent advantage in recovering screen content from these altered observations. External illumination changes the local contrast of screen-related responses.

Method	Source	0			6			12			18			24			30		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Transformer-Based Restoration																			
AST	CVPR'24	19.391	0.705	0.467	15.878	0.631	0.537	13.329	0.572	0.595	11.858	0.525	0.643	10.644	0.495	0.679	9.888	0.479	0.698
AdaIR	ICLR'25	21.714	0.762	0.387	14.165	0.479	0.657	11.084	0.363	0.749	10.677	0.340	0.774	10.320	0.339	0.796	10.028	0.346	0.813
Restormer	CVPR'22	22.601	0.777	0.345	12.367	0.440	0.708	10.796	0.388	0.771	10.261	0.370	0.801	10.088	0.363	0.812	9.931	0.356	0.826
SFHformer	ECCV'24	8.817	0.394	0.807	8.402	0.373	0.815	8.056	0.353	0.824	7.796	0.333	0.833	7.577	0.315	0.839	7.388	0.298	0.844
Uformer	CVPR'22	19.204	0.707	0.464	18.218	0.689	0.483	14.682	0.608	0.548	12.587	0.558	0.600	11.421	0.522	0.640	10.397	0.494	0.673
DehazeFormer	TIP'23	19.366	0.693	0.479	17.212	0.641	0.523	13.778	0.560	0.617	11.804	0.517	0.673	10.518	0.484	0.707	9.417	0.468	0.729
Retinexformer	ICCV'23	19.567	0.673	0.455	9.083	0.341	0.755	8.823	0.336	0.763	8.454	0.337	0.766	8.466	0.362	0.751	8.369	0.375	0.752
Convolution-Based Restoration																			
ConvIR	TPAMI'24	20.271	0.728	0.449	16.390	0.611	0.524	13.087	0.536	0.617	11.405	0.501	0.675	10.606	0.486	0.708	9.985	0.474	0.726
IRNeXt	ICML'23	19.956	0.723	0.454	17.097	0.624	0.521	13.629	0.547	0.610	11.553	0.499	0.672	10.366	0.478	0.700	9.435	0.465	0.718
MIRNetv2	TPAMI'23	18.237	0.692	0.492	16.674	0.626	0.526	12.576	0.533	0.631	10.880	0.499	0.673	9.833	0.479	0.698	9.052	0.465	0.714
DnCNN	TIP'17	8.883	0.347	0.724	8.724	0.352	0.736	8.456	0.353	0.753	7.715	0.338	0.764	7.139	0.340	0.768	7.187	0.336	0.772
U-Net	MICCAI'15	9.184	0.240	0.806	9.150	0.239	0.808	9.036	0.240	0.813	8.927	0.243	0.807	8.782	0.247	0.801	8.584	0.248	0.794
pix2pix	CVPR'17	10.324	0.482	0.664	9.875	0.462	0.676	9.711	0.447	0.681	9.463	0.444	0.685	9.115	0.444	0.697	8.656	0.441	0.711
DeepSN-Net	TPAMI'25	10.489	0.295	0.779	9.931	0.286	0.786	9.385	0.277	0.792	8.919	0.272	0.797	8.600	0.269	0.798	8.342	0.268	0.799
Hybrid / Degradation-Specific Restoration																			
DEA-Net	TIP'24	20.651	0.750	0.424	19.306	0.715	0.461	16.892	0.651	0.532	14.593	0.597	0.595	12.824	0.547	0.649	11.383	0.505	0.685
DExNet	TPAMI'25	20.005	0.732	0.438	18.684	0.694	0.471	15.796	0.609	0.541	13.127	0.530	0.611	11.689	0.487	0.661	10.759	0.469	0.691
DarkIR	CVPR'25	9.221	0.175	0.867	9.189	0.149	0.879	9.132	0.139	0.889	9.054	0.130	0.898	8.951	0.123	0.907	8.824	0.116	0.915
DocRes	CVPR'24	21.145	0.749	0.400	15.074	0.533	0.591	11.766	0.394	0.711	10.506	0.351	0.759	9.940	0.336	0.783	9.892	0.336	0.797
HVI-CIDNet	CVPR'25	19.692	0.718	0.472	11.543	0.429	0.733	10.468	0.403	0.756	9.958	0.394	0.763	9.501	0.390	0.772	9.129	0.391	0.775
INFWIDE	TIP'23	14.618	0.545	0.640	11.221	0.457	0.709	10.035	0.414	0.738	9.414	0.393	0.752	9.311	0.390	0.757	9.066	0.386	0.766
Reflection-Evidence Routed Representation Learning																			
RERL	-	26.699	0.868	0.189	22.855	0.789	0.268	20.240	0.708	0.345	15.649	0.570	0.494	12.510	0.469	0.609	10.431	0.409	0.693

Table 20: Screen recovery as one-sided ambient-light intensity increases. Red text and shading identify RERL; blue text and shading identify the highlighted comparison methods.

CDERL favors traces supported by the reflected evidence and decoder state, and DAR varies processing where external gradients make the local feature condition less reliable. The main screen organization remains recoverable across the tested lighting conditions.

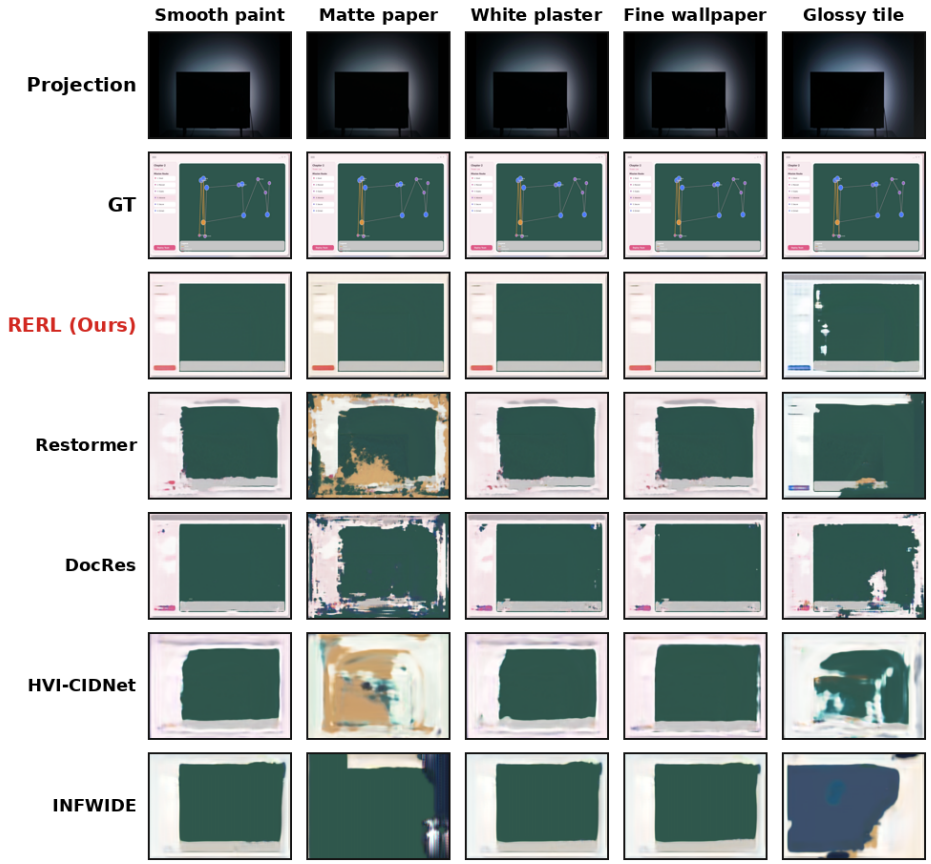

Figure 19: Qualitative recovery across the reflecting-surface materials.

One-sided illumination changes the reflection unevenly across the image. At intensity 18, RERL achieves 15.649 dB PSNR and 0.494 LPIPS, compared with 14.593 dB and 0.595 for DEA-Net, the strongest alternative in both metrics at this setting. The comparison shows recovery from an observation with spatially uneven illumination.

This experiment directly addresses the purpose of spatial guidance: adapting reconstruction where unrelated illumination changes local evidence. CDERL organizes reflected cues, and DAR combines an observation-dependent expert mixture with a spatial correction that can vary across the illuminated reflection.

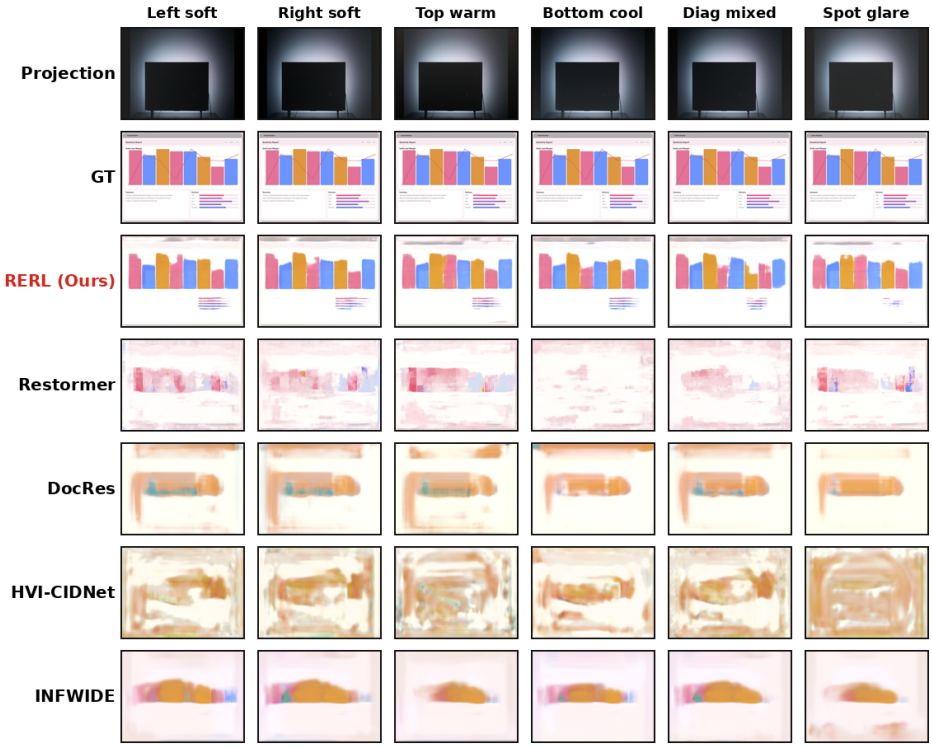

Figure 20: Qualitative recovery under different ambient-light directions and colour conditions.

The material visualization (Figure 19) shows the recovered screen as the reflection spot changes smoothness, texture, and local contrast. RERL preserves the main interface structure across the shown surfaces; comparison outputs show stronger surface artifacts in the more difficult material conditions.

The visual result explains the material table through evidence organization. The screen contribution changes appearance with the surface, and CDERL retains compatible luminance and contour responses and DAR adjusts the transformation path to their altered condition.

The ambient-light visualization (Figure 20) shows recovered screen organization under external illumination. RERL retains the dominant screen regions across left, right, warm, cool, mixed, and glare conditions; comparison methods more often follow the altered surface brightness.

Ambient light changes both the reflection appearance and the interference pattern. CDERL uses feature agreement to organize luminance, colour, and structural cues, and DAR supplies spatially varying compensation where the external gradient dominates.

The one-sided-light visualization (Figure 21) shows recovery as additional illumination increases from zero to a strong level. RERL keeps the main scene organization through the intensity sequence, while other outputs acquire more irregular surface and glare responses.

The intensity trend matches Table 20. As external light weakens local screen evidence, CDERL retains the remaining supported organization and DAR combines image-level expert reweighting with spatially guided correction in the affected regions.

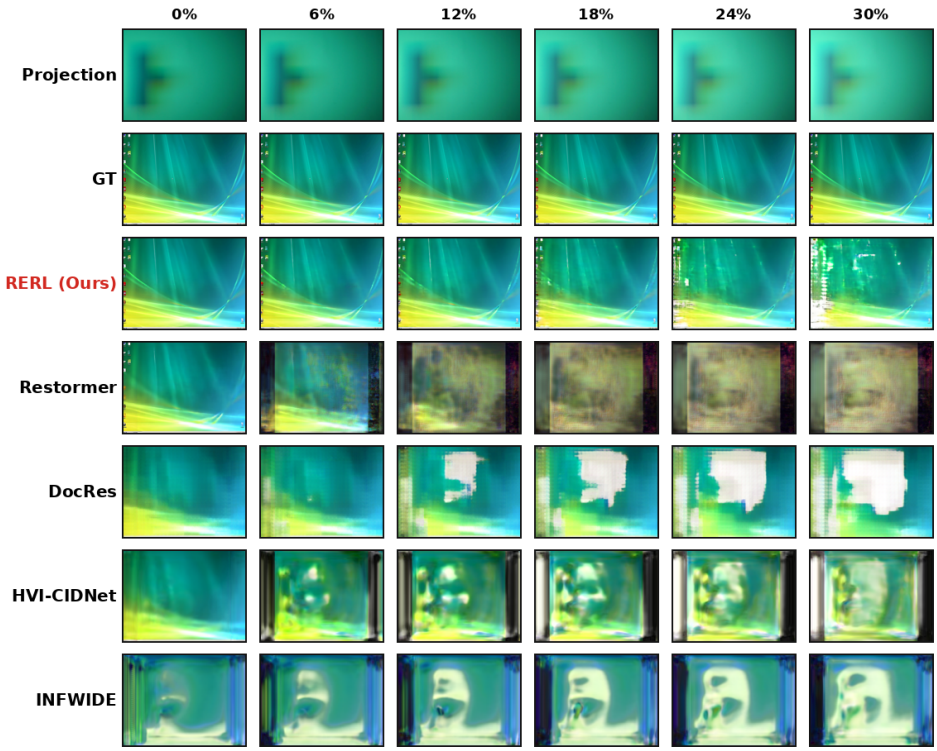

Figure 21: Qualitative recovery as one-sided ambient-light intensity increases.

F.6 Wall Reflectance and Surround-View Curves

The observation-condition results extend recovery analysis to simulated wall-reflectance settings of 35%–85% and orbit-angle changes. These curves examine how recovery follows changes in surface response and observation geometry.

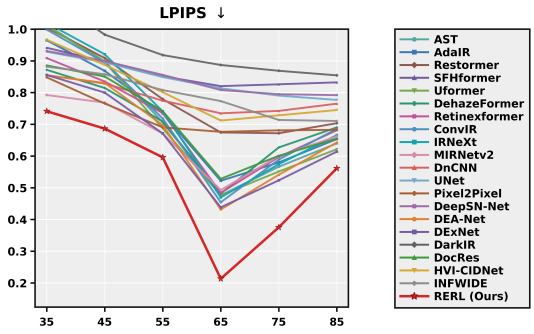

Figure 22: Screen LPIPS across simulated wall-reflectance settings from 35% to 85%.

The wall-reflectance plot (Figure 22) shows perceptual recovery as the reflected optical response changes. RERL has the lowest plotted LPIPS across the conditions, retaining a substantial perceptual separation from the comparison methods.

The result is consistent with the evidence interpretation. Changes in the reflected response modify how local traces appear, and RERL keeps screen-supported organization available before DAR adapts the transformation path.

Figure 23 extends the viewpoint analysis across additional orbit-angle settings. RERL remains the leading plotted method in PSNR and SSIM and the lowest in LPIPS as the position of the reflected observation changes.

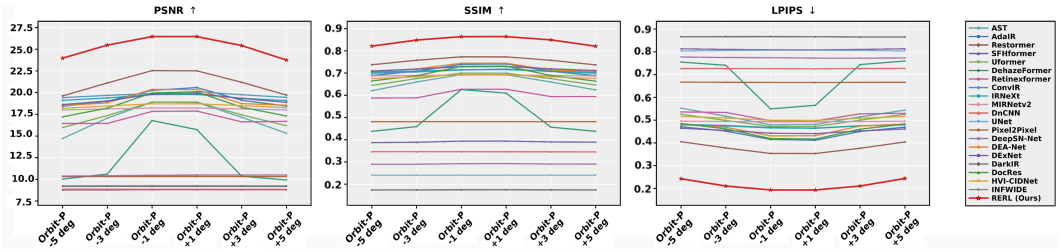

Figure 23: Screen PSNR, SSIM, and LPIPS across the surround-view angle conditions.

The orbit-angle sweep complements the tilt comparison by showing how recovery follows changes in the position of the reflected observation. Together, these comparisons examine reconstruction as the spatial arrangement of the observed reflection changes.

G Feature and Evidence Diagnostics

G.1 Feature Evolution

The feature-evolution visualization (Figure 24) traces where reflected evidence becomes organized in the RERL recovery path. It follows the input projection, encoder levels, middle representation, alignment, routing, token bridge, decoder, output projection, and refinement.

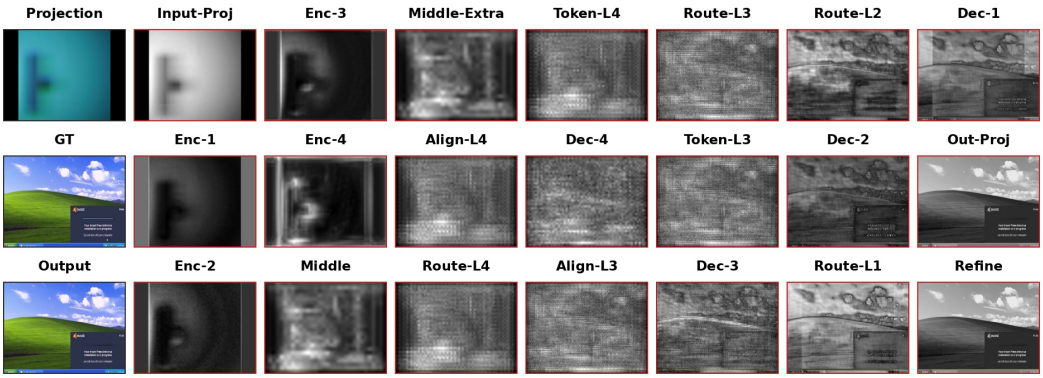

Figure 24: Feature evolution across the encoder, alignment, routing, token, decoder, and refinement stages.

Figure 24 follows the transition from the reflected light pattern to the recovered screen. Early states retain the appearance of the observation, while the final projection and refinement produce the screen's regional layout. The intermediate stages show where feature organization changes along this recovery path, connecting the observation, internal representations, and reconstructed content.

The figure is a mechanism diagnostic. It complements the quantitative ablations by showing that the evidence representation changes through the recovery path before the final screen readout.

G.2 Representation Comparisons

Figure 25 compares screen reconstructions produced by alternative representation modules in a common U-Net-style backbone. Direct skip, Non-local, CBAM-style, Window-attention, MDTA, and NAFBlock bridges are compared with CDERL, showing how representation choice changes the reconstructed screen content under the same reflected observations.

The same-backbone visualization shows how the representation changes the evidence available to the decoder. CDERL produces a more coherent screen-related organization across the previews, and generic representations retain more diffuse spot appearance or surface responses.
The image-level comparison complements Table 6. All alternatives use the same U-Net-style recovery backbone. Their reconstruction differences link representation choice to preservation of screen organization, including regional layout and colour relationships.

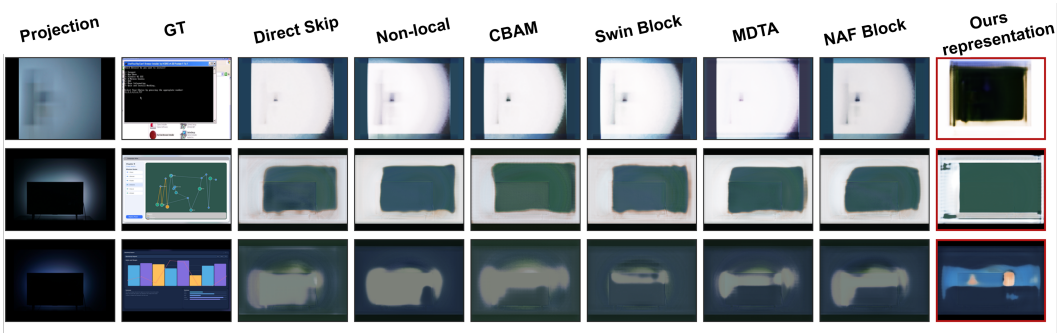

Figure 25: Screen reconstructions from Direct skip, Non-local, CBAM-style, Window-attention, MDTA, and NAFBlock bridges, and CDERL, within a common recovery backbone.

The second representation diagnostic (Figure 26) traces the difference across layers with a shared recovery structure. The panels show distinct feature organizations for the generic representation and CDERL from early encoder states through bottleneck and decoder states.

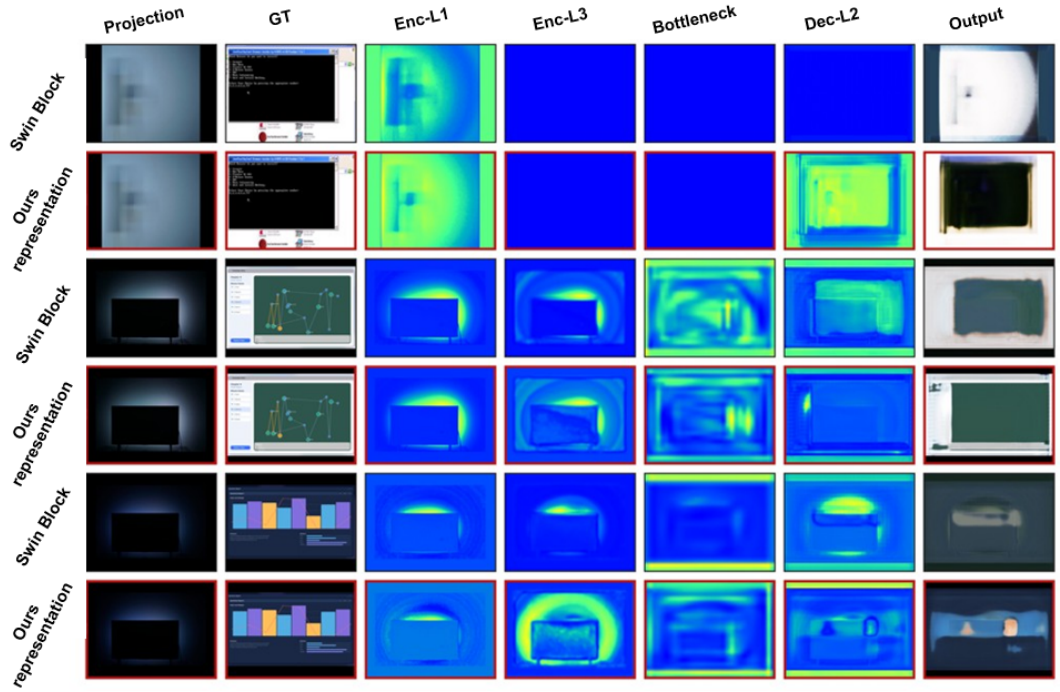

Figure 26: Layer features and final reconstructions for a window-attention bridge and CDERL in the same recovery backbone. Each pair of rows uses the same reflected observation and target screen.

The diagnostic supports a representation-level interpretation of the main result. CDERL changes how reflected evidence is carried through the network, so later decoding receives a different evidence representation before DAR applies its adaptive transformations.

H Additional Qualitative Comparisons

H.1 Evidence Organization Across Conditions

The condition visualizations show how RERL preserves screen structure across independent changes to the reflection spot. Figures 11, 12, 13, 14, 19, and 20 show this behavior for brightness, noise, distance, viewpoint, material, and ambient illumination.

Across these figures, RERL retains broad luminance distribution, major colour regions, and principal contours even when the input appearance changes substantially. The figures provide qualitative

support for the feasibility claim: reflected spots retain enough organized information for coarse hidden-screen recovery across the conditions.

The shared visual pattern follows the architecture. CDERL gives priority to mutually supported reflected evidence, and DAR adapts recovery transformations as attenuation, mixing, glare, noise, and viewpoint changes modify the local feature state.

H.2 Blur and Routing Diagnostics

The blur and routing visualizations show how RERL preserves screen organization as local contours weaken and expert paths change. Figures 15, 16, and 10 provide the visual and curve evidence.

The blur curves compare reconstruction metrics as the blur condition changes, while the routing visualization shows that the full DAR path produces more coherent screen regions than common allocation strategies. Together, these results connect quantitative reconstruction quality with visible differences under condition-dependent transformation.

I Concluding Evidence Chain

The appendix results support one coherent explanation of hidden-screen recovery. Propagation, diffuse reflection, surface mixing, and imaging weaken screen emission and retain coarse luminance, colour, spatial layout, and major contours. CDERL aligns encoder-side evidence with decoder states and transports reusable evidence prototypes through a learned spatial mask. DAR estimates spatial degradation and combines an image-level expert mixture with spatial guidance to reconstruct the organized evidence.

The controlled tables, same-backbone interference tests, component ablations, and perturbation results support this ordering. The qualitative figures show the same behavior in the image domain: RERL preserves screen-supported organization as the reflection spot changes in brightness, noise, distance, viewpoint, blur, material, ambient illumination, and optical condition. Together, these results establish the feasibility of recovering coarse hidden-screen brightness distribution, regional layout, and major contours from wall reflections captured by an ordinary camera through evidence organization and degradation-aware recovery.