FROM WALL GLOWS TO RECOVERABLE SCREENS: HIERARCHICAL RECOVERABILITY-GUIDED ACCRETIVE REPRESENTATION LEARNING FOR INDIRECT SCREEN RECONSTRUCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Can the content of a hidden screen be recovered from an ordinary camera image of the diffuse light it casts onto a visible wall? Spatial propagation and wall reflection mix screen regions and attenuate color, layout, structure, and detail by different amounts at different locations, so each wall glow retains a different set of usable cues. A uniform image-to-image mapping leaves the recoverability and reconstruction strength of these traces unresolved. We introduce Observability-Guided Accretive Representation Learning (OARL) to estimate which screen information the current wall glow supports and use that support to guide reconstruction. Calibrated Hierarchical Recoverability Representation Learning (CHRRL) constructs a shared evidence field, organizes it into color, layout, structure, and detail, and predicts a spatial gate for each type. During training, paired and visually similar nonmatching screens calibrate the global evidence token to distinguish the displayed content from similar alternatives. Observability-Stratified Accretive Reconstruction (OSAR) uses these estimates to continuously weight each stage's residual update at each location. Reconstruction follows the fixed color, layout, structure, and detail sequence, with later stages inheriting all preceding reconstruction states. At inference, the complete path uses the current wall-glow image. OARL achieves the best reported values in all 15 category–metric comparisons across five screen-content categories. Experiments under varied illumination, camera geometry, wall appearance, and image perturbations, including the seven-position camera-translation series, show recovery of screen colors, regional layouts, and principal structures. Component and controlled representation comparisons support recoverability-guided evidence selection and inherited reconstruction. Together, these results provide initial experimental evidence that wall-reflected light recorded by an ordinary camera can support recovery of the main content of a hidden screen under the tested indoor conditions.

1 INTRODUCTION

Indirect screen reconstruction asks how much of a hidden display's content can be recovered from the light it casts onto a visible wall. In our closed-room setting, an ordinary camera records an RGB photograph containing the screen back and the surrounding diffuse wall glow, while the display face is not directly visible. Spatial propagation and wall reflection spread, mix, and attenuate the emitted light, compressing colors, region layout, boundaries, text, icons, and textures into smooth variations. The task is to recover dense screen content from this passive setting within non-line-of-sight imaging (Faccio et al., 2020). We investigate how the cues retained in the observed wall glow can support the recovery of screen colors, regional layouts, and principal structures.

The wall acts as an uneven information channel whose retained cues vary by content and location. A broad color field can remain clear in one region, while a boundary or local intensity transition carries the stronger cue elsewhere. Conventional restoration networks (Zamir et al., 2022) usually encode the observation into a shared representation and apply a uniform decoding policy. Hidden-screen recovery calls for two linked decisions: estimate the recoverability of each information type at each

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.

location, then use that support to regulate successive reconstruction updates. Solving these two decisions allows strong traces to anchor the recovery of weakly supported text, icons, and textures.

Figure 1 illustrates the acquisition setting and the resulting inverse problem.

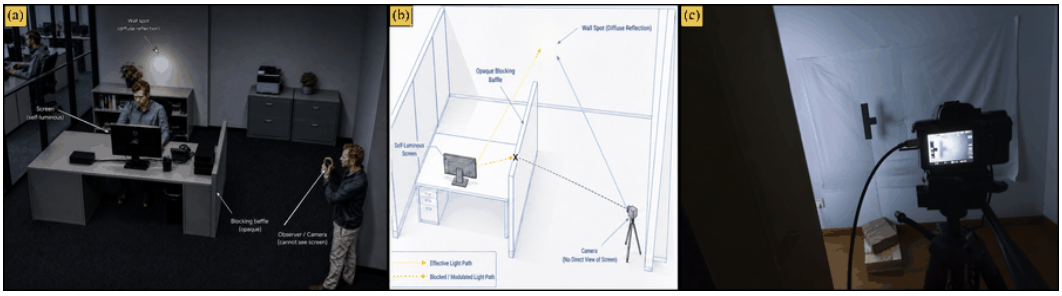


Figure 1: Task setup for reconstructing a hidden screen from its wall glow. The camera views the diffuse wall reflection while the screen remains hidden.

We propose Observability-Guided Accretive Representation Learning (OARL) to recover hidden-screen content by letting the unevenly retained wall-glow cues regulate successive reconstruction updates. OARL first separates color, layout, structure, and detail and learns a spatial gate for each information type. During training, paired screens and visually similar nonmatching screens calibrate the global evidence representation; the shared evidence field and joint objective pass this signal to the gate predictors. OSAR then accumulates reconstruction updates along a fixed color-to-layout-to-structure-to-detail sequence. This order supplies the dependency structure, while the four observation-dependent gates continuously scale each stage's residual at every location. OSAR carries all preceding reconstruction states into later stages, allowing reliable color and layout cues to guide the recovery of weaker structure and detail. At inference, the complete path uses the current wall-glow photograph. Together, the experiments provide initial evidence for recovering the main content of a hidden screen from wall-reflected light recorded by an ordinary camera under the tested indoor conditions.

The main contributions are threefold:

(1) We formulate indirect screen reconstruction as observation-dependent recovery: the model first determines which information types and spatial locations the current wall glow supports, then uses that support to regulate reconstruction.

(2) We introduce Calibrated Hierarchical Recoverability Representation Learning to produce type-specific spatial gates, and Observability-Stratified Accretive Reconstruction to translate those gates into controlled residual updates with explicit state inheritance.

(3) We demonstrate consistent reconstruction gains across five screen-content categories. Component, representation, learning-strategy, inheritance, and forward-path comparisons trace the two-part mechanism, while the camera-translation series tests its stability as the observed glow changes.

2 Related Work

OARL connects non-line-of-sight computational imaging with learned image restoration through a focused question: which dense screen information remains usable after diffuse transport, and how should it control reconstruction. OARL uses the recoverability of color, layout, structure, and detail at each location to determine successive screen state formation and reuse.

Non-Line-of-Sight Computational Imaging. Classical NLOS methods recover hidden scene properties by exploiting carefully structured indirect measurements (Faccio et al., 2020). Velten et al. (2012) use ultrafast time-of-flight measurements and computational inversion to recover hidden three-dimensional shape from diffuse reflections; O'Toole et al. (2018) combine confocal scanning with light-cone-transform inversion to recover hidden shape and albedo from time-resolved multiply scattered light.

Passive approaches show that cues from ordinary cameras can reveal hidden scenes. Saunders et al. (2019) recover an occluder and a partially hidden scene from ray-optics penumbrae in a conventional photograph. **?** assume a moving hidden scene, estimate its unknown occluder from diffuse-surface shadows, and recover a 2D RGB scene by regularized deconvolution. With paired wall-glow/screen supervision in our indoor display–wall setup, OARL uses spatial recoverability gates and inherited reconstruction to recover color regions, layouts, and boundaries from one wall-glow photograph of static screen content (Figure 3).

Architecture-Based Image Restoration. Modern restoration networks learn powerful mappings for visible images corrupted by noise, blur, rain, or low illumination. Restormer uses multi-Dconv head transposed attention for high-resolution denoising, deblurring, and deraining (Zamir et al., 2022), while Uformer develops attention-based feature processing (Wang et al., 2022). NAFNet uses multiplicative gating as an alternative to conventional nonlinear activations for denoising and deblurring (Chen et al., 2022).

Cui et al. (2023) use content-adaptive frequency decomposition and modulation for multiple restoration tasks, and Luo et al. (2024) use an image controller and degradation-conditioned features to adapt a unified model across degradations. These methods improve feature processing or task adaptation when the degraded image directly corresponds to the target. A wall glow mixes distant screen regions and preserves information types unevenly. OARL thus complements restoration architecture design with an explicit, spatially resolved estimate of usable color, layout, structure, and detail cues, carrying it into every progressive update.

3 METHOD

3.1 OVERALL ARCHITECTURE

OARL estimates which screen information the wall glow supports, then uses it to control where and how strongly reconstruction proceeds. The RGB input image results from spatial propagation and diffuse reflection, which mixes distant screen regions and suppresses many high-frequency traces. The surviving information varies by semantic type and position, so the model constructs four spatially resolved recoverability gates for color, layout, structure, and detail.

Figure2 shows the progressive reconstruction path. Calibrated Hierarchical Recoverability Representation Learning (CHRRL) predicts separate location-specific recoverability gates for trustworthy color, layout, structure, and detail evidence under indirect light transport. Observability-Stratified Accretive Reconstruction (OSAR) follows this fixed order, inheriting earlier states as context while observation-dependent gates control each update's spatial support and magnitude. OARL thus links observation-specific recoverability to evidence inheritance at every location. Let H and W denote input height and width, C evidence-feature width, and D auxiliary-token dimension.

The model first converts the observed wall glow $y = [y_{c,h,w}] \in \mathbb{R}^{3\times H\times W}$ into a preliminary screen representation $z = [z_{c,h,w}] \in \mathbb{R}^{3\times H\times W}$. This image-shaped estimate supplies a common spatial reference for subsequent support estimation and reconstruction. Its learned nonlinear spatial relation over $\mathcal{D} = [0, H] \times [0, W]$ is

$$z_{c,h,w} = \chi_c\left(\beta^{(0)}_{c,h,w} + \sum_{r=1}^{3}\iint_{\mathcal{D}} \mathsf{W}^{(0)}_{cr}(h, w; \xi, \eta) y_r(\xi, \eta)\,\mathrm{d}\xi\mathrm{d}\eta\right),\ c = 1, 2, 3. \tag{1}$$

Here, $\mathsf{W}^{(0)}$ describes the learned spatial association between the wall observation and the preliminary screen value, $\beta^{(0)}$ is the learned offset, and χ_c is the bounded channel readout. The integral becomes a finite spatial summation on the discrete grid. Computed entirely from y, z provides the screen-space reference against which CHRRL identifies recoverable information.

Evidence guides reconstruction: z anchors gates to screen coordinates, f retains local/contextual support, CHRRL produces four maps m, and OSAR scales colocated stage-specific residuals. Reliable traces initialize states, contextualizing weaker structure/detail. Wall-glow recoverability enters updates through semantic path. Progressive coupling distinguishes regional informativeness, limiting weak or mixed-evidence amplification. Fixed order; evidence governs update strength.

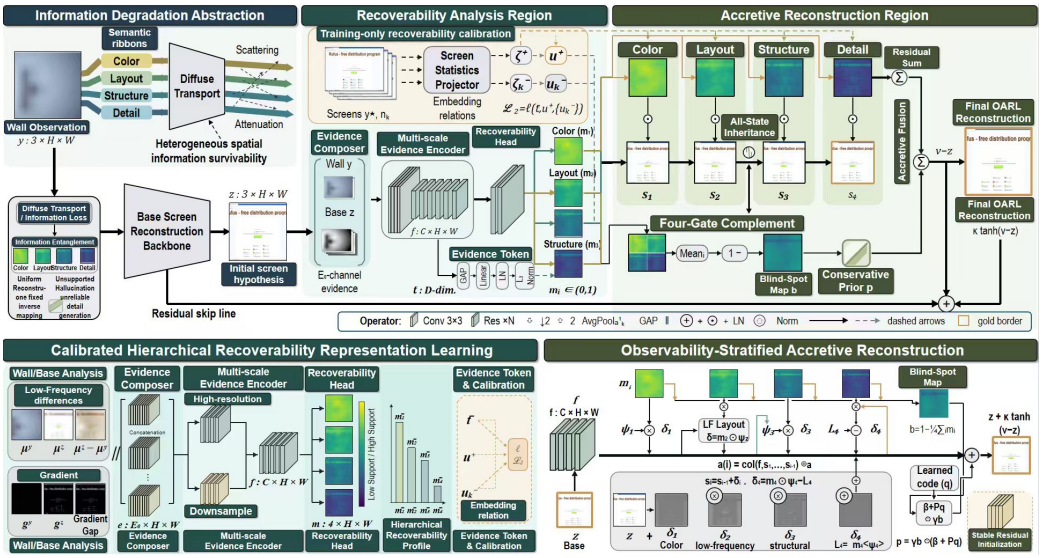


Figure 2: OARL first estimates observation-dependent recoverability gates for color, layout, structure, and detail, then uses them to regulate inherited residual updates along the accretive semantic backbone.

3.2 Calibrated Hierarchical Recoverability Representation Learning

CHRRL estimates the recoverability of different information types at each location after wall reflection. It compares the wall observation y with the preliminary estimate z, builds a shared representation from chromatic, contextual, and gradient cues, and predicts a distinct spatial gate for color, layout, structure, and detail. These gates regulate the evidence used by each reconstruction stage.

CHRRL first exposes broad organization and sharp changes through complementary local statistics. Both images are linearly rescaled between the prescribed endpoints and clipped to $[0, 1]$, producing $\breve{y}$ and $\breve{z}$. For either normalized image, local averaging summarizes coarse layout and gradient magnitude captures edge cues. With $\diamond \in \{y, z\}$ and local neighborhood $\mathcal{B}_\varrho(h, w)$, these statistics are

$$\mu^{\diamond}_{c,h,w} = \frac{1}{|\mathcal{B}_\varrho|} \iint_{\mathcal{B}_\varrho(h,w)} \breve{\diamond}_c(\xi, \eta)\, \mathrm{d}\xi \mathrm{d}\eta, \quad g^{\diamond}_{h,w} = \frac{1}{3} \sum_{c=1}^{3} \sqrt{\|\nabla \breve{\diamond}_{c,h,w}\|_2^2 + \varepsilon_g}. \tag{2}$$

These paired statistics align coarse organization with local transitions that remain informative after wall reflection. The evidence descriptor $e_{h,w}$ concatenates the normalized RGB values, local means, their signed and absolute differences, and both gradient summaries. Residual local and pooled contextual branches then form

$$f_{\nu,h,w} = \varphi_\nu \left(\beta^{(f)}_\nu + \sum_{r=1}^{E_0} A^{(f)}_{\nu r} e_{r,h,w} + \sum_{r=1}^{E_0} \iint_{\mathcal{B}_{\mathrm{loc}}} \mathsf{W}^{(\mathrm{loc})}_{\nu r} e_r \, \mathrm{d}\xi \, \mathrm{d}\eta + \sum_{r=1}^{E_0} \iint_{\mathcal{B}_{\mathrm{ctx}}} \mathsf{W}^{(\mathrm{ctx})}_{\nu r} \bar{e}_r \, \mathrm{d}\xi \, \mathrm{d}\eta \right). \tag{3}$$

Here, $\bar{e}_r$ is pooled evidence. The local branch preserves spatial changes, while the contextual branch evaluates them over a wider wall region, producing the evidence field $f \in \mathbb{R}^{C \times H \times W}$ from y and z. The evidence field gives every location a basis for estimating recoverable content.

Observability-conditioned attention connects spatially related traces before CHRRL estimates type-specific recoverability. We denote this operator by $\oplus_a$ and use the tag $\varpi \in \{\mathrm{g}, \mathrm{s}\}$ to distinguish gate extraction from stage inheritance. For a field $\mathfrak{r}$ over lattice $\mathcal{P}$, the attention output is

$$[\mathfrak{r} \oplus^{(i)}_{a,\varpi}]_{\nu,h,w} = \sum_{\alpha=1}^{A_{i,\varpi}} \sum_{(\xi,\eta)\in\mathcal{P}} \frac{\exp\left(D^{-1/2}_{i,\varpi} \mathfrak{r}^{\top}_{:,h,w} Q^{(i,\varpi,\alpha)} \mathfrak{r}_{:,\xi,\eta} \right)}{\sum_{(\zeta,\omega)\in\mathcal{P}} \exp\left(D^{-1/2}_{i,\varpi} \mathfrak{r}^{\top}_{:,h,w} Q^{(i,\varpi,\alpha)} \mathfrak{r}_{:,\zeta,\omega} \right)} \sum_{\mu=1}^{E_{i,\varpi}} V^{(i,\varpi,\alpha)}_{\nu\mu} \mathfrak{r}_{\mu,\xi,\eta}. \tag{4}$$

The compatibility matrix Q determines spatial information exchange, while V projects the weighted evidence. Here, $A_{i,\varpi}$ is the number of attention heads, $D_{i,\varpi}$ is the attention-scaling dimension, and $E_{i,\varpi}$ is the input feature width. Attention combines spatially related traces so each gate reflects local and contextual consistency. Gate extraction uses $E_{i,\mathrm{g}} = C$, whereas stage i uses $E_{i,\mathrm{s}} = C+3(i-1)$.

For color, layout, structure, and detail, the recoverability gate is

$$m_{i,h,w} = \left\{ 1 + \exp\left[-\beta_i^{(m)} - \sum_{\nu=1}^{C} \iint_{\mathcal{B}_m} \mathsf{W}_{i\nu}^{(m)}(\xi,\eta)[f \oplus_{a,\mathrm{g}}^{(i)}]_{\nu,h-\xi,w-\eta}\, \mathrm{d}\xi \mathrm{d}\eta \right] \right\}^{-1}. \tag{5}$$

The four maps form $m \in (0,1)^{4\times H\times W}$ and continuously control the corresponding semantic updates.

The comparison branch summarizes observation-derived evidence for correspondence supervision. Spatial averaging and affine projection first form the global token τ; layer normalization then produces t:

$$\tau_\upsilon = \beta_\upsilon^{(t)} + \sum_{\nu=1}^{C} W_{\upsilon\nu}^{(t)} \frac{1}{HW} \sum_{h=1}^{H}\sum_{w=1}^{W} f_{\nu,h,w}, \quad t_\upsilon = \theta_\upsilon^{(t)} + \lambda_\upsilon^{(t)} \frac{\tau_\upsilon - \overline{\tau}}{\sqrt{\varsigma_t^2 + \varepsilon_t}}. \tag{6}$$

During training, a shared screen-token projection maps the paired and nonmatching screens to auxiliary tokens, followed by stabilized normalization and correspondence loss:

$$\begin{aligned} &\zeta_\upsilon(a) = \beta_\upsilon^{(s)} + \sum_{j=1}^{J} W_{\upsilon j}^{(s)} \left[\frac{1}{HW} \sum_{h=1}^{H}\sum_{w=1}^{W} \vartheta_j(a_{:,h,w}) \right], \quad \zeta^+ = \zeta(y^\star), \quad \zeta_k^- = \zeta(n_k), \\ &\mathfrak{s}_\rho(a) = \frac{a}{\max\{\|a\|_2, \rho\}}, \quad \bar{t} = \mathfrak{s}_\rho(t), \quad u^+ = \mathfrak{s}_\rho(\zeta^+), \quad u_k^- = \mathfrak{s}_\rho(\zeta_k^-), \quad \mathcal{L}_2 = \ell\big(\bar{t}, u^+, \{u_k^-\}_{k=1}^K\big). \end{aligned} \tag{7}$$

This objective calibrates the global wall-derived token by distinguishing its paired screen from visually similar nonmatching screens. At inference, the wall glow produces the evidence field, recoverability gates, and reconstruction.

3.3 Observability-Stratified Accretive Reconstruction

OSAR uses the four recoverability gates to weight residual updates as it accumulates a shared screen reconstruction. Its semantic backbone follows color, layout, structure, and detail, so later stages inherit earlier states. The initial state is $s_{0,c,h,w} = z_{c,h,w}$.

At stage $i \in \{1,2,3,4\}$, the inherited descriptor stacks the shared feature with all previously formed states, and $\oplus_a$ converts it into the stage-specific attention field:

$$\mathfrak{r}_{h,w}^{(i)} = \mathrm{col}(f_{:,h,w}, s_{1,:,h,w}, \ldots, s_{i-1,:,h,w}), \quad \mathfrak{a}_{\nu,h,w}^{(i)} = [\mathfrak{r}^{(i)} \oplus_{a,\mathrm{s}}^{(i)}]_{\nu,h,w}. \tag{8}$$

For the first stage, the state portion is empty. Each later descriptor contains all preceding states, and the layout state additionally uses a pooled summary $\bar{\mathfrak{a}}^{(2)}$ over $\mathcal{B}_{\mathrm{pool}}$. The stage candidate is defined as

$$\psi_{i,c,h,w} = \varphi_{i,c}\left(\beta_c^{(i)} + \sum_\nu \iint_{\mathcal{B}_i} \mathsf{W}_{c\nu}^{(i)} \mathfrak{a}_\nu^{(i)}\, \mathrm{d}\xi \mathrm{d}\eta + \mathbf{1}_{\{i=2\}} \sum_\nu \iint_{\mathcal{B}_{\mathrm{lay}}} \mathsf{W}_{c\nu}^{(\mathrm{lay})} \bar{\mathfrak{a}}_\nu^{(2)}\, \mathrm{d}\xi \mathrm{d}\eta \right). \tag{9}$$

The first integral transforms the current-stage descriptor into a channel-adaptive candidate, while the indicator activates an additional pooled layout contribution only at the second stage. This design allows the layout stage to incorporate spatially aggregated evidence before the candidate is inherited by subsequent reconstruction stages. The cumulative state is then written as

$$s_{i,c,h,w} = z_{c,h,w} + \sum_{j=1}^{i} m_{j,h,w} \left[\psi_{j,c,h,w} - \mathbf{1}_{\{j=4\}} \langle \psi_{4,c} \rangle_{\mathcal{B}_\phi(h,w)} \right]. \tag{10}$$

Subtracting the detail candidate's local average emphasizes spatial variation beyond the low-frequency content established by earlier stages. The gate m_j scales each residual before it is accumulated into the inherited state. OSAR then regulates residual strength as inherited states accumulate.

The gate complement summarizes locations with lower aggregate recoverability and modulates the learned prior correction:

$$b_{h,w} = 1 - \frac{1}{4}\sum_{i=1}^{4} m_{i,h,w}, \quad p_{c,h,w} = \gamma b_{h,w}\left(\beta^{(p)}_{c,h,w} + \sum_{r=1}^{C} P^{(p)}_{c,h,w,r} q^{(p)}_{r}\right). \tag{11}$$

The learned correction supplies a prior-guided residual in weakly supported regions, with its contribution scaled continuously by the gate complement. The bounded output is

$$v_{c,h,w} = s_{4,c,h,w} + p_{c,h,w}, \quad \hat{y}_{c,h,w} = z_{c,h,w} + \kappa \tanh(v_{c,h,w} - z_{c,h,w}). \tag{12}$$

The hyperbolic tangent bounds the offset from z to $(-\kappa, \kappa)$, while $\hat{y} \in \mathbb{R}^{3\times H\times W}$ combines gated evidence inheritance with the learned correction.

Training jointly optimizes reconstruction fidelity and correspondence-aware evidence regularization:

$$\mathcal{L}_1 = d(\hat{y}, y^\star), \quad \mathcal{L} = \Delta(\mathcal{L}_1, \mathcal{L}_2). \tag{13}$$

Here, d measures reconstruction discrepancy, ℓ compares paired and nonmatching screens through the global evidence token, and Δ combines the two objectives to balance pixel-level reconstruction accuracy with correspondence-aware discrimination in the learned evidence representation.

4 Experiments and Results

We evaluate wall-glow screen recovery and recoverability-guided evidence inheritance on approximately 12,400 paired images from five categories: Figure, Screen, Game, OSLock, and Browser. We test whether wall glow supports recognizable screen content, then examine how recoverability estimation and inherited states contribute to reconstruction. Training uses paired samples; inference receives the current glow. We report PSNR, SSIM, and LPIPS against representative restoration methods under one reconstruction setting, with controlled sweeps over illumination, sensor noise, camera geometry, wall appearance, visibility, blur, and camera shake. Table 1 summarizes main results.

Table 1: Quantitative comparison on Figure, Screen, Game, OSLock, and Browser using PSNR, SSIM, and LPIPS.

Method	Source	Figure			Screen			Game			OSLock			Browser		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
CNN-based General Image Restoration																
UNet (Ronneberger et al., 2015)	MICCAI'15	11.815	0.566	0.667	9.184	0.240	0.806	10.285	0.471	0.758	12.180	0.487	0.762	7.793	0.339	0.793
DnCNN (Zhang et al., 2017)	TIP'17	14.204	0.574	0.664	8.883	0.347	0.724	8.738	0.427	0.726	11.818	0.479	0.722	8.650	0.440	0.693
NAFNet (Chen et al., 2022)	ECCV'22	17.729	0.648	0.622	21.617	0.759	0.379	18.160	0.659	0.576	15.242	0.562	0.664	17.028	0.666	0.461
DGUNet (Mou et al., 2022)	CVPR'22	18.038	0.695	0.623	18.962	0.696	0.482	17.888	0.660	0.595	16.547	0.624	0.651	16.933	0.666	0.500
DExNet (Huang et al., 2025)	TPAMI'25	18.602	0.721	0.621	20.005	0.732	0.438	17.629	0.659	0.595	15.518	0.608	0.652	16.903	0.668	0.484
ConvIR (Cui et al., 2024)	TPAMI'24	18.540	0.715	0.629	20.271	0.728	0.449	17.642	0.651	0.601	14.911	0.593	0.687	15.261	0.607	0.584
IRNeXt (Cui et al., 2023)	ICML'23	17.479	0.668	0.633	19.956	0.723	0.454	18.136	0.668	0.574	16.134	0.608	0.678	15.986	0.632	0.549
DeepSN-Net (Deng et al., 2025)	TPAMI'25	14.401	0.499	0.692	10.489	0.295	0.779	10.868	0.444	0.739	12.836	0.423	0.771	11.845	0.445	0.691
MIRNetv2 (Zamir et al., 2023)	TPAMI'23	18.340	0.713	0.654	18.237	0.692	0.492	17.679	0.656	0.612	15.211	0.604	0.658	16.431	0.659	0.513
Transformer/Hybrid General Image Restoration																
Restormer (Zamir et al., 2022)	CVPR'22	18.921	0.710	0.551	22.601	0.777	0.345	17.368	0.644	0.568	16.754	0.607	0.633	16.342	0.662	0.469
Uformer (Wang et al., 2022)	CVPR'22	17.586	0.619	0.626	19.204	0.707	0.464	17.547	0.647	0.614	15.149	0.565	0.674	16.616	0.659	0.504
SFHformer (Jiang et al., 2024)	ECCV'24	13.205	0.597	0.713	8.817	0.394	0.807	10.010	0.462	0.760	12.461	0.509	0.740	11.408	0.469	0.766
PromptIR (Potlapalli et al., 2023)	NeurIPS'23	18.757	0.691	0.620	21.564	0.755	0.380	17.784	0.660	0.571	15.342	0.594	0.637	16.816	0.670	0.462
AdaIR (Cui et al., 2025)	ICLR'25	17.930	0.669	0.623	21.714	0.762	0.387	17.436	0.639	0.575	15.168	0.573	0.646	16.725	0.661	0.468
AST (Zhou et al., 2024)	CVPR'24	17.891	0.639	0.632	19.391	0.705	0.467	17.456	0.645	0.616	15.072	0.537	0.692	16.243	0.654	0.512
Task-specific, All-in-One, and Generative Restoration																
OneRestore (Guo et al., 2024)	ECCV'24	12.529	0.586	0.704	10.409	0.464	0.714	10.313	0.506	0.747	12.237	0.522	0.721	8.986	0.439	0.748
DocRes (Zhang et al., 2024)	CVPR'24	18.772	0.718	0.611	21.145	0.749	0.400	17.528	0.647	0.579	16.357	0.609	0.647	16.237	0.653	0.474
DarkIR (Feijoo et al., 2025)	CVPR'25	4.185	0.038	0.929	9.221	0.175	0.867	4.465	0.041	0.936	8.787	0.095	0.804	5.047	0.054	0.915
Retinexformer (Cai et al., 2023)	ICCV'23	17.895	0.690	0.619	19.567	0.673	0.455	17.925	0.656	0.592	14.600	0.559	0.651	16.889	0.655	0.492
DehazeFormer (Song et al., 2023)	TIP'23	18.917	0.709	0.616	19.366	0.693	0.479	17.774	0.649	0.607	14.873	0.537	0.669	16.567	0.651	0.516
DEA-Net (Chen et al., 2024)	TIP'24	18.179	0.715	0.642	20.651	0.750	0.424	17.729	0.658	0.588	15.962	0.607	0.674	16.587	0.663	0.494
Pix2Pix (Isola et al., 2017)	CVPR'17	15.476	0.615	0.638	10.324	0.482	0.664	11.140	0.450	0.692	12.405	0.518	0.730	10.504	0.421	0.676
Ours (Full)	–	**19.141**	**0.732**	**0.489**	**25.558**	**0.852**	**0.221**	**18.423**	**0.699**	**0.536**	**16.918**	**0.642**	**0.627**	**18.209**	**0.715**	**0.430**

OARL leads all 15 category–metric comparisons in Table 1. Zamir et al.'s Restormer (CVPR 2022) uses an efficient Transformer to capture long-range pixel interactions for high-resolution image restoration (Zamir et al., 2022). On the present wall-glow benchmark, OARL reaches 25.558/0.852/0.221 PSNR/SSIM/LPIPS on Screen versus Restormer's 22.601/0.777/0.345. Compared with Restormer, OARL also increases OSLock PSNR from 16.754 to 16.918 dB. The lead spans all five content types, showing that the improvement is not tied to a specific screen category. Across Figure, Screen, Game, OSLock, and Browser, OARL consistently improves distortion-oriented metrics while simultaneously reducing perceptual error. Such agreement across PSNR, SSIM, and LPIPS suggests that the method preserves both global screen structure and fine visual details more reliably than existing methods.

4.1 Comparative Evaluation

Figure 3 compares the recovery of screen organization, semantic regions, layouts, icons, boundaries, and textures across the five content types. OARL uses CHRRL's type-specific support and OSAR's inherited states to guide this recovery.

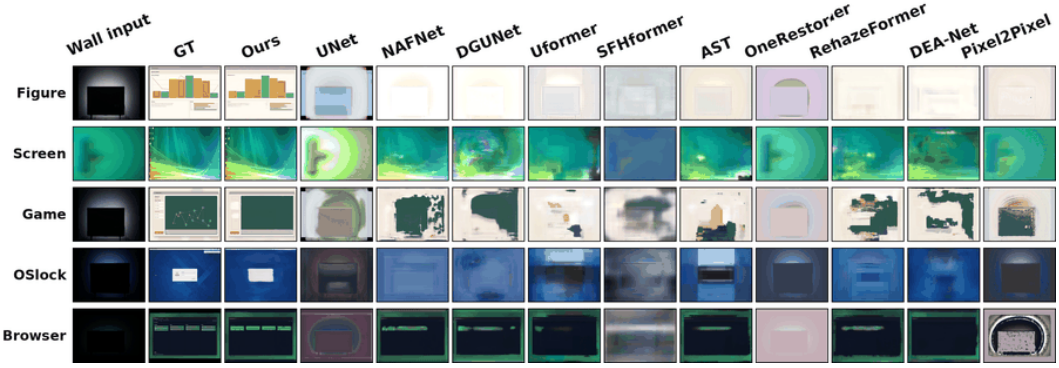


Figure 3: Qualitative comparison of OARL with representative restoration methods on different screen-content categories.

4.2 Ablation Study

The component ablation tests OARL's two decisions: estimating differentiated recoverability and carrying established evidence forward. Full OARL attains best or tied-best values in all 15 category–metric comparisons in Table 2 at the displayed precision. On Screen, Full reports 25.5575 / 0.8522 / 0.2214 PSNR / SSIM / LPIPS. The independent w/o HierRecover variant reports 23.8990 / 0.8267 / 0.2798, and the independent w/o Evidence variant reports 22.8692 / 0.7941 / 0.3402. Across the five categories, these independent comparisons support complementary contributions from CHRRL's differentiated recoverability and OSAR's evidence inheritance.

Table 2: Component ablation across five content categories. Each entry reports PSNR↑ / SSIM↑ / LPIPS↓.

Method	OSLock			Figure			Screen			Game			Browser		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
HierRecover Variants															
w/o HierRecover	16.73	0.6390	0.6360	18.91	0.7267	0.4993	23.90	0.8267	0.2798	18.08	0.6936	0.5758	17.99	0.7080	0.4619
HR → LiteGate	16.13	0.6313	0.6517	19.02	0.7287	0.6230	20.26	0.6326	0.5159	18.23	0.6672	0.6594	17.19	0.6359	0.5765
HR → BlurGate	16.25	0.6378	0.6512	19.09	0.7275	0.6658	19.14	0.6314	0.5216	18.29	0.6691	0.6602	17.22	0.6375	0.5796
HR → ConvNet	16.55	0.6328	0.6451	19.11	0.7252	0.6397	19.12	0.7105	0.4971	18.17	0.6863	0.6143	11.40	0.6921	0.5056
Evidence Inheritance Variants															
w/o Evidence	16.40	0.6324	0.6362	19.08	0.7244	0.5618	22.87	0.7941	0.3402	18.31	0.6898	0.5667	18.15	0.7033	0.4672
EI → CompactHead	16.67	0.6383	0.6589	19.09	0.7264	0.6764	16.58	0.5967	0.6046	18.12	0.6663	0.6785	16.76	0.6209	0.6456
EI → BlurHead	16.22	0.6405	0.6540	19.14	0.7260	0.6733	18.67	0.6260	0.5480	18.13	0.6656	0.6775	17.18	0.6298	0.5968
EI → ResNet	16.02	0.6249	0.6487	18.39	0.7321	0.6661	17.23	0.6159	0.5552	17.75	0.6670	0.6423	17.33	0.6648	0.5360
Full	**16.92**	**0.6420**	**0.6272**	**19.14**	**0.7321**	**0.4891**	**25.56**	**0.8522**	**0.2214**	**18.42**	**0.6991**	**0.5356**	**18.21**	**0.7153**	**0.4297**

The structural comparisons in Figure 4 show how OSAR turns established states into a stronger reconstruction base. Inheritance variants replace full state transfer with no inheritance, parallel inheritance, weighted inheritance, or local inheritance. Forward variants remove the path or adopt reverse, random, or swapped orders. Across five displayed content types, Full retains screen organization

and local content more consistently, showing that explicit state inheritance and the color-to-layout-to-structure-to-detail dependency help weaker cues build on recovered evidence.

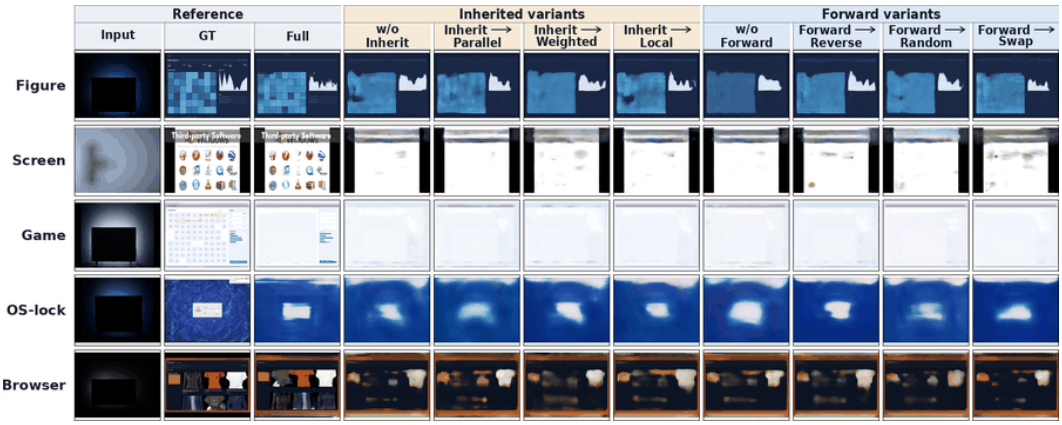


Figure 4: Structural comparison of the Full configuration with inherited-state and forward-path variants.

4.3 Camera-Translation Robustness

The camera-translation series shows OARL remains stable as viewpoint changes the observed glow. At the normal position, OARL reaches about 19.14 dB PSNR, 0.732 SSIM, and 0.489 LPIPS, and leads all three metrics at all seven positions in Figure 5. PSNR and SSIM peak near 0%, where LPIPS is lowest; the curves change smoothly toward endpoints, demonstrating stable Figure-category reconstruction throughout the translation range. For each changed glow, CHRRL recomputes recoverability gates while OSAR preserves the fixed evidence-inheritance path.

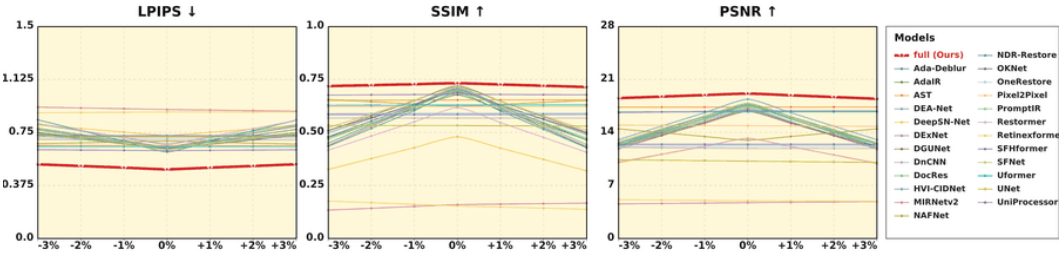


Figure 5: PSNR, SSIM, and LPIPS on the Figure category under camera z-translations from -3% to +3% at 1% increments

The visual results in Figure 6 reinforce the quantitative trend: OARL maintains recognizable chart organization throughout the -3% to +3% range and most closely matches the target near the normal position. OARL's reconstruction remains organized as the available wall-glow cues shift, whereas several alternatives remain dominated by the low-frequency glow appearance.

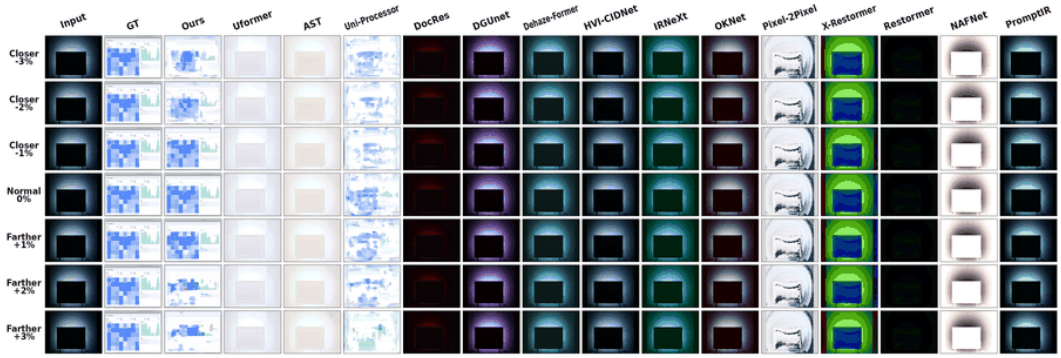


Figure 6: Qualitative reconstructions of one Figure-category example under seven camera z-translations

4.4 Mechanism Validation

The representation comparison evaluates whether CHRRL retains recovery-relevant screen evidence more effectively than the compared signal- or prior-guided representations. Under the same U-Net recovery architecture, OARL representation (Ours-Rep in Table 3) ranks first in all 15 category–metric comparisons. Its largest PSNR margins occur on Game, where it achieves 14.97 dB versus 10.28 dB for the next-best representation, and on Screen, where it reaches 13.08 dB versus 9.09 dB. The advantage also extends to OSLock, where OARL representation increases PSNR from 11.95 to 12.04 dB. Holding the recovery architecture fixed supports comparison of these representations within a shared reconstruction framework.

Table 3: Information-representation comparison using the same U-Net recovery architecture. Controls are defined in Appendix C.1.

Representation	Origin	Figure			Screen			Game			OSLock			Browser		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Signal-Based Representations																
RGB-Rep	This study	12.33	0.5772	0.7060	7.97	0.3060	0.8364	10.22	0.4764	0.7785	11.27	0.4539	0.7679	9.09	0.4198	0.7811
Edge-Rep	This study	12.53	0.5735	0.6712	8.34	0.4030	0.8079	10.28	0.4761	0.7225	11.73	0.4870	0.7551	9.19	0.4236	0.7335
Frequency-Rep	This study	12.12	0.5728	0.7166	8.81	0.3930	0.8520	9.89	0.4641	0.7743	11.48	0.4536	0.7777	9.16	0.4199	0.7604
Prior-Guided Representations																
Retinex-Rep	This study	11.72	0.4976	0.7133	9.09	0.4125	0.8281	10.10	0.4555	0.7608	11.91	0.4708	0.7891	8.76	0.3745	0.7964
Degradation-Rep	This study	12.21	0.5719	0.7144	8.84	0.4189	0.8100	10.14	0.4878	0.7833	11.50	0.4795	0.7618	10.36	0.4696	0.7161
Hybrid-Cue-Rep	This study	12.54	0.5718	0.6874	9.01	0.3928	0.8005	10.07	0.4276	0.7469	11.95	0.4707	0.7794	8.35	0.4211	0.7662
Ours-Rep	–	**14.63**	**0.5940**	**0.6459**	**13.08**	**0.4686**	**0.7245**	**14.97**	**0.5534**	**0.6901**	**12.04**	**0.4890**	**0.7481**	**12.63**	**0.4985**	**0.6726**

The learning-strategy comparison evaluates OARL's complete learning configuration against the listed alternatives. Under the same U-Net recovery architecture, OARL learning (Ours Learn. in Table 4) leads all 15 comparisons. On Screen, it achieves 13.77 dB PSNR, 0.5326 SSIM, and 0.6764 LPIPS, compared with the best alternative values of 12.66 dB, 0.4609, and 0.6965 for the respective metrics. The advantage remains consistent on OSLock, including 13.49 versus 13.46 dB in PSNR and 0.7465 versus 0.7511 in LPIPS. Together, the two comparisons examine how the representation supplies recovery-relevant evidence and how the learning configuration uses that evidence within the shared recovery architecture.

Table 4: Learning-strategy comparison using the same U-Net recovery architecture. Controls are defined in Appendix C.1.

Learning	Origin	Figure			Screen			Game			OSLock			Browser		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Supervision-Based Learning																
Direct Learn.	This study	18.41	0.6885	0.6254	12.66	0.4609	0.7311	16.60	0.6347	0.6335	13.42	0.5008	0.7835	15.44	0.6247	0.5713
Perceptual Learn.	This study	17.88	0.6499	0.5448	11.78	0.4141	0.6965	16.78	0.6017	0.5532	13.38	0.4896	0.7511	15.46	0.6080	0.5065
Deep Sup. Learn.	This study	18.15	0.6682	0.5625	12.15	0.4404	0.7258	16.75	0.6195	0.5849	13.45	0.5013	0.7711	15.48	0.6235	0.5329
Contrastive Learn.	This study	18.31	0.6602	0.5529	11.98	0.4226	0.7206	16.73	0.6120	0.5705	13.12	0.4706	0.7598	15.65	0.6101	0.5186
Distillation Learn.	This study	17.99	0.6729	0.5765	12.41	0.4491	0.7158	16.56	0.6187	0.5938	13.26	0.4865	0.7622	15.62	0.6206	0.5360
Adaptive Learning Strategies																
Curriculum Learn.	This study	18.41	0.6764	0.5589	12.28	0.4381	0.7175	16.68	0.6153	0.5750	13.46	0.5005	0.7628	15.43	0.6189	0.5284
Uncertainty Learn.	This study	18.47	0.6745	0.5543	12.10	0.4179	0.7231	16.70	0.6178	0.5814	13.40	0.4932	0.7651	15.23	0.6138	0.5225
Recurrent Learn.	This study	18.39	0.6803	0.5680	12.14	0.4355	0.7307	16.65	0.6193	0.5887	13.39	0.4830	0.7605	15.58	0.6184	0.5339
Task-Guided Learn.	This study	18.19	0.6851	0.5647	11.71	0.4426	0.7282	16.70	0.6247	0.5853	13.42	0.5163	0.7601	15.33	0.6214	0.5321
Ours Learn.	–	**18.61**	**0.7136**	**0.5368**	**13.77**	**0.5326**	**0.6764**	**17.01**	**0.6603**	**0.5470**	**13.49**	**0.5718**	**0.7465**	**16.10**	**0.6581**	**0.4947**

5 Conclusion

OARL recovers hidden-screen content from the diffuse glow on a visible wall. CHRRL learns type-specific spatial recoverability gates from shared evidence, with paired-versus-nonmatching global-token calibration during training. OSAR scales residual updates with observability-dependent gates, propagating reliable evidence through inherited states. OARL leads reported quantitative results across five categories. Component, representation, learning-strategy, state-inheritance, and reconstruction-order comparisons link gains to recoverability-guided selection and state reuse. Separate stress tests show lower wrong-target attraction for the enhanced OARL representation than alternatives under false-contour interference. Controlled illumination, sensor noise, camera geometry, wall appearance, visibility, blur, and camera shake tests demonstrate recovery across tested conditions. These findings provide initial experimental evidence for ordinary-camera recovery of hidden-screen content from wall-reflected light under tested indoor conditions.

REFERENCES

Daniele Faccio, Andreas Velten, and Gordon Wetzstein. Non-line-of-sight imaging. *Nature Reviews Physics*, 2(6):318–327, 2020. doi: 10.1038/s42254-020-0174-1.

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.

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

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

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.

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

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

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

Ziwei Luo, Fredrik K. Gustafsson, Zheng Zhao, Jens Sjolund, and Thomas B. Schon. Controlling vision-language models for multi-task image restoration. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=t3vnnLeajU.

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

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.

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

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.

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, 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, 2023. URL https://proceedings.mlr.press/v202/cui23d.html.

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.

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.

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

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

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.

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.

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

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.

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.

Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12470–12479, 2023. doi: 10.1109/ICCV51070.2023.01149.

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.

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.

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. 1125–1134, 2017. doi: 10.1109/CVPR.2017.632.

APPENDIX

Mechanism and Additional Evidence for Observability-Guided Accretive Representation Learning

This appendix provides the complete formulation and experimental analysis of OARL for passive, single-frame indoor screen reconstruction from an indirect wall glow. Section A describes the acquisition process, training and inference protocol, and dataset. Section B derives the two OARL modules. Section C evaluates photometric, geometric, surface, and motion robustness and examines the evaluated representation and reconstruction configurations. OARL denotes the complete reconstruction model. The controlled comparisons evaluate OARL representation and OARL learning configurations within a shared U-Net recovery architecture. The representation stress tests use an enhanced representation configuration, distinguished from the main representation comparison. In existing plot labels, 'Ours' or 'Full' identifies the proposed configuration within that experiment; it does not imply identical architectures or checkpoints across experiments.

A ADDITIONAL SCENE INFORMATION

This section specifies the acquisition setting, supervision protocol, and evaluation data underlying OARL. The camera records a single RGB image of the screen's diffuse wall reflection from a position with no direct view of the display. Light transport redistributes screen information before camera acquisition and preserves color, layout, structure, and detail unevenly across the wall. OARL models this spatially and semantically varying evidence through observation-dependent reconstruction.

A.1 WALL-GLOW FORMATION AND UNEVEN INFORMATION LOSS

The observed wall glow is formed through spatial propagation and diffuse reflection. A display faces a nearby wall and emits spatially varying RGB radiance determined by its current content. After propagating through the room, the emitted light reaches the wall, where reflection redistributes its energy over a broader spatial extent before the camera records it. The resulting single RGB wall-glow photograph contains the smooth intensity and chromatic variations used for reconstruction.

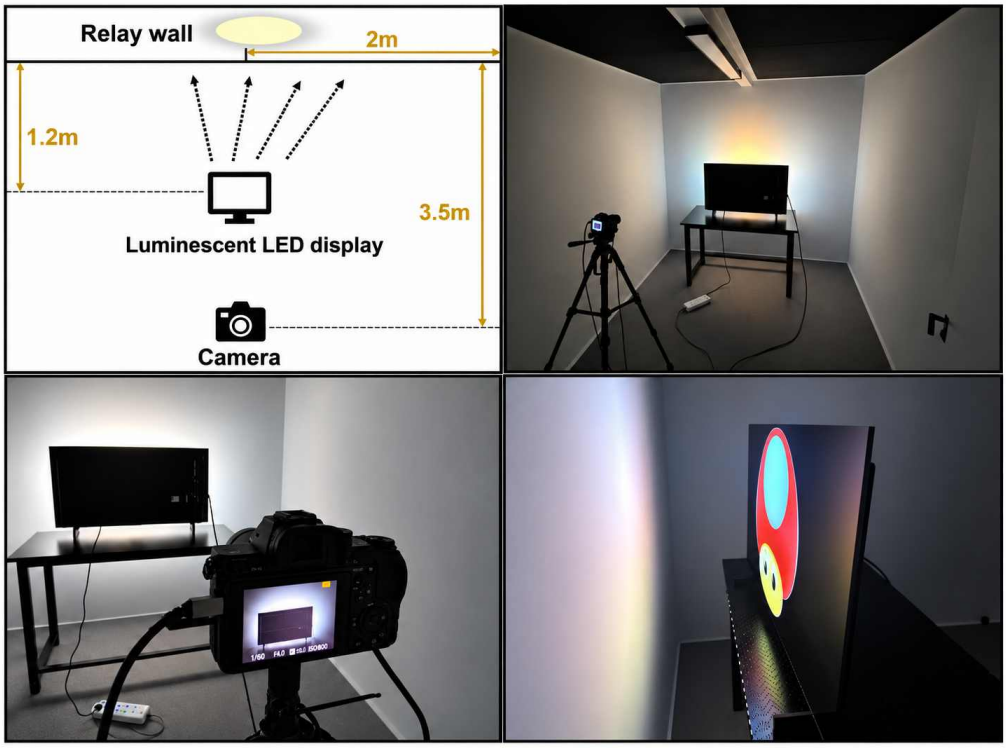
This transport process induces a many-to-many spatial relation between the hidden screen and the observed wall. Light from one screen location spreads over an extended wall region, and each wall measurement aggregates contributions from multiple screen regions. Spatial propagation and reflection consequently mix neighboring screen signals and attenuate their spatial variations before acquisition. Broad chromatic regions and coarse spatial organization retain measurable low-frequency variations, whereas spatial mixing weakens small boundaries, characters, icons, and local textures. The retained evidence therefore varies with both information type and spatial position.

The wall image is thus a nonuniform information field. Different wall locations provide different levels of support for color, layout, structure, and detail, and this support changes with the current observation. A region with a clear chromatic response may contain weak boundary or texture evidence, whereas a localized intensity transition can preserve structural information under a weak absolute color response. OARL captures these differences with four spatially resolved gates that regulate the corresponding reconstruction updates.

Room geometry, wall reflectance, display emission, camera response, and environmental illumination jointly determine the wall-glow field. OARL learns the resulting observation-to-screen relation and uses its gates to quantify the relative reconstruction support available at each location and semantic stage.

A.2 ACQUISITION SETUP AND DATASET COLLECTION

The dataset was collected in a real indoor reflection environment, as illustrated in Figure 1. A 60-inch self-emissive LED display was positioned facing a white diffuse-reflecting wall, with a screen-to-wall distance of approximately 0.5 m. Light emitted by the display directly illuminated the wall and produced a spatially distributed reflected glow whose intensity and chromatic pattern varied with the displayed content. This configuration provides a real passive indirect-imaging setting in



Supplementary Figure 1: Real-world acquisition setup used for constructing the paired wall-glow dataset. The figure shows the geometric relationship among the relay wall, self-emissive LED display, and camera, together with representative photographs of the actual indoor acquisition environment. The display faces the diffuse wall and produces a content-dependent reflected glow, while the camera is positioned toward the rear side of the display and records the screen-back silhouette together with the surrounding wall reflection.

which the screen information is observed through the wall-reflected light field rather than through direct frontal imaging of the display.

A Sony A7S II camera was used as the observation device and was positioned approximately along the direction facing the rear side of the display. Consequently, the recorded image does not directly observe the displayed screen content. Instead, each captured observation contains the silhouette of the screen back together with the reflected illumination distributed over the surrounding wall. The acquisition examples in Figure 1 further show the complete experimental environment, the camera-side observation configuration, and the spatial relationship between the illuminated wall and the display.

For each acquisition, the wall-glow observation was recorded together with the actual content shown on the display. The displayed contents include webpages, desktop interfaces, figures and charts, user-interface elements, and password-like character strings, providing screen images with substantially different spatial layouts, colors, structures, and fine-scale details. Each dataset sample therefore consists of a captured composite observation containing the screen-back region and wall glow, paired with the corresponding ground-truth screen image.

The reconstruction task is defined under a single-frame passive observation setting. During inference, only one captured composite image is provided to the model, and the objective is to recover the specific content displayed on the hidden screen. The paired screen image is used only as supervision during training and as the reference during quantitative evaluation. This acquisition protocol preserves the indirect nature of the task while providing explicit observation–target correspondence for learning the wall-glow-to-screen reconstruction mapping.

A.3 Training and Inference Contract

OARL uses paired wall-glow and screen images during optimization. The paired screen provides reconstruction supervision, and visually similar nonmatching screens provide the auxiliary comparison signal used by Calibrated Hierarchical Recoverability Representation Learning. This comparison teaches the global evidence representation to distinguish target-consistent evidence from visually similar alternatives.

At inference, a single RGB wall photograph drives the complete reconstruction path. OARL computes the preliminary image-shaped representation, shared evidence features, and four observation-dependent spatial gates from this photograph and the learned parameters. The reconstruction then follows the fixed color, layout, structure, and detail sequence, with the gates controlling the spatial support and magnitude of every update. Each later stage inherits the states computed from the same observation, preserving single-frame deployment.

The global evidence representation is also derived from the current wall photograph. During training, paired and nonmatching screen representations supervise this global token and thereby shape the shared evidence field and gate predictors. During deployment, the final reconstruction is computed directly from the wall glow and learned model parameters.

This division gives OARL a clear information contract: screen-level correspondence guides representation learning during optimization, whereas deployment requires a single wall-glow image.

A.4 Dataset Coverage and Evaluation

The reported evaluation considers five screen-content categories—**Figure**, **Screen**, **Game**, **OSLock**, and **Browser**—under the same single-wall-glow reconstruction protocol. The categories contain different forms of visual organization, including plots and thin graphics, dense desktop or application layouts, natural and rendered imagery, sparse operating-system interfaces, text, icons, and multi-region webpage structures. The reported collection contains approximately 1,250 Figure pairs, 3,800 Screen pairs, 1,250 Game pairs, 1,400 OSLock pairs, and 4,700 Browser pairs. Each pair consists of a wall-glow observation and its corresponding displayed screen, while only the wall-glow image is provided as input during inference.

Reconstruction fidelity is evaluated by comparing the predicted screen with its paired reference using peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), and learned perceptual image patch similarity (LPIPS). Higher PSNR and SSIM indicate closer pixel-level and structural agreement, respectively, whereas lower LPIPS indicates smaller perceptual feature distance. The same metrics and wall-glow-to-screen input–output contract are used across the five content categories, allowing the reported comparisons to examine reconstruction behavior over screen contents with different spatial organizations and frequency characteristics.

Together, the five categories cover plots and thin graphics, dense desktop and application layouts, natural and rendered imagery, sparse operating-system interfaces, text, icons, and multi-region webpages. The robustness experiments further vary illumination, sensor noise, camera geometry, wall appearance, visibility, blur, and camera shake, providing a broad evaluation of the observation-to-screen reconstruction process.

B Derivation of the Mechanism

This section derives the complete OARL computation from the indirect wall observation to the reconstructed screen. OARL first constructs an image-shaped preliminary estimate that provides a common screen-space reference. It then extracts observation-dependent evidence, separates that evidence into spatial gates for color, layout, structure, and detail, and uses the gates to regulate a fixed-order accretive reconstruction. All deployed quantities are computed from the current RGB wall-glow image and learned parameters, while paired and nonmatching screens supply supervision during training.

B.1 Observation Abstraction and Preliminary Recovery

Let $y \in \mathbb{R}^{3\times H\times W}$ denote the RGB wall-glow observation and $y^\star \in \mathbb{R}^{3\times H\times W}$ denote its paired screen image during training and evaluation. The wall photograph is produced after screen radiance propagates through the indoor scene, reaches the wall, undergoes diffuse redistribution, and is integrated by the camera. We summarize this acquisition process with a condition-dependent observation operator.

Task-Level Wall Observation

The relation between the hidden screen and the observed wall glow is written as

$$y_{c,h,w} = [\mathcal{P}_\eta(y^\star)]_{c,h,w} + \epsilon_{c,h,w}, \qquad c = 1, 2, 3. \tag{14}$$

Here, $\mathcal{P}_\eta$ denotes an acquisition-dependent transport operator associated with capture condition η, while $\epsilon \in \mathbb{R}^{3\times H\times W}$ collects residual acquisition variations. The operator $\mathcal{P}_\eta$ captures the spatial redistribution, attenuation, and mixing produced under capture condition η: one screen location contributes to an extended wall region, and each measured wall location aggregates contributions from multiple screen locations.

Equation (14) provides the task-level relation between the hidden screen and its indirect observation. OARL learns the inverse relation from paired training observations and directly reconstructs the screen from the wall photograph at inference.

Directly mapping this highly mixed observation to the final screen requires joint interpretation of coarse chromatic responses, broad spatial organization, local boundaries, and weak fine-scale structures. OARL introduces an intermediate image-shaped representation that supplies a common screen-space reference before information-type-specific evidence is estimated.

Preliminary Screen Estimate

The preliminary representation is defined as

$$z_{c,h,w} = \chi_c\left(\beta^{(0)}_{c,h,w} + \sum_{r=1}^{3}\iint_{\mathcal{D}} \mathsf{W}^{(0)}_{cr}(h, w; \xi, \eta) y_r(\xi, \eta)\,\mathrm{d}\xi\mathrm{d}\eta\right). \tag{15}$$

The coefficients $\mathsf{W}^{(0)}$ describe a learned spatial association between the observed wall field and the preliminary screen-space estimate, while $\beta^{(0)}$ provides the corresponding learned offset and χ_c denotes the bounded channel readout. On a discrete image grid, the spatial integral is implemented by the corresponding finite aggregation over sampled locations.

The representation $z \in \mathbb{R}^{3\times H\times W}$ provides the spatially organized starting point for the four reconstruction stages. OARL uses it as a common reference from which color, layout, structure, and detail evidence are evaluated separately and accumulated under different spatial controls.

Because z is computed from y and learned parameters, it preserves the single-image inference path. The same representation participates in evidence construction, initializes the accumulated reconstruction state, and anchors the final bounded readout, thereby establishing a consistent coordinate and appearance space for both OARL modules.

B.2 Calibrated Hierarchical Recoverability Representation Learning

Indirect transport preserves different components of screen content with different strengths. Broad color fields remain visible through low-frequency wall variations, whereas local boundaries, text strokes, icons, and textures produce weaker traces whose strength also varies spatially. Calibrated Hierarchical Recoverability Representation Learning explicitly estimates this information-type and spatial variation.

The module constructs a shared evidence representation and derives four continuous spatial gates for color, layout, structure, and detail. These learned control variables quantify the relative support available under the current observation and directly regulate the corresponding reconstruction stages.

Evidence-Feature Construction

The evidence representation begins with a location-wise descriptor $e_{h,w} \in \mathbb{R}^{E_0}$. This descriptor combines normalized appearance values from the wall observation and preliminary estimate with local spatial statistics. In particular, it contains normalized RGB measurements, locally averaged responses, signed discrepancies, absolute discrepancies, and gradient summaries. These components expose complementary aspects of the observation: RGB responses retain chromatic information, local means emphasize coarse organization, differences between observation-derived quantities identify inconsistencies, and gradient statistics summarize local structural variation.

The resulting evidence descriptor is transformed as

$$f_{\nu,h,w} = \varphi_\nu\left(\beta_\nu^{(f)} + \sum_{r=1}^{E_0} A_{\nu r}^{(f)} e_{r,h,w} + \sum_{r=1}^{E_0} \iint_{\mathcal{B}_{\mathrm{loc}}} \mathsf{W}_{\nu r}^{(\mathrm{loc})} e_r \,\mathrm{d}\xi\mathrm{d}\eta + \sum_{r=1}^{E_0} \iint_{\mathcal{B}_{\mathrm{ctx}}} \mathsf{W}_{\nu r}^{(\mathrm{ctx})} \bar{e}_r \,\mathrm{d}\xi\mathrm{d}\eta\right). \tag{16}$$

The first transformation preserves the response associated with the current spatial position. The local aggregation over $\mathcal{B}_{\mathrm{loc}}$ relates this response to its immediate neighborhood, whereas the contextual aggregation over $\mathcal{B}_{\mathrm{ctx}}$ uses the pooled quantity $\bar{e}_r$ to connect local evidence with a wider region. Their combination produces

$$f \in \mathbb{R}^{C\times H\times W},$$

which serves as the shared observation-evidence field used by subsequent type-specific branches.

This combination of spatial scales is useful in the present setting because the wall glow contains cues at different effective resolutions. A broad chromatic region is not characterized by the same spatial support as a thin boundary or text stroke. Restricting evidence extraction to a single local neighborhood would therefore provide an incomplete description of the observation, while global aggregation alone could suppress weak localized traces. Equation (16) retains both local and contextual responses in a common representation before the four information types are separated.

Since the descriptor e is formed from y and z, and z itself is derived from y, the complete evidence field f is observation-derived at inference and encodes the wall patterns used by the four type-specific branches.

Type-Specific Spatial Recoverability Gates

OARL applies separate observability-conditioned transformations to the shared evidence field for the four semantic information types. Let $\oplus_a$ denote the observability-conditioned attention process defined in the main method. For type $i \in \{1,2,3,4\}$, the corresponding spatial gate is

$$m_{i,h,w} = \left\{1 + \exp\left[-\beta_i^{(m)} - \sum_{\nu=1}^{C} \iint_{\mathcal{B}_m} \mathsf{W}_{i\nu}^{(m)}(\xi,\eta)[f\oplus_{a,\mathrm{g}}^{(i)}]_{\nu,h-\xi,w-\eta} \,\mathrm{d}\xi\mathrm{d}\eta\right]\right\}^{-1}. \tag{17}$$

The four indices correspond respectively to color, layout, structure, and detail, yielding

$$m = [m_1, m_2, m_3, m_4] \in (0,1)^{4\times H\times W}.$$

The use of four maps rather than a single shared mask allows OARL to represent two forms of variation simultaneously. First, different information types can receive different support at the same spatial location. A wall region may, for example, retain a reliable chromatic response while providing only weak support for local high-frequency detail. Second, the same information type can

receive different support at different locations because indirect propagation and wall reflection redistribute screen evidence nonuniformly.

The sigmoid form provides continuous recoverability weights for the reconstruction updates. During reconstruction, each m_i is broadcast over the three RGB channels and scales the residual associated with its corresponding semantic stage. A larger value therefore permits a larger observation-conditioned contribution, whereas a smaller value suppresses an update that receives weaker support from the learned evidence representation.

Optimized jointly with reconstruction supervision, the four maps differentiate reconstruction support across semantic type and spatial location. Their continuous values determine the contribution admitted from each stage and establish a direct operational link between the retained wall evidence and the reconstructed screen.

Global Evidence Representation

Spatial gates regulate local reconstruction, while training uses a compact representation of the complete wall evidence. The shared feature field is globally averaged and projected into a D-dimensional representation,

$$\tau_\upsilon = \beta_\upsilon^{(t)} + \sum_{\nu=1}^{C} W_{\upsilon\nu}^{(t)} \left(\frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} f_{\nu,h,w} \right), \qquad t_\upsilon = \theta_\upsilon^{(t)} + \lambda_\upsilon^{(t)} \frac{\tau_\upsilon - \overline{\tau}}{\sqrt{\varsigma_t^2 + \varepsilon_t}}. \tag{18}$$

where

$$\overline{\tau} = \frac{1}{D} \sum_{j=1}^{D} \tau_j, \qquad \varsigma_t^2 = \frac{1}{D} \sum_{j=1}^{D} (\tau_j - \overline{\tau})^2. \tag{19}$$

The resulting vector

$$t = [t_1, \ldots, t_D]^\top \in \mathbb{R}^D$$

summarizes the observation-derived evidence over the complete wall image and drives the auxiliary calibration process during training. The four spatial gates retain the positional information required for local reconstruction control, while t provides the corresponding global representation.

This design separates local reconstruction control from global comparison supervision. The spatial maps m_i regulate the four reconstruction stages, whereas t supports comparison between the current wall evidence and screen-level representations during training.

Matched and Nonmatching Evidence Calibration

To distinguish target-consistent evidence from visually similar alternatives, OARL compares the global wall-evidence representation with the paired target and a set of nonmatching screens during training.

Let $\zeta(\cdot)$ denote the screen-side projection into the same auxiliary representation space. The stabilized representations are defined as

$$\bar{t} = \mathfrak{s}_\rho(t), \qquad u^+ = \mathfrak{s}_\rho(\zeta(y^\star)), \qquad u_k^- = \mathfrak{s}_\rho(\zeta(n_k)), \qquad \mathcal{L}_2 = \ell\big(\bar{t}, u^+, \{u_k^-\}_{k=1}^K\big). \tag{20}$$

with

$$\mathfrak{s}_\rho(a) = \frac{a}{\max\{\|a\|_2, \rho\}}, \qquad \rho > 0. \tag{21}$$

Here, $y^\star$ denotes the paired screen and $n_k \in \mathbb{R}^{3\times H\times W}$ denotes a visually similar but nonmatching screen. The stabilized normalization keeps the representation well-defined for vectors with very

small norm. The comparison functional ℓ encourages the wall-derived representation to remain more compatible with its paired screen than with nonmatching alternatives.

This auxiliary relation injects screen-level correspondence into the shared evidence field during optimization. At inference, the learned field and its gates are computed directly from the wall photograph, and the reconstruction retains its single-image deployment path.

The comparison functional ℓ pulls the wall-derived token toward its paired screen representation and separates it from the nonmatching representations. Through joint optimization, correspondence supervision shapes the shared evidence field, while reconstruction supervision trains the spatial gates and reconstruction stages to use that evidence.

B.3 Observability-Stratified Accretive Reconstruction

The four spatial gates determine how strongly observation-derived evidence affects reconstruction. Observability-Stratified Accretive Reconstruction integrates this evidence through a fixed semantic progression from color to layout, structure, and detail, while adapting every stage's contribution to the current observation.

Every input follows the same coarse-to-fine semantic progression, and the learned gates determine where and how strongly each update is applied. Later stages inherit the states produced earlier, progressively adding new evidence to the established reconstruction. Fixed semantic dependencies and observation-dependent gate control therefore operate together throughout the four-stage path.

Initial Accumulated State

The reconstruction process begins from the preliminary estimate,

$$s_0 = z. \tag{22}$$

Thus,

$$s_0 \in \mathbb{R}^{3\times H\times W}$$

contains only quantities obtained from the current wall observation. The four semantic stages refine a shared image state rather than generating four independent screen predictions. This initialization also establishes a direct connection between preliminary recovery and the subsequent accretive process: every semantic update is expressed relative to the same screen-space reference.

Fixed Color-to-Detail Accretion

At stage i, the shared evidence feature is combined with all reconstruction states that have already been obtained. The inherited descriptor is

$$\mathfrak{r}^{(i)} = \mathrm{col}(f, s_1, \ldots, s_{i-1}), \qquad \mathfrak{a}^{(i)} = \mathfrak{r}^{(i)} \oplus^{(i)}_{a,\mathrm{s}}, \qquad \delta_i = m_i \odot \widetilde{\psi}_i, \qquad s_i = z + \sum_{j=1}^{i} \delta_j. \tag{23}$$

For $i = 1, 2, 3, 4$, the stages correspond respectively to color, layout, structure, and detail. At the first stage, the state portion of $\mathfrak{r}^{(1)}$ is empty and the update is generated from the observation evidence alone. At later stages, the descriptor contains increasingly rich reconstruction context. Layout is therefore conditioned on the color state already formed, structure inherits both color and layout information, and detail is generated with access to all preceding states.

The attention field $\mathfrak{a}^{(i)}$ is transformed into a stage-specific candidate ψ_i. The layout stage additionally uses pooled spatial context because layout describes larger-scale region organization. The detail stage emphasizes spatial variation by subtracting the candidate's local average,

$$\widetilde{\psi}_i = \psi_i - \mathbf{1}_{\{i=4\}} \langle \psi_4 \rangle_{\mathcal{B}_\phi}. \tag{24}$$

For the first three stages, $\widetilde{\psi}_i = \psi_i$. For $i = 4$, the local average $\langle \psi_4 \rangle_{\mathcal{B}_\phi}$ is subtracted so that the detail update emphasizes spatial variation beyond the low-frequency content already represented by earlier color, layout, and structure stages.

The pointwise product $m_i \odot \widetilde{\psi}_i$ then connects recoverability estimation with reconstruction. Each gate is broadcast across the RGB channels and regulates the spatial magnitude of the corresponding residual. The candidate update therefore represents what the current stage proposes to add, whereas m_i controls how much of that proposal is admitted at each location.

The accumulated definition

$$s_i = z + \sum_{j=1}^{i} \delta_j$$

preserves all preceding residuals explicitly. Each later stage operates on the existing state and uses the inherited representation to determine its update. This establishes the "accretive" reconstruction process: observation-regulated color, layout, structure, and detail components progressively form one shared screen estimate.

The fixed color-to-layout-to-structure-to-detail order defines the dependency structure of the deployed reconstruction. Within this backbone, the learned spatial gates adapt the magnitude and location of every update to the current observation.

Recoverability-Guided Prior Correction

The four gates also provide a compact indication of locations where the current observation gives relatively weak support across the semantic hierarchy. Rather than introducing a hard threshold, OARL computes the continuous complement of their average,

$$b_{h,w} = 1 - \frac{1}{4} \sum_{i=1}^{4} m_{i,h,w}, \qquad p_{c,h,w} = \gamma b_{h,w} \left(\beta^{(p)}_{c,h,w} + \sum_{r=1}^{C} P^{(p)}_{c,h,w,r} q^{(p)}_r \right). \tag{25}$$

The map

$$b \in (0,1)^{H \times W}$$

increases smoothly as the mean gate response decreases and continuously identifies locations that require a stronger learned correction.

The vector $q^{(p)} \in \mathbb{R}^C$ contains learned prior parameters and is expanded into an image-shaped correction by $P^{(p)}$. The scalar $\gamma > 0$ controls the correction scale, while b determines how strongly the correction participates at each spatial location. The resulting prior contribution varies continuously across the screen and concentrates on weakly supported regions.

The term p is a learned residual correction generated from the gate complement and prior vector. A high value of b indicates weak aggregate support at that position and admits a proportionally stronger correction. The complete term is generated internally from the observation-conditioned gates and learned parameters.

Bounded Inference Readout

After the four semantic updates have been accumulated, the final state is combined with the continuous correction,

$$v_{c,h,w} = s_{4,c,h,w} + p_{c,h,w}, \qquad \hat{y}_{c,h,w} = z_{c,h,w} + \kappa \tanh(v_{c,h,w} - z_{c,h,w}). \tag{26}$$

The intermediate quantity v combines the accumulated reconstruction with the recoverability-guided prior correction. The final output

$$\hat{y} \in \mathbb{R}^{3\times H\times W}$$

is expressed as a bounded offset from the preliminary representation z. Because

$$-1 < \tanh(\cdot) < 1,$$

the element-wise deviation satisfies

$$-\kappa < \hat{y}_{c,h,w} - z_{c,h,w} < \kappa.$$

The bounded readout constrains the final correction relative to the common preliminary reference. It anchors the prediction to the observation-derived representation while incorporating the complete sequence of gated semantic updates and the continuous prior correction.

Every quantity required by Eq. (26) is computed from the current wall image and learned parameters. The preliminary estimate z, the accumulated state s_4, the gates m_i, the recoverability-complement map b, and the correction p together form the complete single-frame reconstruction path at inference.

JOINT TRAINING OBJECTIVE

The reconstruction mechanism is optimized using both image-level reconstruction supervision and the auxiliary comparison signal introduced in Eq. (20). Their roles are summarized as

$$\mathcal{L}_1 = d(\hat{y}, y^\star), \qquad \mathcal{L} = \Delta(\mathcal{L}_1, \mathcal{L}_2)\,. \tag{27}$$

The functional d measures the discrepancy between the reconstructed screen $\hat{y}$ and its paired target $y^\star$, while $\mathcal{L}_2$ provides the auxiliary matched-versus-nonmatching comparison defined above. The operator Δ denotes their combination during optimization. Together, these two signals serve different roles: the reconstruction term supervises the final image output, whereas the auxiliary comparison constrains the observation-derived evidence representation with respect to paired and nonmatching screen content.

The paired target and nonmatching screens optimize the reconstruction and evidence representations during training. Once trained, OARL directly produces $\hat{y}$ from the current wall-glow observation.

Equation (27) unifies image reconstruction with matched-versus-nonmatching evidence calibration. The two objectives jointly optimize output fidelity and target-specific evidence selection, directly coupling the global representation learned by CHRRL with the final screen reconstruction produced by OSAR.

C ADDITIONAL EXPERIMENTAL RESULTS

The experiments in this section are arranged by the source of variation in the wall observation. Photometric tests probe illumination and sensor changes; geometric tests probe camera position, orientation, distance, and partial visibility; surface tests probe wall texture, material, and ambient light; and motion tests probe blur and camera shake. The final group (C.21–C.25) examines evidence selection and accumulation in the specified representation, learning, refinement, and stage-transfer configurations. Each experiment keeps the remaining acquisition and evaluation settings fixed, reports the same screen-domain metrics (LPIPS, SSIM, and PSNR) unless a mechanism-specific metric is stated, and uses the displayed qualitative panels together with the corresponding metric curves to connect visual behavior with aggregate measurements. The robustness evidence is consequently read as a progression: the observation changes, the reconstructed screen changes over the tested range, and the final comparisons examine how the evaluated configurations use the remaining evidence.

C.1 MECHANISM-ORIENTED READING OF THE MAIN DIAGNOSTICS

The representation controls in Table 3 are constructed in this study: RGB uses the original RGB observation; Edge repeats a normalized Sobel magnitude across channels; Retinex uses normalized log reflectance relative to a 41×41 local mean; Frequency mixes low-pass and high-pass responses; Degradation concatenates darkness, gradients, and residual luminance; Hybrid-Cue combines reflectance luminance, gradients, and a weighted recoverability cue. The separate diagnostic label OursMax denotes RGB/low-light/reflectance fusion with recoverability-dependent gain, edge, and residual hints; it is distinct from Hybrid-Cue and is not a maximum-response aggregation.

The learning controls implement the following study-defined training rules. Direct uses the base reconstruction objective; Perceptual adds luminance alignment; Deep Supervision adds multi-scale reconstruction and Laplacian alignment; Contrastive adds paired-content discrimination and a wrong-pair margin; Distillation uses a target-derived smoothing proxy; Curriculum varies multi-scale and Laplacian weights with training progress; Uncertainty reweights reconstruction errors; Recurrent uses a smoothed-prediction stability proxy; and Task-Guided weights reconstruction using observation evidence. The Distillation and Recurrent labels therefore denote loss proxies, not a separate learned teacher or a recurrent network. The same control definitions apply to the one-epoch comparison.

Tables 2–4 examine component, representation, and learning contributions. The inheritance and ordering comparison and the camera-translation diagnostics examine reconstruction behavior under controlled changes. Read together, these results motivate the controlled stress tests that follow: the Representation tests ask whether the model selects screen-consistent evidence, whereas the Learning tests ask whether that evidence is accumulated in the proposed order and carried into later stages. The five mechanism tests are therefore mapped to the two claims as follows: REP-PK1 and REP-PK3 probe evidence selection under false and weak contours; LRN-PK1 probes consistency with a frozen screen-to-wall forward operator; LRN-PK2 probes performance under a matched one-epoch budget; and LRN-PK4 probes ordered inheritance from strong to weak cue regions.

PHOTOMETRIC AND SENSOR VARIATION

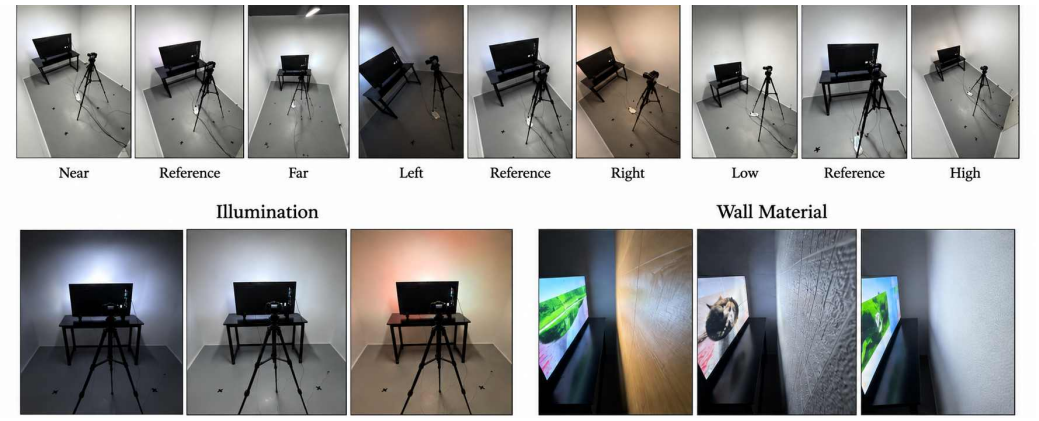
The first set of tests changes the intensity or direction of light reaching the wall and the noise level at capture. These perturbations alter the strength and contrast of the indirect evidence while leaving the displayed screen unchanged. The resulting experiments measure OARL's ability to preserve screen-level organization across photometric and sensor changes.

C.2 REPRESENTATIVE ROBUSTNESS ACQUISITION SETTINGS

The additional robustness baselines include Ada-Deblur (**?**), NDR-Restore (**?**), OKNet (**?**), HVI-CIDNet (**?**), X-Restormer (**?**), and UniProcessor (**?**).

The robustness evaluation combines real indoor wall-glow acquisitions with controlled variations of recorded images to examine changes in illumination, geometry, surface appearance, and local image structure. Representative real-world acquisition configurations are shown in Figure 2, including variations in camera distance, viewing angle, camera height, environmental illumination, and relay-wall material. These changes alter the geometry, spatial distribution, intensity, and surface-dependent appearance of the observed wall glow while preserving the same hidden-screen reconstruction objective.

The acquisition examples and the controlled image variations serve complementary roles. The photographs document physical changes in the observation setting, whereas the image-based sweeps isolate specified changes in an existing wall observation. Their interpretation therefore follows the factor varied in each experiment: a change in camera geometry alters the recorded spatial arrangement, while an intensity or noise perturbation changes the available signal without defining a new target screen. The reconstruction is evaluated against the corresponding screen reference, so agreement is assessed in screen space rather than by similarity to the perturbed wall image. This distinction connects the acquisition setting to the reported robustness results and separates preservation of target content from reproduction of observation-side appearance. The following experiments specify the tested factor and its range before presenting the corresponding reconstructions and metric responses.



Supplementary Figure 2: Representative real-world acquisition settings used in the robustness evaluation. The examples illustrate variations in camera distance, viewing angle, camera height, environmental illumination, and relay-wall material. Three representative configurations are shown for each physical factor to visualize the corresponding changes in the acquisition environment. The complete perturbation ranges are specified in the corresponding robustness experiments.

The photographs provide representative examples of the major physical perturbation settings used in the robustness study rather than an exhaustive visualization of every tested condition. The complete perturbation ranges and parameter settings are specified separately in the corresponding robustness experiments below.

C.3 Reconstruction under illumination changes

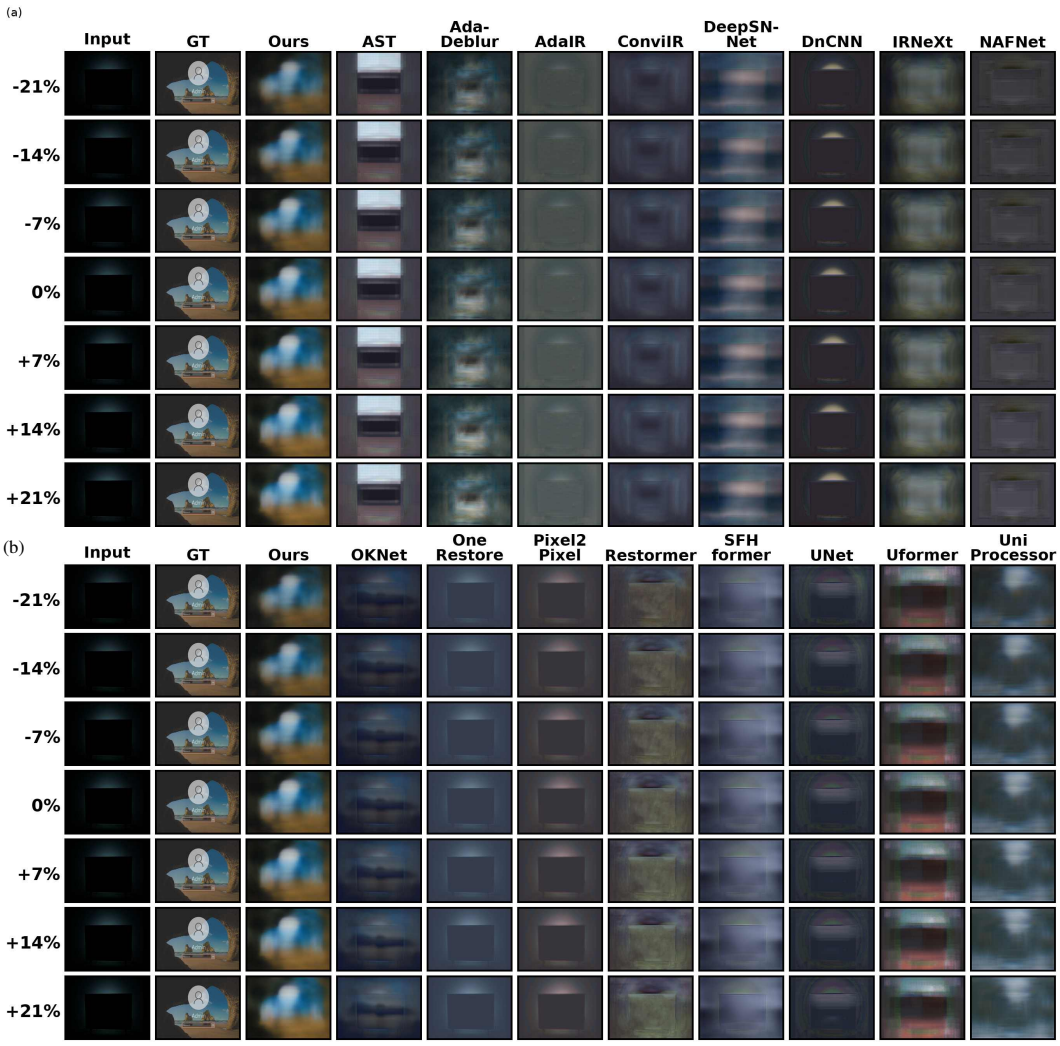
We vary global illumination by -21%, -14%, -7%, 0%, $+7\%$, $+14\%$, and $+21\%$ on OSLock, Screen, Figure, and Game, keeping all other imaging parameters fixed.

Each dataset is shown with two baseline groups in vertically stacked panels (a) and (b), followed by its quantitative results and analysis. Rows specify the perturbation levels; columns show the input, ground truth, and compared reconstructions. We report LPIPS (lower is better), SSIM, and PSNR (higher is better).

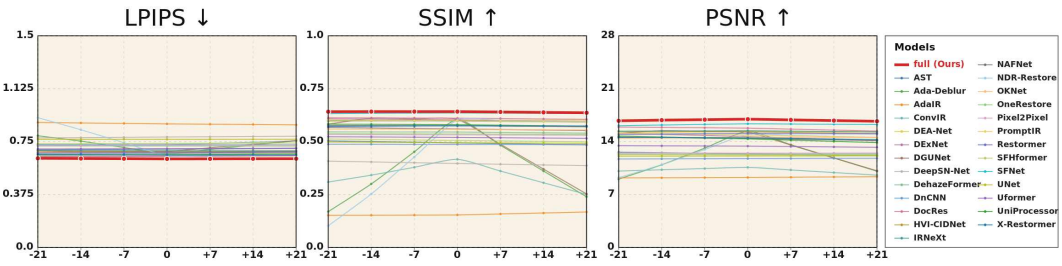
The illumination sweep examines whether changes in wall-glow intensity alter the recovered screen content. The nominal condition provides a reference within each category, and the darker and brighter conditions test departures from that reference while preserving the reconstruction target. A screen-consistent response should retain the positions and relationships of the target regions across these conditions. Matching the overall brightness of the wall observation alone is insufficient: the reconstructed foreground, background, and internal boundaries must remain consistent with the displayed screen. The qualitative panels therefore show how changes in the indirect input affect the organization of the output, rather than treating a brighter reconstruction as evidence of improved recovery.

The four content categories expose different aspects of this response. OSLock tests the arrangement of broad background regions around a localized central feature. Screen tests separation of a compact application window from the surrounding desktop. Figure tests the relative positions, heights, and color ordering of grouped bars. Game tests the continuity of a large playing field and its boundary with the adjacent interface. These structures provide target-specific references for assessing the same illumination perturbation. Their preservation is visible at the level of regional layout even when fine details remain difficult to resolve, linking the qualitative comparison to the main reconstruction objective.

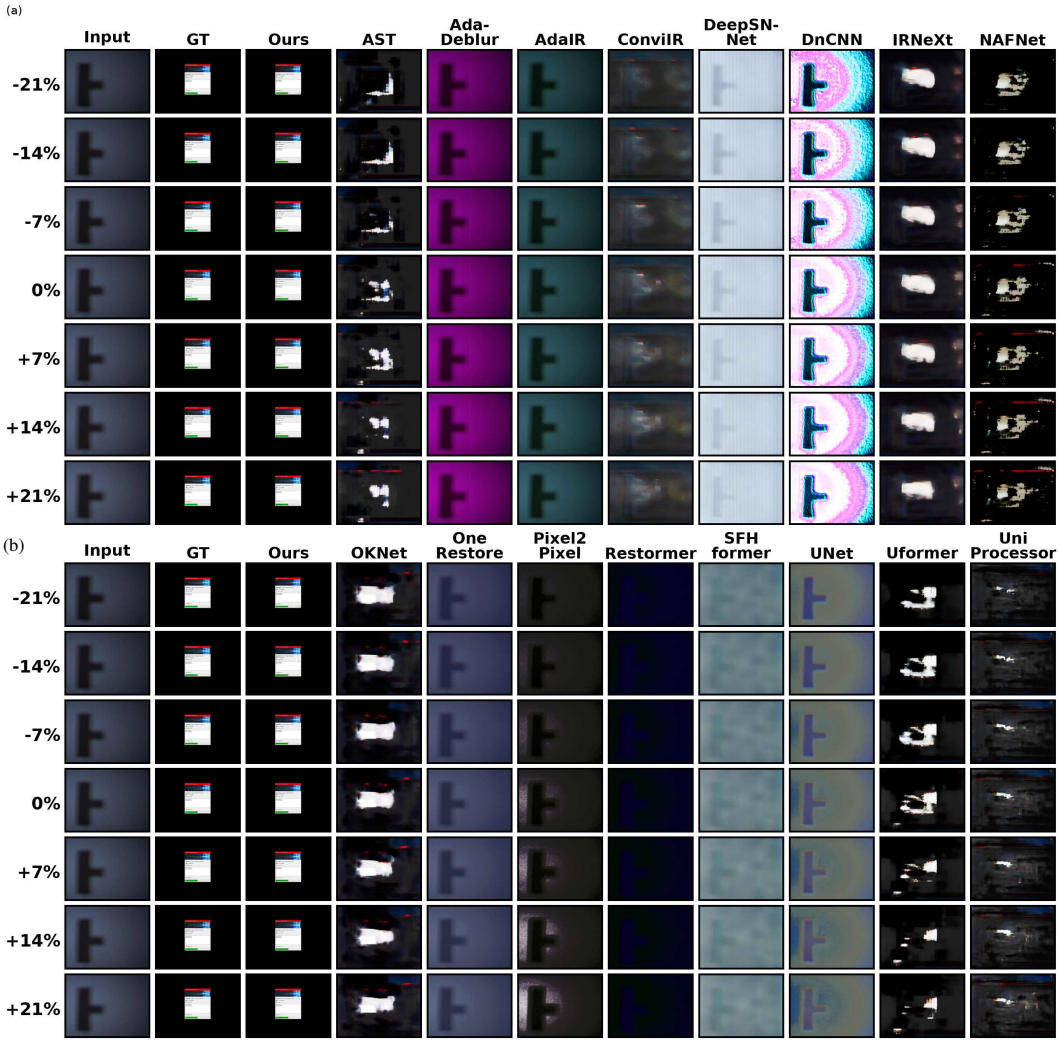
The metric curves complement these examples by reporting reconstruction quality over the tested illumination range. PSNR measures pixel-level agreement, SSIM reflects structural similarity, and LPIPS measures perceptual discrepancy. Their responses are considered jointly with the panels because a single aggregate measure does not identify which screen region is retained or lost. The relevant comparison includes both the ordering of methods at a given illumination level and the change in each method across the sweep. The following results relate these quantitative responses to the observed foreground placement, graphical structure, and regional boundaries.



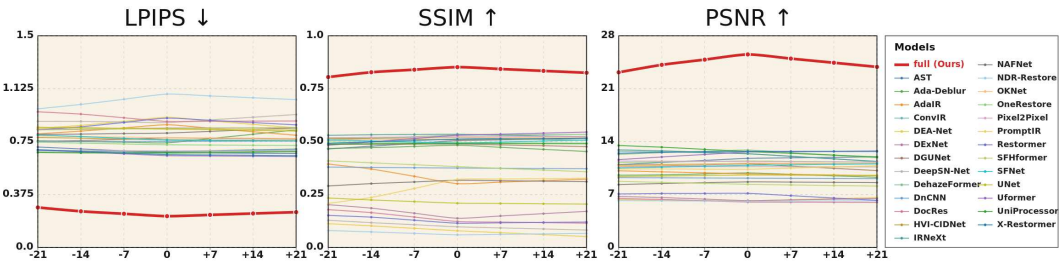
Supplementary Figure 3: OSLock under global illumination changes. Panels (a) and (b) show the two baseline groups. Rows correspond to seven illumination levels from -21% to $+21\%$; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



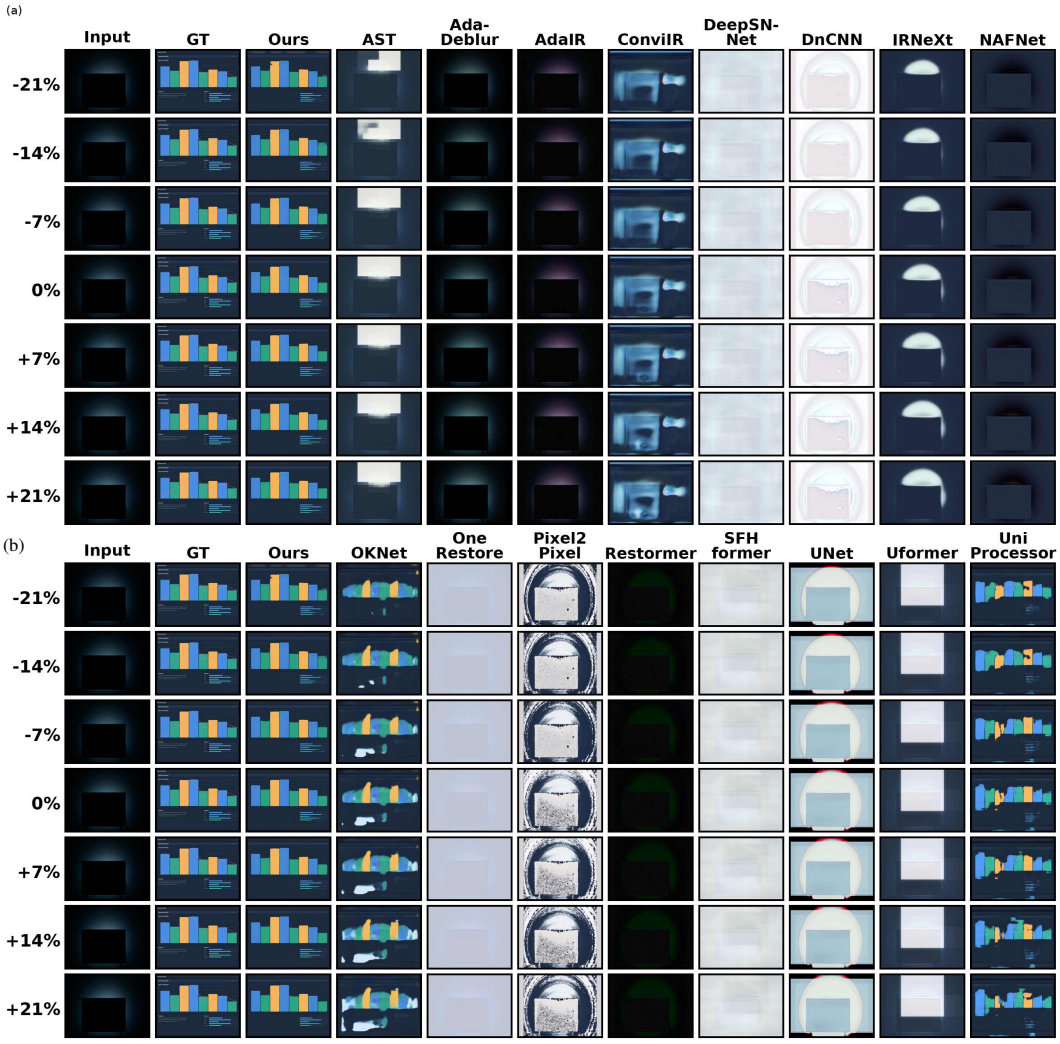
On OSLock, OARL retains the blue background, central profile region, and warm lower scenery across the illumination sweep. Several competing outputs become dark or nearly uniform, or introduce broad color distortions. OARL preserves the arrangement of the distinctive screen regions at both illumination extremes. Its nearly flat LPIPS, SSIM, and PSNR curves confirm that this structural advantage extends across the evaluated illumination range. The central light region remains separated from the lower scenery instead of becoming a single broad response, giving OARL a more faithful foreground-to-background arrangement throughout the sweep.



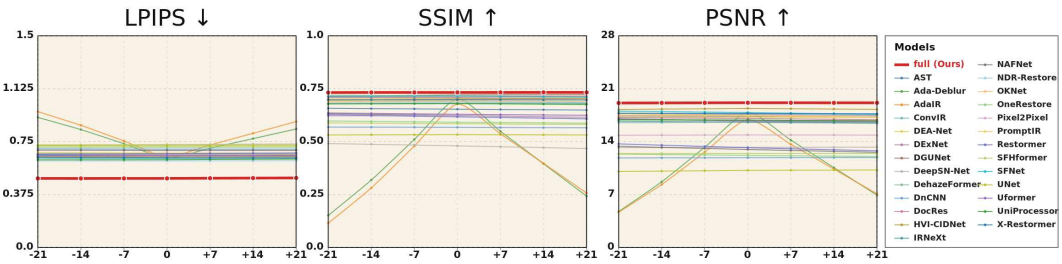
Supplementary Figure 4: Screen under global illumination changes. Panels (a) and (b) show the two baseline groups. Rows correspond to seven illumination levels from -21% to $+21\%$; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



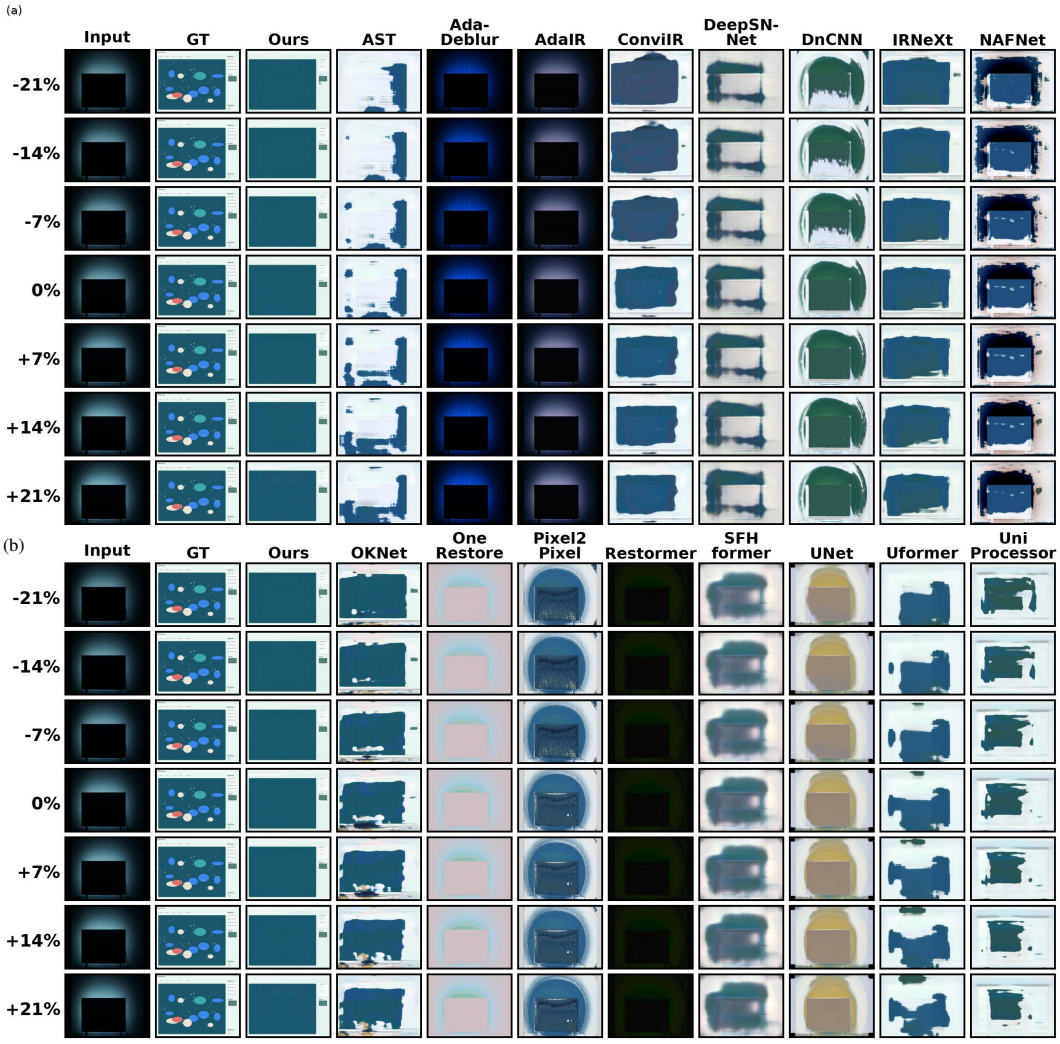
On Screen, OARL recovers the compact bright window against the dark desktop, including its colored upper strip and light interior. Several baselines retain the large input silhouette or spread tinted responses beyond the target window. OARL maintains a clearer separation between the foreground application and its surroundings. PSNR and SSIM peak near nominal illumination, where LPIPS reaches its minimum; OARL retains lower LPIPS and higher SSIM and PSNR at both ends of the sweep. These results demonstrate reliable recovery of a localized interface under both reduced and increased illumination.



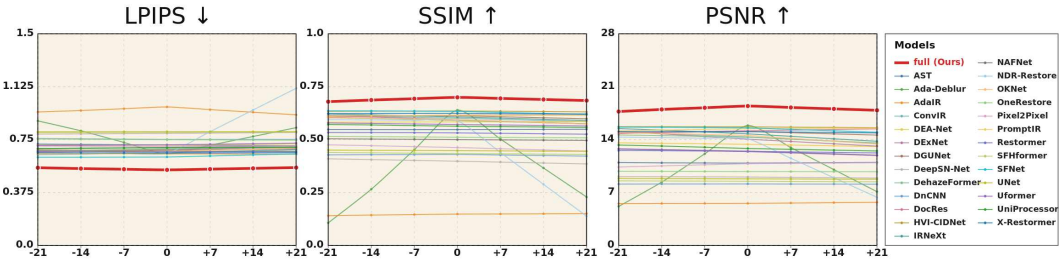
Supplementary Figure 5: Figure under global illumination changes. Panels (a) and (b) show the two baseline groups. Rows correspond to seven illumination levels from -21% to $+21\%$; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



The Figure example contains grouped blue, green, and yellow bars on a dark plotting area. OARL preserves their relative heights, spacing, and color order throughout the illumination sweep. Several competing reconstructions replace the chart with uniform blocks or unrelated contours, while others merge the bars into broad patches. The OARL curves remain nearly flat and lead the three plotted metrics, including at the darkest and brightest settings. The agreement between bar-level organization and aggregate quality demonstrates strong robustness for structured graphical content under global illumination changes.



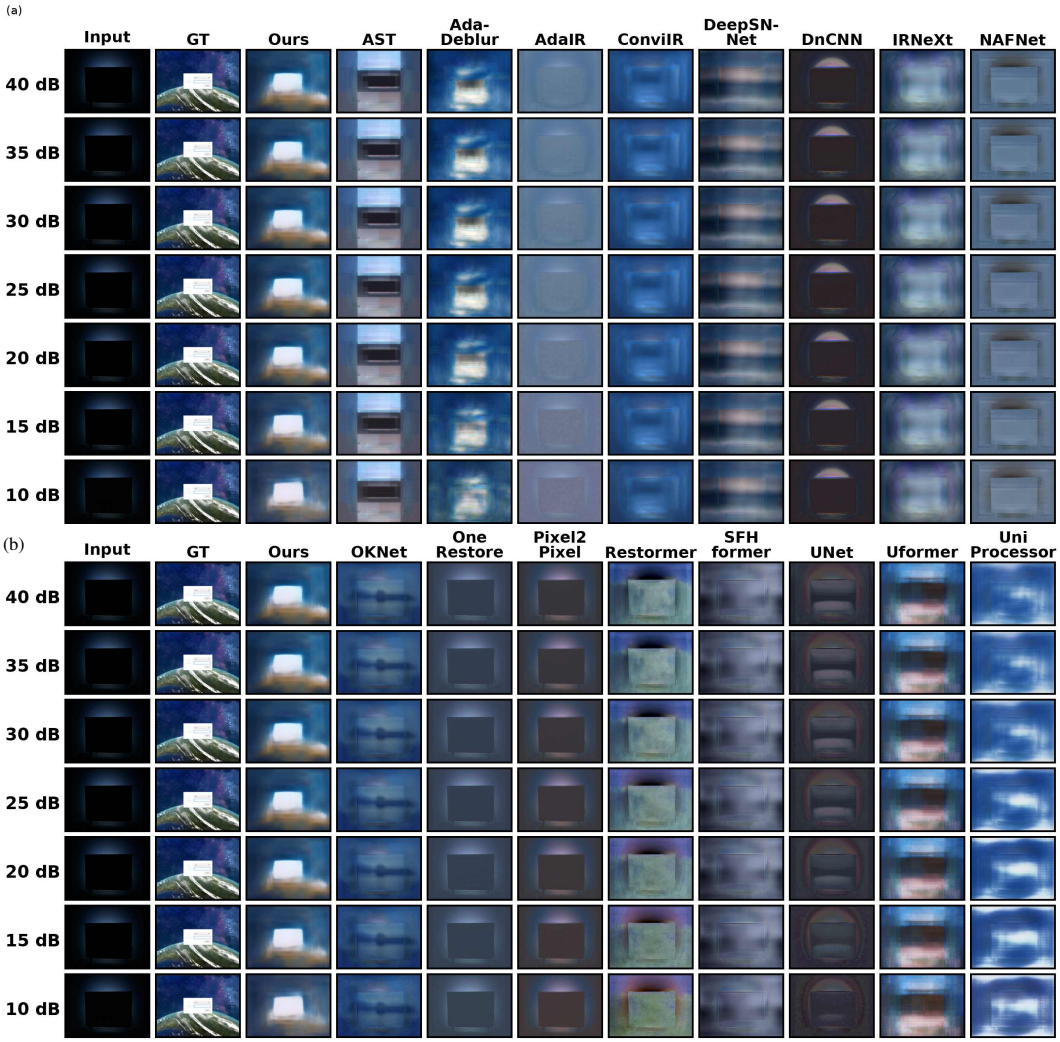
Supplementary Figure 6: Game under global illumination changes. Panels (a) and (b) show the two baseline groups. Rows correspond to seven illumination levels from -21% to $+21\%$; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



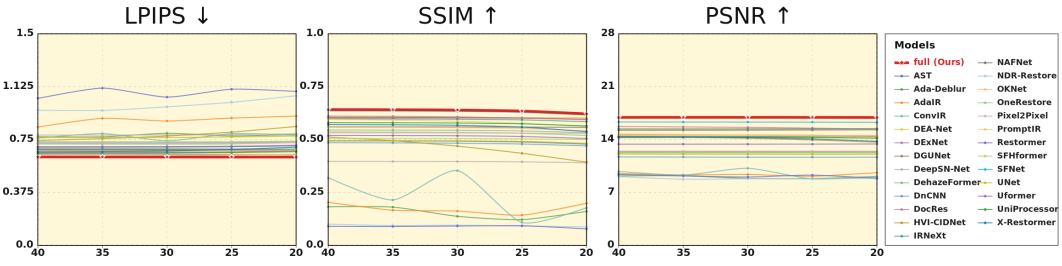
On Game, OARL preserves the broad teal playing field and its division from the narrow light sidebar across the illumination sweep. Several competing outputs introduce holes, distorted borders, or large regions whose colors depart from the reference. OARL produces a more coherent field with a cleaner surrounding interface. Its LPIPS remains low, while SSIM and PSNR show only modest changes around nominal illumination. The persistent metric advantage at both illumination extremes supports the visual result: the main screen regions remain well organized as the strength of the wall glow changes.

C.4 Reconstruction under camera noise

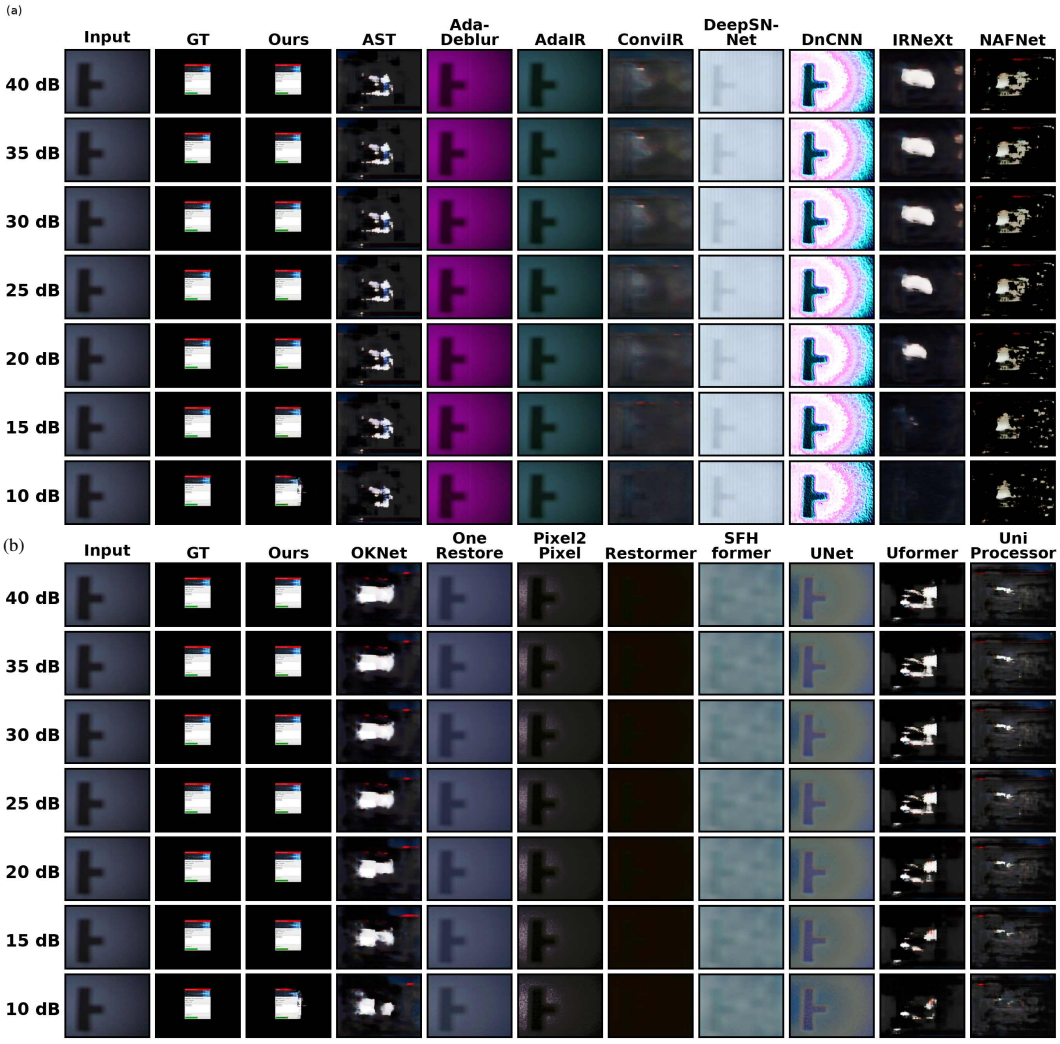
Added Gaussian noise introduces local variation alongside the screen-dependent wall response. We add Gaussian camera noise at SNRs of 40, 35, 30, 25, 20, 15, and 10 dB, with other acquisition settings fixed. The quantitative plots cover 40–20 dB.



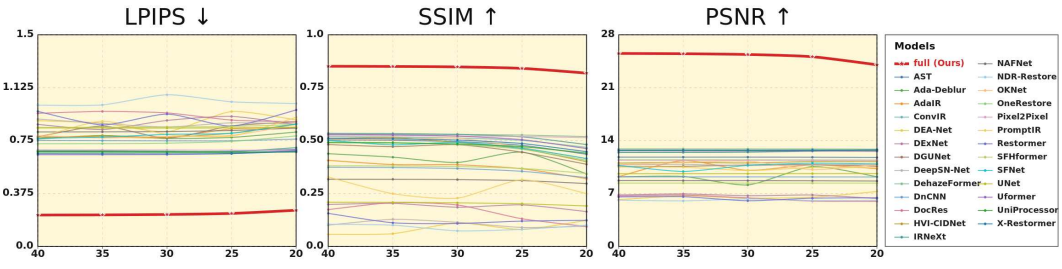
Supplementary Figure 7: OSLock under camera noise. Panels (a,b) compare the baseline groups; rows show perturbation levels. Below: LPIPS (↓), SSIM and PSNR (↑). Curves: 40–20 dB.



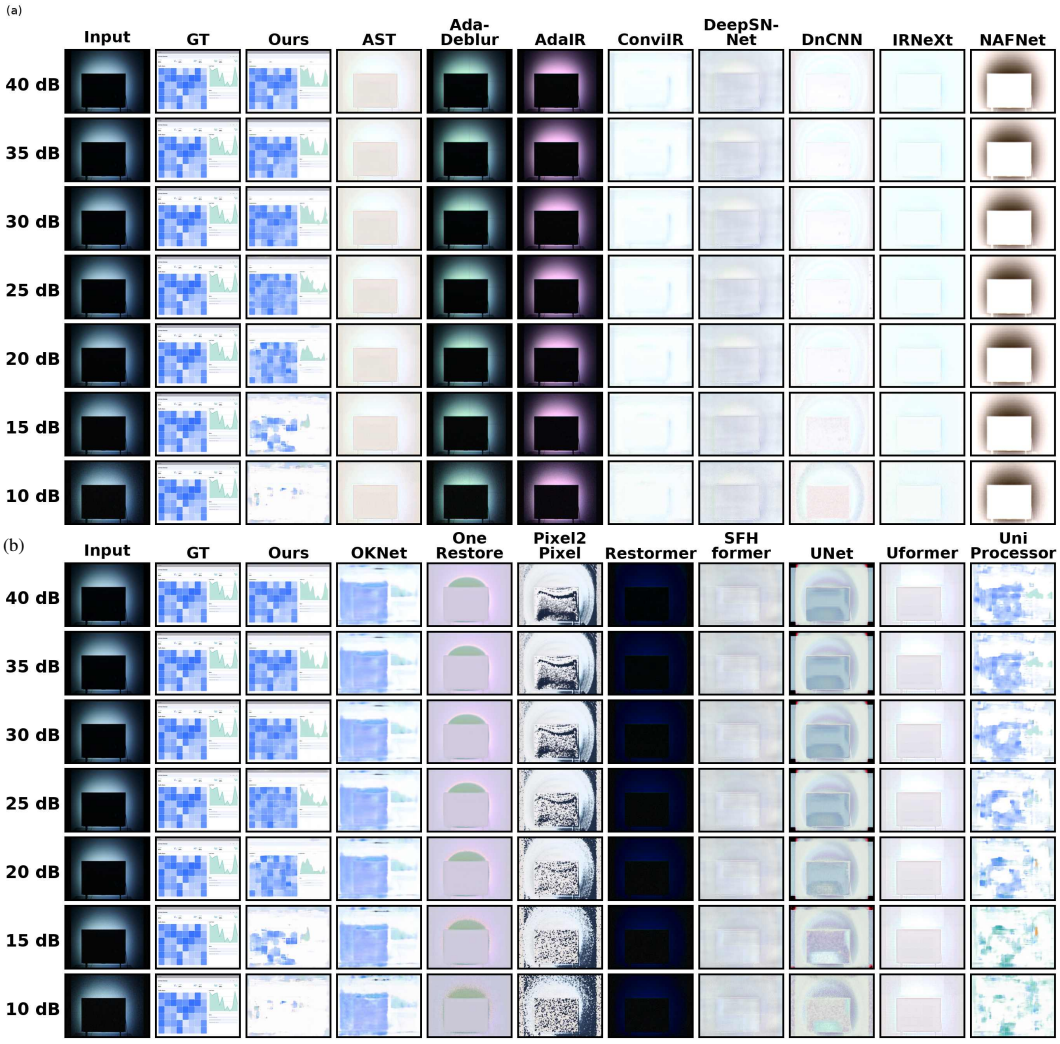
On OSLock, OARL preserves the central bright dialog and the blue-to-warm organization of the mountain background as noise increases. Several baselines produce diffuse bands or dark blocks. The low LPIPS and leading SSIM and PSNR over the plotted SNR range support OARL's stronger recovery of the dominant screen structure.



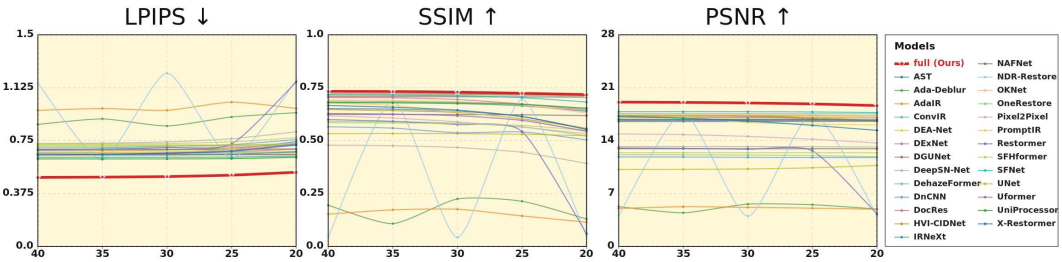
Supplementary Figure 8: Screen under camera noise. Panels (a) and (b) show the two baseline groups. Rows correspond to seven SNRs from 40 to 10 dB; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset. Curves: 40–20 dB.



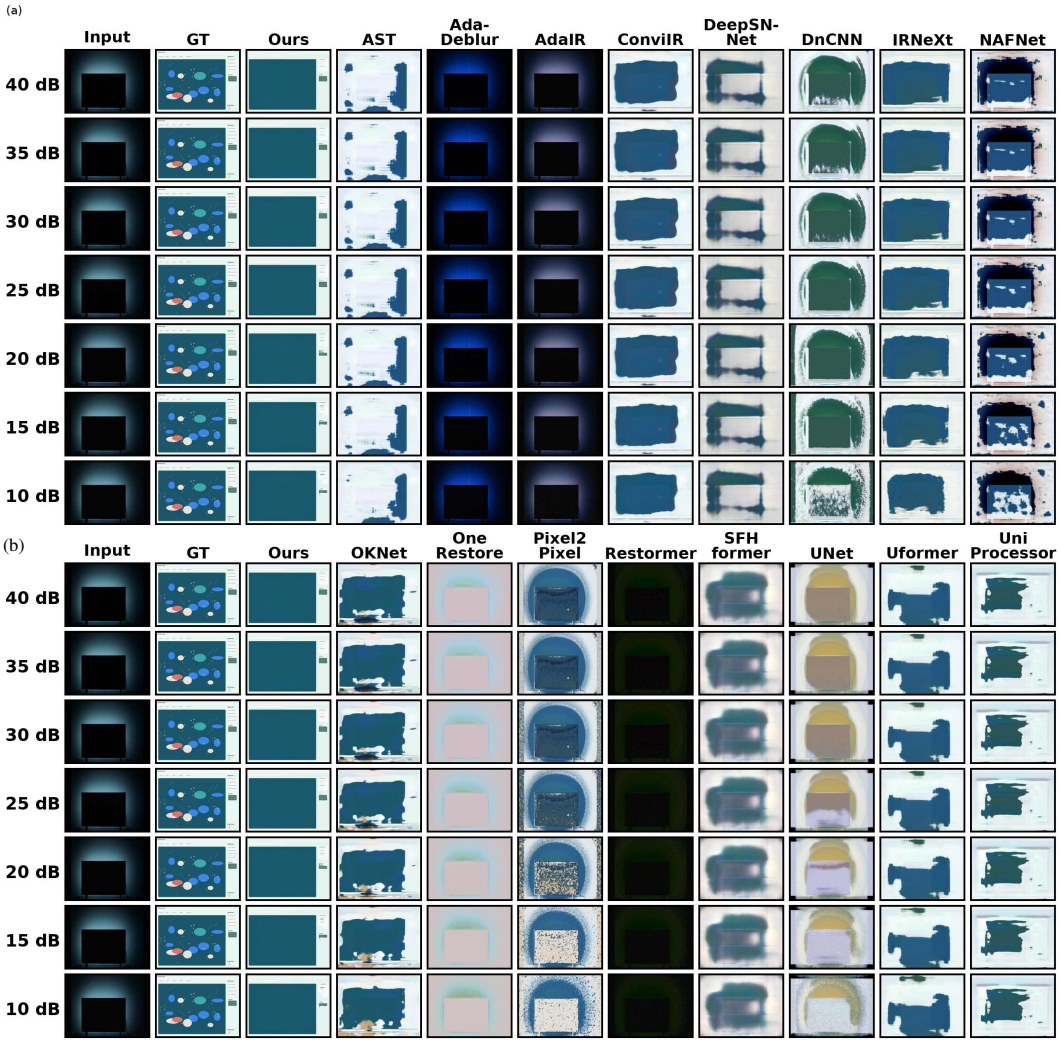
The Screen example places a small bright application window within a dark desktop. OARL maintains the window position, rectangular outline, and foreground contrast as the observation becomes noisier. Several competing methods disperse the bright response across the background or reproduce the input-side silhouette. The quantitative curves show a substantial and persistent advantage in LPIPS, SSIM, and PSNR over the plotted 40–20 dB range. Together with the localized window recovered in the qualitative panels, these results demonstrate effective suppression of acquisition noise while retaining the spatial organization of the target interface.



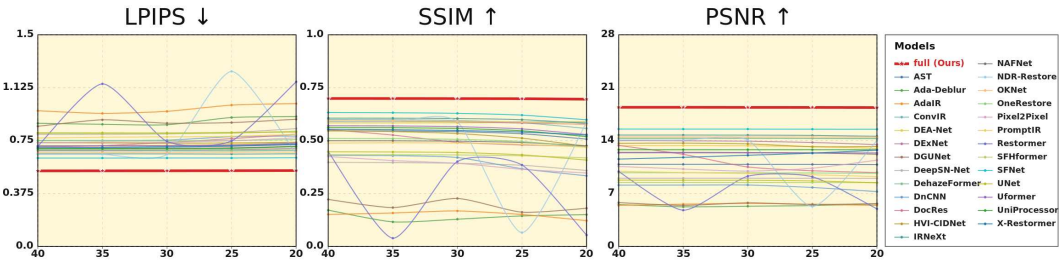
Supplementary Figure 9: Figure under camera noise. Panels (a) and (b) show the two baseline groups. Rows correspond to seven SNRs from 40 to 10 dB; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset. Curves: 40–20 dB.



On Figure, OARL recovers the blue matrix on the left and the separate light plot region on the right with clear spatial organization at the higher and intermediate SNRs. Several baselines instead produce nearly white outputs or diffuse rectangular patches. Across the plotted 40–20 dB range, OARL retains the lowest LPIPS and highest SSIM and PSNR, with gradual changes as noise increases. This metric separation agrees with the stronger recovery of the two graphical regions and demonstrates OARL's advantage in reconstructing structured content from noisy indirect observations.



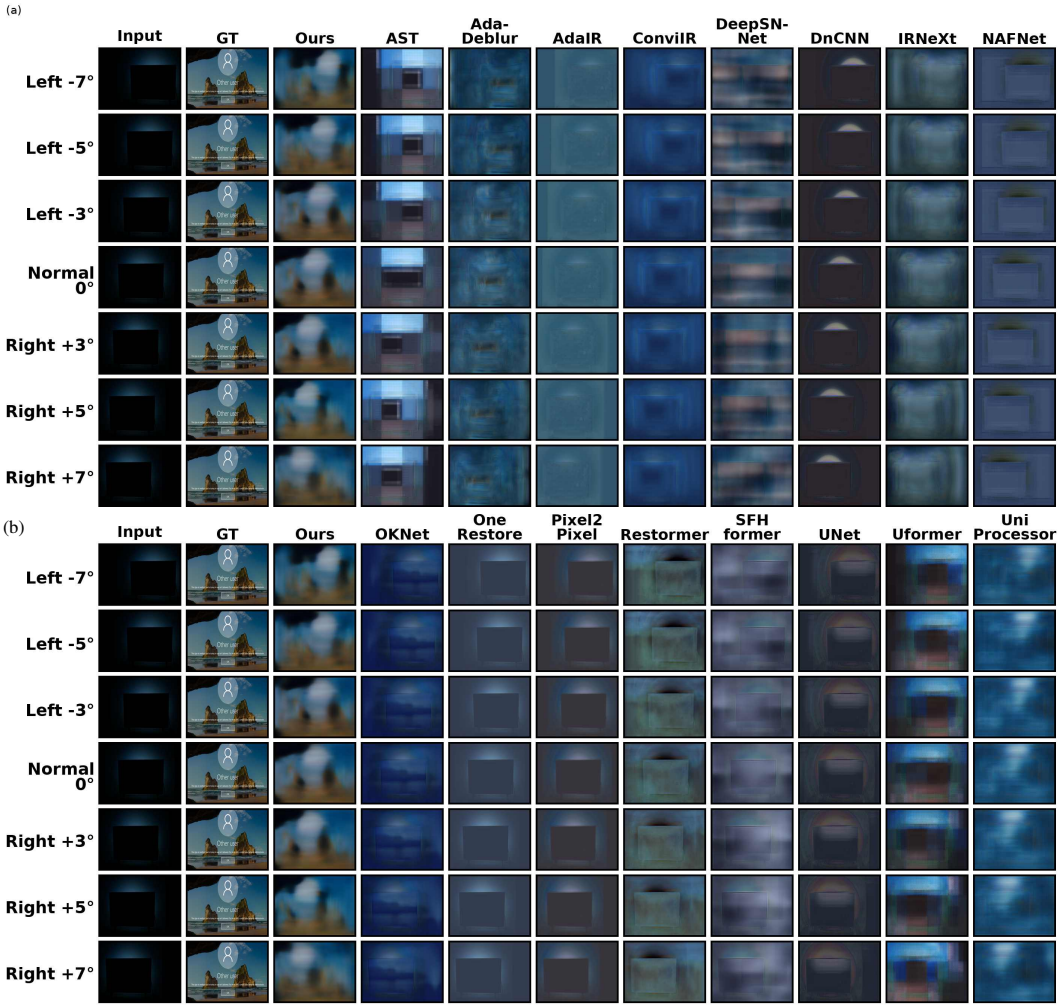
Supplementary Figure 10: Game under camera noise. Panels (a) and (b) show the two baseline groups. Rows correspond to seven SNRs from 40 to 10 dB; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset. Curves: 40–20 dB.



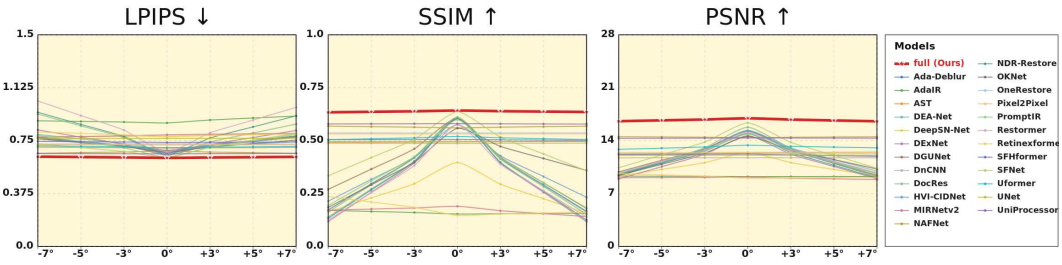
On Game, OARL keeps the teal playing field contiguous and separates it from the light sidebar as camera noise increases. Several competing outputs introduce large gaps, dark contours, or irregular color patches around the field. The cleaner regional layout is accompanied by nearly constant OARL metric curves over the plotted SNR range. In particular, the persistent SSIM and PSNR advantage indicates that the coherent large-scale organization is reflected in the aggregate evaluation. These results demonstrate robust recovery of the game interface despite the increased noise in the wall observation.

C.5 Robustness to horizontal camera translation

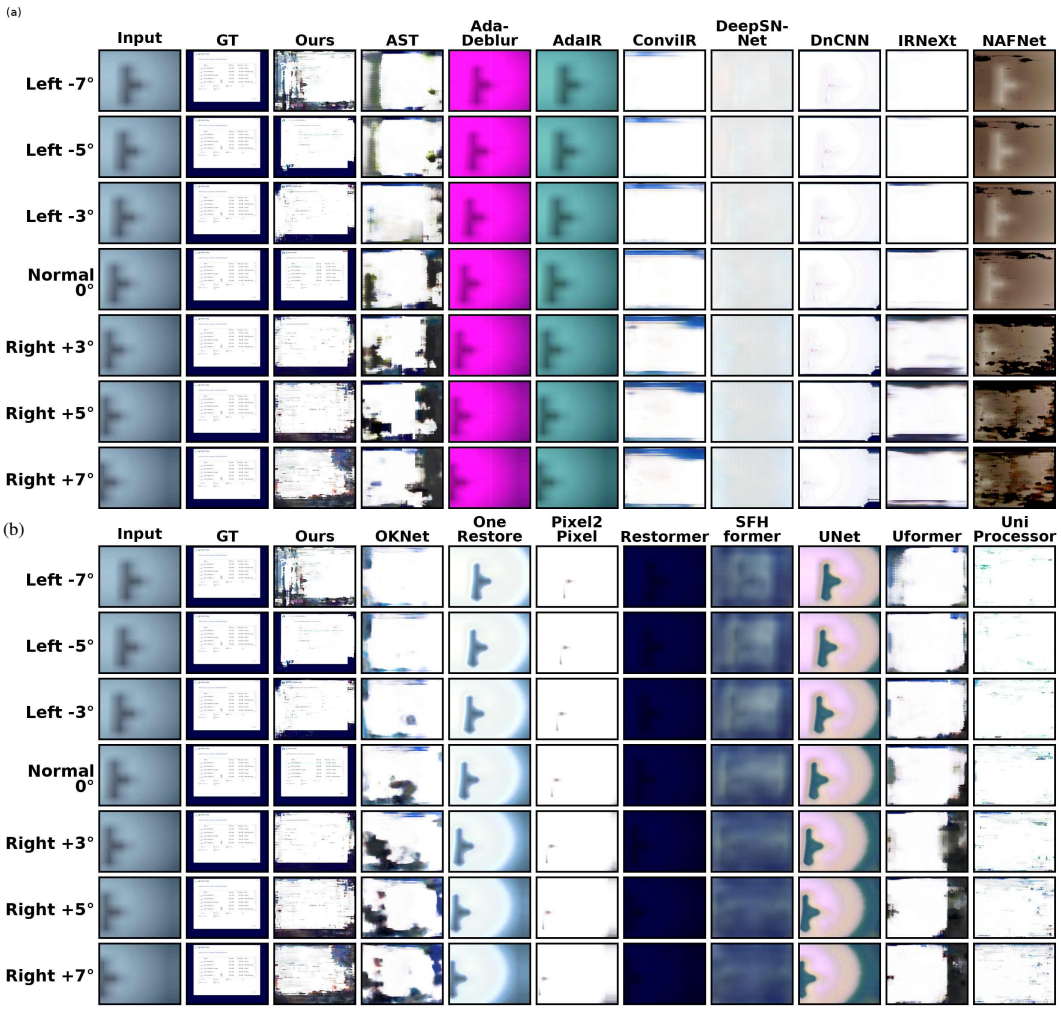
We test horizontal camera offsets of $-7°$, $-5°$, $-3°$, $0°$, $+3°$, $+5°$, and $+7°$. This geometric sweep changes the observed glow while keeping the displayed screen and remaining acquisition settings fixed.



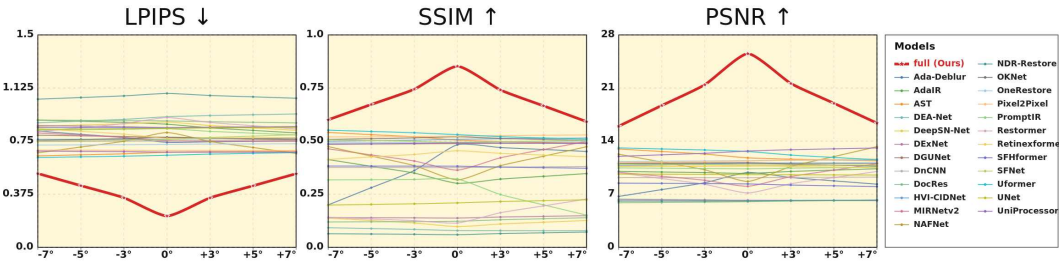
Supplementary Figure 11: OSLock under horizontal camera translation. Panels (a,b) compare the baseline groups; rows show perturbation levels. Below: LPIPS (↓), SSIM and PSNR (↑).



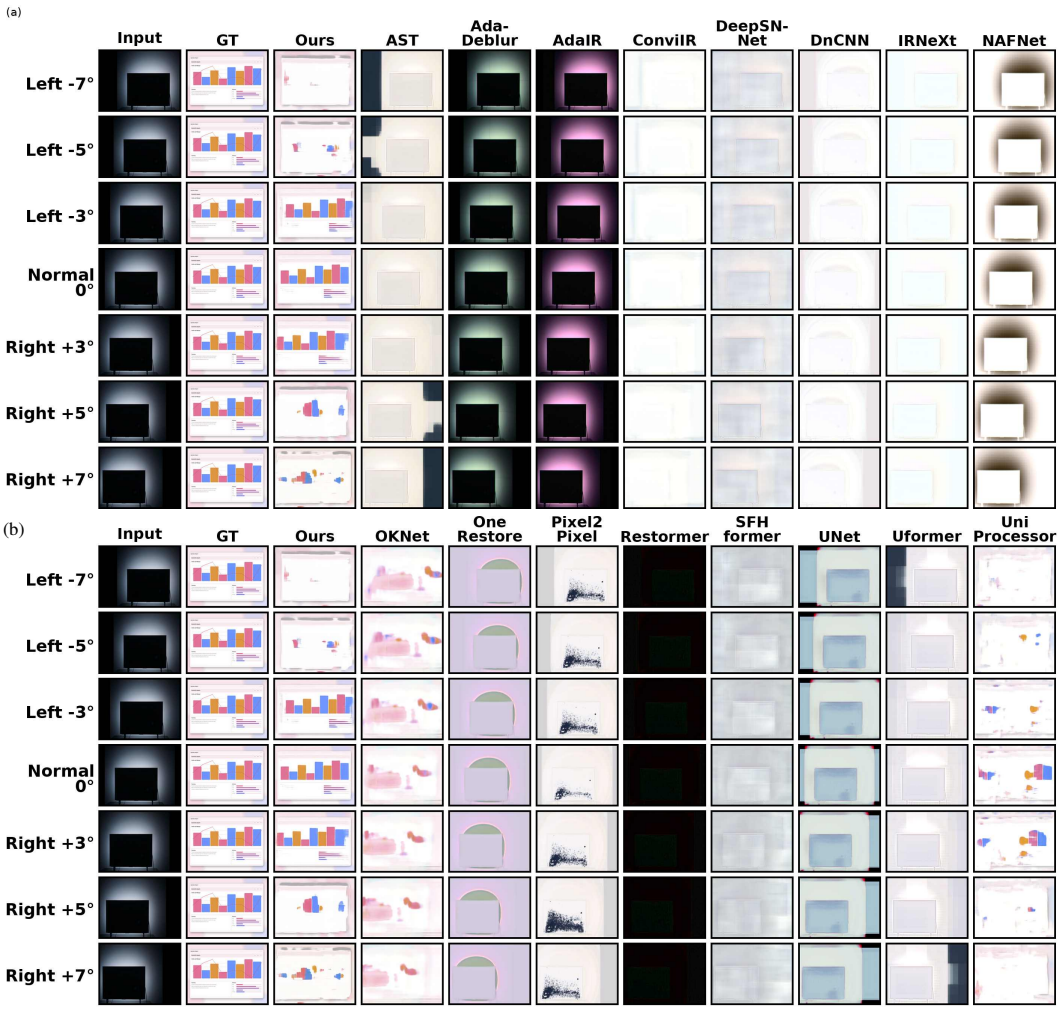
On OSLock, OARL retains the central profile region and the blue upper and warm lower portions of the screen across horizontal positions. Competing methods introduce stronger banding or dark regions. Its balanced metric profiles on either side of the reference position demonstrate reliable recovery of the screen's broad organization under translation. The main bright region remains at its target location as horizontal viewpoint changes alter the recorded wall-glow pattern.



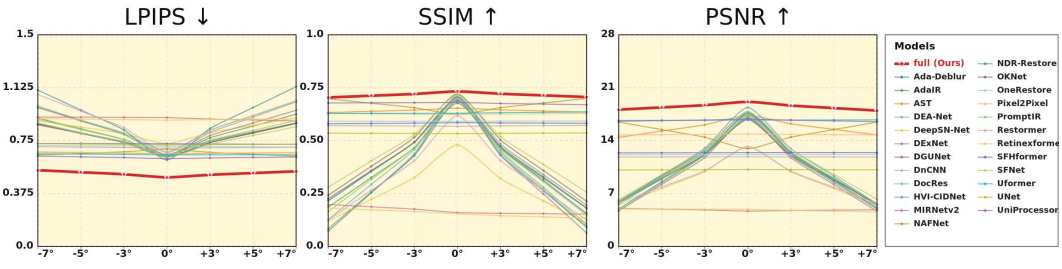
Supplementary Figure 12: Screen under horizontal camera translation. Panels (a) and (b) show the two baseline groups. Rows correspond to seven horizontal offsets from $-7°$ to $+7°$; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



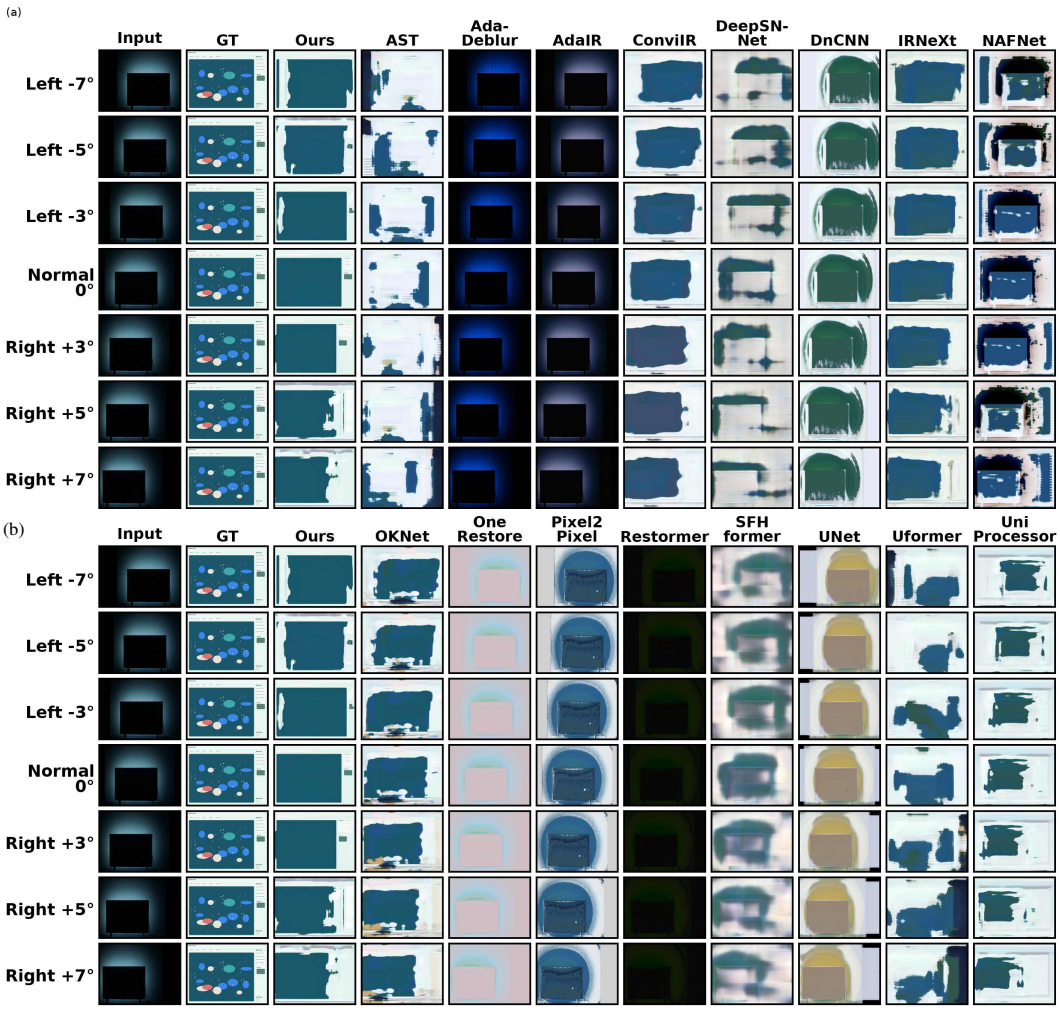
On Screen, OARL recovers the bright table window most clearly near the reference position, with a more faithful rectangular footprint and internal organization than the competing outputs. Several baselines remain largely uniform or preserve broad input-related patterns. As the camera moves to either side, OARL's LPIPS increases and its SSIM and PSNR decrease smoothly while retaining the leading ranking. The corresponding responses on the negative and positive sides establish a consistent geometric advantage, linking the strong window reconstruction near the reference view to superior aggregate performance throughout the sweep.



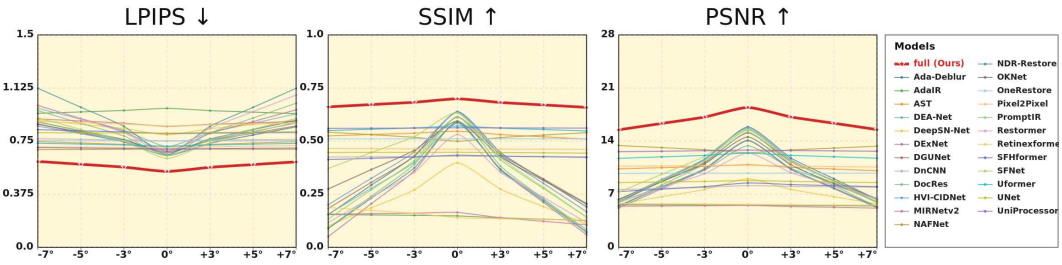
Supplementary Figure 13: Figure under horizontal camera translation. Panels (a) and (b) show the two baseline groups. Rows correspond to seven horizontal offsets from $-7°$ to $+7°$; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



The Figure panels show a colored bar chart on a light background. OARL reconstructs the distinct bars and their ordering especially clearly at the central offsets, whereas several competing methods return pale fields or broad illumination contours. The quantitative response remains favorable across the full sweep: OARL has lower LPIPS and higher SSIM and PSNR on both sides of the reference position. Its wider structural advantage at the outer offsets demonstrates stronger tolerance to changes in the spatial arrangement of the wall glow, complementing the clear bar recovery around the nominal view. The light canvas and colored bars provide complementary cues to this retained graphical organization.



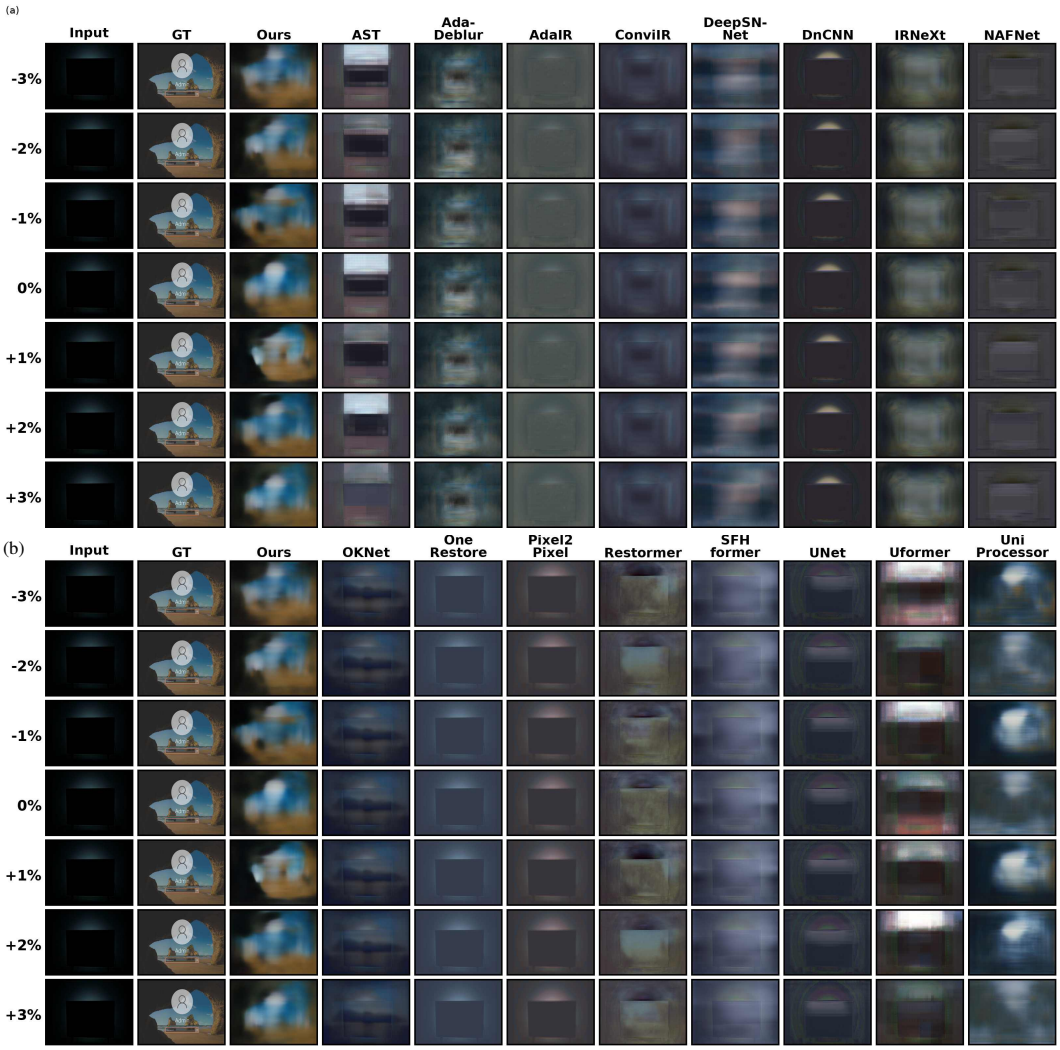
Supplementary Figure 14: Game under horizontal camera translation. Panels (a) and (b) show the two baseline groups. Rows correspond to seven horizontal offsets from $-7°$ to $+7°$; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



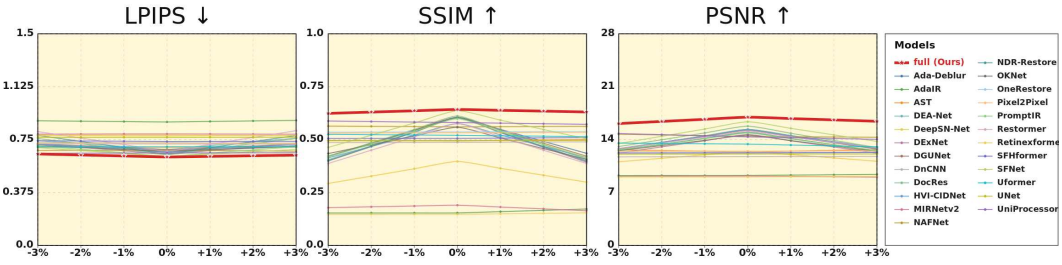
On Game, OARL reconstructs a coherent teal field together with the light surrounding interface. Several competing methods break the field into patches, introduce broad gaps, or replace the rectangular layout with curved contours. OARL's regional structure is clearest around the central camera positions, and the metric curves retain their advantage as the camera moves in either direction. The smooth increase in LPIPS and gradual reduction in SSIM and PSNR away from the reference view demonstrate a balanced response to horizontal translation while preserving stronger reconstruction quality across the tested range.

C.6 Robustness to vertical camera translation

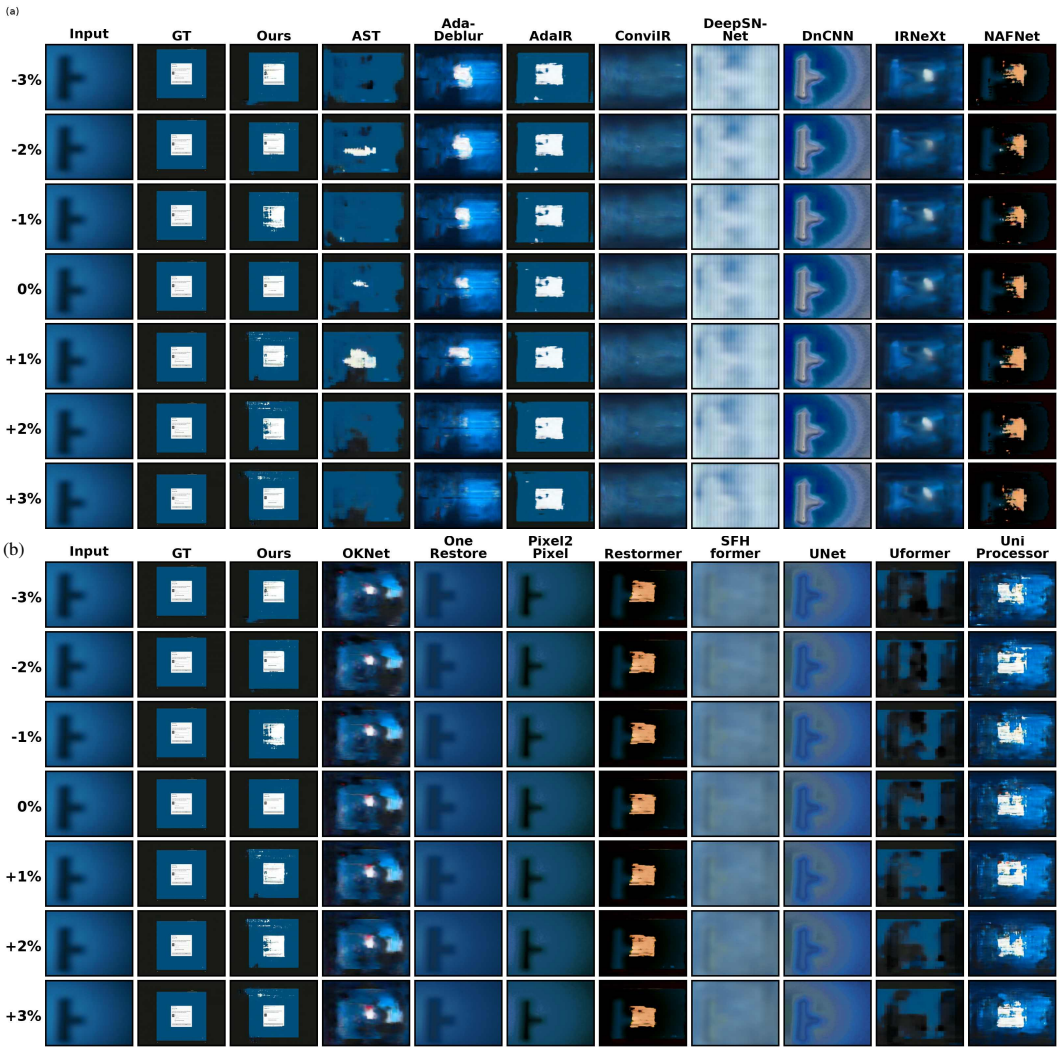
We vary camera height by relative offsets of -3%, -2%, -1%, 0%, $+1\%$, $+2\%$, and $+3\%$. The displayed screen, projection setup, and other imaging conditions remain fixed.



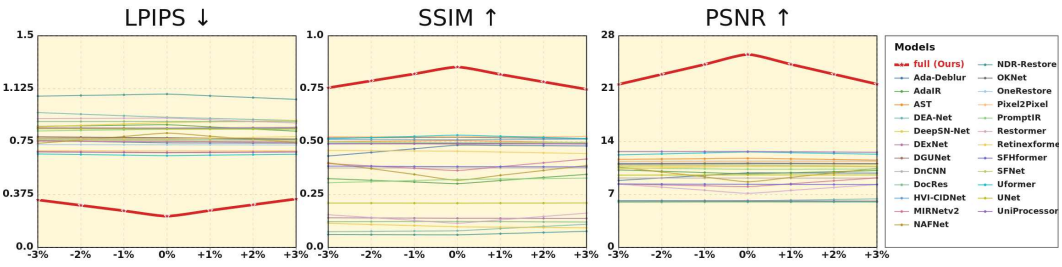
Supplementary Figure 15: OSLock under vertical camera translation. Panels (a,b) compare the baseline groups; rows show perturbation levels. Below: LPIPS (↓), SSIM and PSNR (↑).



On OSLock, OARL retains the central profile region and the main blue and warm background regions across camera heights. Several baselines respond with broad horizontal bands or dark blocks. OARL's low LPIPS and gently varying SSIM and PSNR demonstrate stable recovery of this sparse screen organization on both sides of the reference height.



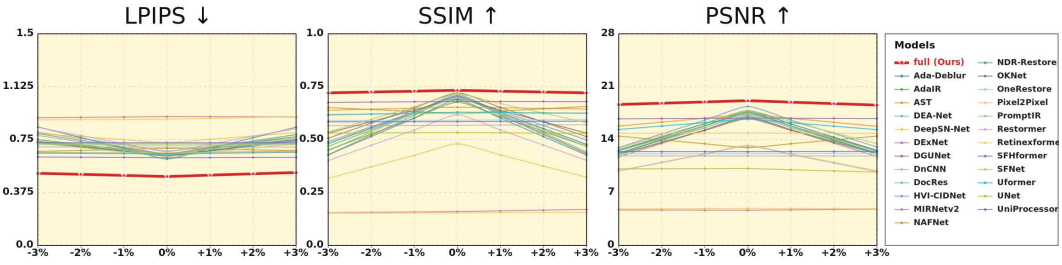
Supplementary Figure 16: Screen under vertical camera translation. Panels (a) and (b) show the two baseline groups. Rows correspond to seven relative heights from -3% to $+3\%$; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



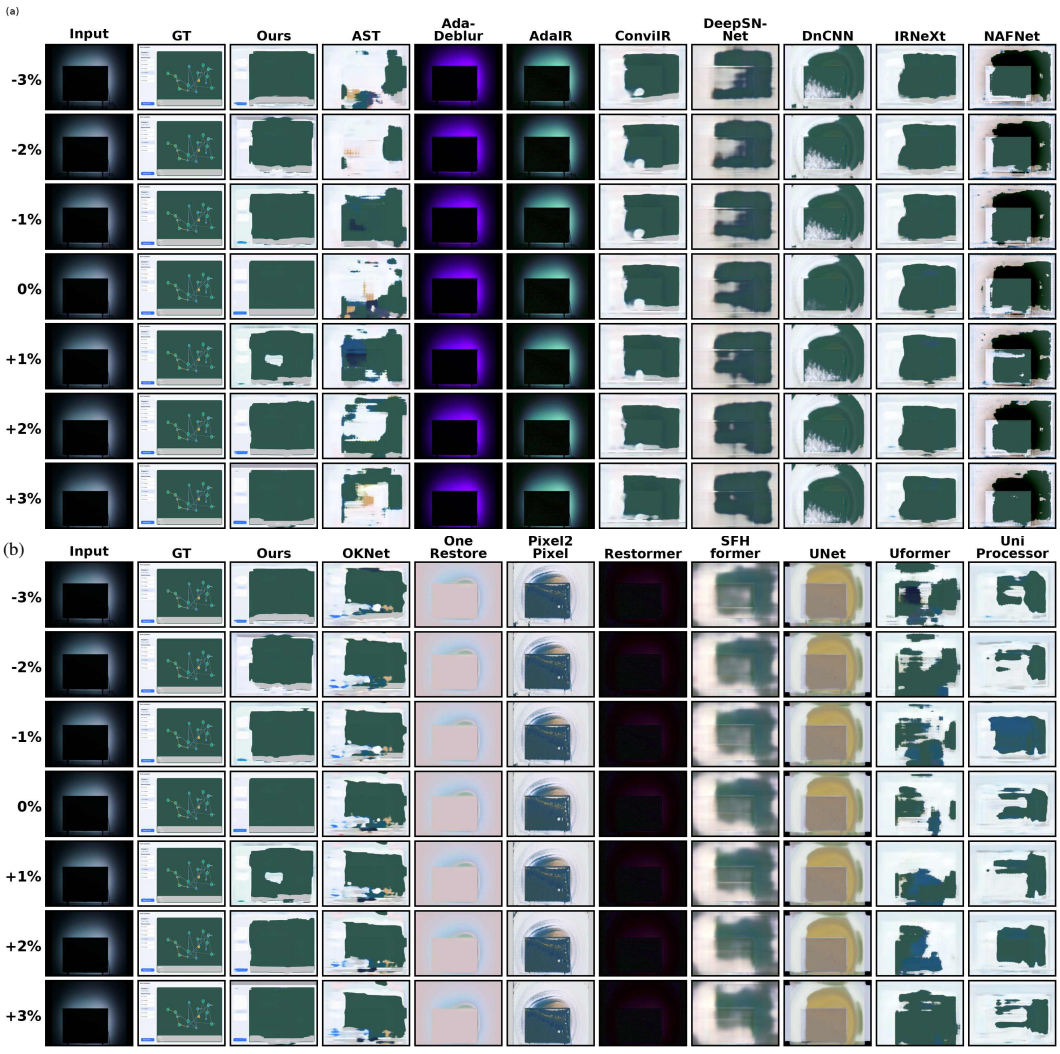
The Screen example contains a small bright dialog within a blue desktop region. OARL localizes this foreground element and preserves its contrast with the surrounding screen across the height sweep. Several competing outputs spread brightness across a much larger area or replace the target layout with a tinted silhouette. The OARL metric curves vary smoothly around the nominal height and remain clearly separated from the baselines at both boundary offsets. These results demonstrate that the foreground location and overall interface organization remain more reliably recoverable with OARL as the camera moves vertically.



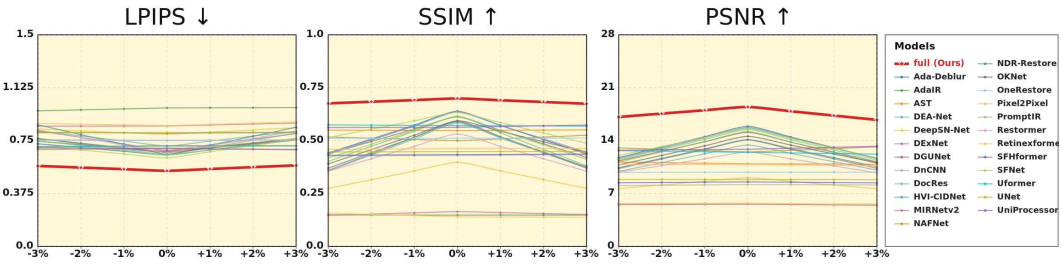
Supplementary Figure 17: Figure under vertical camera translation. Panels (a) and (b) show the two baseline groups. Rows correspond to seven relative heights from -3% to $+3\%$; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



On Figure, OARL retains the green and orange bar groups against the light chart background. The recovered colored regions preserve the ordering and relative prominence of the bars more clearly than the nearly uniform fields produced by several baselines. Its SSIM curve remains close to the nominal value across the height sweep, and the LPIPS and PSNR curves retain the same favorable ranking. This agreement demonstrates that the bar structure remains better organized under vertical displacement, with balanced reconstruction quality for camera positions above and below the reference height.



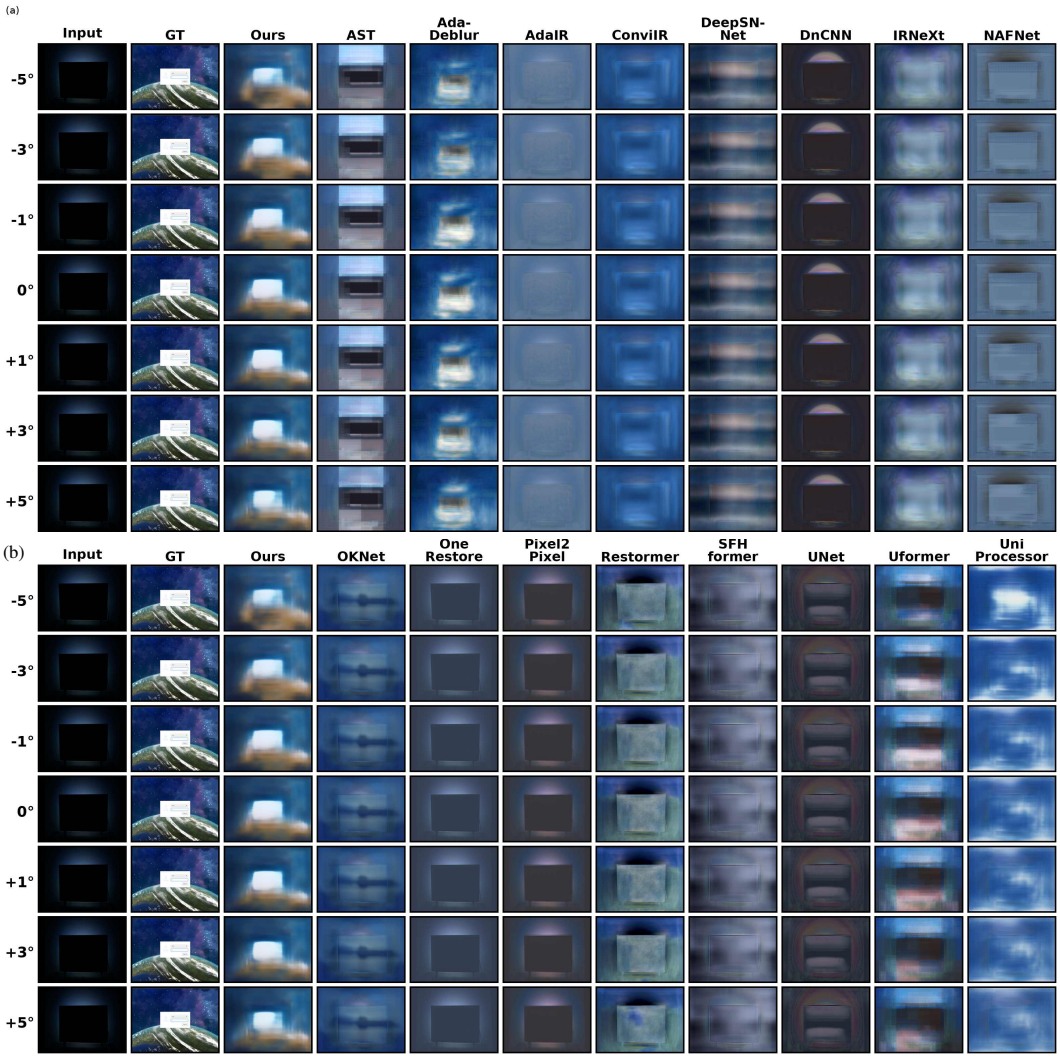
Supplementary Figure 18: Game under vertical camera translation. Panels (a) and (b) show the two baseline groups. Rows correspond to seven relative heights from -3% to $+3\%$; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



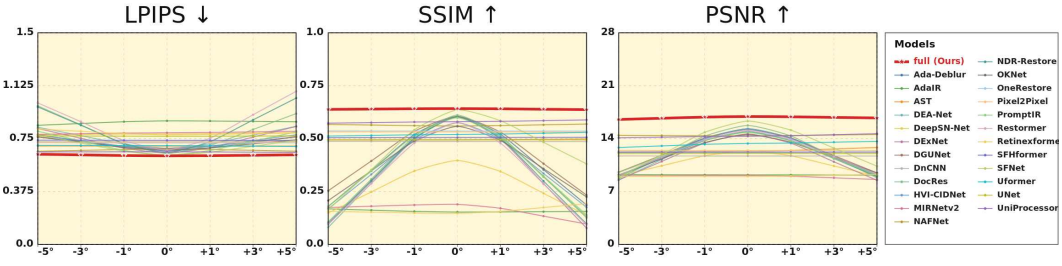
On Game, the target combines a green playing region with a light control strip on the left. OARL preserves this large-scale division and produces a more contiguous field than several competing methods, which introduce broad holes or displace the interface boundary. The quantitative curves remain smooth over the vertical offsets, with lower LPIPS and higher SSIM and PSNR throughout the comparison. Their similar response on either side of the nominal position supports the visual evidence of stable regional alignment, demonstrating robust reconstruction as camera height changes.

C.7 Robustness to vertical viewing-angle changes

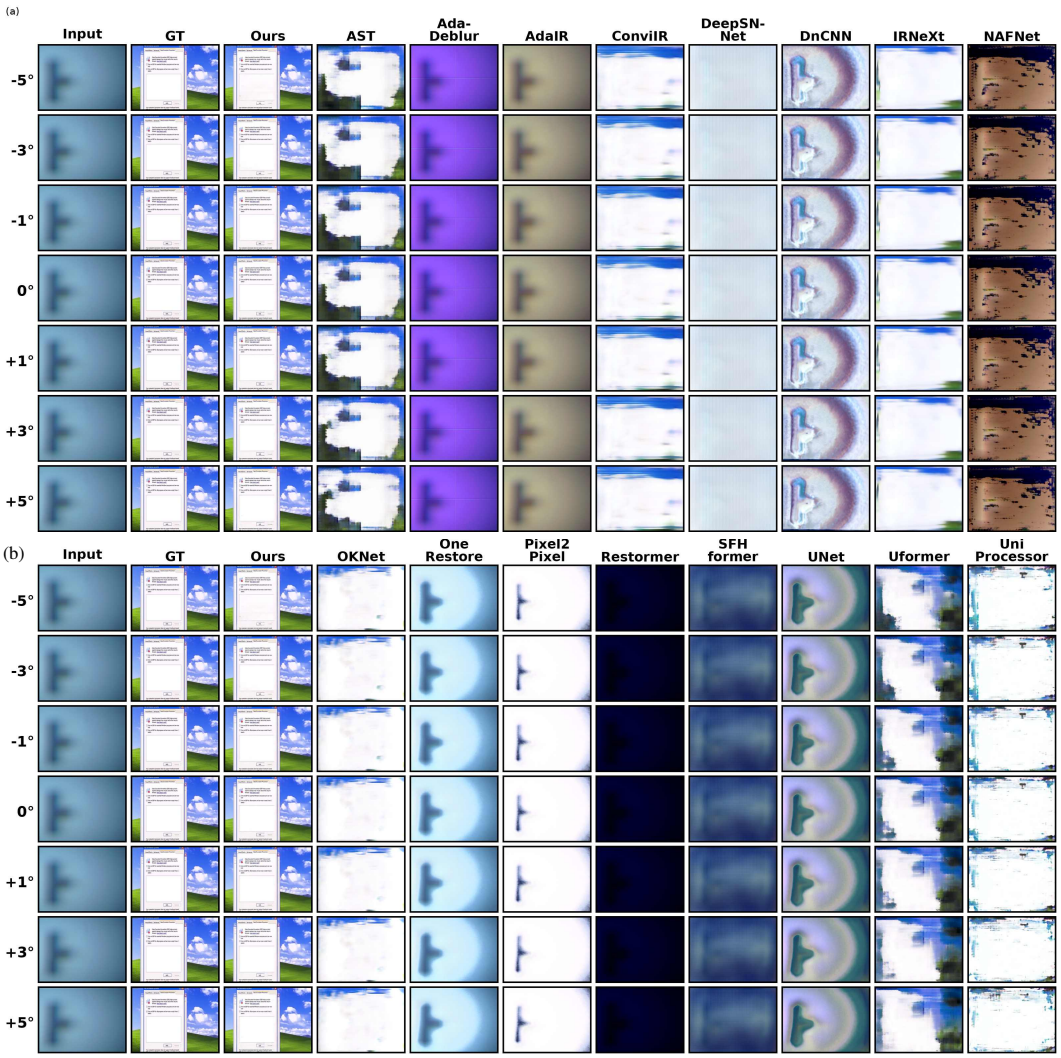
We vary camera pitch over $-5°$, $-3°$, $-1°$, $0°$, $+1°$, $+3°$, and $+5°$, holding the camera distance and all remaining projection and imaging parameters constant.



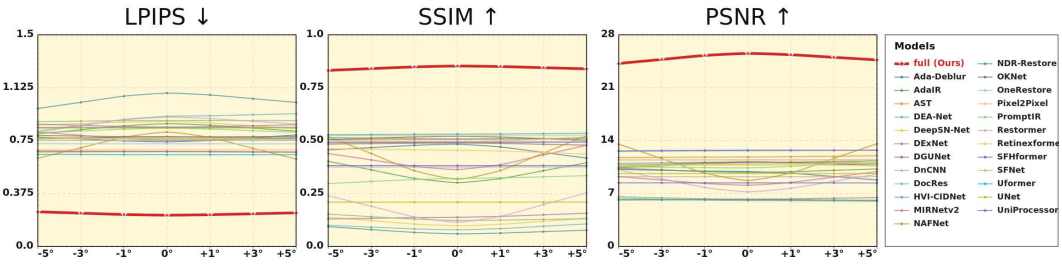
Supplementary Figure 19: OSLock under vertical viewing-angle changes. Panels (a,b) compare the baseline groups; rows show perturbation levels. Below: LPIPS (↓), SSIM and PSNR (↑).



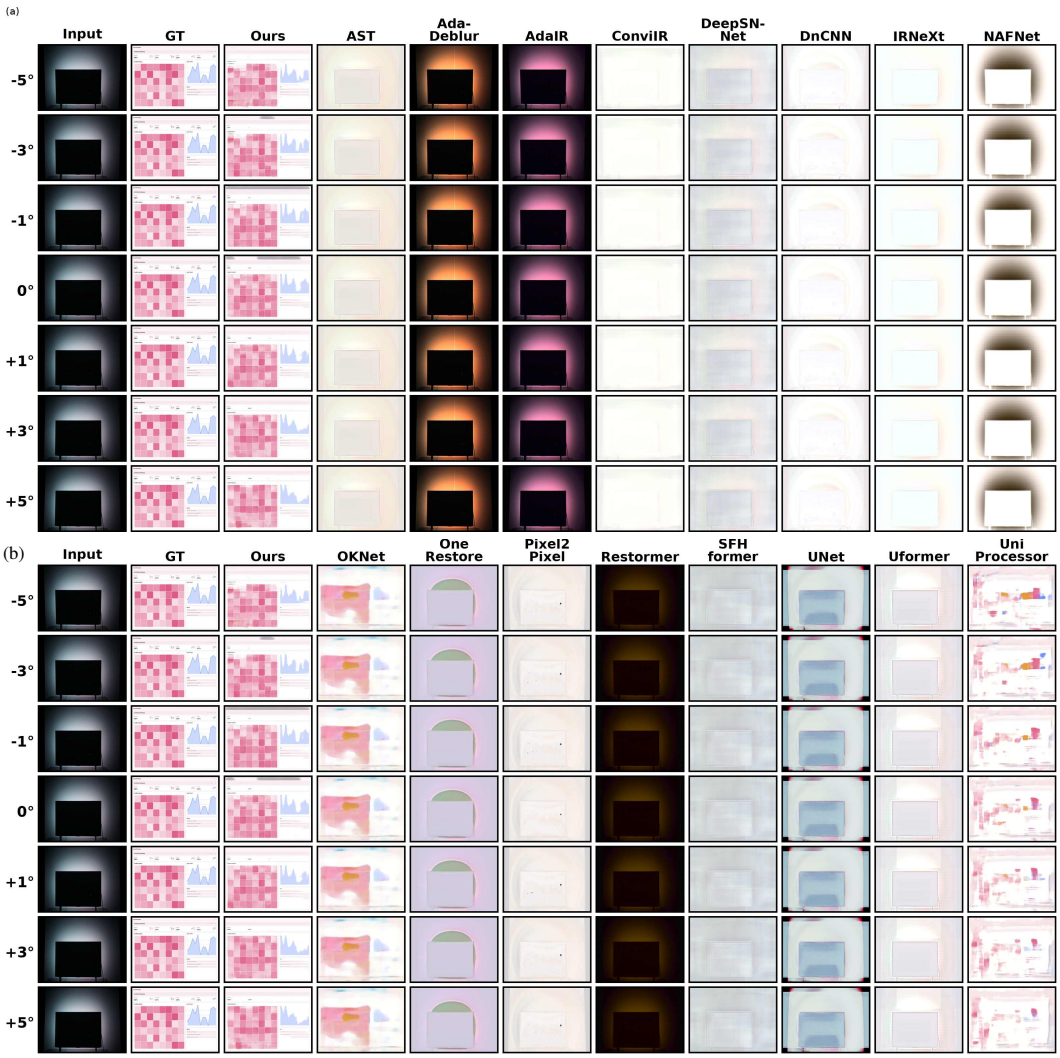
On OSLock, OARL recovers the central bright dialog above the warm mountain region across the pitch sweep. Several baselines substitute broad bands or dark rectangular structures. The nearly flat OARL metric profiles preserve the leading ranking in both pitch directions, demonstrating strong tolerance to vertical changes in the observed wall glow.



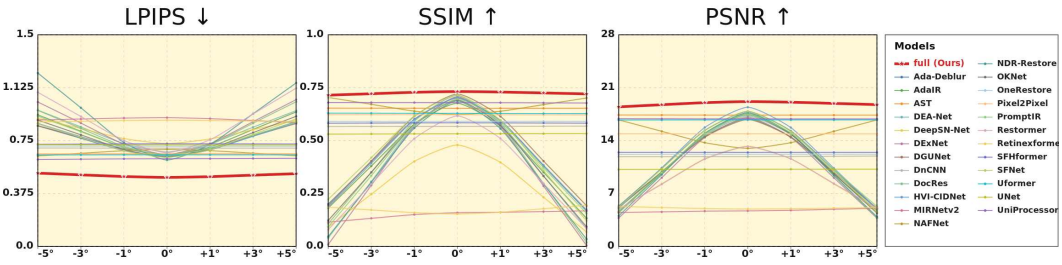
Supplementary Figure 20: Screen under vertical viewing-angle changes. Panels (a) and (b) show the two baseline groups. Rows correspond to seven pitch angles from $-5°$ to $+5°$; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



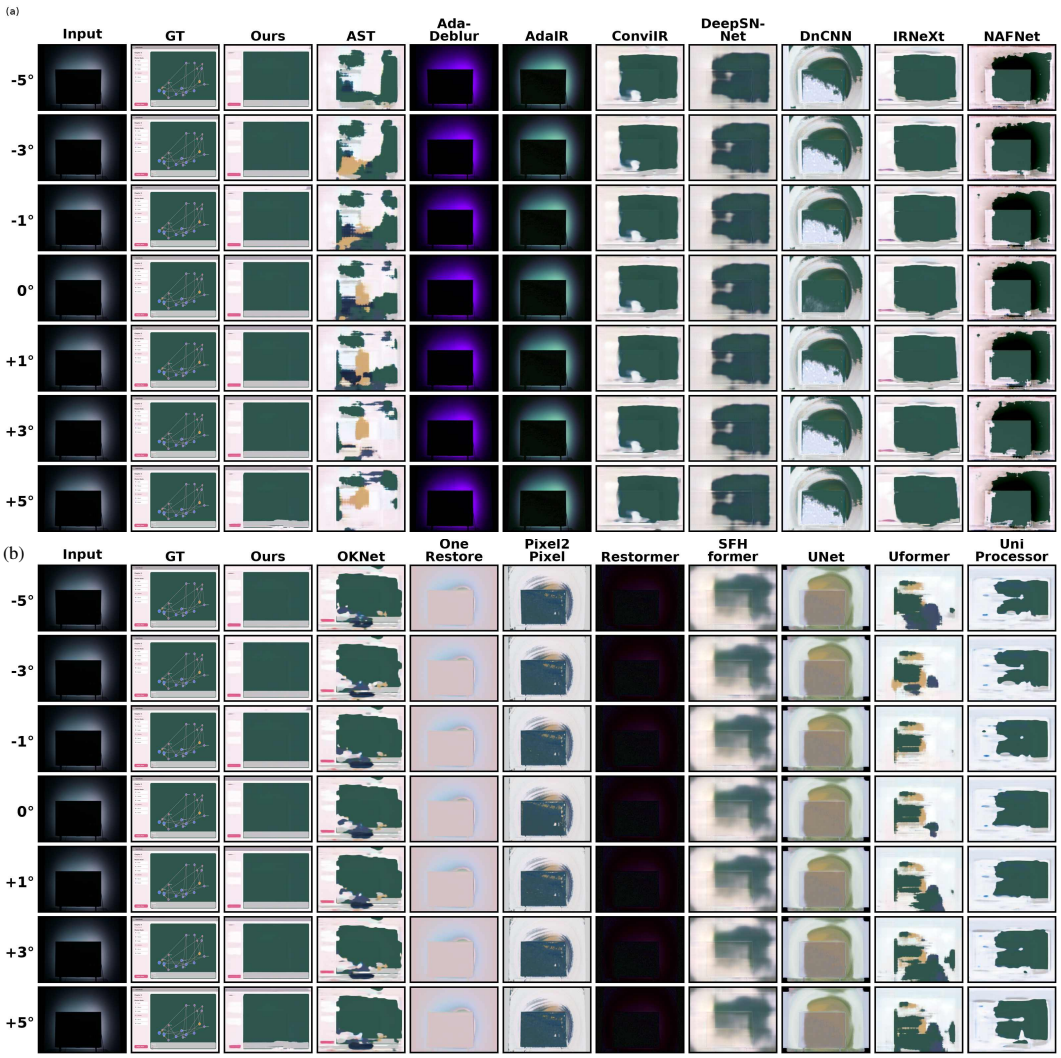
On Screen, OARL preserves the white foreground window on the left and the blue-sky and green-hill wallpaper on the right. This arrangement remains recognizable across both positive and negative pitch angles. Several baselines merge the window into the wallpaper or produce large tinted regions. OARL's LPIPS stays low and its SSIM and PSNR remain well above the competing curves throughout the sweep. The small change from the nominal view agrees with the stable window-to-wallpaper boundary, demonstrating reliable recovery of a composite desktop layout under camera pitch variation.



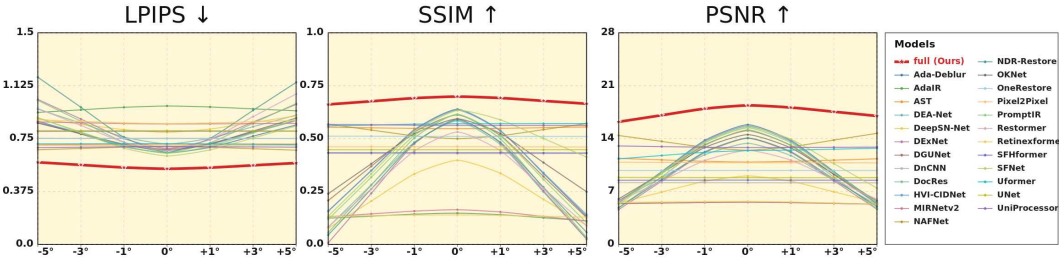
Supplementary Figure 21: Figure under vertical viewing-angle changes. Panels (a) and (b) show the two baseline groups. Rows correspond to seven pitch angles from -5° to $+5^\circ$; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



The Figure example combines a red and pink matrix with a separate blue plot region on a light background. OARL preserves the matrix pattern and the division between the two graphical components across pitch angles. Several competing outputs lose this organization and become nearly uniform or retain broad input-related contours. The OARL metric curves remain close to their reference values on both sides of the sweep, while several baselines show much larger structural changes. These results demonstrate strong preservation of distinct chart regions under vertical viewpoint perturbations.



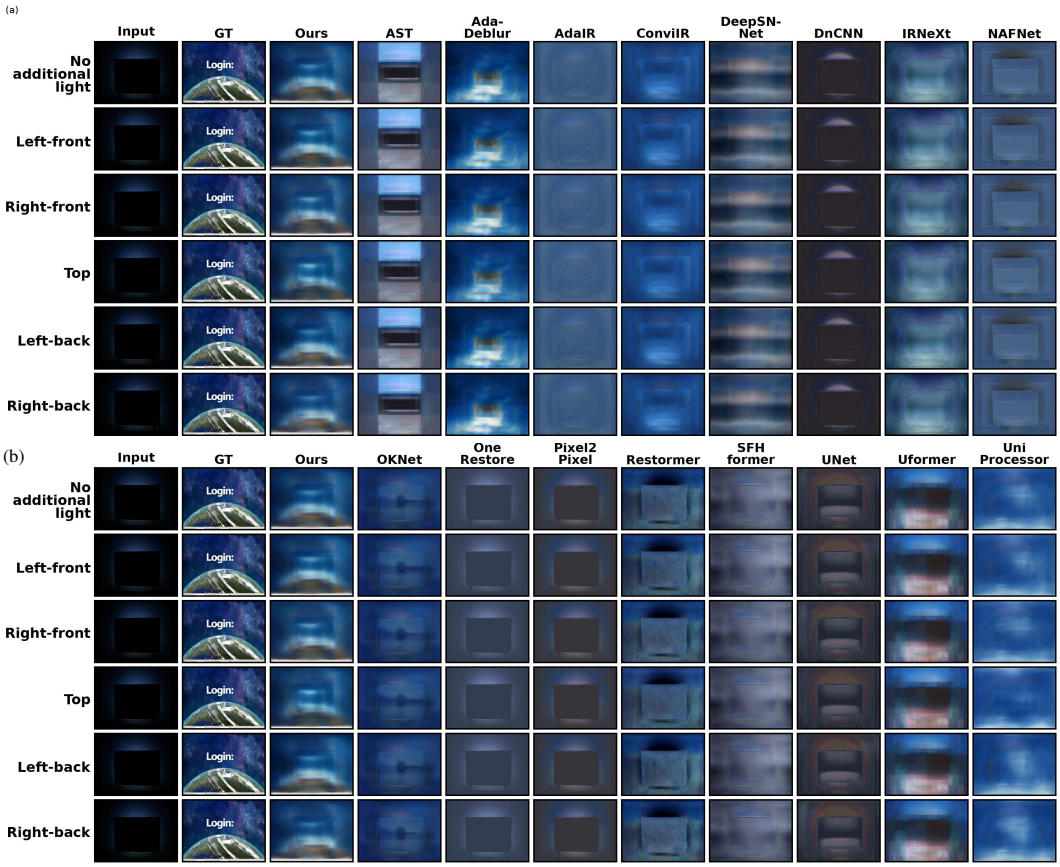
Supplementary Figure 22: Game under vertical viewing-angle changes. Panels (a) and (b) show the two baseline groups. Rows correspond to seven pitch angles from $-5°$ to $+5°$; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



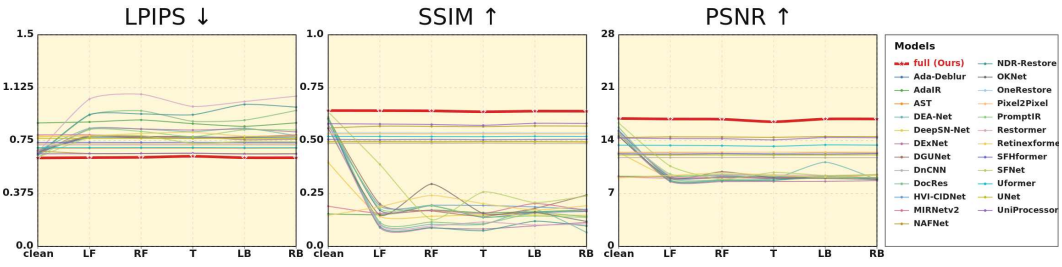
On Game, OARL maintains the dark green field, the light pink control strip on the left, and their rectangular arrangement under pitch changes. Several competing outputs contain wide gaps or curved structures that disrupt the field boundary. OARL's lower LPIPS and higher SSIM and PSNR persist at both ends of the angular range. Their balanced variation around the nominal pitch agrees with the continuity of the recovered playing region. The comparison demonstrates that OARL preserves the main interface geometry more effectively as the camera tilts upward or downward.

C.8 Robustness to directional illumination

Directional lighting changes the spatial distribution of illumination while the target screen remains fixed. We compare no additional light with left-front, right-front, top, left-back, and right-back illumination. These six conditions change light direction while preserving the displayed screen, projection geometry, and evaluation protocol.



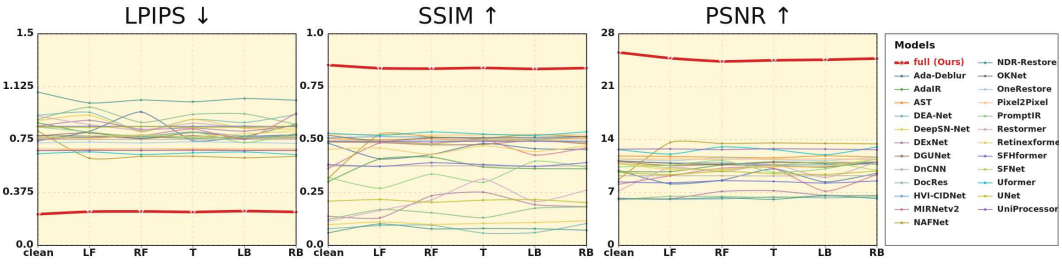
Supplementary Figure 23: OSLock under directional illumination. Panels (a) and (b) show the two baseline groups. Rows correspond to no additional light, left-front, right-front, top, left-back, and right-back lighting; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



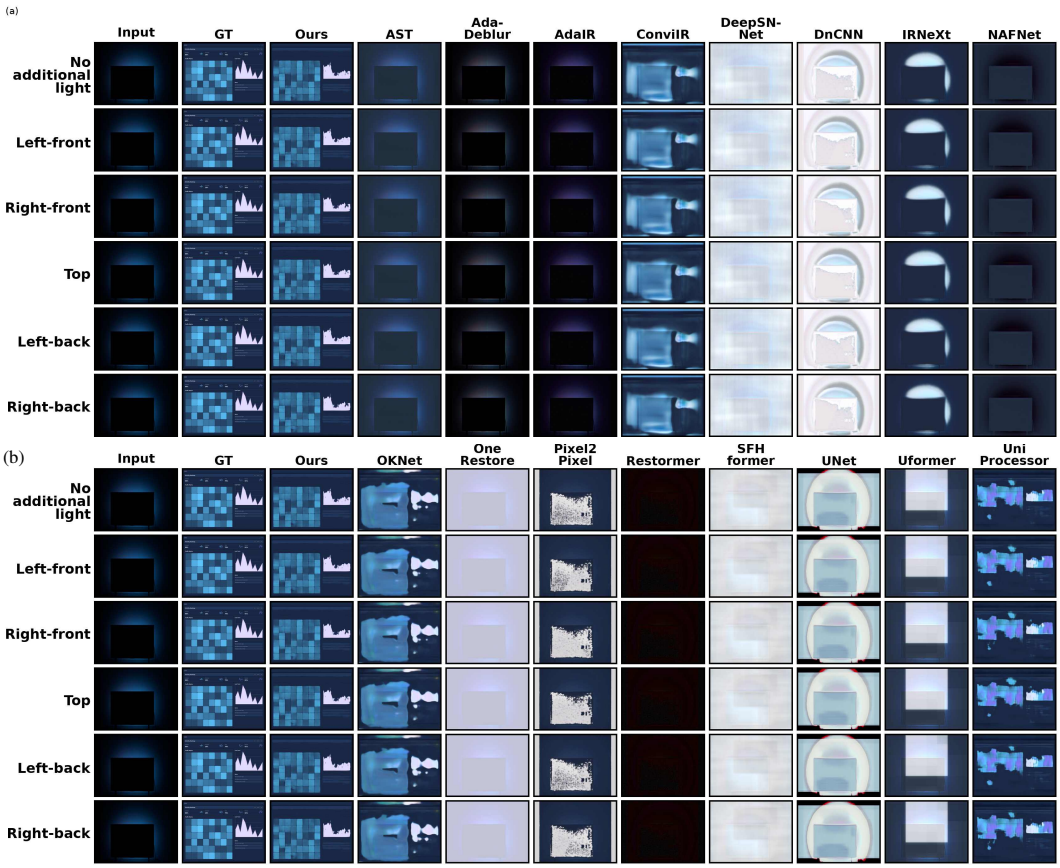
On OSLock, OARL preserves the blue upper background and the bright curved mountain region near the bottom as the light source changes direction. Several baselines respond with broad horizontal bands, dark blocks, or displaced bright patches. The recovered brightness boundary remains more closely tied to the target screen in OARL's outputs. Its LPIPS and SSIM curves show little directional variation, and PSNR retains a clear lead across the six settings. The agreement between this regional structure and the metric profiles demonstrates reliable lock-screen recovery under spatially varying illumination.



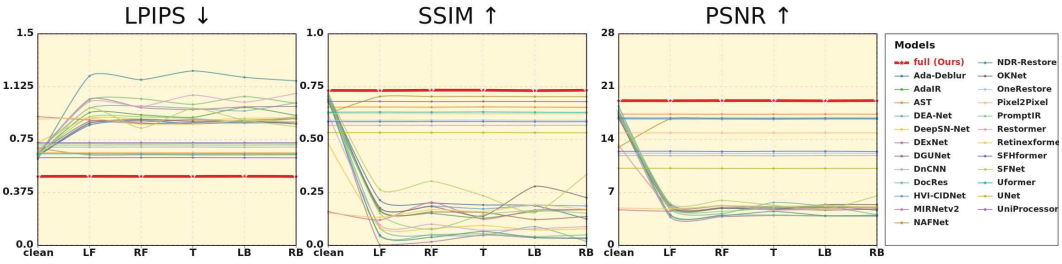
Supplementary Figure 24: Screen under directional illumination. Panels (a) and (b) show the two baseline groups. Rows correspond to no additional light, left-front, right-front, top, left-back, and right-back lighting; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



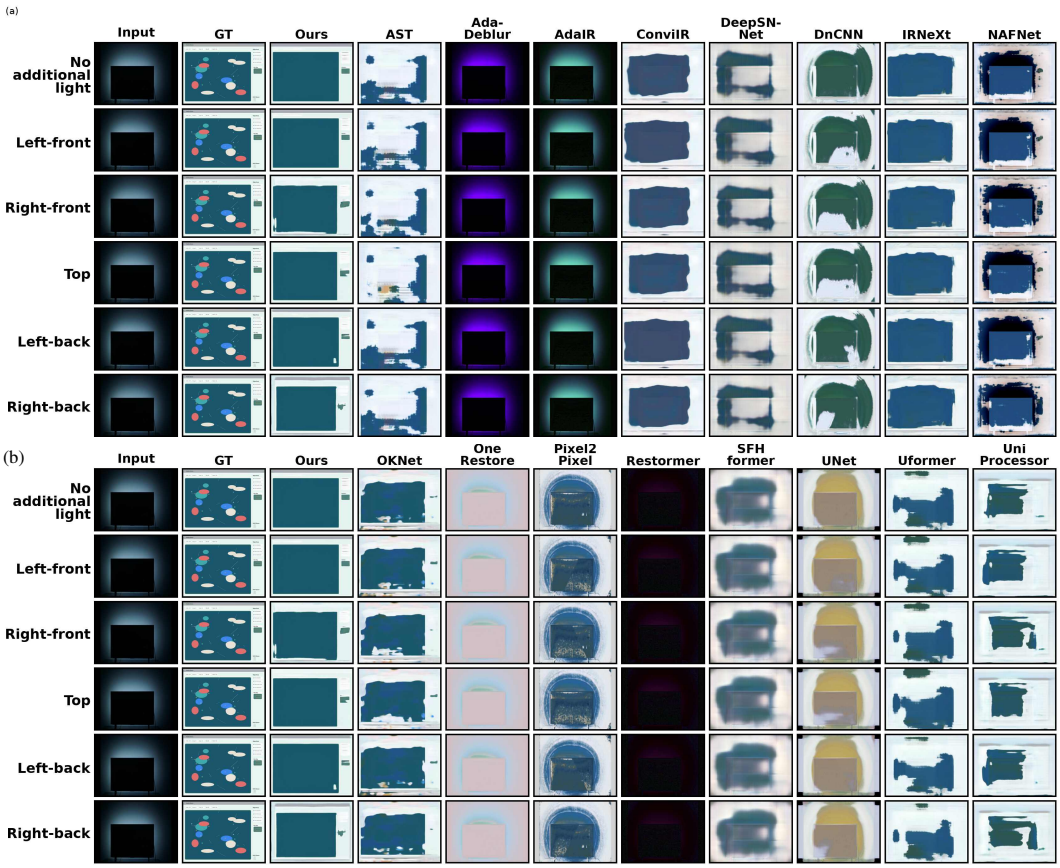
On Screen, OARL retains the central bright application window, its red upper strip, and the dark desktop around it under each lighting direction. The placement of the window and the division of its light interior remain more faithful than the diffuse or strongly tinted outputs of several baselines. This advantage extends from front lighting to top and rear lighting, where the observation changes substantially. The quantitative curves show consistently low LPIPS together with high SSIM and PSNR; their small directional variation agrees with the stable foreground layout. These results demonstrate that OARL reconstructs the localized application content reliably as the illumination distribution changes, maintaining a clear distinction between the target window and the surrounding screen. The clean dark area around the application further distinguishes OARL from methods that spread the foreground response into the desktop. The preserved light interior and colored upper edge provide complementary evidence of window localization, supporting the quantitative advantage across the different source directions.



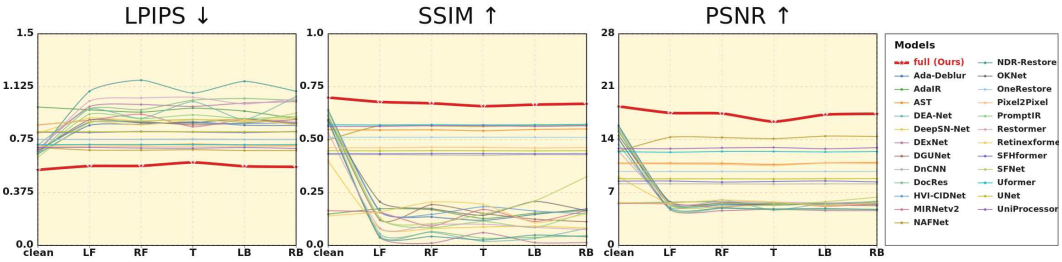
Supplementary Figure 25: Figure under directional illumination. Panels (a) and (b) show the two baseline groups. Rows correspond to no additional light, left-front, right-front, top, left-back, and right-back lighting; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



The Figure example contains a blue matrix on the left and a pale violet plot on the right against a dark background. OARL retains the matrix cells and the separation of the two graphical regions across the lighting directions. Several competing outputs instead emphasize broad rectangular patches or curved illumination contours. The OARL curves remain nearly constant in all three metrics, whereas several baselines show a pronounced structural decline once directional light is added. The retained grid pattern therefore agrees with a persistent aggregate advantage. These results demonstrate that the spatial arrangement of the chart remains recoverable with OARL under front, overhead, and rear illumination, with consistent preservation of the target's contrasting graphical regions. The blue cells occupy the same principal plotting region as in the target, and the pale plot remains separated from the matrix. This spatial relationship gives a direct visual counterpart to the stable SSIM response: the reconstructed graphical components retain their distinct positions as the source moves between lighting directions.



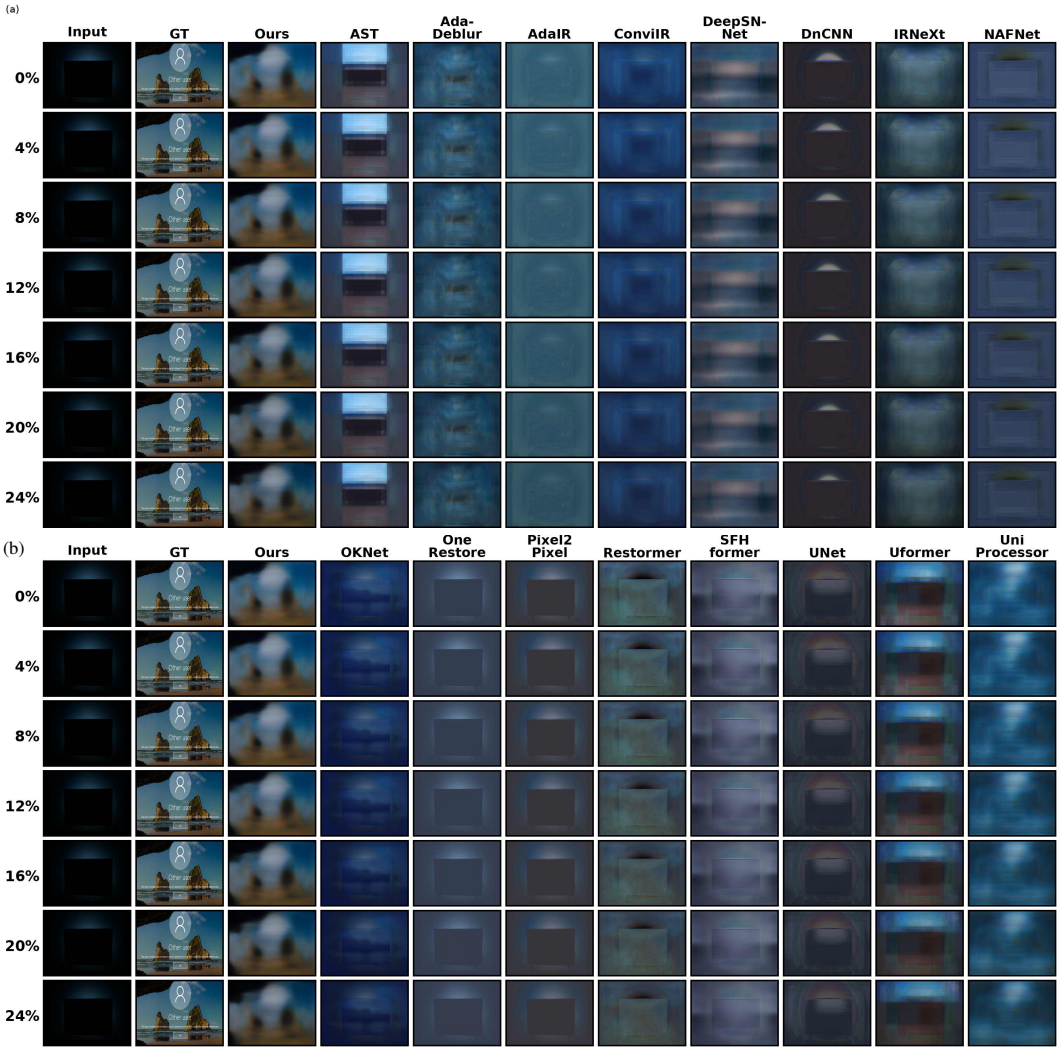
Supplementary Figure 26: Game under directional illumination. Panels (a) and (b) show the two baseline groups. Rows correspond to no additional light, left-front, right-front, top, left-back, and right-back lighting; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



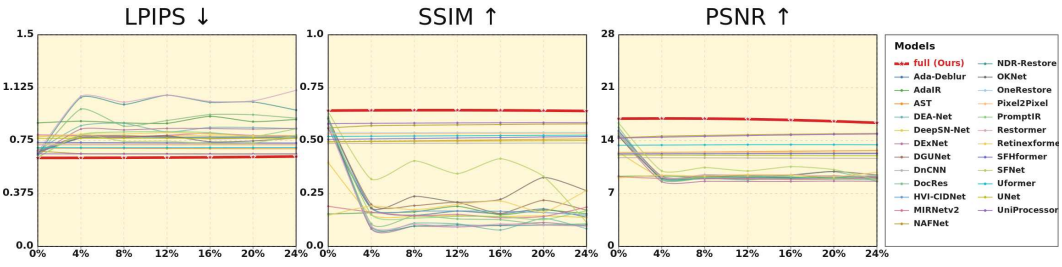
On Game, OARL recovers the broad teal field and the light sidebar as distinct regions under all six lighting conditions. Several competing outputs introduce large holes within the field or distort its outer contour, weakening the rectangular organization of the interface. OARL maintains a cleaner field-to-sidebar boundary as the illumination direction changes. Its SSIM and PSNR remain above the baseline curves, while LPIPS stays at a favorable level throughout the comparison. The top-light condition produces a visible metric response, yet OARL retains the same leading ranking. This agreement between coherent regional reconstruction and aggregate quality demonstrates robust recovery of the principal game layout under changes in the spatial distribution of incident light. The light sidebar remains concentrated along the side of the playing field, giving the reconstructed interface a more faithful overall footprint. The comparison with methods that introduce broad interior gaps is particularly clear in this example, where the continuity of the teal region directly exposes differences in recovered screen organization.

C.9 Robustness to fine-grained brightness attenuation

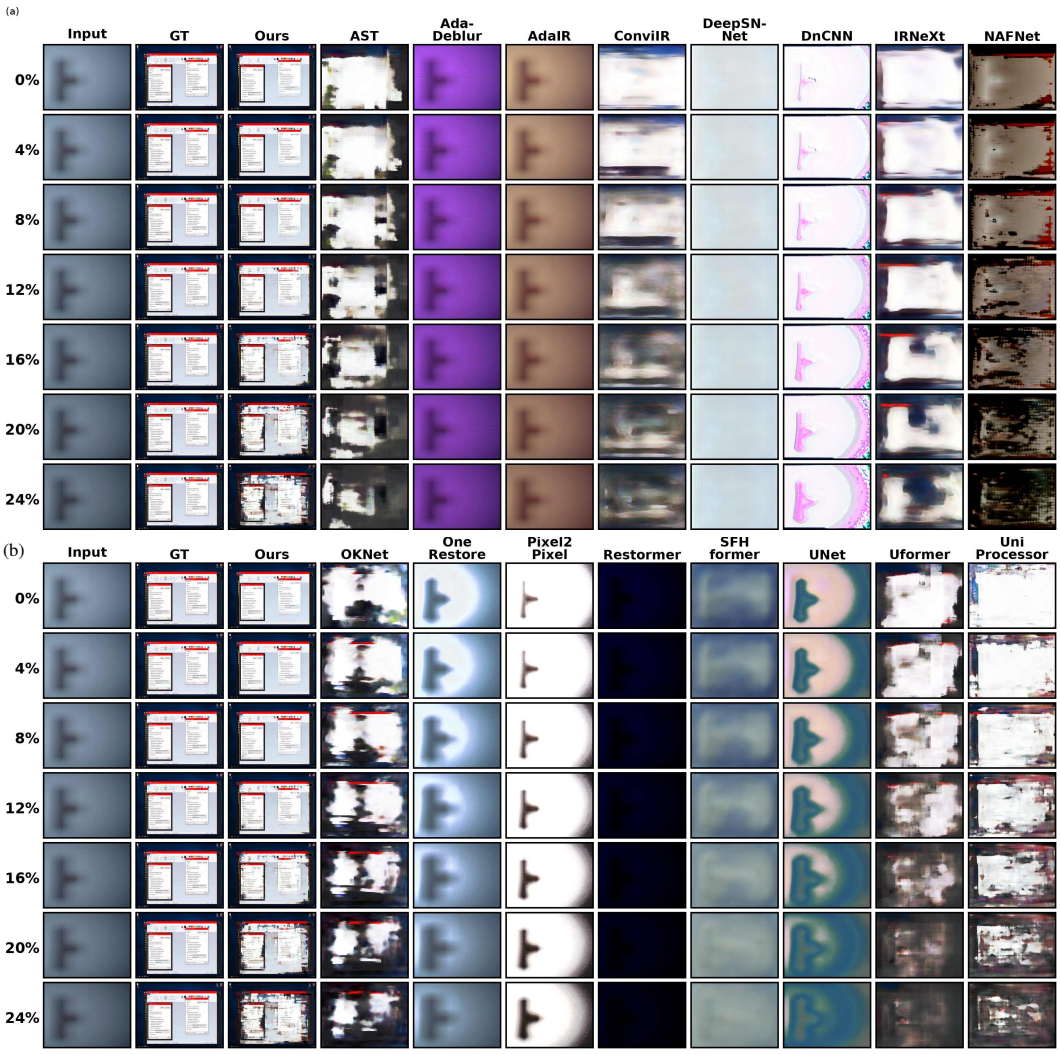
Progressive darkening tests how much screen organization remains recoverable as the recorded glow weakens. We attenuate input brightness by 0%, 4%, 8%, 12%, 16%, 20%, and 24%. Larger values indicate stronger darkening; the displayed content and all other acquisition factors remain fixed.



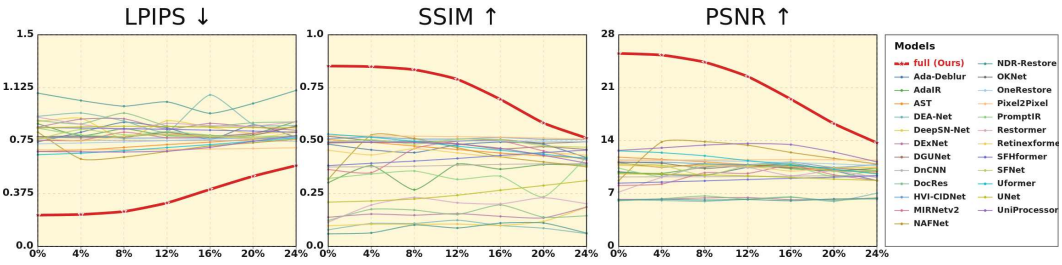
Supplementary Figure 27: OSLock under progressive brightness attenuation. Panels (a,b) compare the baseline groups; rows show perturbation levels. Below: LPIPS (↓), SSIM and PSNR (↑).



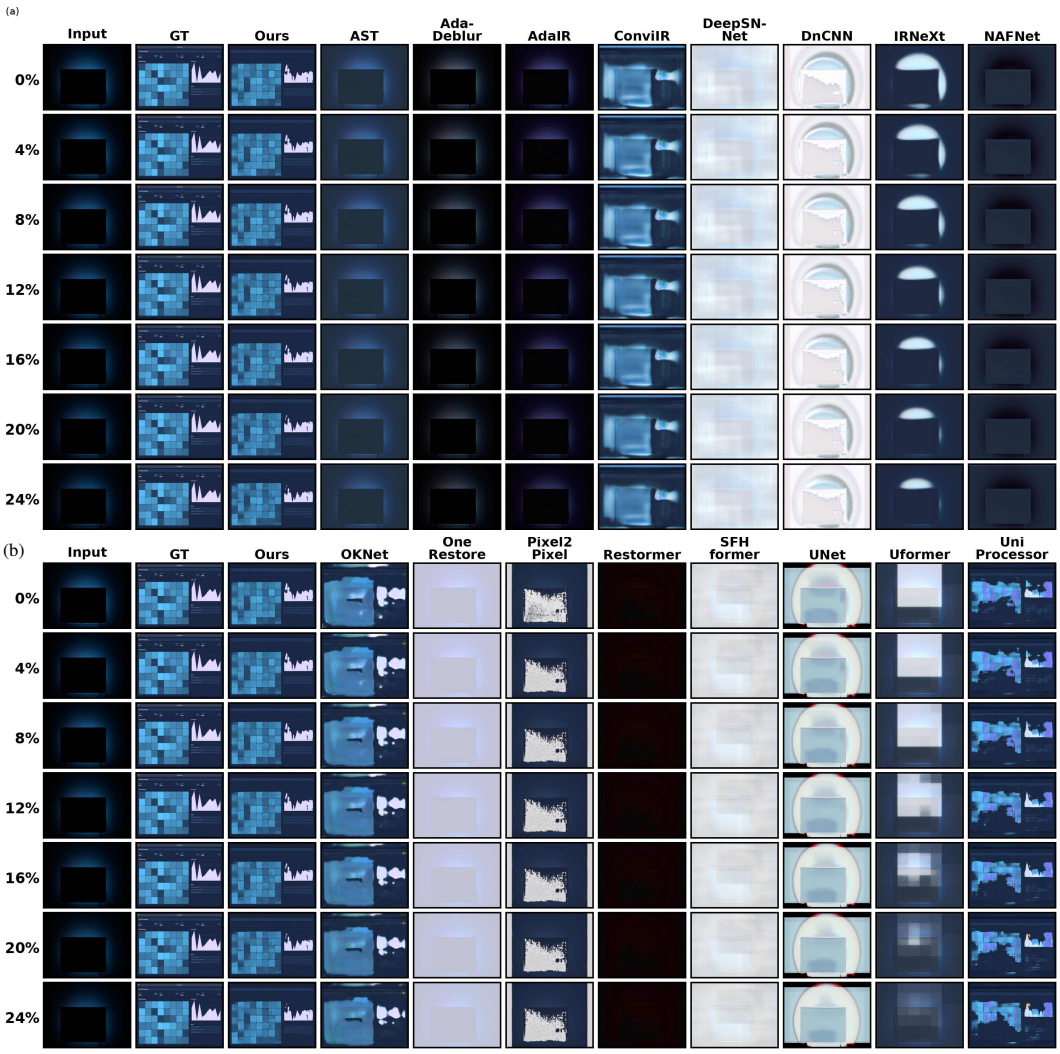
On OSLock, OARL preserves the profile region and the blue-to-warm background organization as the input is darkened. Several competing methods return broad bands or nearly uniform regions. Its stable LPIPS and SSIM and gradually changing PSNR show that the dominant lock-screen structure remains reliably represented across the attenuation range.



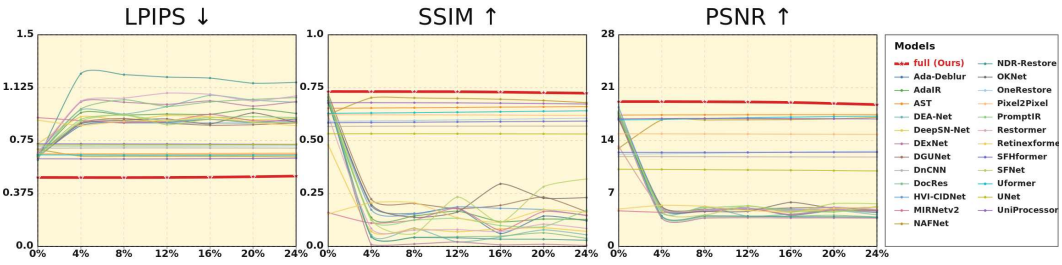
Supplementary Figure 28: Screen under progressive brightness attenuation. Panels (a) and (b) show the two baseline groups. Rows correspond to seven attenuation levels from 0% to 24%; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



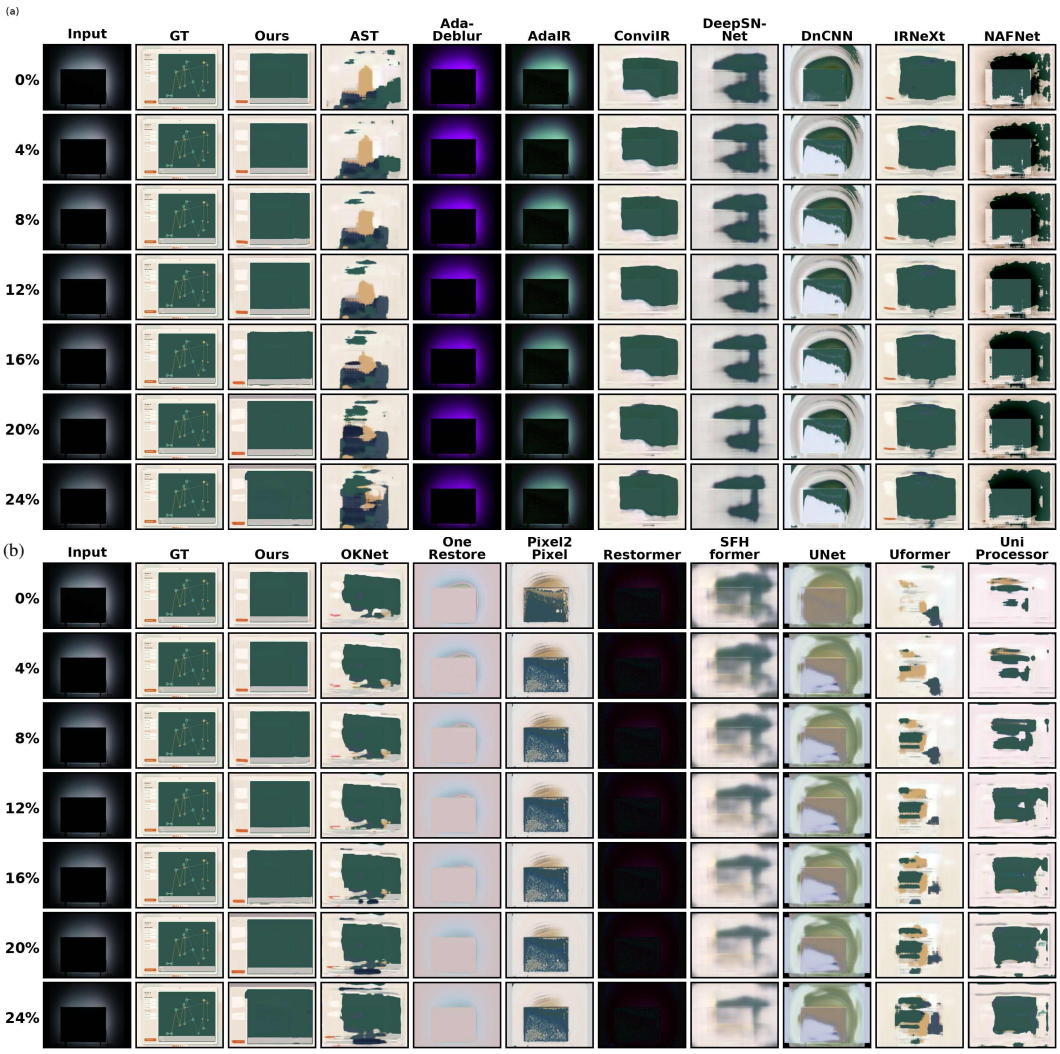
The Screen panels contain two tall, bright application regions with a colored upper border. OARL recovers their vertical arrangement clearly at low and intermediate attenuation, with stronger foreground separation than the diffuse outputs of several baselines. Its LPIPS increases and SSIM and PSNR decrease as the observation becomes darker, while the curves retain a clear comparative advantage over most of the sweep. The strong reconstruction of the paired application regions agrees with this aggregate response, demonstrating OARL's ability to retain useful interface structure as the available photometric signal is reduced.



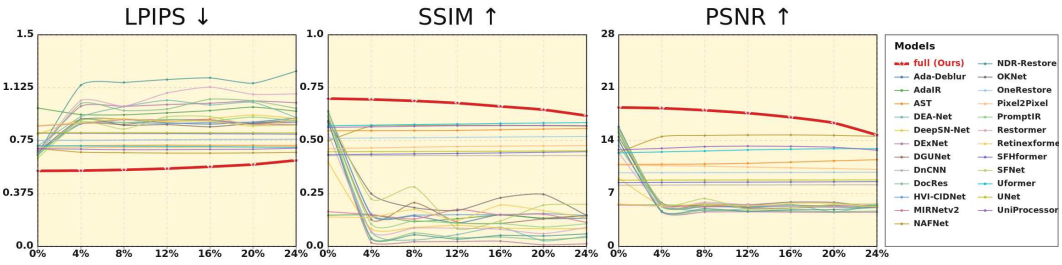
Supplementary Figure 29: Figure under progressive brightness attenuation. Panels (a) and (b) show the two baseline groups. Rows correspond to seven attenuation levels from 0% to 24%; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



On Figure, OARL preserves the blue matrix and the separate violet plot region against the dark background as brightness is reduced. Several competing methods lose the graphical organization and return large uniform blocks or unrelated contours. The matrix pattern remains distinguishable in OARL's outputs across the attenuation levels. Its LPIPS, SSIM, and PSNR curves also remain nearly flat and retain the leading ranking, in contrast to the substantial structural decline of several baselines after darkening. The visual and quantitative comparisons jointly demonstrate strong robustness of the recovered chart structure to reduced input brightness.



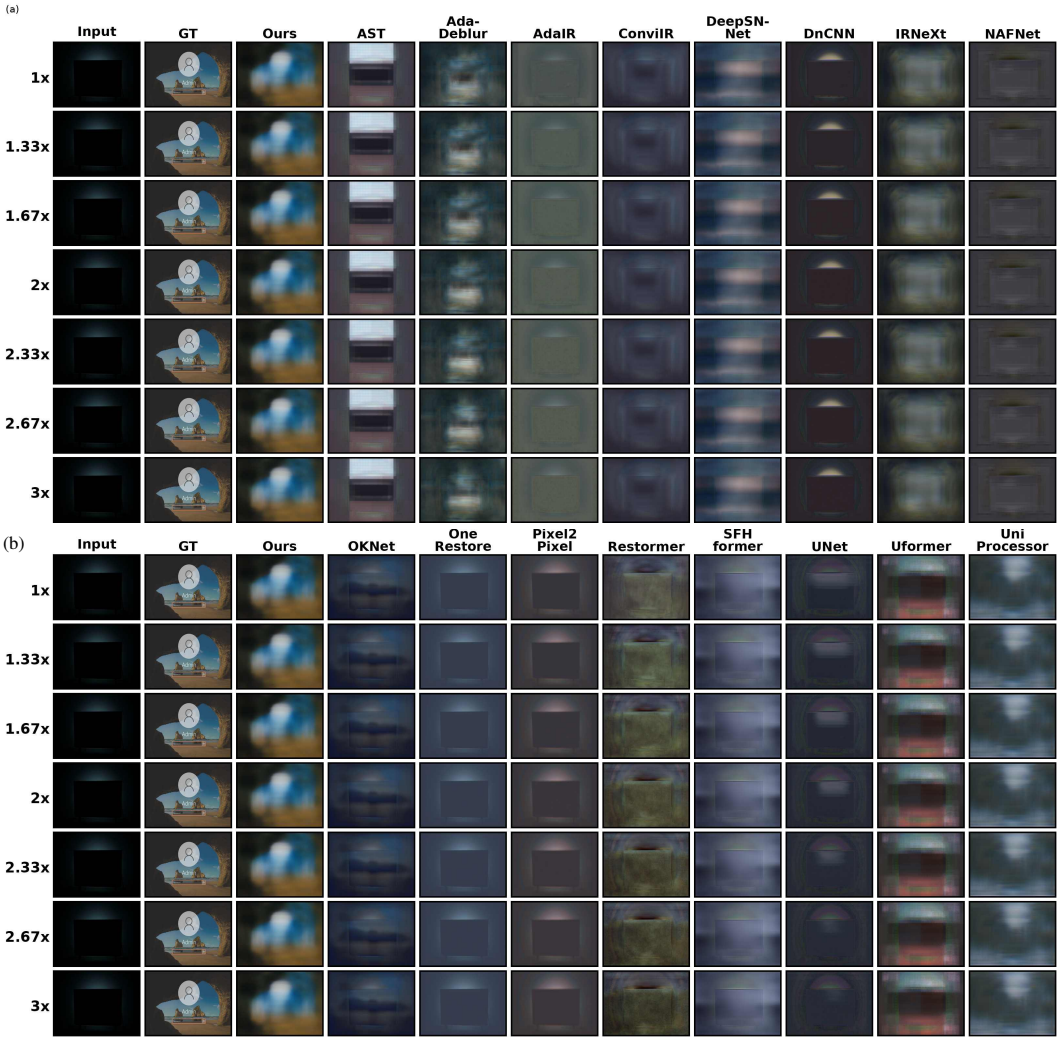
Supplementary Figure 30: Game under progressive brightness attenuation. Panels (a) and (b) show the two baseline groups. Rows correspond to seven attenuation levels from 0% to 24%; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



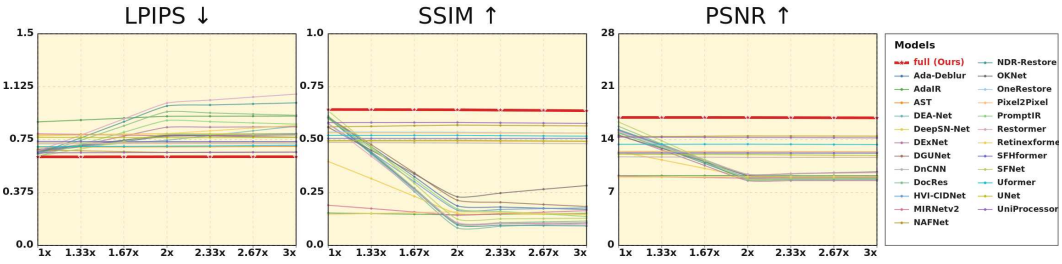
On Game, OARL reconstructs the green field together with the cream control strip on the left and the colored lower tab. Several competing methods introduce broad gaps within the field or merge it with the surrounding interface. OARL preserves a more coherent regional division as the observation is darkened. Its metric curves change gradually with attenuation and remain favorable at the stronger settings, linking the retained screen layout to superior aggregate quality. These results demonstrate robust reconstruction of the dominant game regions when the strength of the indirect observation is progressively reduced.

C.10 Robustness to camera-to-wall distance changes

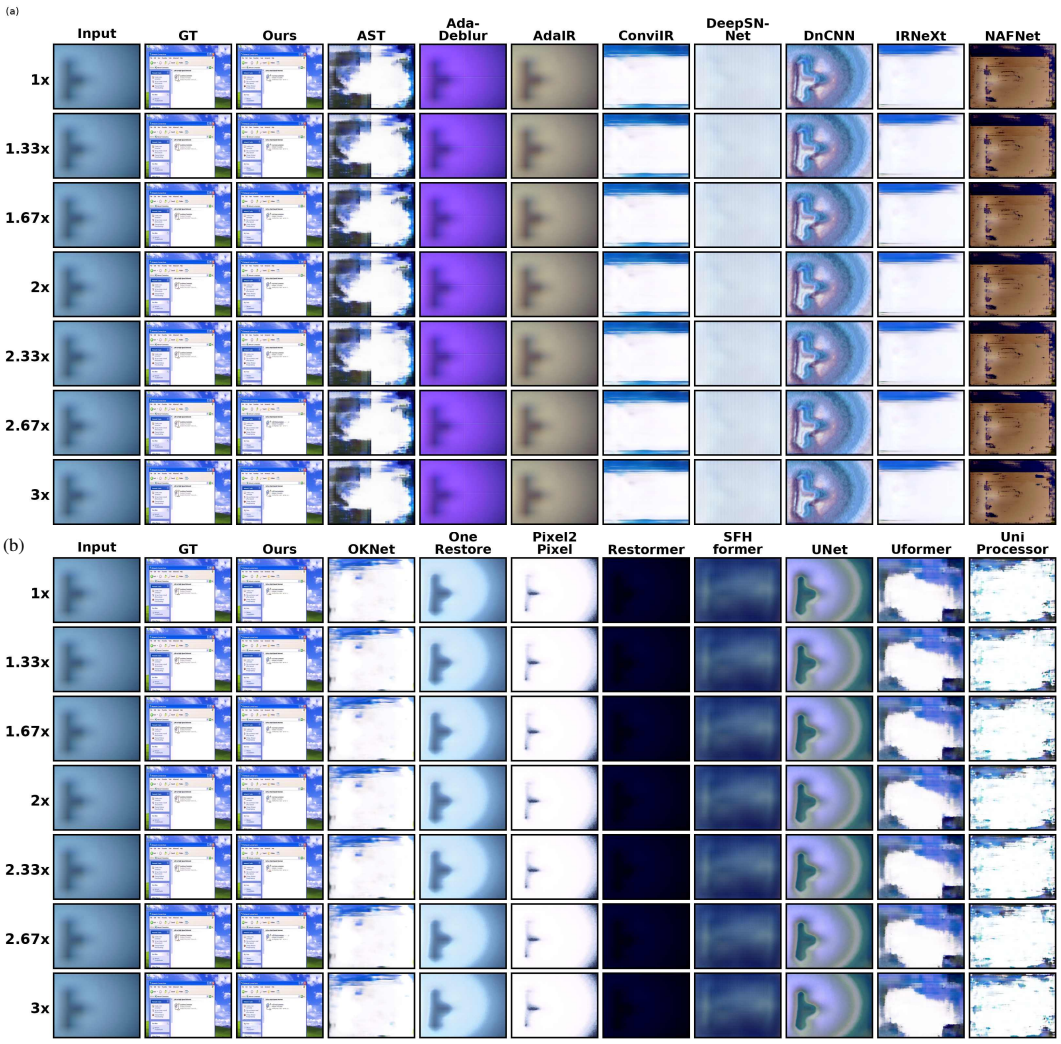
Changing camera-to-wall distance tests recovery across observation scales. We vary camera-to-wall distance by factors of 1, 1.33, 1.67, 2, 2.33, 2.67, and 3 relative to the reference distance. The displayed screen and reconstruction protocol remain unchanged.



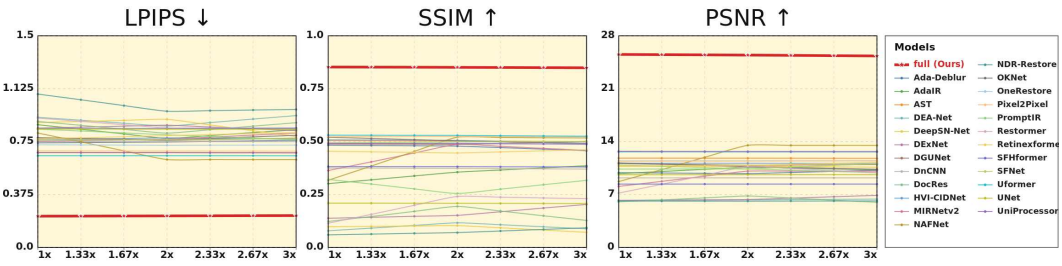
Supplementary Figure 31: OSLock under camera-to-wall distance changes. Panels (a,b) compare the baseline groups; rows show perturbation levels. Below: LPIPS (↓), SSIM and PSNR (↑).



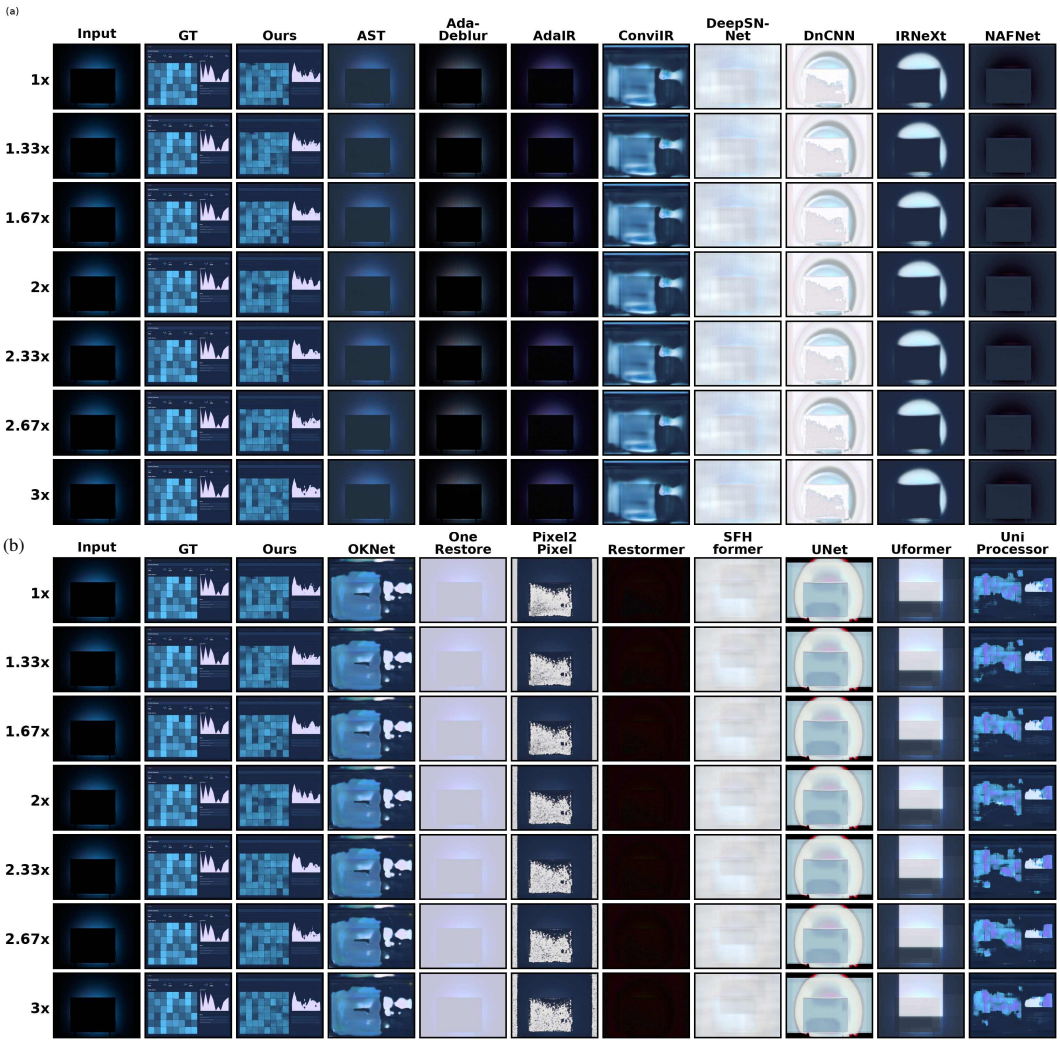
On OSLock, OARL retains the central profile region and the main background color bands across camera-to-wall distances. Several baselines introduce stronger rectangular or banded distortions. The nearly constant OARL metric curves preserve their advantage through the farthest setting, demonstrating reliable recovery of the lock-screen layout despite changes in observation scale.



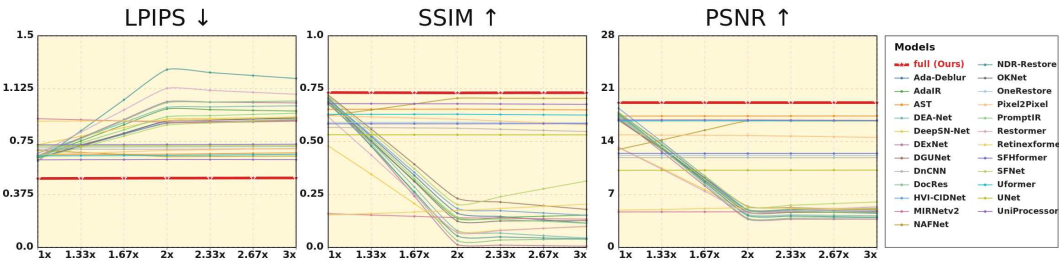
Supplementary Figure 32: Screen under camera-to-wall distance changes. Panels (a) and (b) show the two baseline groups. Rows correspond to seven distance multipliers from $1\times$ to $3\times$; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



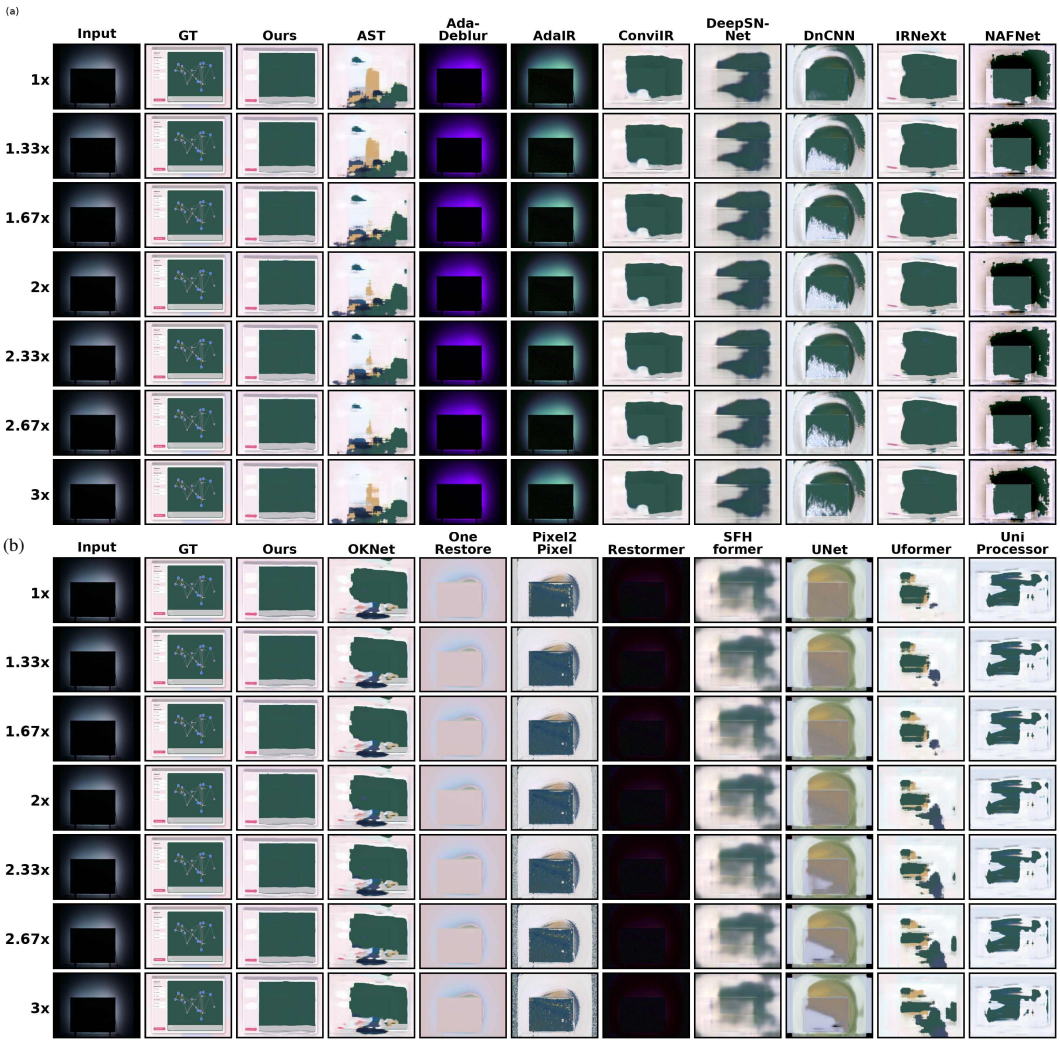
On Screen, OARL recovers the bright application window, its blue navigation strip, and the surrounding blue-sky and green-hill wallpaper across the distance sweep. Several competing methods blend these components into diffuse patches or lose the foreground window. OARL's LPIPS and SSIM curves remain nearly constant, and PSNR retains a substantial lead through the farthest distance. The preserved foreground-to-background arrangement agrees with these profiles. Together, the results demonstrate robust recovery of a composite desktop interface under the changes in observation scale and signal distribution induced by camera-to-wall distance.



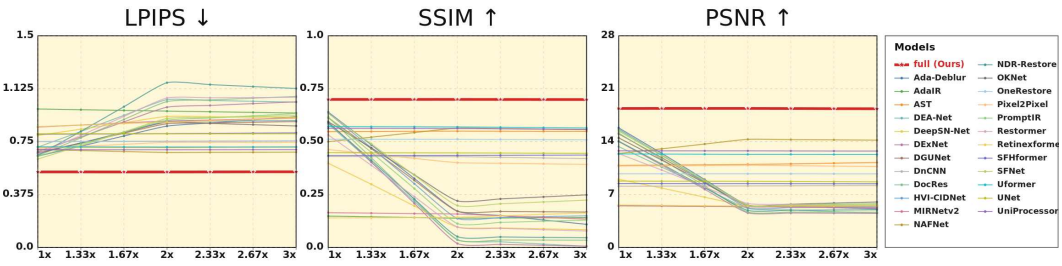
Supplementary Figure 33: Figure under camera-to-wall distance changes. Panels (a) and (b) show the two baseline groups. Rows correspond to seven distance multipliers from $1\times$ to $3\times$; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



The Figure panels show a dark blue matrix alongside a light violet plot. OARL preserves the matrix pattern and the boundary between the two chart regions across the distance settings. Several baselines instead reproduce broad blocks or illumination contours and show a marked metric decline as distance increases. The nearly flat OARL curves retain low LPIPS and high SSIM and PSNR throughout the comparison, including the threefold distance setting. This consistent quantitative lead supports the visual recovery of the chart's internal organization under distance-induced changes in the wall observation.



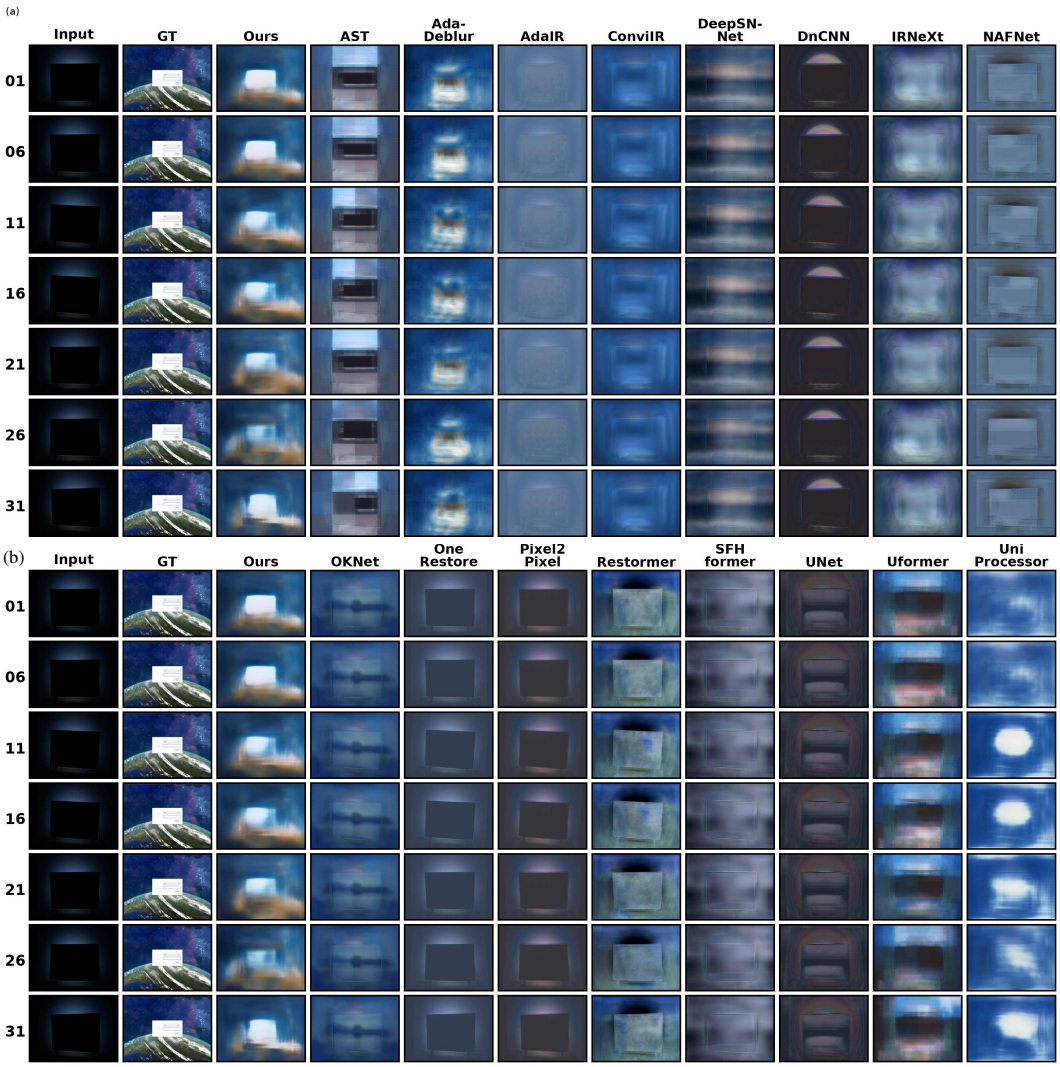
Supplementary Figure 34: Game under camera-to-wall distance changes. Panels (a) and (b) show the two baseline groups. Rows correspond to seven distance multipliers from $1\times$ to $3\times$; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



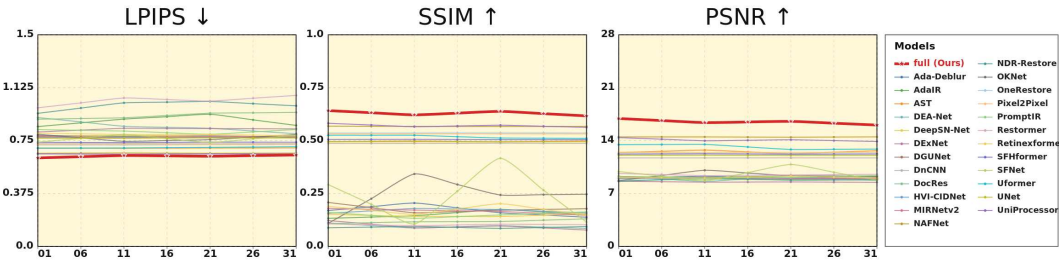
On Game, OARL retains the dark green playing field, the pink control region on the left, and the narrow colored lower boundary across camera-to-wall distances. Several competing outputs contain large holes or replace the target geometry with curved or fragmented regions. OARL maintains a cleaner rectangular field and a more faithful division from the interface controls. Its three metric profiles remain nearly constant over the sweep, while several baselines show a pronounced decline at larger distances. These results establish strong preservation of the main game layout as the camera moves farther from the wall.

C.11 Robustness under joint viewpoint changes

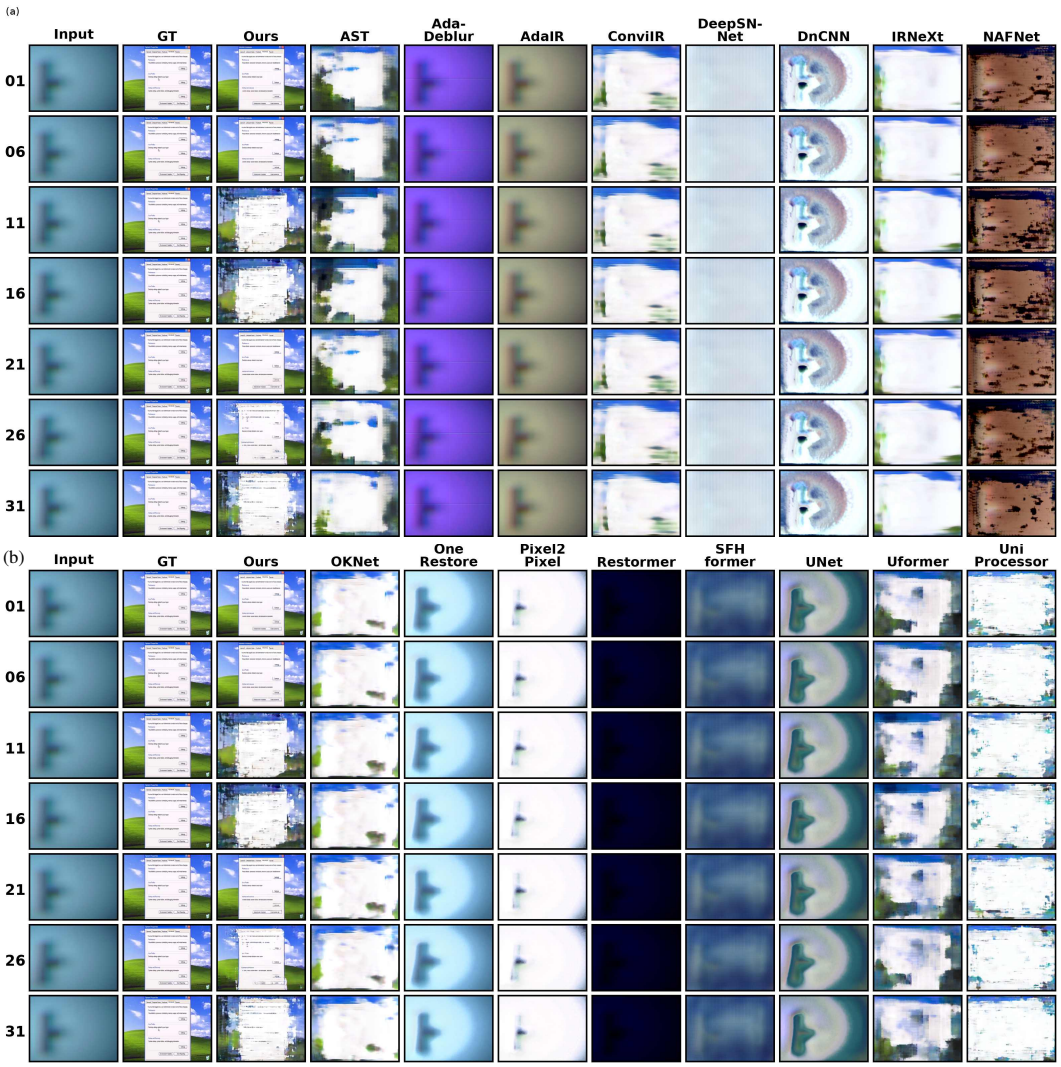
We evaluate seven combined-viewpoint settings—01, 06, 11, 16, 21, 26, and 31—denoted Combined views 1–7, to test recovery under joint viewpoint changes. The aliases follow the original condition order, and the model and evaluation protocol are fixed.



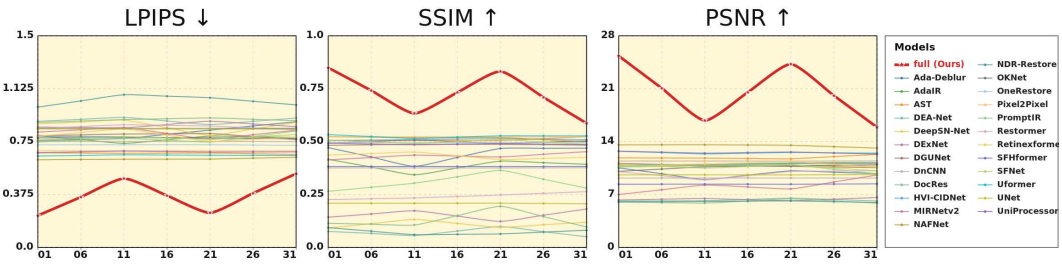
Supplementary Figure 35: OSLock under joint viewpoint changes. Panels (a,b) compare the baseline groups; rows show perturbation levels. Below: LPIPS (↓), SSIM and PSNR (↑).



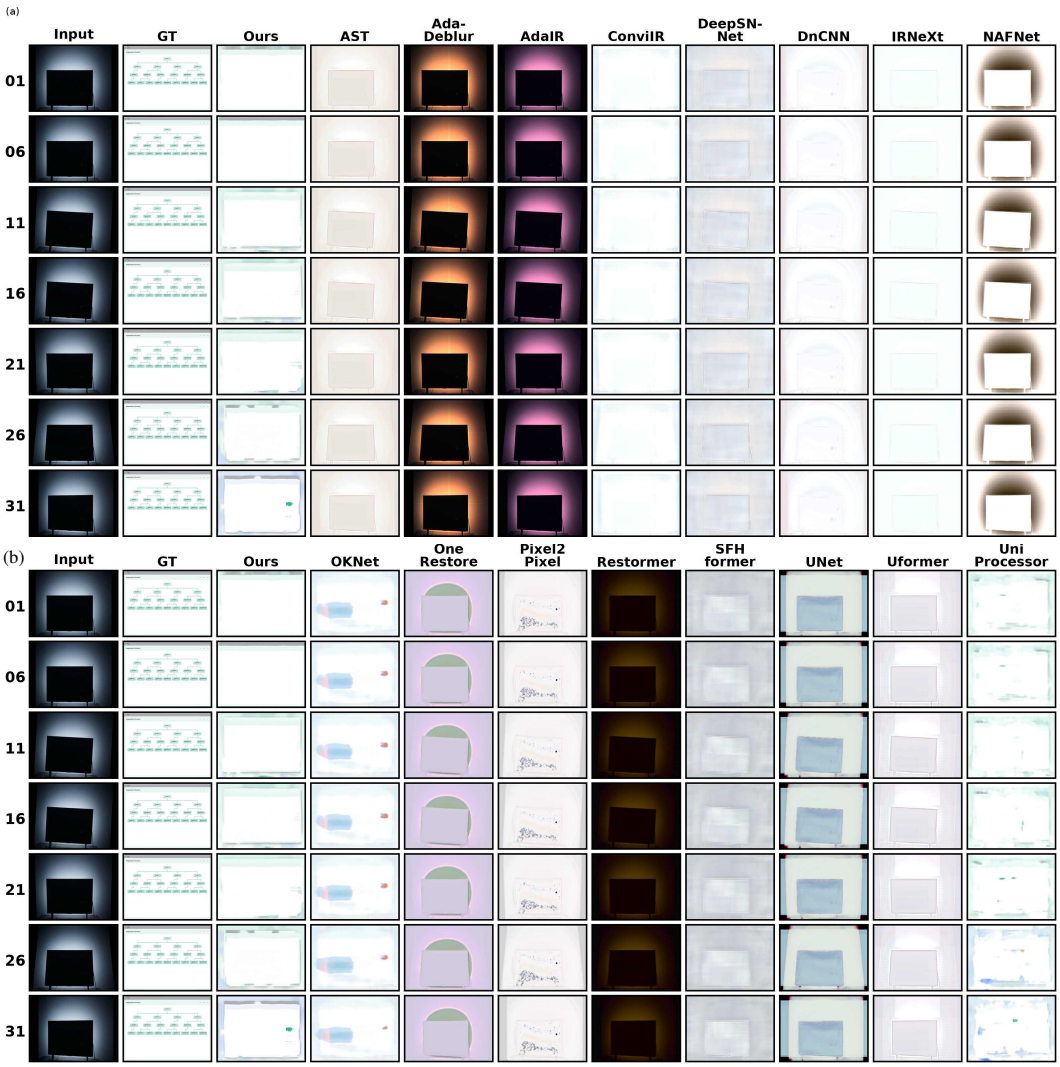
On OSLock, OARL retains the bright central dialog and the blue and warm background regions across the composite views. Several baselines produce unrelated bands or dark blocks. Its favorable LPIPS, SSIM, and PSNR profiles demonstrate robust recovery of the principal screen organization when multiple viewpoint factors change together.



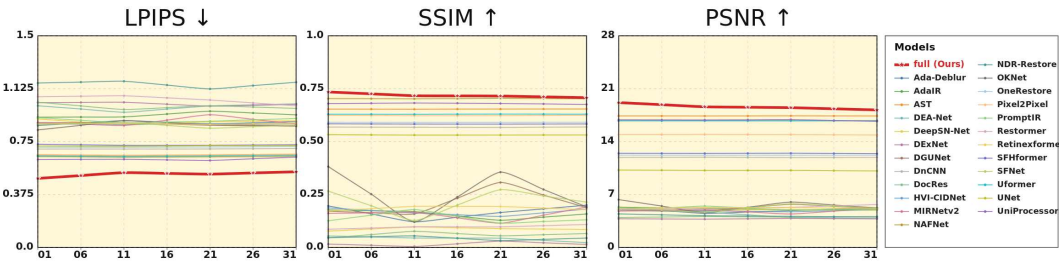
Supplementary Figure 36: Screen under joint viewpoint changes. Panels (a) and (b) show the two baseline groups. Rows correspond to composite viewpoints 01, 06, 11, 16, 21, 26, and 31; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



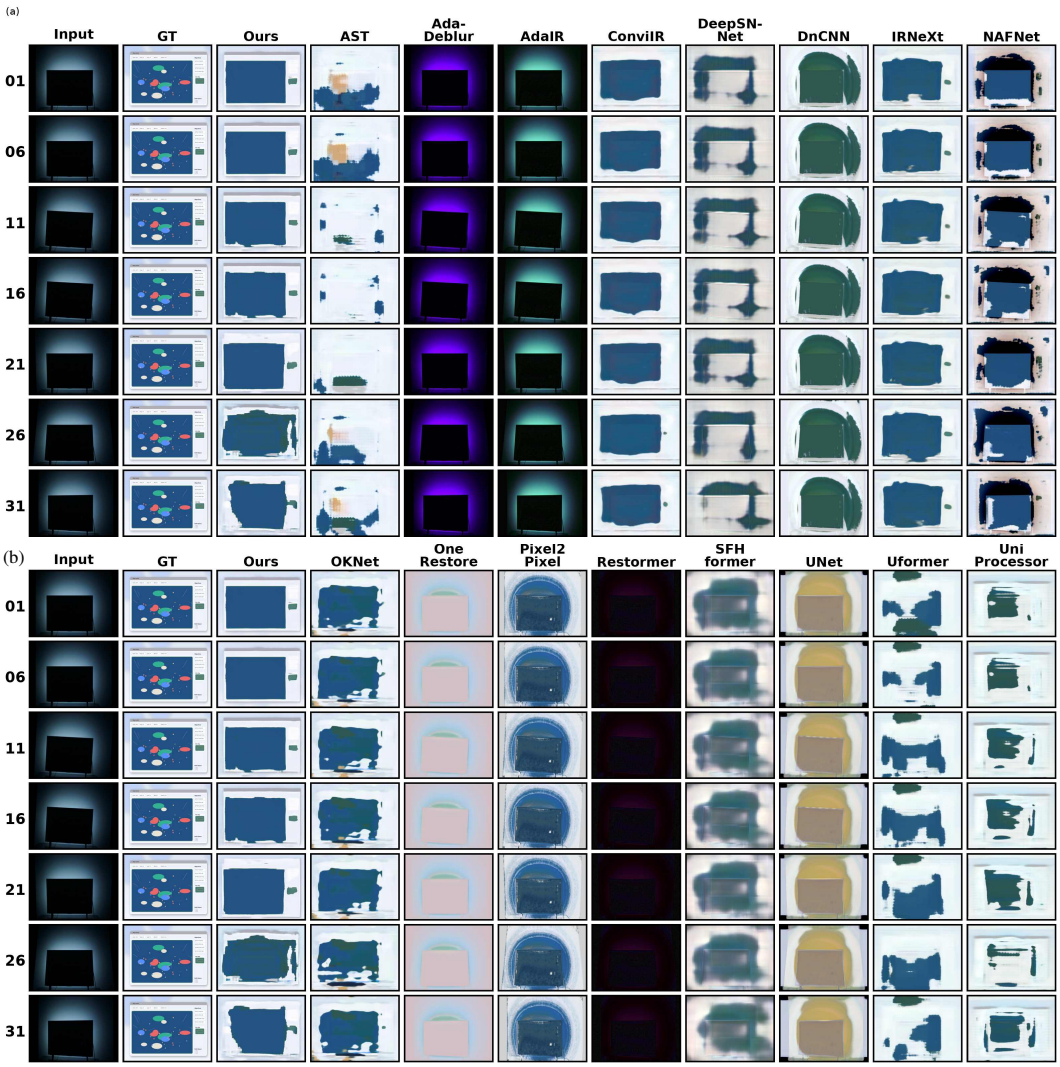
On Screen, OARL reconstructs the bright application window within the landscape desktop most clearly at settings 01, 06, and 21 (Combined views 1, 2, and 5). Several competing outputs lose the window arrangement in diffuse or strongly tinted regions. The OARL curves respond to the different composite views while retaining lower LPIPS and higher SSIM and PSNR throughout the sweep. Their recovery around setting 21 agrees with the clearer foreground window in the qualitative panel. These results demonstrate a consistent comparative advantage when multiple geometric factors vary together, preserving stronger interface reconstruction across the evaluated acquisition settings.



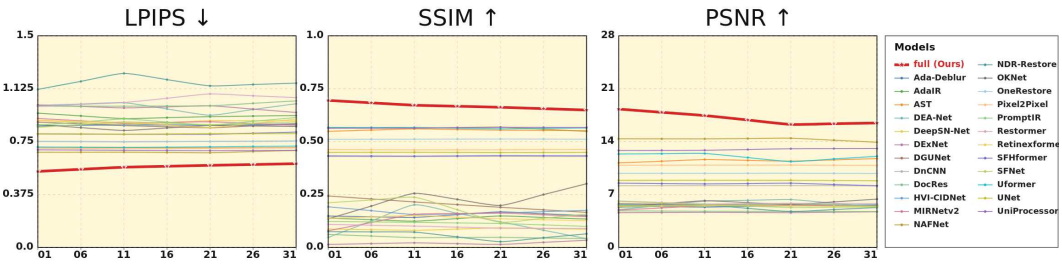
Supplementary Figure 37: Figure under joint viewpoint changes. Panels (a) and (b) show the two baseline groups. Rows correspond to composite viewpoints 01, 06, 11, 16, 21, 26, and 31; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



On Figure, OARL recovers the light chart canvas and its surrounding frame under the composite viewpoint settings. Several competing methods emphasize the dark input silhouette or generate broad illumination contours that depart from the target's light appearance. The quantitative curves provide a clear complementary result: OARL retains the lowest LPIPS and highest SSIM and PSNR across the indexed views. The persistent metric separation establishes stronger reconstruction of this graphical content when several viewpoint factors change simultaneously, extending the geometric robustness demonstrated by the individual translation and angle sweeps.



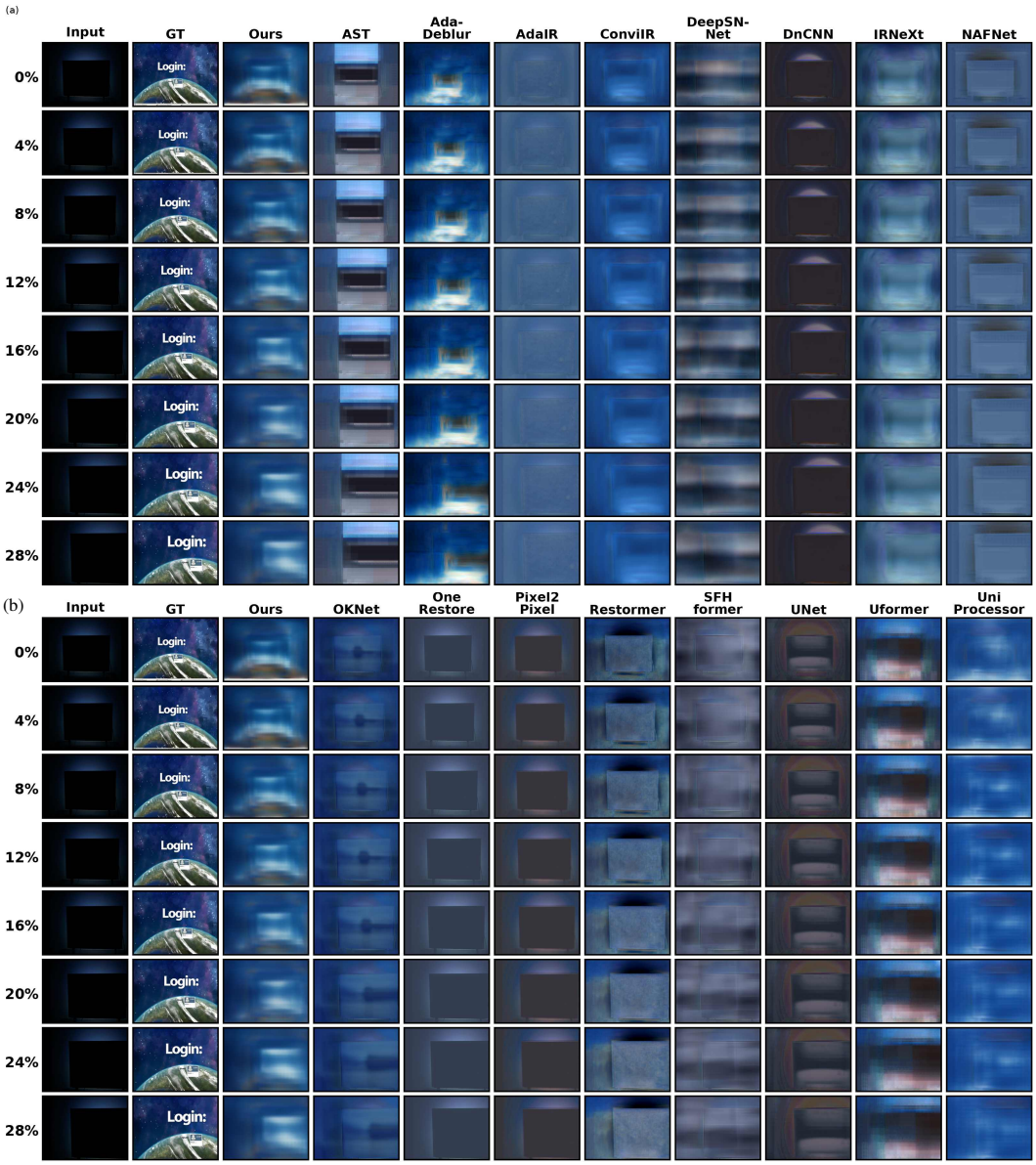
Supplementary Figure 38: Game under joint viewpoint changes. Panels (a) and (b) show the two baseline groups. Rows correspond to composite viewpoints 01, 06, 11, 16, 21, 26, and 31; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



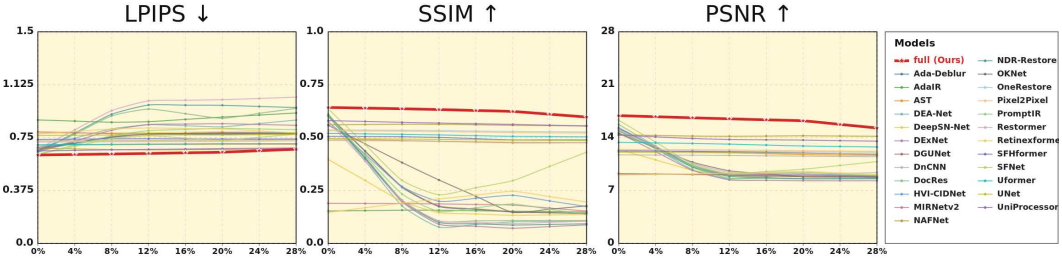
On Game, OARL preserves the blue playing region within the light interface frame across the composite viewpoints. Several baselines break the field into disconnected patches or introduce broad internal gaps. OARL maintains a more contiguous central region and a clearer separation from the surrounding controls. Its LPIPS and SSIM curves change gradually across the indexed settings, while PSNR retains a leading position under stronger perturbations. The visual continuity of the field agrees with this aggregate advantage, demonstrating robust preservation of the dominant game layout when several geometric changes affect the wall observation together.

C.12 Robustness to progressive right-side cropping

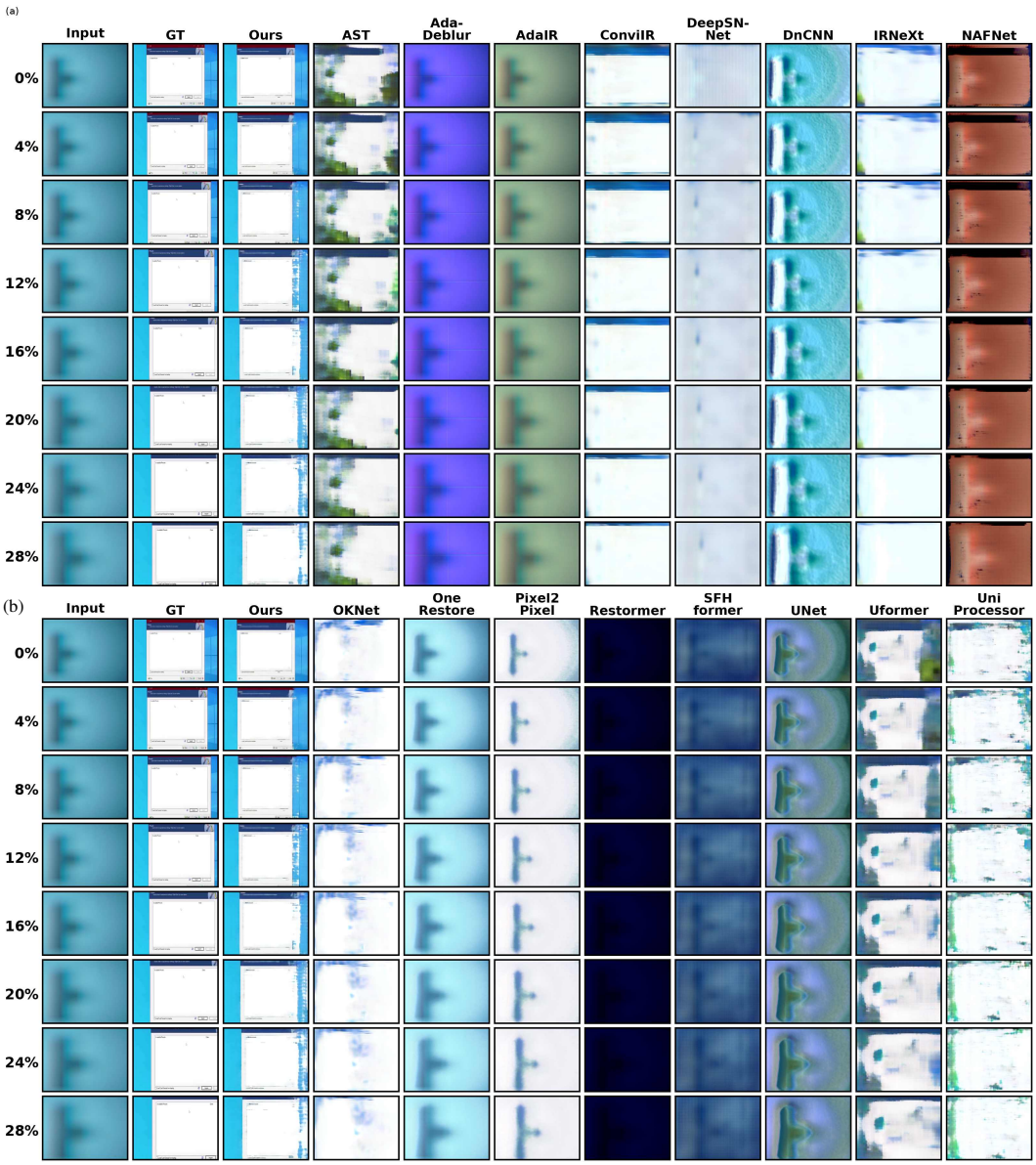
Cropping tests whether the remaining wall-glow regions still support the hidden screen's layout. We remove 0%–28% of the input from the right in 4% steps, keeping other conditions fixed.



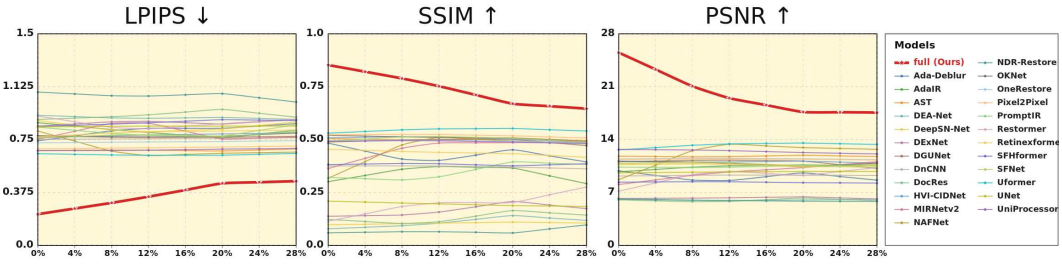
Supplementary Figure 39: OSLock under right-side cropping. Panels (a,b) and dataset metrics.



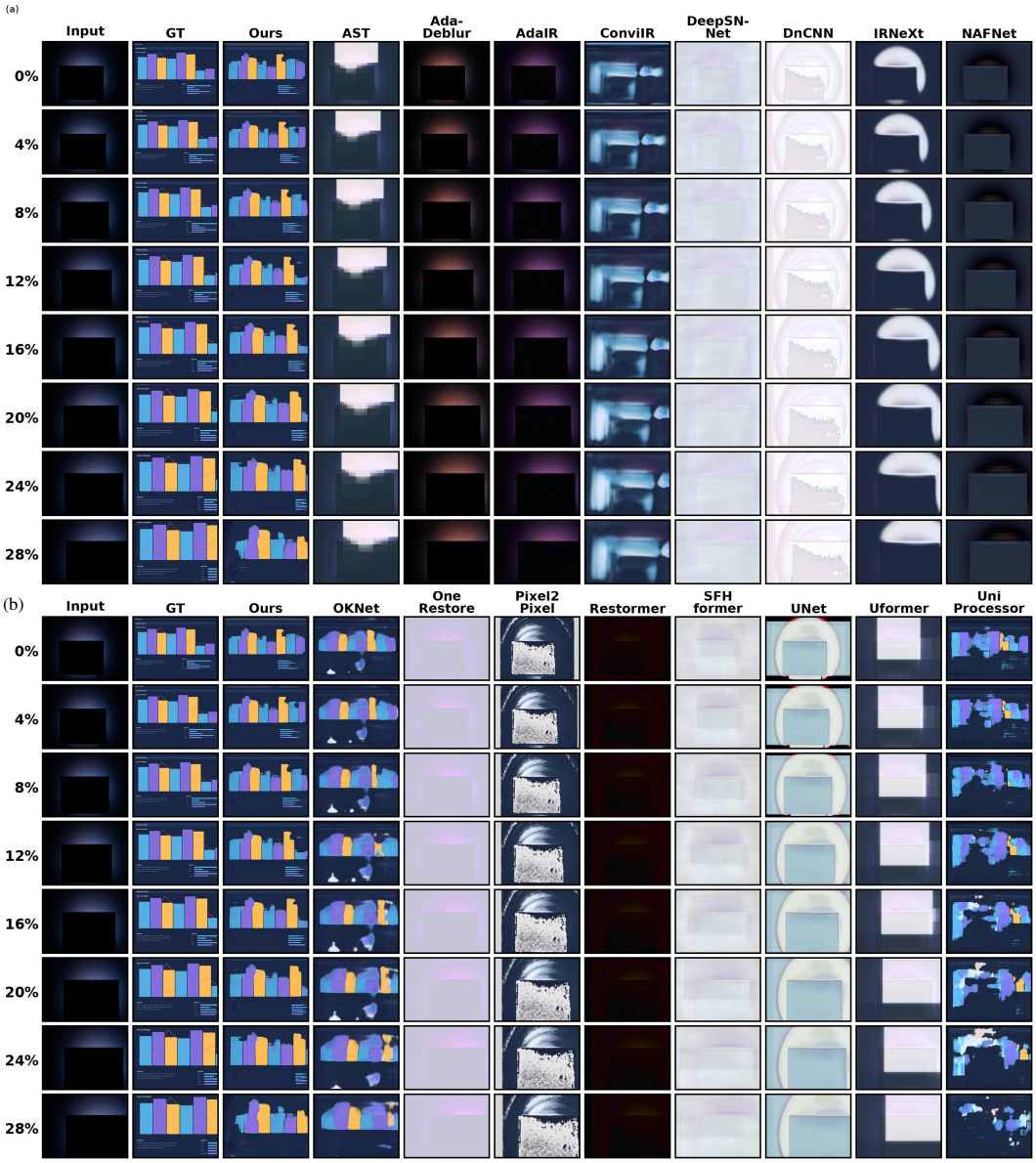
OARL retains the blue background and bright lower mountain region, with leading metrics as right-side observations are removed.



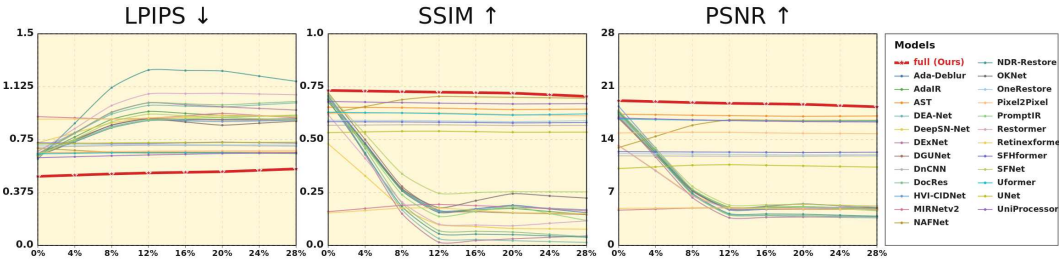
Supplementary Figure 40: Screen under right-side cropping. Panels (a,b) compare the baseline groups; rows show perturbation levels. Below: LPIPS (↓), SSIM and PSNR (↑).



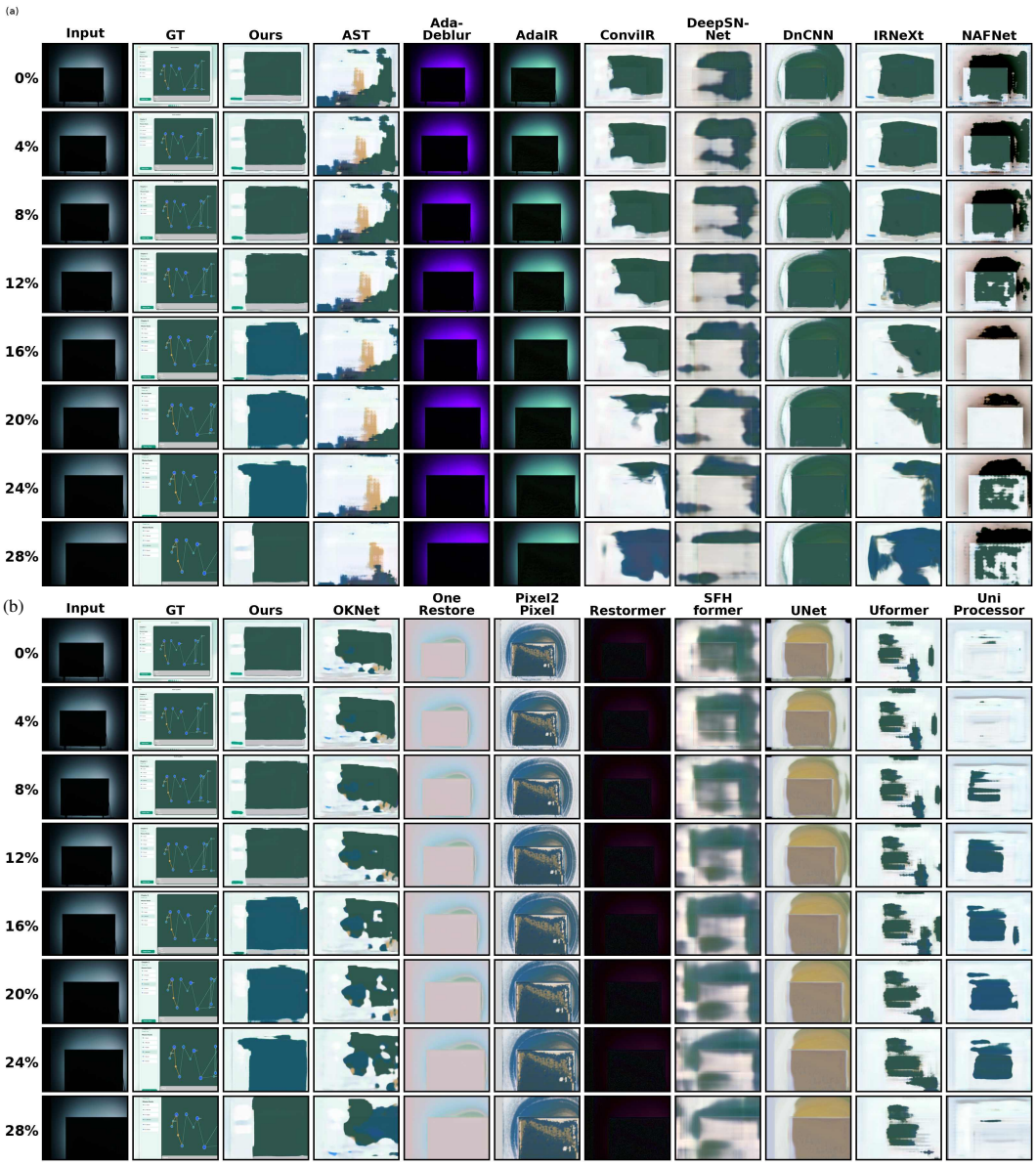
On Screen, OARL preserves the tall bright window and its blue surrounding region as progressively more of the observation is cropped. Several baselines lose the window boundary in diffuse fields. The OARL curves retain lower LPIPS and higher SSIM and PSNR at the largest crop ratios, demonstrating stronger recovery of the foreground layout from the remaining wall-glow evidence.



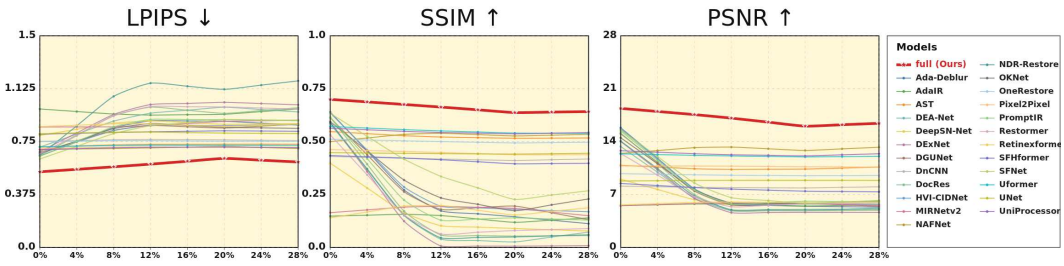
Supplementary Figure 41: Figure under right-side cropping. Panels (a,b) compare the baseline groups; rows show perturbation levels. Below: LPIPS (↓), SSIM and PSNR (↑).



On Figure, OARL retains the sequence of colored bars against the dark chart background as right-side evidence is removed. Several competing outputs instead contain uniform regions or unrelated contours. The gradual OARL metric response, with a persistent lead through the largest crop ratio, demonstrates stronger preservation of chart organization under partial observation.



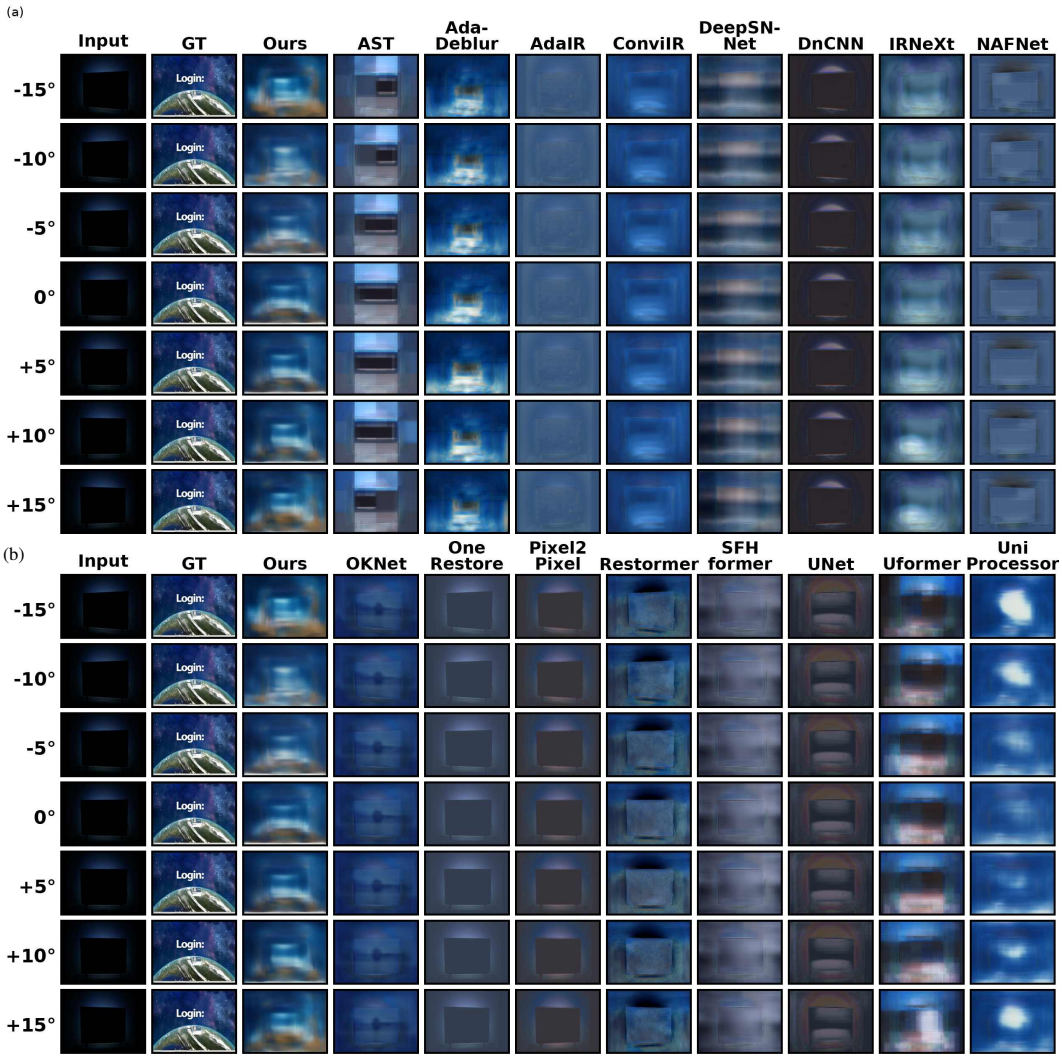
Supplementary Figure 42: Game under right-side cropping. Panels (a,b) compare the baseline groups; rows show perturbation levels. Below: LPIPS (↓), SSIM and PSNR (↑).



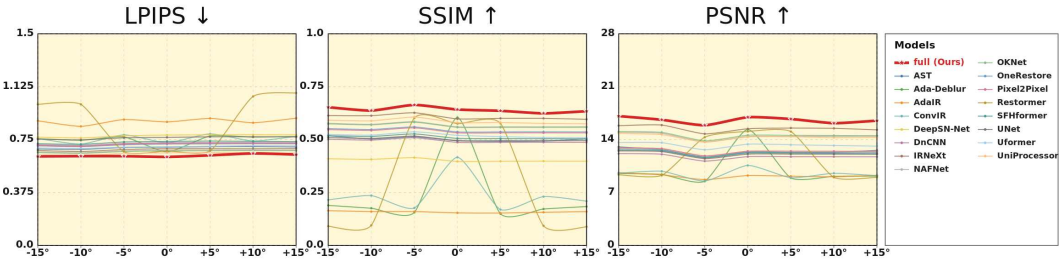
On Game, OARL maintains a contiguous green field beside the light control region as the crop increases. Several baselines develop large holes or fragmented boundaries. Its favorable metric ranking at the larger crop ratios agrees with the more coherent regional layout, demonstrating effective recovery of the main interface structure from a partially observed wall glow.

C.13 Robustness to horizontal camera-offset changes

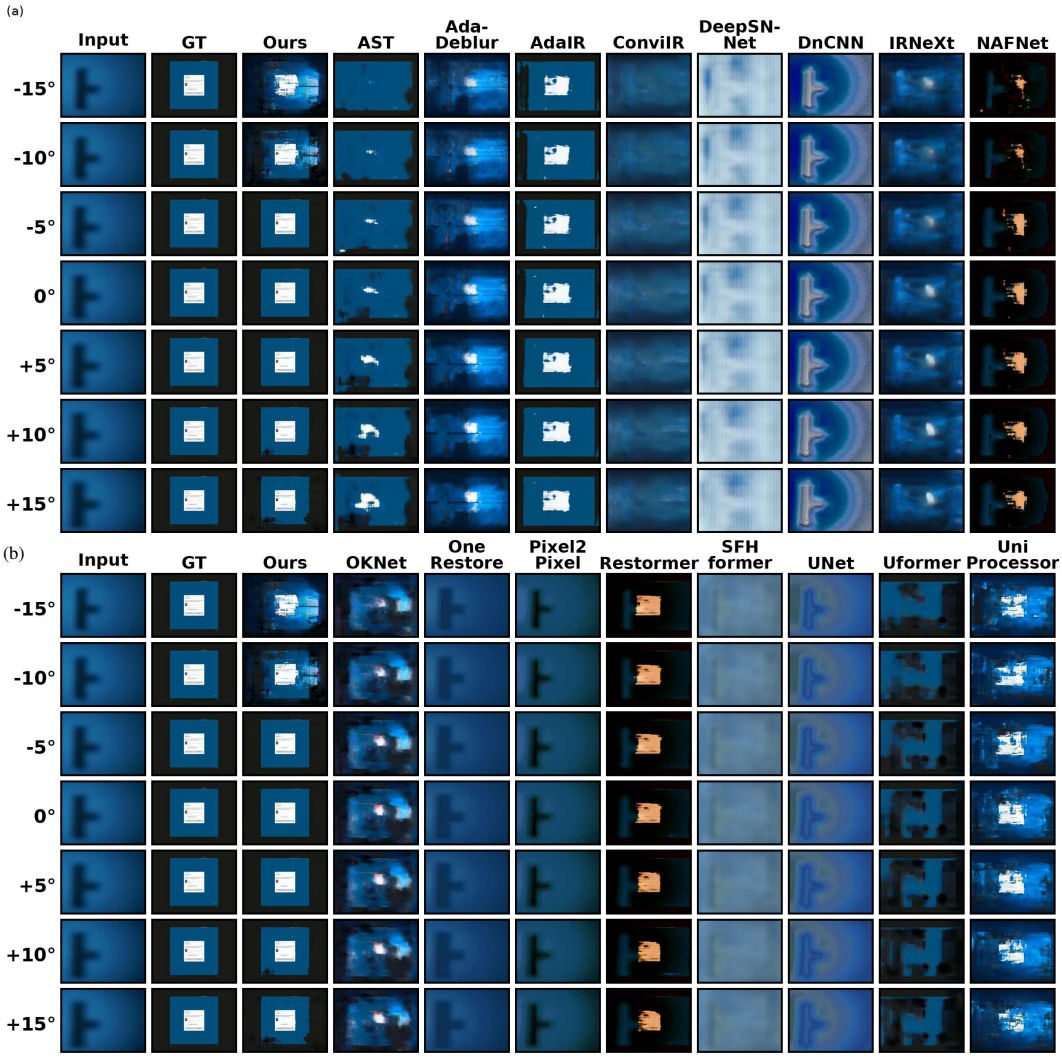
We vary the horizontal camera angle over $-15°$, $-10°$, $-5°$, $0°$, $+5°$, $+10°$, and $+15°$, with the displayed content and all remaining acquisition and evaluation settings fixed.



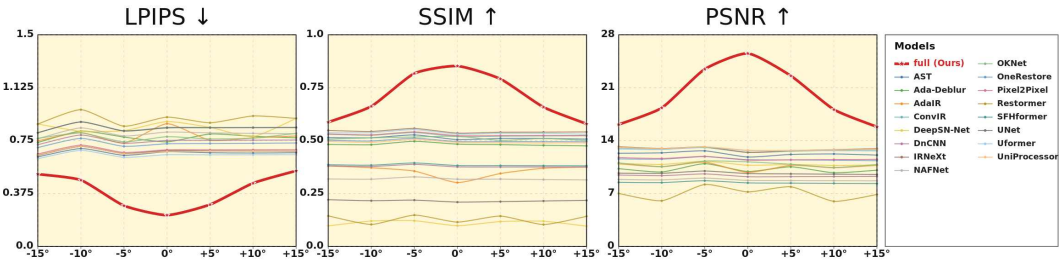
Supplementary Figure 43: OSLock under horizontal angular offsets. Panels (a,b) compare the baseline groups; rows show perturbation levels. Below: LPIPS (↓), SSIM and PSNR (↑).



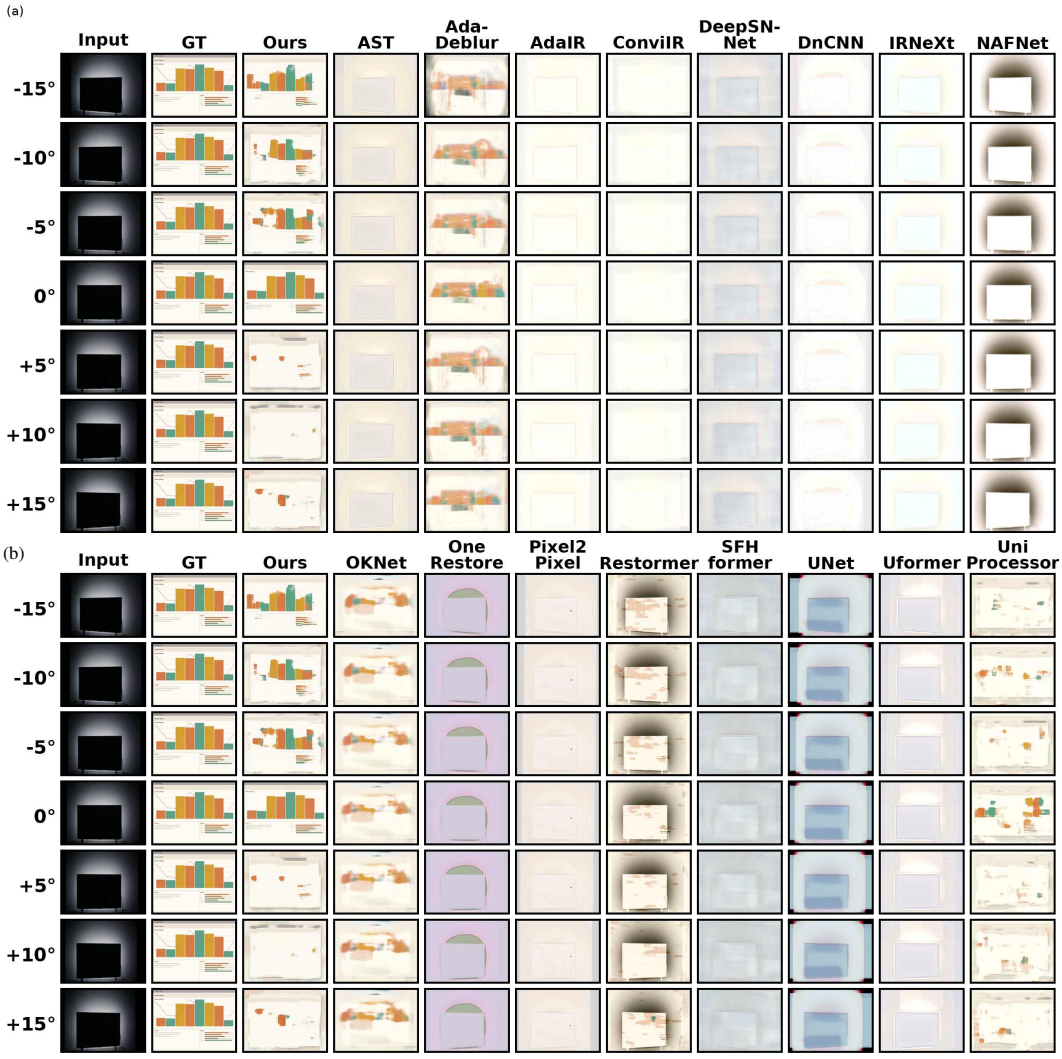
On OSLock, OARL retains the blue upper region and bright lower mountain structure across horizontal viewing angles. Several baselines produce broad bands or dark blocks. Its favorable metric profiles on both sides of the reference view demonstrate reliable recovery of the principal lock-screen appearance under substantial angular offsets.



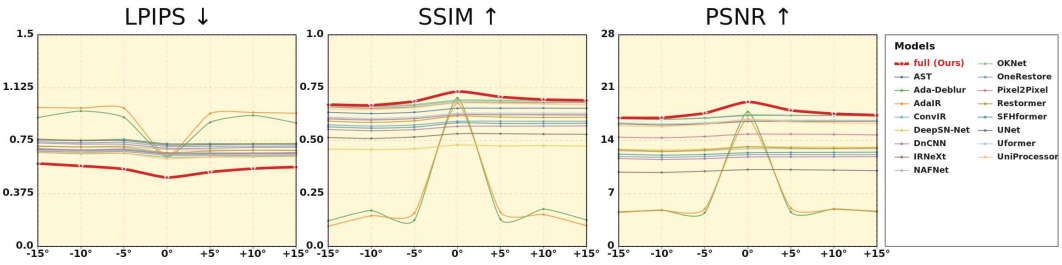
Supplementary Figure 44: Screen under horizontal angular offsets. Panels (a) and (b) show the two baseline groups. Rows correspond to seven horizontal angles from $-15°$ to $+15°$; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



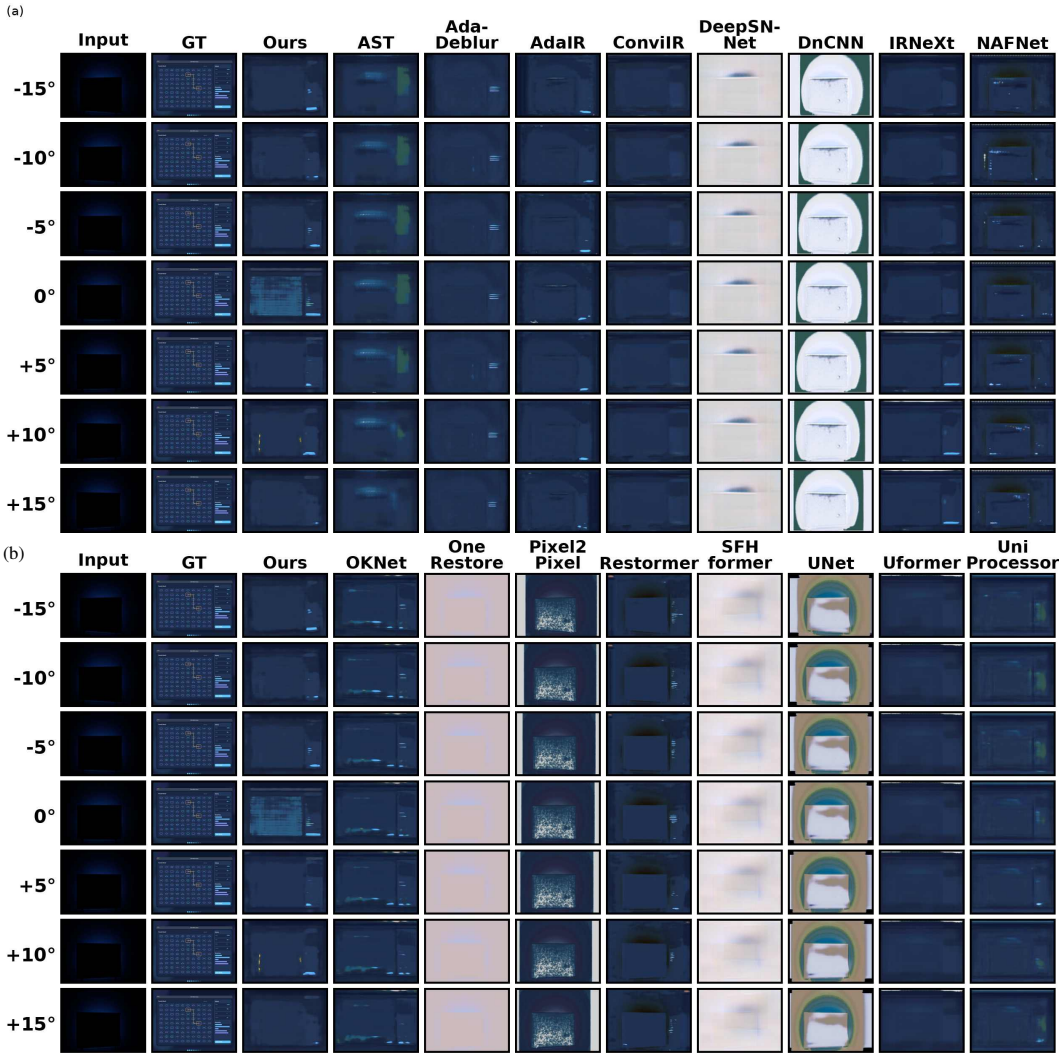
On Screen, OARL localizes the small white dialog within the blue rectangular desktop region. The foreground boundary is particularly clear around the reference view and remains more faithful than the diffuse or fragmented responses of several baselines. Its metric curves improve toward the central angles and change smoothly toward either extreme, retaining lower LPIPS and higher SSIM and PSNR across the comparison. This balanced aggregate response supports OARL's ability to recover the localized interface structure under horizontal viewing-angle changes, with a clear advantage throughout the evaluated angular range.



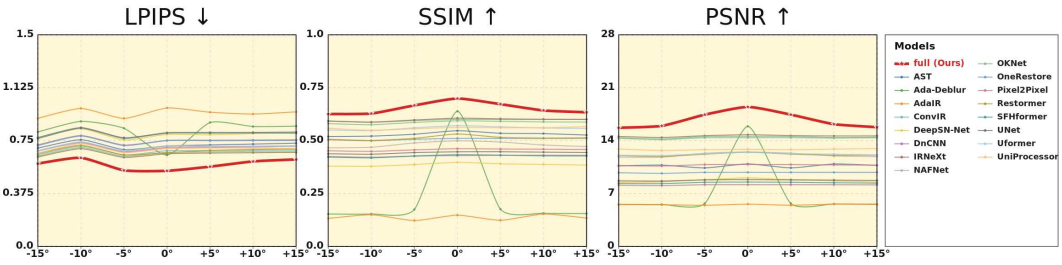
Supplementary Figure 45: Figure under horizontal angular offsets. Panels (a) and (b) show the two baseline groups. Rows correspond to seven horizontal angles from $-15°$ to $+15°$; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



The Figure example contains green and orange bars on a light chart canvas. OARL recovers distinct colored bars most clearly at the nominal and several negative-angle views, whereas several baselines return nearly uniform fields or broad colored patches. Across the complete angular sweep, the OARL curves retain lower LPIPS and higher SSIM and PSNR. Their favorable ranking at both ends of the range establishes a broader structural advantage beyond the reference view. These results demonstrate stronger reconstruction of graphical content as horizontal angular displacement changes the spatial arrangement of the wall glow.



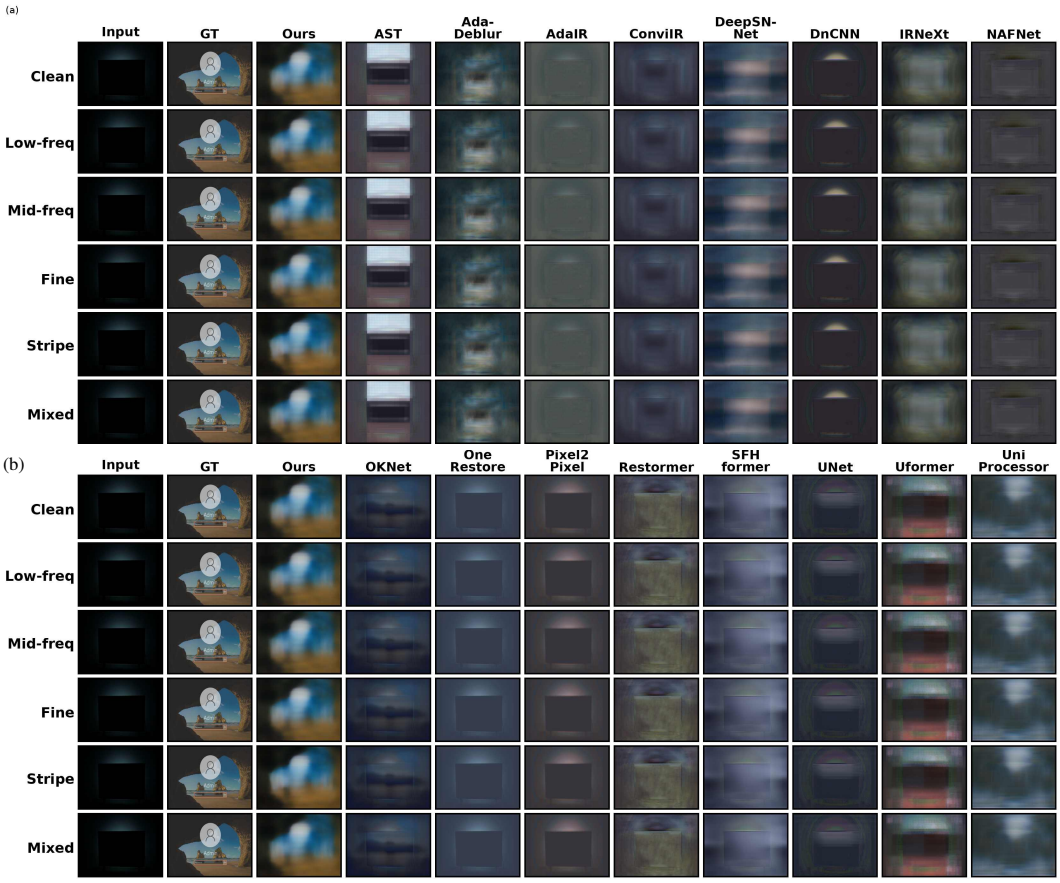
Supplementary Figure 46: Game under horizontal angular offsets. Panels (a) and (b) show the two baseline groups. Rows correspond to seven horizontal angles from $-15°$ to $+15°$; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



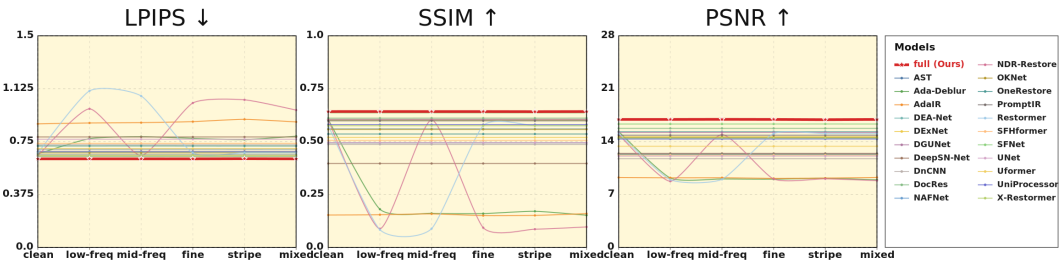
On Game, OARL preserves the dark blue screen region and the bright control area near the lower right. Around the reference view, its reconstruction also retains more of the internal board pattern than several diffuse competing outputs. Across the angular sweep, OARL maintains lower LPIPS and higher SSIM and PSNR, with a gradual response away from the nominal position. The persistent metric advantage supports the stronger recovery of the screen's dark regional organization, demonstrating robust reconstruction under substantial horizontal camera offsets on either side of the reference view.

C.14 Robustness to wall-texture changes

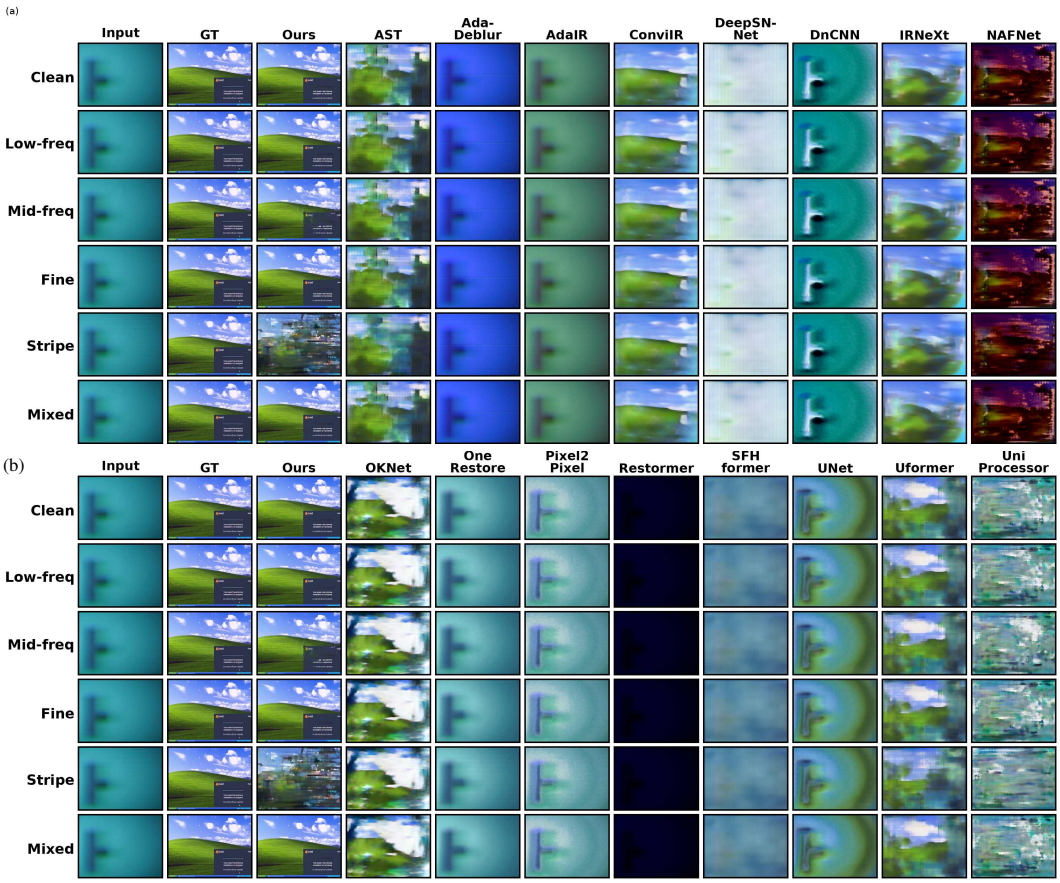
Wall texture changes the background patterns recorded with the screen-dependent glow. We compare clean, low-frequency, mid-frequency, fine, stripe, and mixed wall textures. These surface perturbations change the recorded background while the hidden screen, projection geometry, and imaging setup remain fixed.



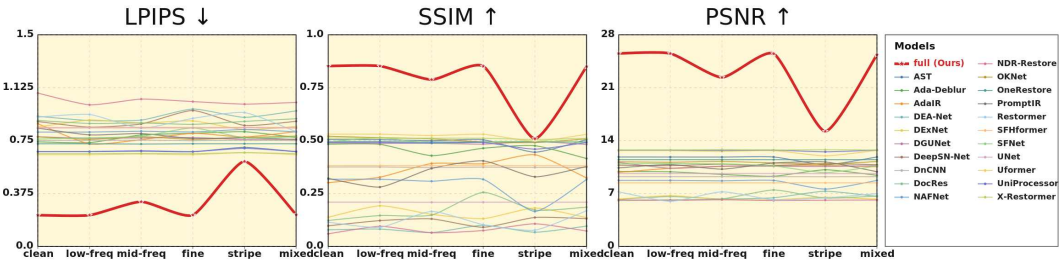
Supplementary Figure 47: OSLock under wall-texture changes. Panels (a) and (b) show the two baseline groups. Rows correspond to clean, low-frequency, mid-frequency, fine, stripe, and mixed wall textures; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



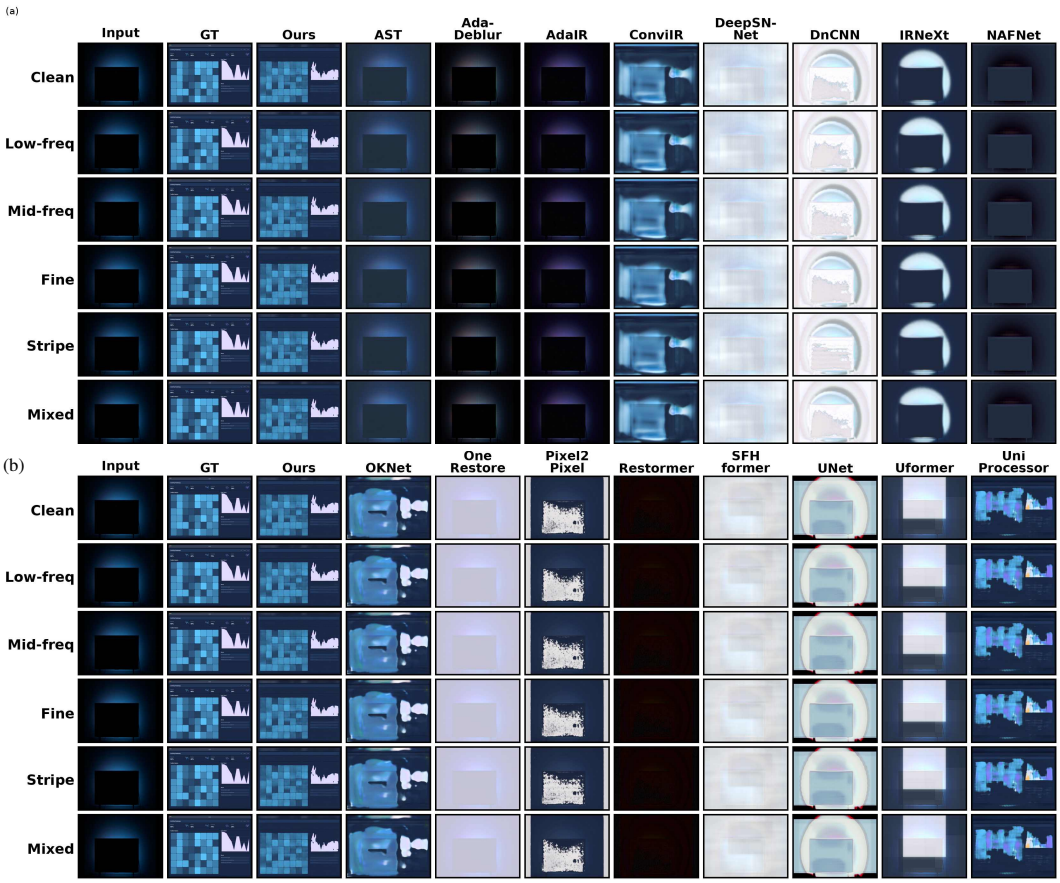
On OSLock, OARL preserves the bright profile region within the blue and warm background across the six wall textures. Several competing outputs introduce broad bands or dark blocks whose organization departs from the target. The recovered profile-to-background arrangement changes little between clean, fine, striped, and mixed surfaces. OARL's nearly constant metric curves remain favorable across the same settings, with low LPIPS and leading SSIM and PSNR. These results demonstrate robust preservation of the principal lock-screen appearance when wall texture changes the background structure of the indirect observation.



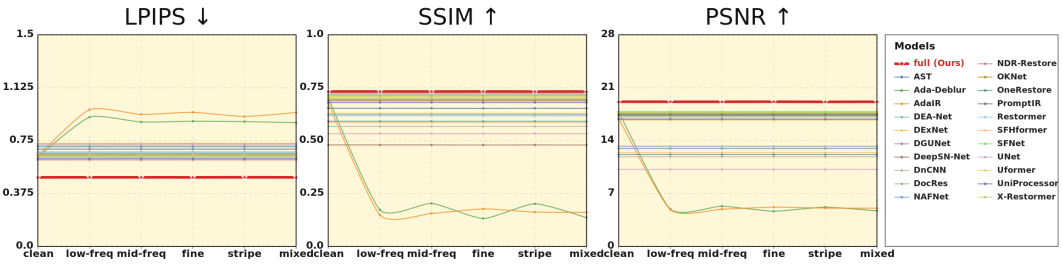
Supplementary Figure 48: Screen under wall-texture changes. Panels (a) and (b) show the two baseline groups. Rows correspond to clean, low-frequency, mid-frequency, fine, stripe, and mixed wall textures; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



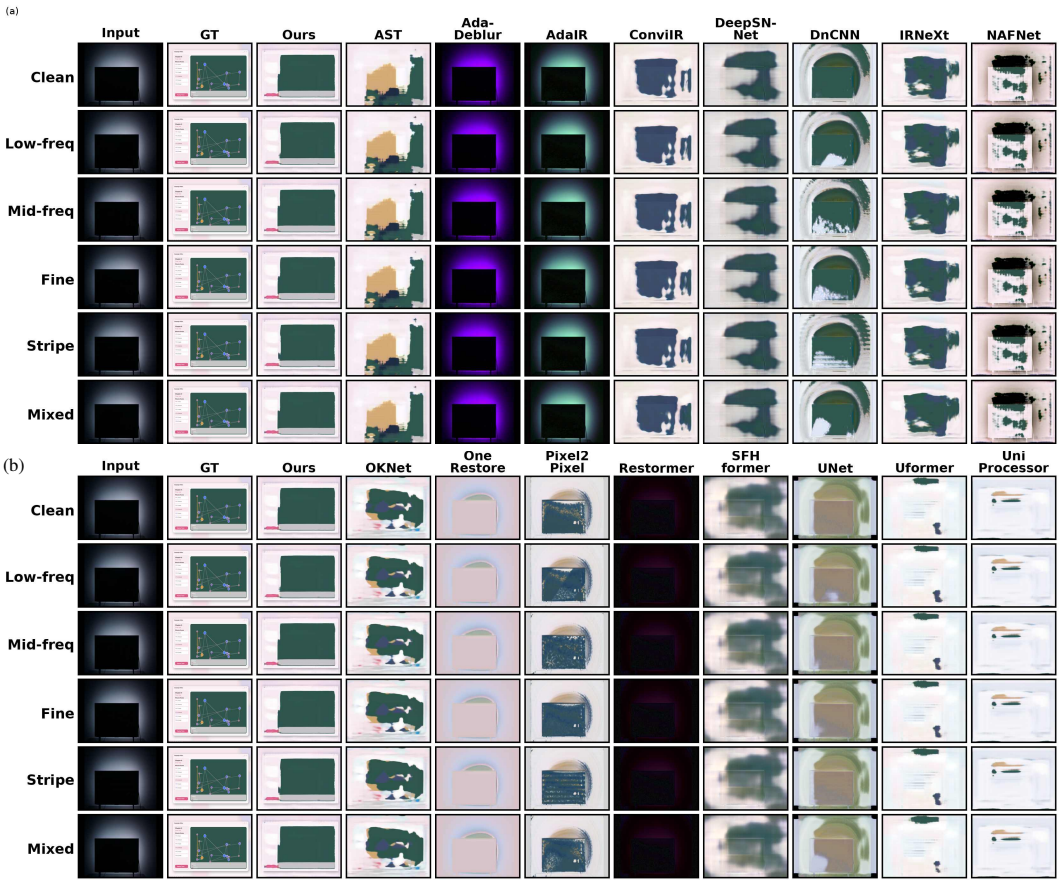
On Screen, OARL recovers the blue-sky and green-hill desktop together with the dark dialog at the lower right. This foreground-to-background arrangement is especially clear on the clean, low-frequency, fine, and mixed surfaces. Several baselines retain a diffuse landscape or introduce strong color distortions. The metric curves distinguish the surface conditions while preserving OARL's leading ranking; in particular, the recovery on the mixed surface is consistent with the clearly reconstructed desktop and dialog. The joint comparison demonstrates stronger adaptation to background texture variation, preserving both the broad wallpaper structure and the localized interface region over the evaluated surfaces. OARL therefore maintains a clear reconstruction advantage as the wall's spatial texture changes. The lower-right dialog provides a localized reference for judging the reconstruction independently of the broad wallpaper colors. OARL places this dark region within the landscape composition more consistently than methods that recover only a diffuse blue-green background, strengthening the visual evidence for its interface reconstruction advantage.



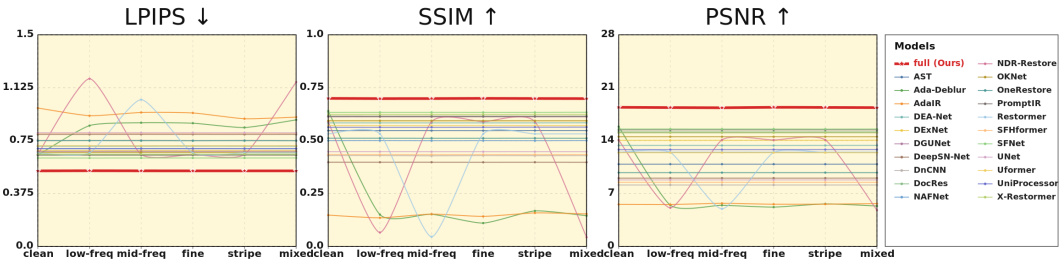
Supplementary Figure 49: Figure under wall-texture changes. Panels (a) and (b) show the two baseline groups. Rows correspond to clean, low-frequency, mid-frequency, fine, stripe, and mixed wall textures; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



On Figure, OARL preserves the blue matrix cells and the pale plot region beside them across all six wall textures. Several competing methods instead reconstruct a uniform dark block or curved contours associated with the input appearance. The contrast between the structured target matrix and the wall texture makes this comparison particularly informative: OARL keeps the graphical pattern organized across low-frequency, fine, striped, and mixed backgrounds. Its LPIPS, SSIM, and PSNR curves remain nearly flat and retain the leading ranking throughout the comparison. This agreement demonstrates strong preservation of target chart structure under surface-pattern changes, with consistent recovery of the matrix-to-plot division and a clear aggregate advantage over the competing reconstructions. The matrix cells also preserve alternating light and dark blue responses within the target region. This localized variation distinguishes the reconstructed chart from a single colored rectangle, while the pale neighboring plot preserves the larger arrangement. Both features reinforce the structural advantage shown by the SSIM comparison.



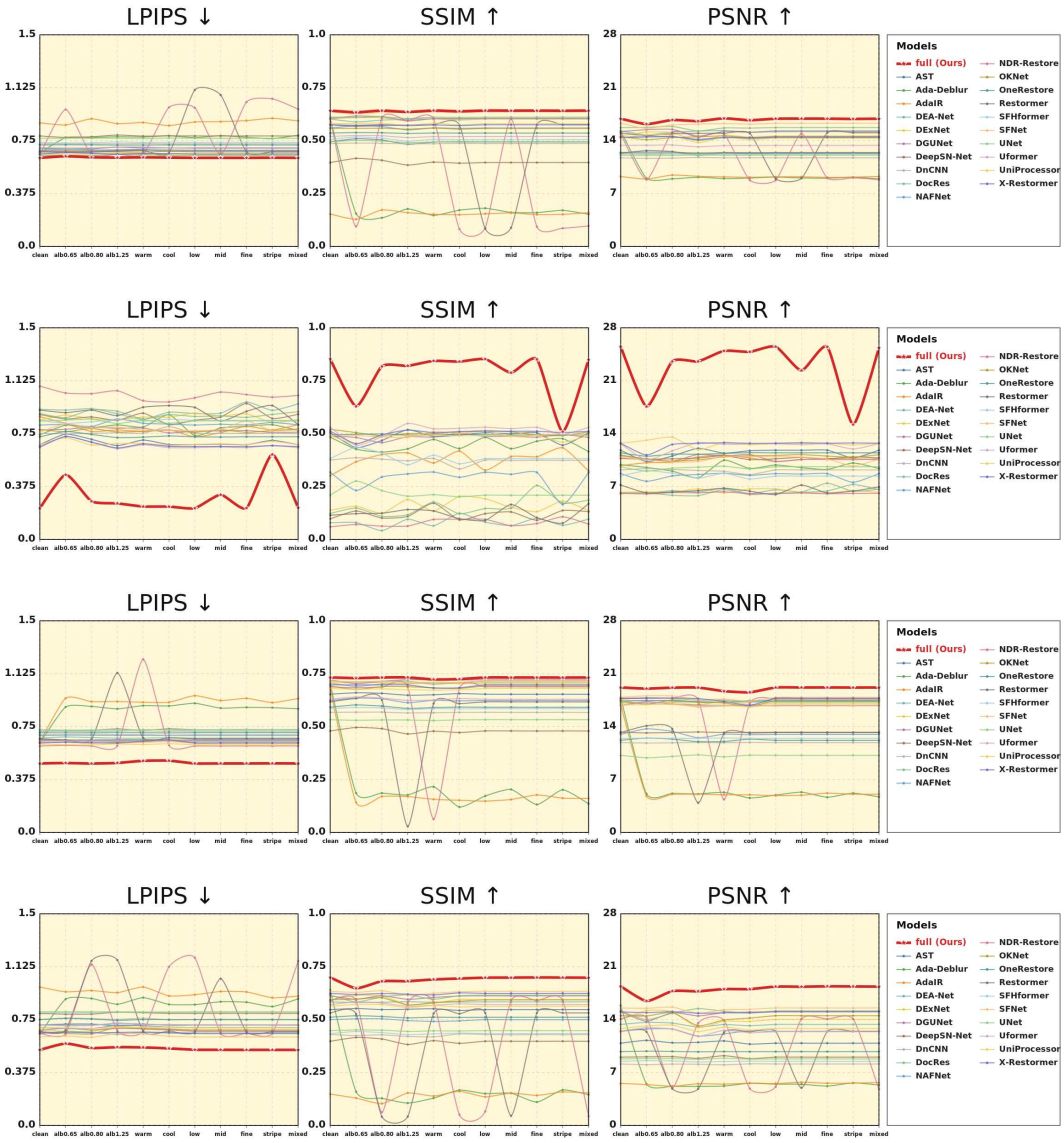
Supplementary Figure 50: Game under wall-texture changes. Panels (a) and (b) show the two baseline groups. Rows correspond to clean, low-frequency, mid-frequency, fine, stripe, and mixed wall textures; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



On Game, OARL retains the dark green field and the light pink control region on the left across wall textures. Several competing outputs introduce holes within the field, shift its border, or fragment the surrounding interface. OARL preserves a cleaner, more contiguous playing region from the clean wall through the striped and mixed conditions. The corresponding metric profiles are nearly constant and remain favorable in LPIPS, SSIM, and PSNR. The contrast between this stable regional reconstruction and the larger baseline fluctuations demonstrates reliable recovery when background texture changes the wall observation. In particular, the consistent field-to-control boundary confirms that the main game organization remains better preserved across the tested surface patterns. The light control strip and narrow colored lower edge supply additional references for the position and extent of the playing field. OARL keeps these large interface regions arranged around a coherent central area, whereas several competing outputs displace the border or create unrelated bright openings within it.

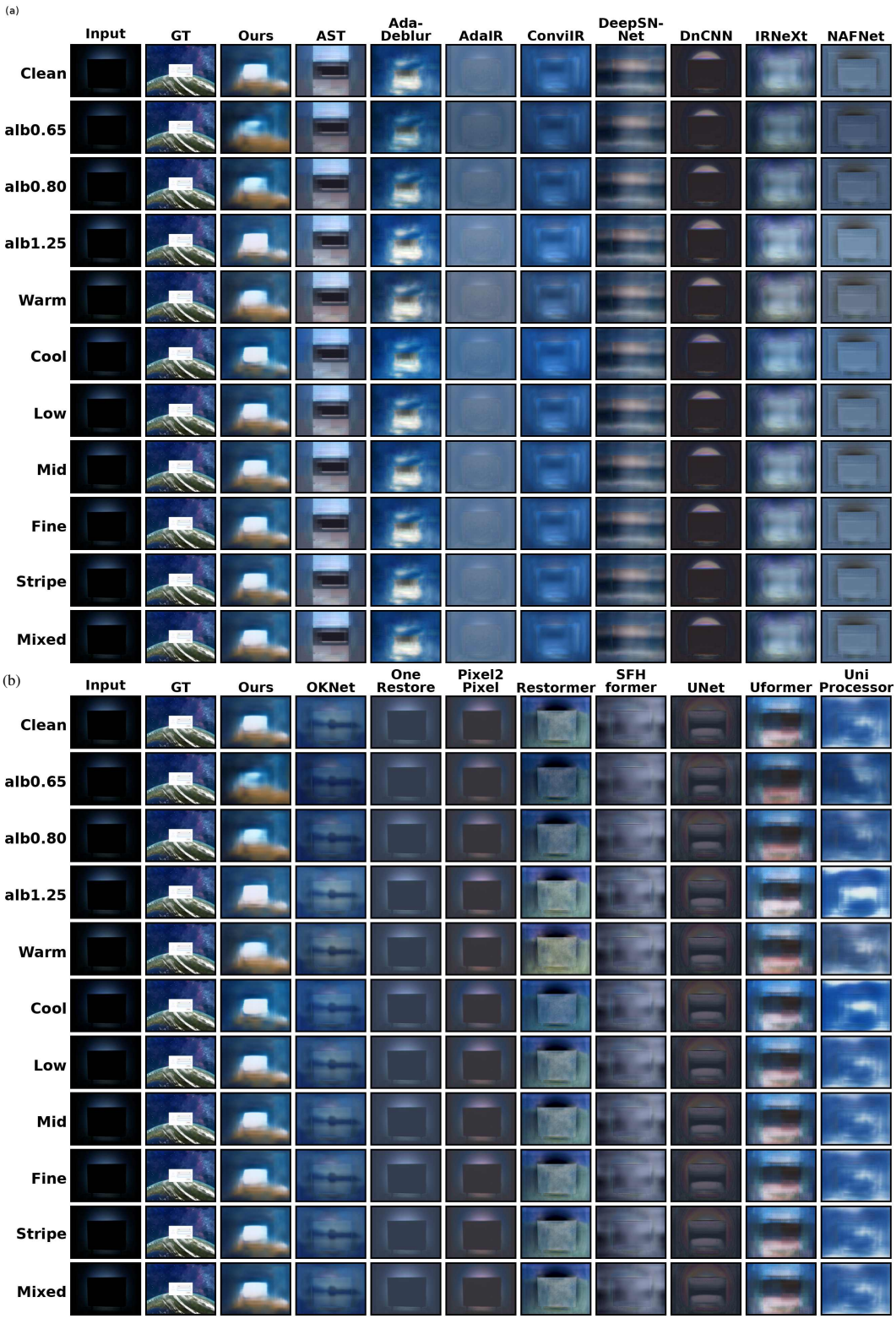
C.15 Reconstruction across surface materials and textures

We vary wall reflectance, color, and texture over eleven settings: clean, alb0.65, alb0.80, alb1.25, warm, cool, low, mid, fine, stripe, and mixed. The hidden screen and projection and imaging setup remain fixed.



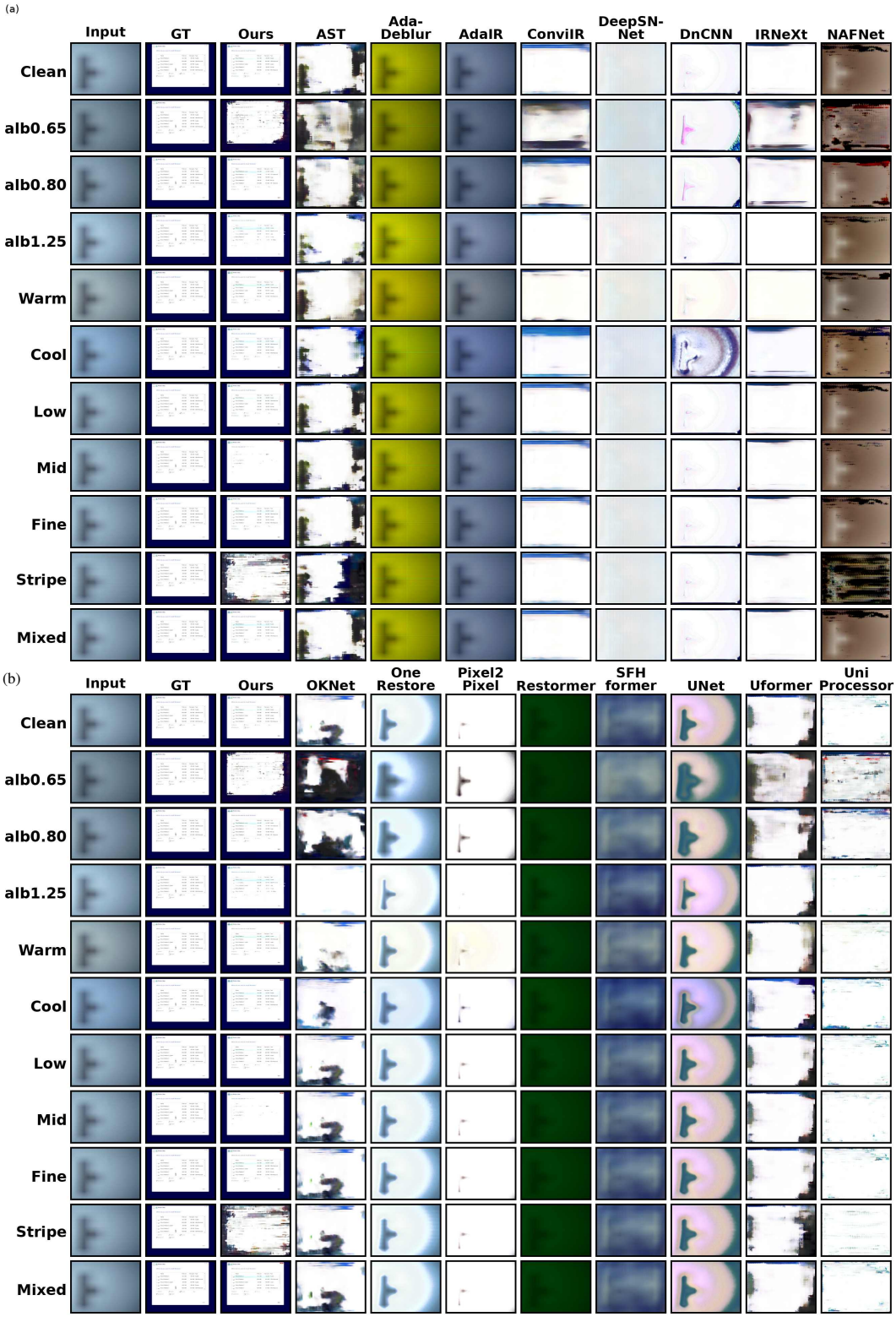
Supplementary Figure 51: Surface-material and texture robustness. Curves are ordered OSLock, Screen, Figure, and Game from top to bottom; each row reports LPIPS, SSIM, and PSNR.

OARL retains a favorable ranking across reflectance, color, and texture changes. OSLock preserves nearly constant perceptual and structural quality; Figure remains strong across albedo and texture settings; Game retains its regional reconstruction advantage. Screen shows a larger response to low albedo and stripes while maintaining the leading quantitative ranking. The following qualitative comparisons connect these dataset-specific responses to the recovered foreground regions, chart bars, and interface boundaries. The albedo conditions test changes in signal strength, the warm and cool surfaces test color response, and the texture conditions test spatial background variation. OARL preserves its advantage across these distinct surface factors, with particularly consistent chart structure on Figure and regional organization on Game. The results demonstrate robust screen reconstruction over the evaluated wall appearances.



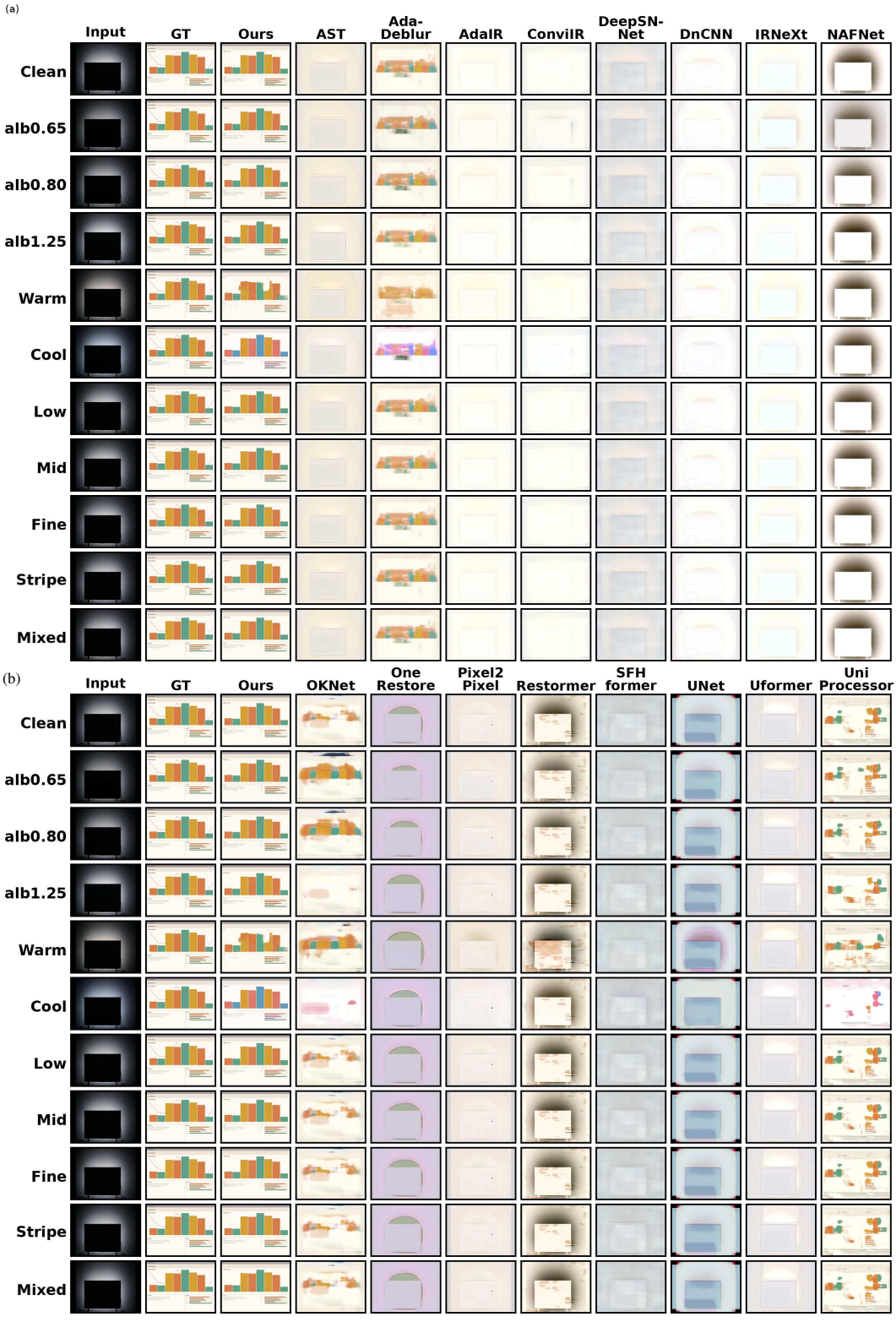
Supplementary Figure 52: OSLock: eleven surface conditions.

OARL preserves the bright dialog and blue-to-warm background arrangement across the surface settings. Several baselines introduce broad bands or dark blocks, while OARL retains a more faithful lock-screen organization.



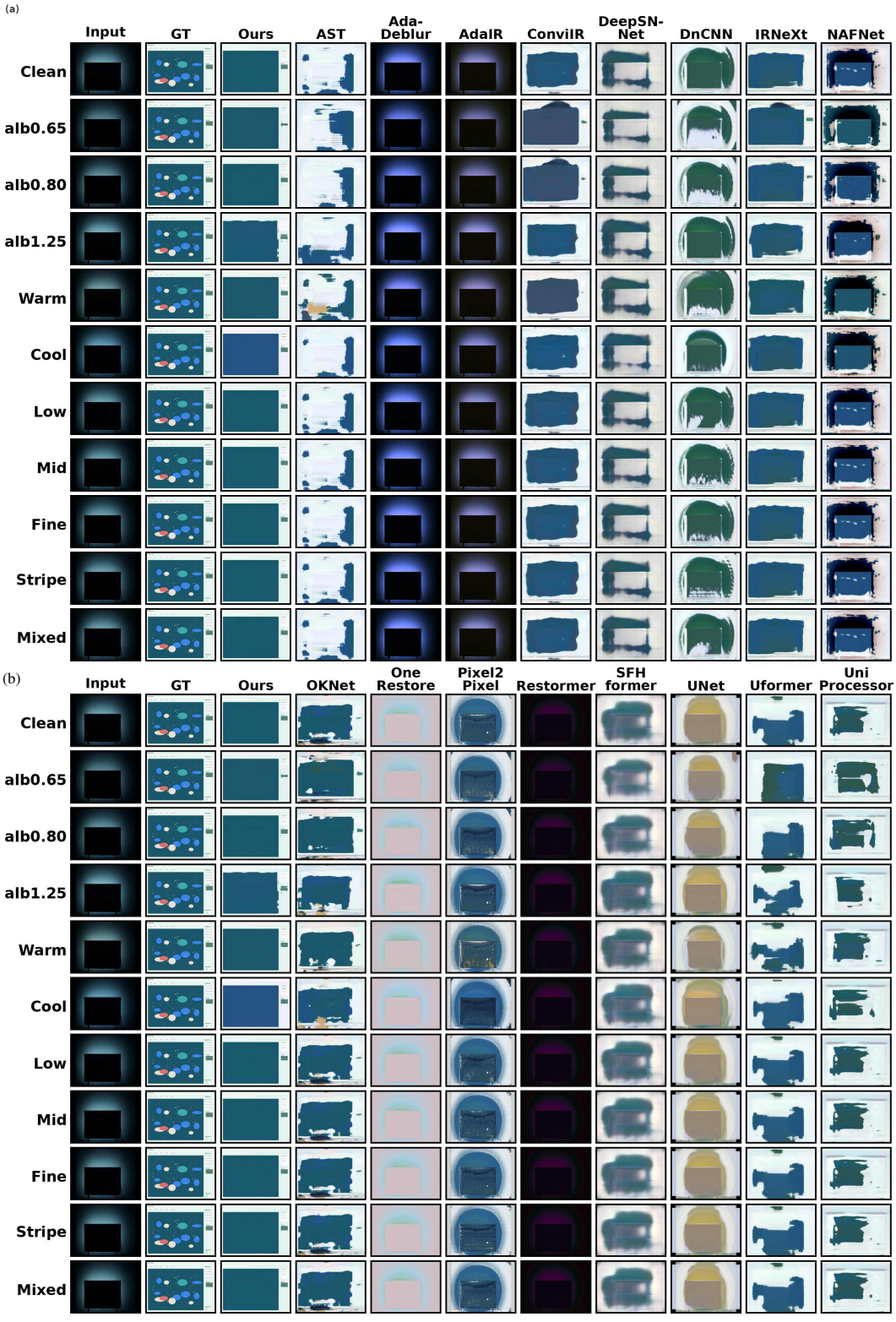
Supplementary Figure 53: Screen: eleven surface conditions.

OARL recovers the white table window and its dark frame clearly across most surface conditions. Several competing outputs merge the window into diffuse fields, highlighting the stronger foreground organization recovered by OARL.



Supplementary Figure 54: Figure: eleven surface conditions.

OARL retains the ordered colored bars across albedo and texture settings and preserves their overall arrangement under warm and cool surfaces. Several baselines return pale fields or mixed patches with weaker chart structure.

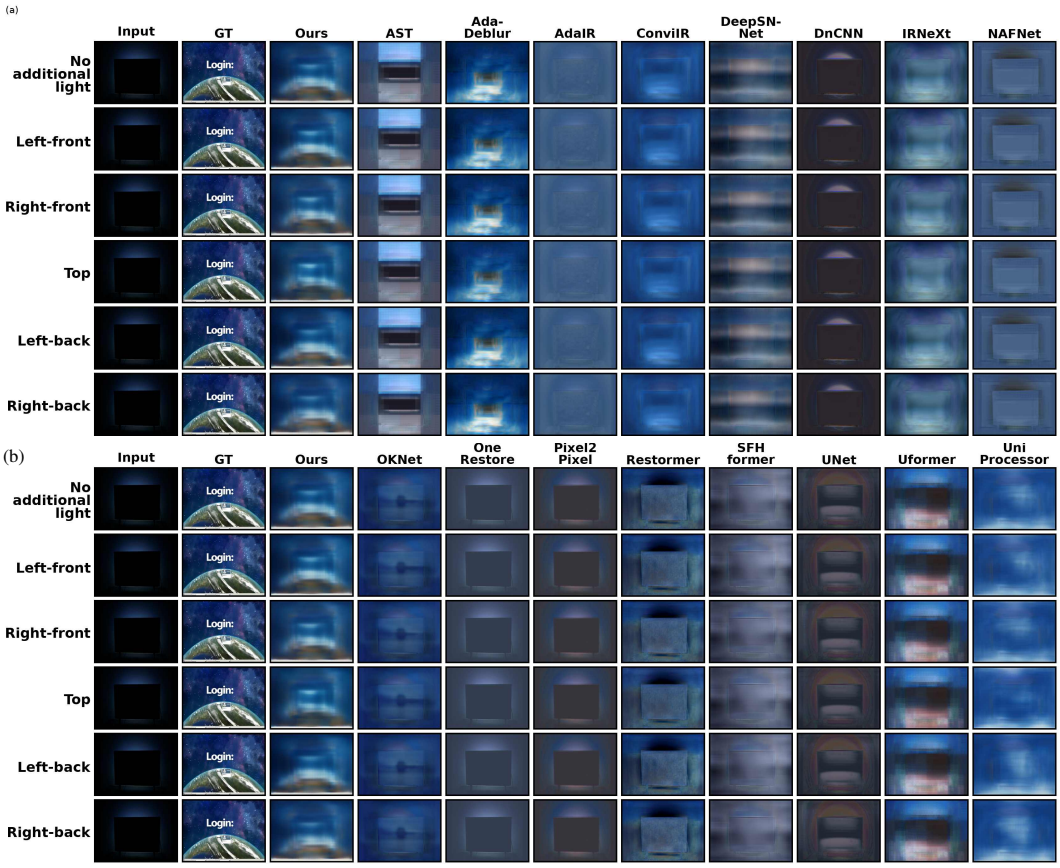


Supplementary Figure 55: Game: eleven surface conditions.

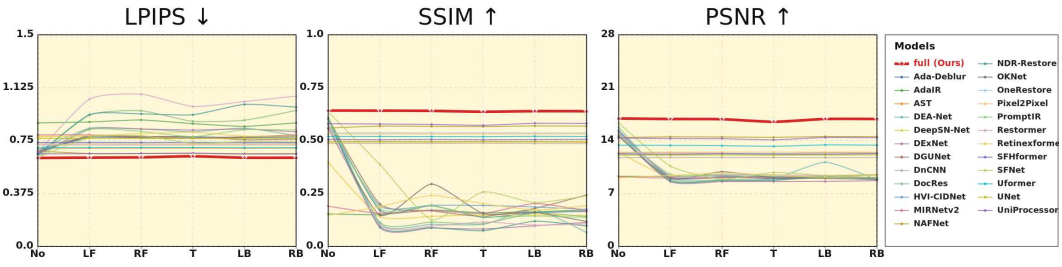
OARL preserves the teal field and narrow light sidebar across material and texture changes. Several competing outputs contain broad holes or distorted borders, highlighting the cleaner division between the playing region and controls.

C.16 Robustness to the direction of additional ambient light

Overlapping windows, ordered chart bars, and adjoining control regions test the recovery of screen organization under directional illumination. We compare no additional ambient light with illumination from the left-front, right-front, top, left-back, and right-back. The displayed screen, projection geometry, and all other imaging parameters remain fixed.



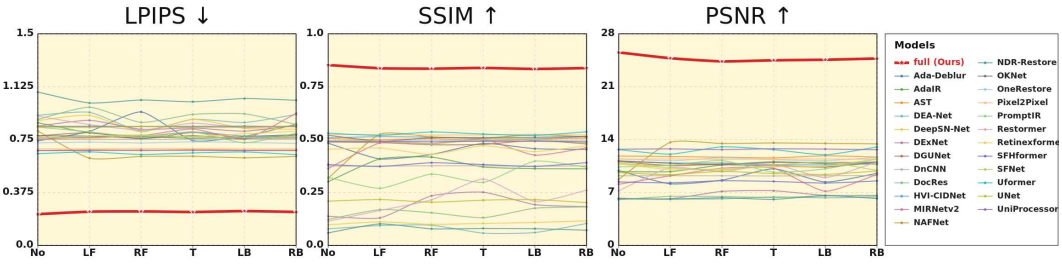
Supplementary Figure 56: OSLock under ambient-light direction changes. Panels (a) and (b) show the two baseline groups. Rows correspond to no additional light, left-front, right-front, top, left-back, and right-back ambient lighting; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



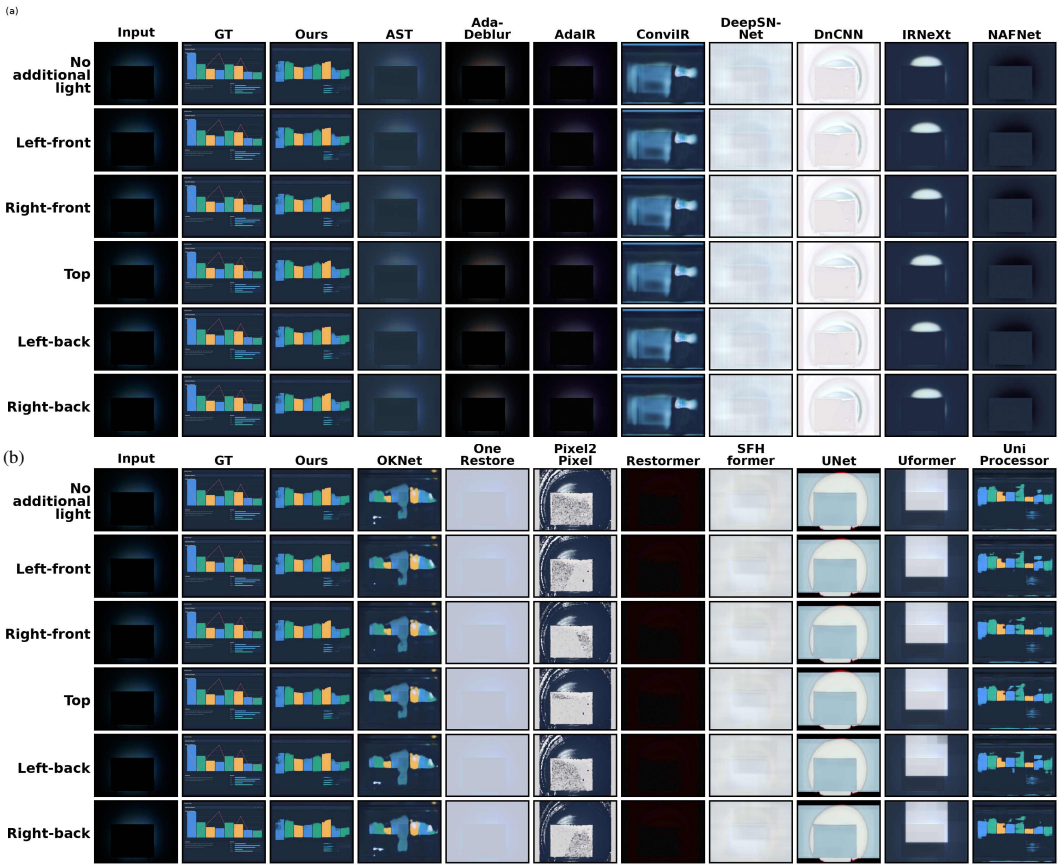
On OSLock, OARL preserves the dark blue upper background and the bright curved mountain region near the bottom across ambient-light directions. Several competing methods return horizontal bands or misplaced rectangular responses. OARL maintains a more faithful division between the upper and lower screen regions as additional light enters from the front, above, or behind. Its LPIPS and SSIM curves remain nearly constant, and PSNR retains a clear advantage across the six conditions. These results demonstrate stable recovery of the dominant lock-screen organization under changes in the direction of background illumination.



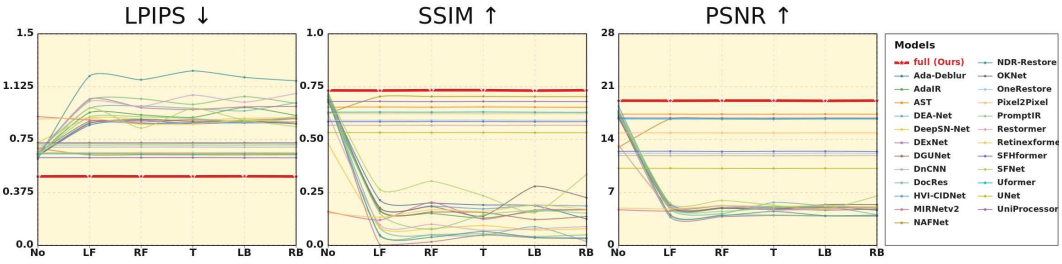
Supplementary Figure 57: Screen under ambient-light direction changes. Panels (a) and (b) show the two baseline groups. Rows correspond to no additional light, left-front, right-front, top, left-back, and right-back ambient lighting; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



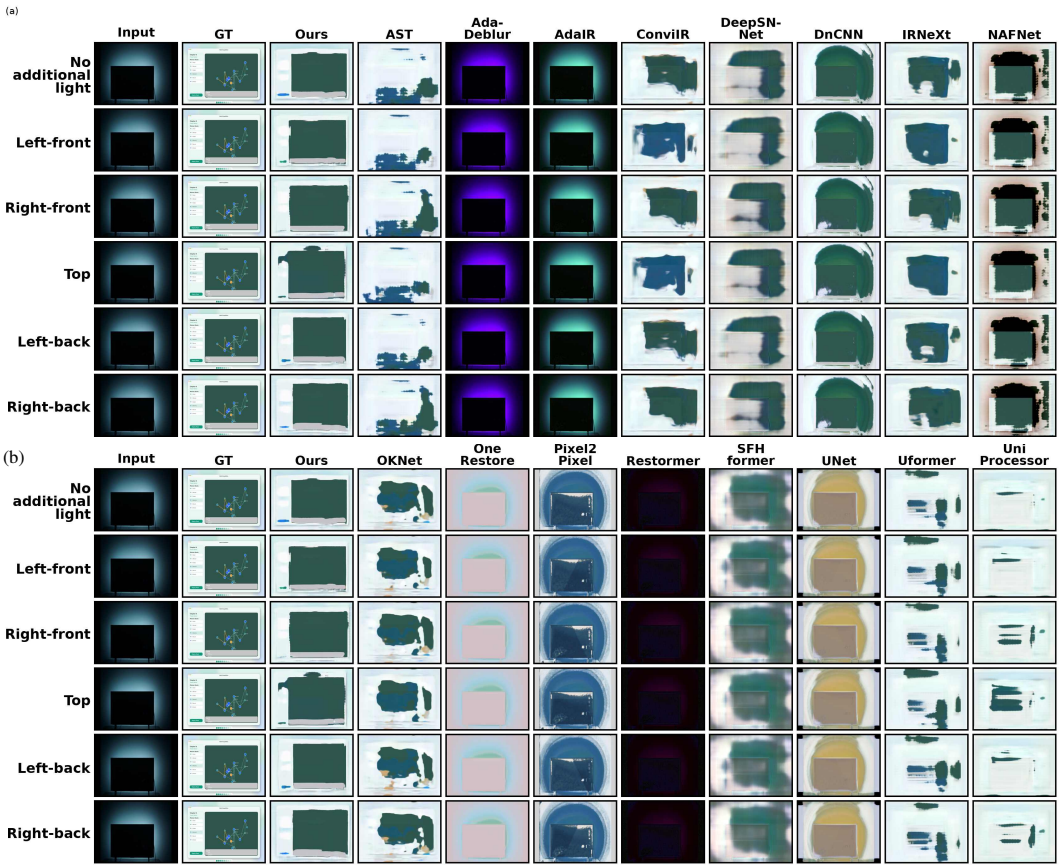
The Screen example contains a dark foreground window over a larger light application region, with a narrow red upper strip. OARL preserves this nested arrangement across ambient-light directions, keeping the dark foreground localized within the light interface. Several competing outputs replace the window structure with diffuse patches, broad tinted fields, or unrelated contours. OARL's low LPIPS and high SSIM and PSNR vary only slightly across the six conditions, including top and rear illumination. The quantitative separation agrees with the consistent alignment of the nested windows. These results demonstrate robust recovery of overlapping interface regions as the direction of additional ambient light changes the appearance of the wall glow. The overlap between the dark foreground and light background windows provides a direct test of layered interface organization. OARL keeps the upper dark region within the larger application footprint and preserves the colored top boundary, linking its low perceptual error to concrete foreground and background relationships visible in the target.



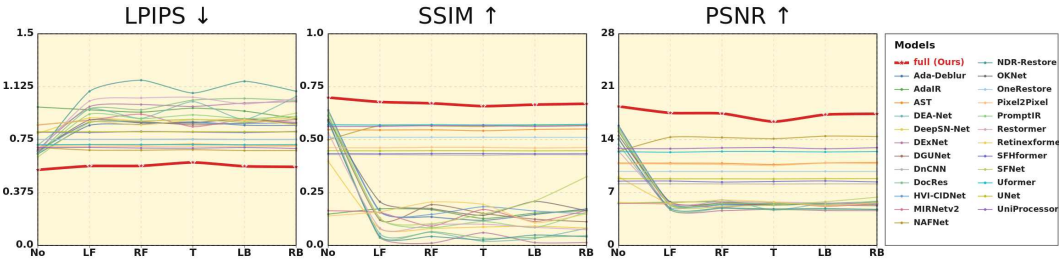
Supplementary Figure 58: Figure under ambient-light direction changes. Panels (a) and (b) show the two baseline groups. Rows correspond to no additional light, left-front, right-front, top, left-back, and right-back ambient lighting; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



On Figure, OARL preserves the sequence of blue, green, and yellow bars on the dark chart canvas across ambient-light directions. The colored bars remain distinct and retain their overall height pattern, whereas several competing outputs become uniform blocks or broad curved regions. The OARL curves stay nearly flat in all three metrics, with a clear advantage after additional light is introduced. The preserved ordering of the bar sequence therefore agrees with the aggregate structural and perceptual results. This comparison demonstrates strong robustness for graphical content under directional background illumination: the chart's main colored structures remain organized across front, top, and rear lighting conditions, maintaining a more faithful target appearance than the competing reconstructions. The surrounding dark canvas remains distinct from the colored bars, making their positions and ordering easier to compare with the reference. Several baselines preserve only the canvas-like block and discard the organized colored response; OARL retains both components, reinforcing its clear advantage in the structural metric across lighting directions.



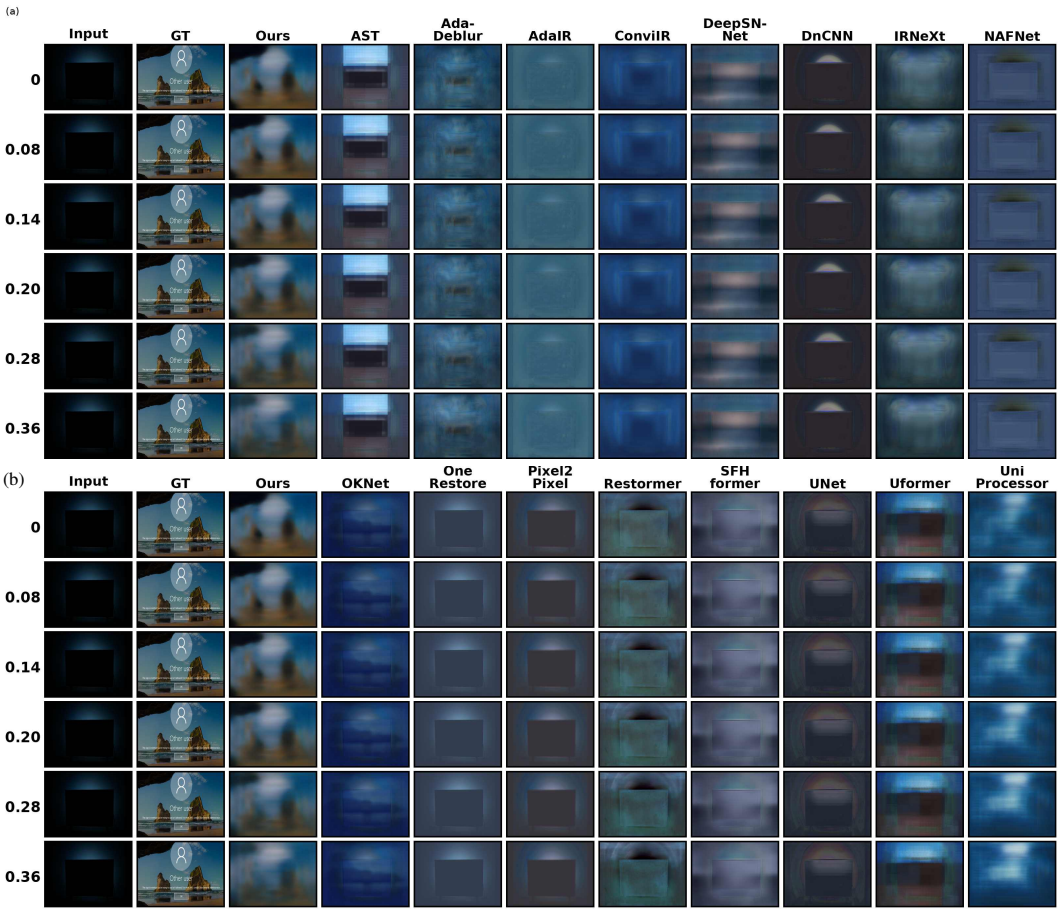
Supplementary Figure 59: Game under ambient-light direction changes. Panels (a) and (b) show the two baseline groups. Rows correspond to no additional light, left-front, right-front, top, left-back, and right-back ambient lighting; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



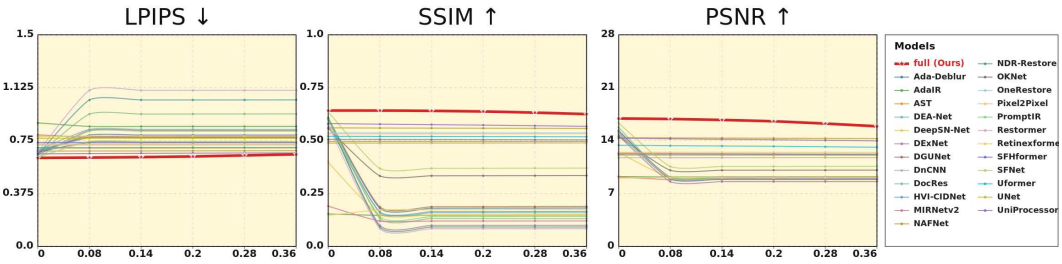
On Game, OARL retains the green central field and its light surrounding interface across the ambient-light directions. Several competing methods create large internal gaps or replace the rectangular field with irregular patches. OARL preserves a more coherent boundary between the field and the control region, including under rear illumination. Its SSIM and PSNR remain above the competing curves and its LPIPS stays favorable throughout the comparison. The variation under top lighting preserves the same overall ranking, agreeing with the stronger regional organization in the qualitative results. Together, these results demonstrate that OARL retains the principal game layout when additional ambient illumination changes direction, with a consistent advantage in both spatial coherence and aggregate reconstruction quality. The control strip stays adjacent to the central field rather than becoming a broad bright opening within it. This regional relationship remains closer to the target in OARL than in outputs with curved or disconnected borders, providing a concrete visual explanation for the persistent structural advantage over the lighting sweep.

C.17 Recovery under synthetic ambient-light modulation

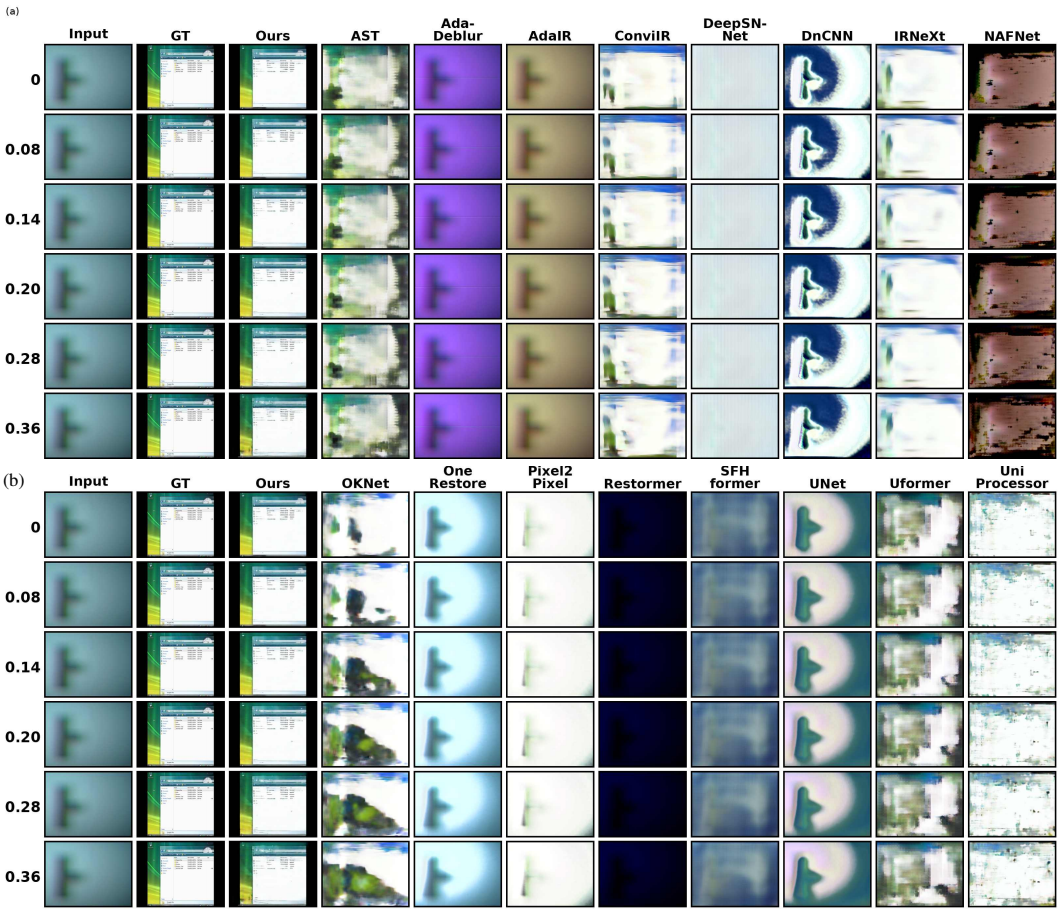
We test synthetic top-directed illumination perturbations of strength 0, 0.08, 0.14, 0.20, 0.28, and 0.36. In linear RGB, the recorded wall image is multiplied by $1 + aF$, with F normalized to $[-1, 1]$ and dimensionless strength a. The figure labels "ambient intensity" denote a, with 0 indicating the unmodified reference; the values are not measured in lux. Screen content, projection geometry, and evaluation remain fixed.



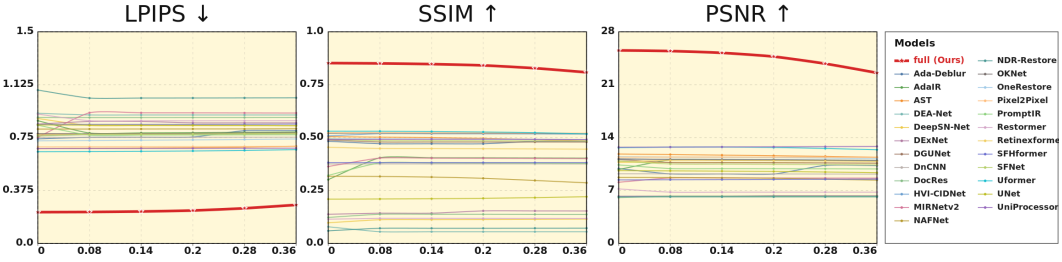
Supplementary Figure 60: OSLock under ambient-light intensity changes. Panels (a) and (b) show the two baseline groups. Rows correspond to ambient intensities of 0, 0.08, 0.14, 0.20, 0.28, and 0.36; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



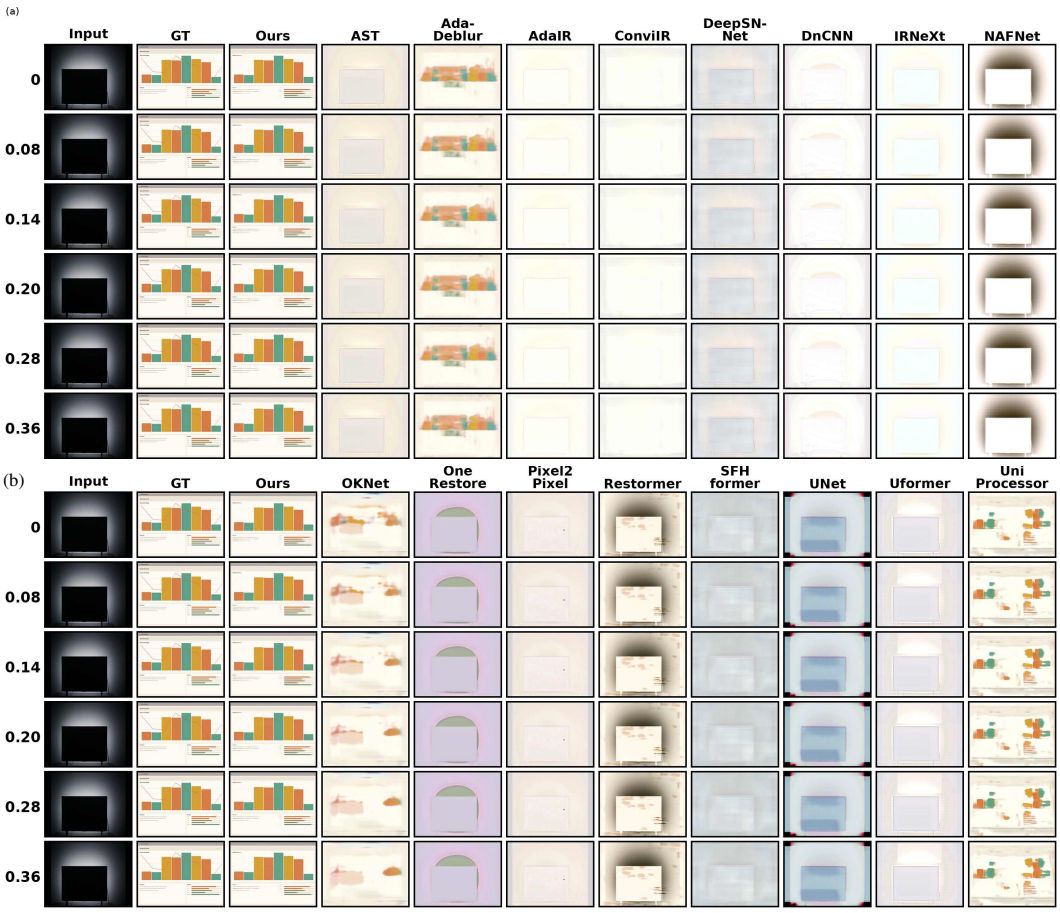
As perturbation strength increases, OARL preserves OSLock's light profile region, blue background, and warm lower scenery. Several baselines produce diffuse horizontal bands or nearly uniform regions. OARL retains the contrast between the upper interface and lower background across all six levels. LPIPS and SSIM remain nearly constant, while PSNR changes gradually and retains the leading ranking. These results support preservation of the lock-screen appearance as synthetic illumination modulation changes the wall observation's brightness and contrast.



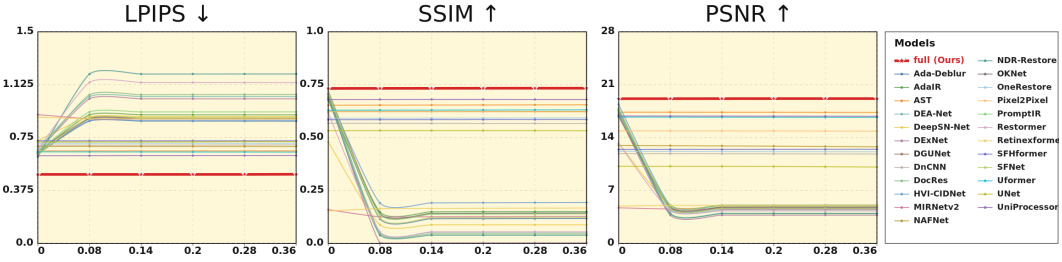
Supplementary Figure 61: Screen under ambient-light intensity changes. Panels (a) and (b) show the two baseline groups. Rows correspond to ambient intensities of 0, 0.08, 0.14, 0.20, 0.28, and 0.36; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



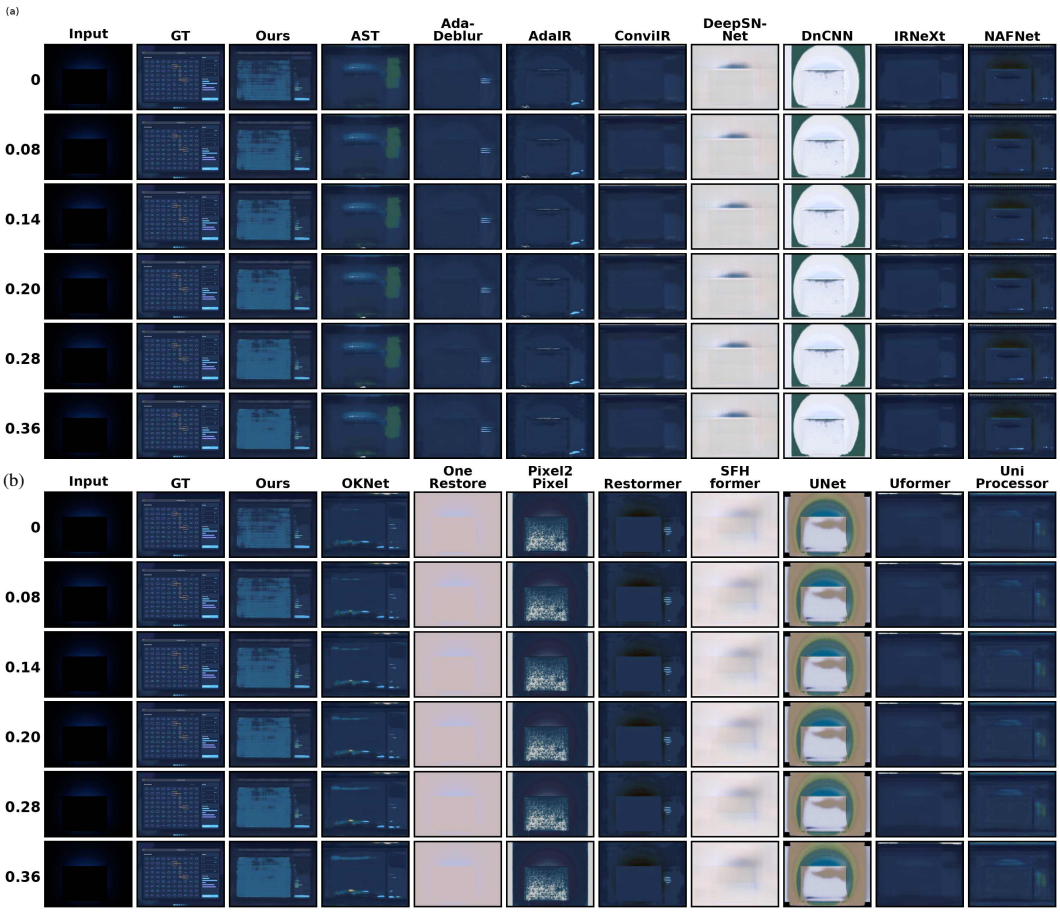
On Screen, OARL preserves the tall white application window, its green upper region, and the colored strip along the left. Several competing methods spread the foreground response across the image or replace the interface with tinted patterns. The window remains clearly localized in OARL's outputs as illumination-perturbation strength increases. Its quantitative curves change gradually over the six levels and remain well separated from the baselines, with lower LPIPS and higher SSIM and PSNR. The metric advantage at the strongest perturbation level agrees with the persistent foreground boundary in the qualitative panel. These results demonstrate robust recovery of a bright, structured interface under synthetic illumination modulation of the indirect observation. The white interior remains bounded by the application frame, and the green upper region stays separated from the surrounding screen. These complementary boundaries distinguish the recovered interface from a diffuse bright patch, reinforcing the agreement between foreground localization and the strong structural metric as perturbation strength increases.



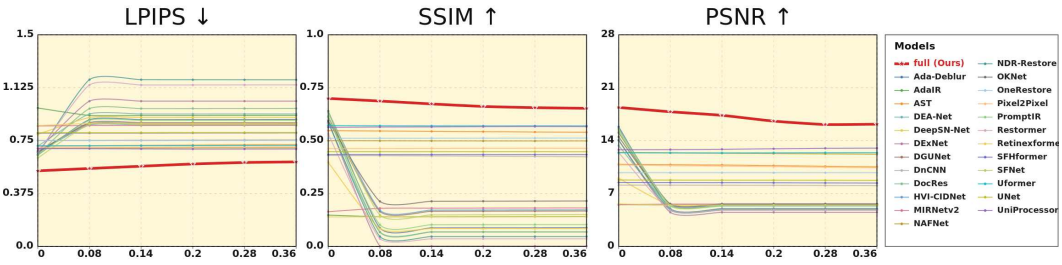
Supplementary Figure 62: Figure under ambient-light intensity changes. Panels (a) and (b) show the two baseline groups. Rows correspond to ambient intensities of 0, 0.08, 0.14, 0.20, 0.28, and 0.36; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



On Figure, OARL retains the green and orange bars against the light chart background as illumination-perturbation strength increases. The bar ordering and overall height pattern remain clearly represented across the six levels, whereas several competing outputs become nearly uniform or merge the colored structures into broad patches. OARL's three metric profiles remain nearly flat, maintaining low LPIPS and high SSIM and PSNR throughout the perturbation sweep. This agreement between distinct bar structure and aggregate quality demonstrates effective preservation of the target chart under synthetic illumination modulation. The comparison is consistent from the unmodified reference to the strongest setting, establishing a clear advantage in reconstructing organized graphical content as perturbation strength increases. The warm and cool bar colors remain separated within the light canvas, preserving a direct correspondence between the reconstructed groups and the target. This color ordering complements the stable relative heights and supports the favorable perceptual comparison under the strongest tested perturbation.



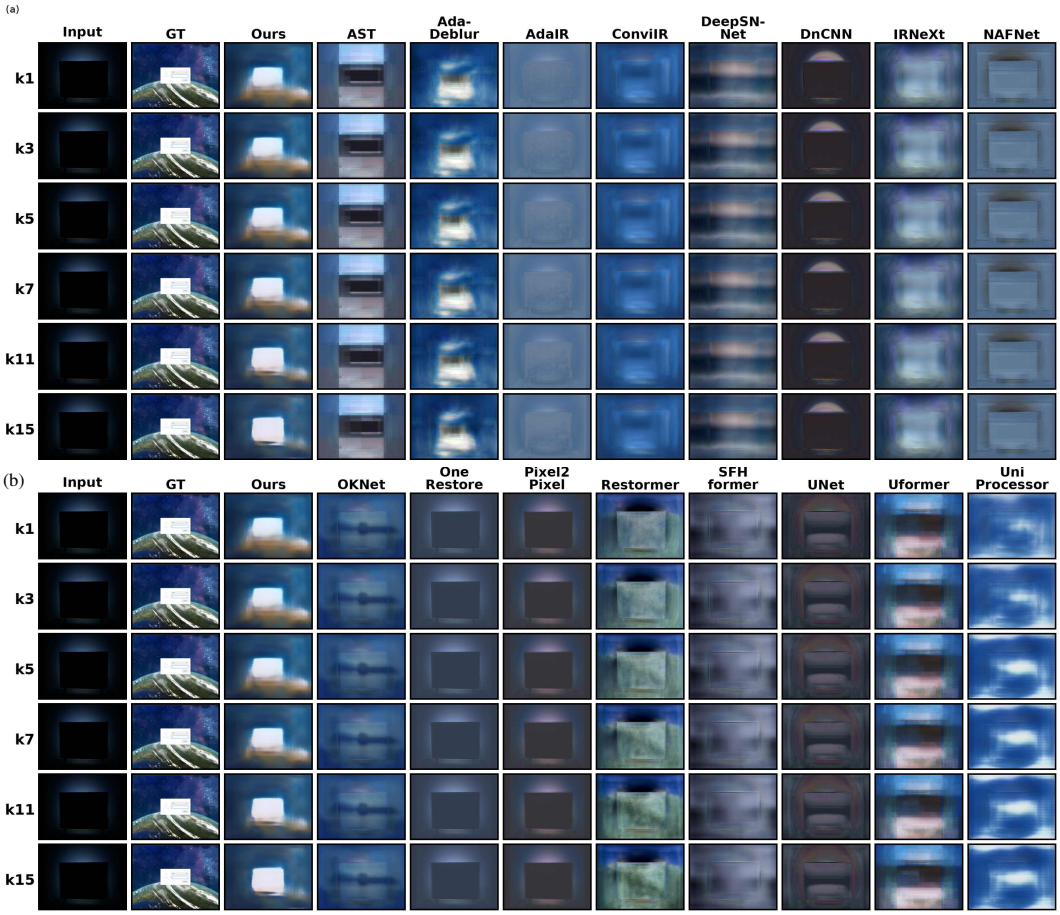
Supplementary Figure 63: Game under ambient-light intensity changes. Panels (a) and (b) show the two baseline groups. Rows correspond to ambient intensities of 0, 0.08, 0.14, 0.20, 0.28, and 0.36; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



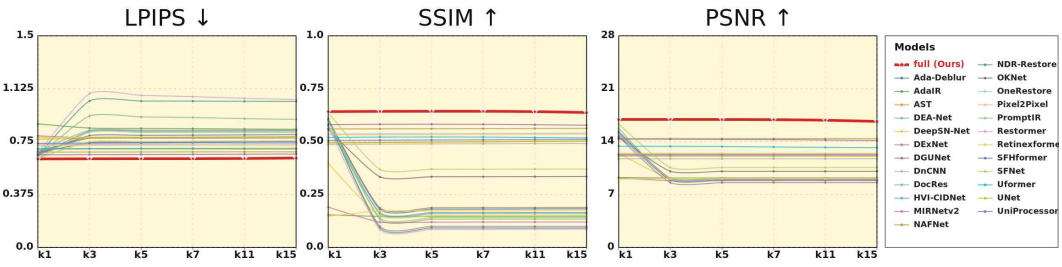
The Game example contains a dark blue board with a brighter control region at the lower right. OARL preserves the board's rectangular organization and recovers more of its internal pattern at the lower perturbation levels than several diffuse competing outputs. As illumination-perturbation strength increases, the OARL metric curves change gradually and retain the leading ranking across LPIPS, SSIM, and PSNR. The persistent aggregate advantage complements the stronger recovery of the dark screen region and its surrounding controls. These results demonstrate that OARL maintains superior reconstruction quality for a low-contrast game interface when synthetic illumination modulation changes the contrast of the wall observation over the full evaluated perturbation range. The brighter control area provides an additional positional reference within the predominantly dark interface. OARL keeps this region associated with the board footprint, while several competing methods replace the scene with broad smooth fields. The comparison therefore highlights preservation of the main screen organization as well as the favorable metric ranking.

C.18 Robustness to increasing motion blur

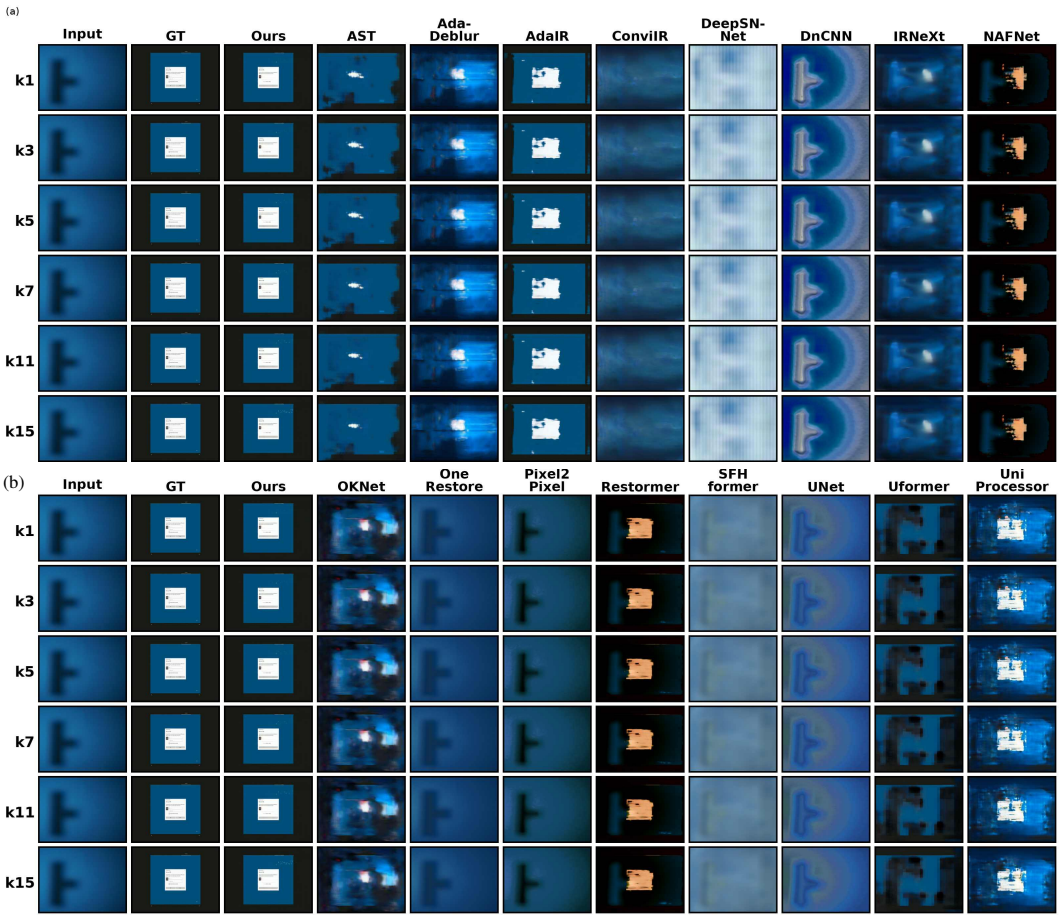
Horizontal motion blur tests recovery from broader cues as local wall-glow transitions are smoothed. Unit-sum horizontal kernels of 1, 3, 5, 7, 11, and 15 pixels (k1–k15) are applied to the recorded wall image; k1 leaves it unchanged. Screen content, geometry, and evaluation remain fixed.



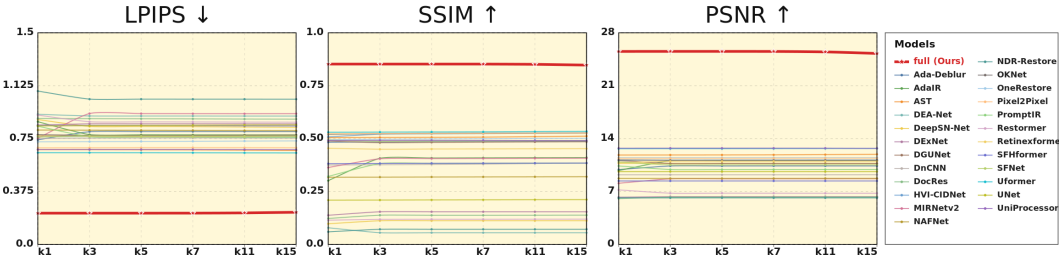
Supplementary Figure 64: OSLock under increasing motion blur. Panels (a) and (b) show the two baseline groups. Rows correspond to blur kernels k1, k3, k5, k7, k11, and k15; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



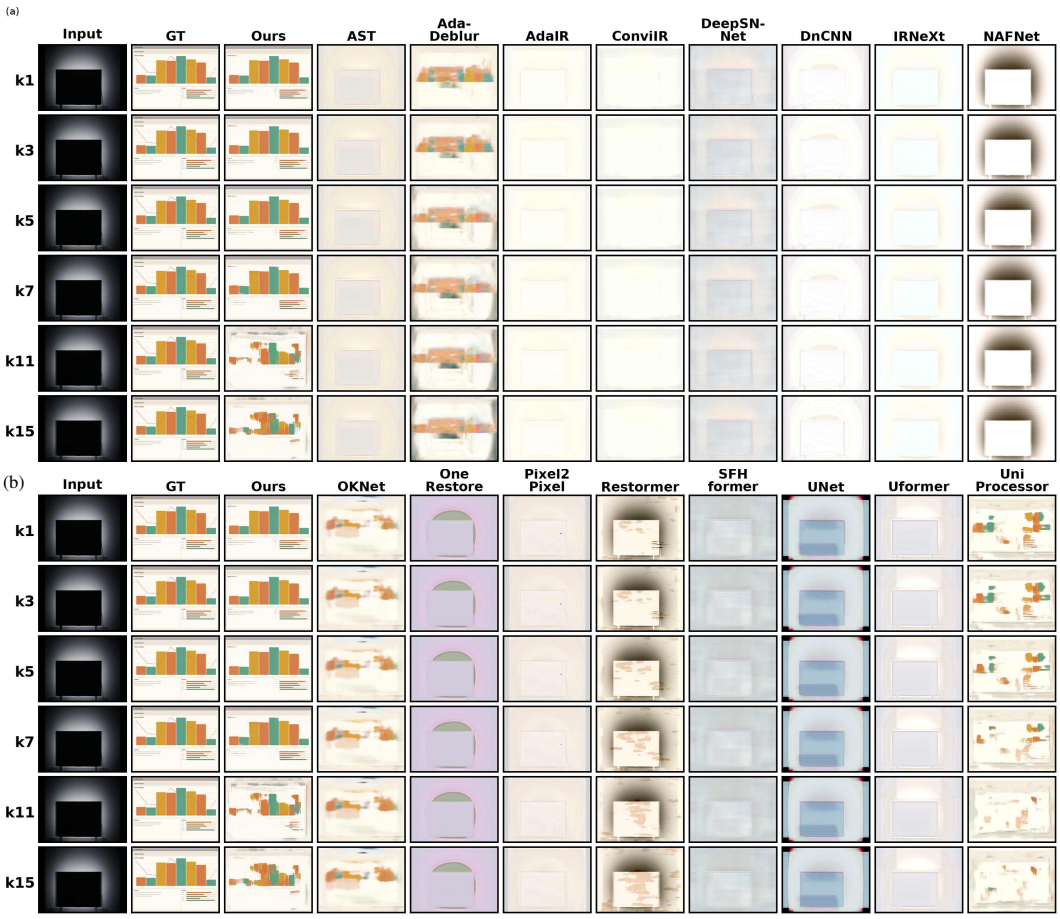
On OSLock, OARL retains the central bright dialog above the warm mountain background as the blur kernel grows. Several baselines introduce broad bands or dark rectangular structures. The dialog remains localized within the blue upper region in OARL's outputs, preserving the dominant screen arrangement through the strongest blur setting. Its nearly flat LPIPS and SSIM curves and leading PSNR agree with this visual continuity. The results demonstrate robust recovery of the principal foreground and background regions when motion blur removes local variation from the wall observation.



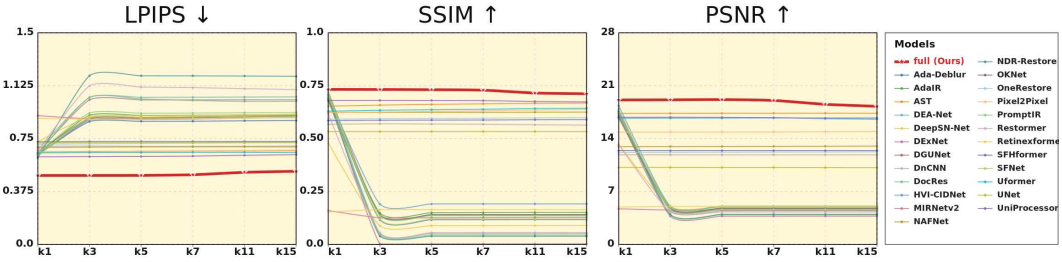
Supplementary Figure 65: Screen under increasing motion blur. Panels (a) and (b) show the two baseline groups. Rows correspond to blur kernels k1, k3, k5, k7, k11, and k15; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



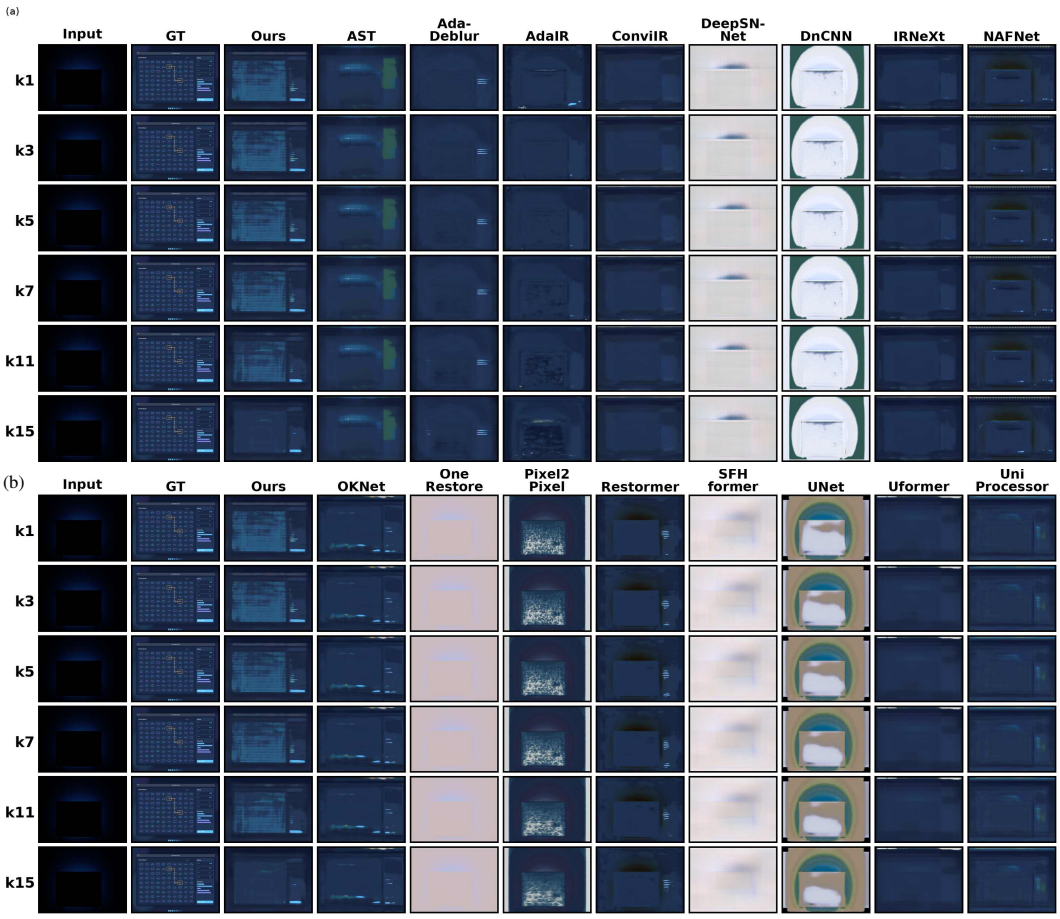
On Screen, OARL recovers the small white dialog inside the blue desktop region with consistent placement and a clear rectangular boundary from k1 through k15. Several competing methods produce fragmented bright responses or diffuse fields that do not preserve this localized interface. The OARL metric curves remain nearly constant and retain a substantial advantage in LPIPS, SSIM, and PSNR over the full kernel range. This consistency is especially clear at k11 and k15, where the recovered dialog remains distinct from its surroundings. The visual and quantitative evidence demonstrates reliable reconstruction of compact foreground content under strong motion blur, maintaining the target's spatial organization as the indirect input becomes increasingly smoothed. The dark area outside the blue desktop region remains distinct from the white foreground dialog, preserving the nested contrast structure of the target. This three-region arrangement supplies a clear visual reference across kernel sizes and reinforces the aggregate evidence for stable foreground recovery under motion blur.



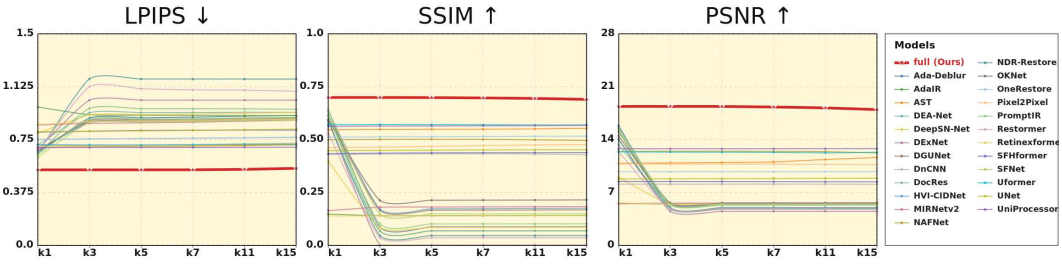
Supplementary Figure 66: Figure under increasing motion blur. Panels (a) and (b) show the two baseline groups. Rows correspond to blur kernels k1, k3, k5, k7, k11, and k15; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



The Figure example shows green and orange bars on a light chart canvas. OARL preserves the distinct bars and their relative height pattern particularly clearly from k1 through k7, while several competing methods return pale fields or broad mixed-color patches. The OARL curves retain the lowest LPIPS and highest SSIM and PSNR through k11 and k15, with a gradual response at the stronger kernels. This persistent aggregate advantage complements the more accurate bar organization under low and intermediate blur. Together, the results demonstrate stronger recovery of structured graphical content as motion blur suppresses local transitions in the wall observation, extending the visual advantage around the clearer chart conditions to the full evaluated kernel sweep. The light chart canvas remains separate from the colored bar groups, so the reconstructed content has a closer graphical organization than outputs dominated by broad illumination patterns. This correspondence between the target canvas and the recovered bar region supplies an additional visual basis for OARL's favorable perceptual ranking.



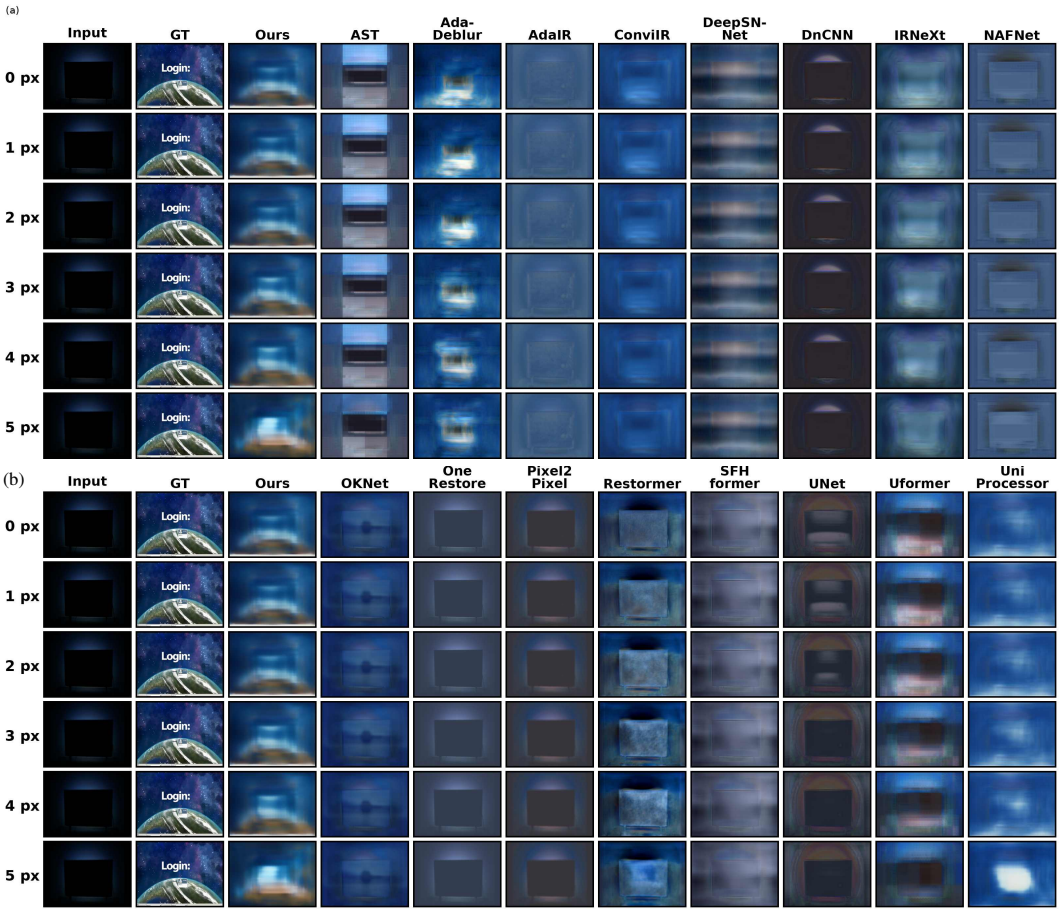
Supplementary Figure 67: Game under increasing motion blur. Panels (a) and (b) show the two baseline groups. Rows correspond to blur kernels k1, k3, k5, k7, k11, and k15; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



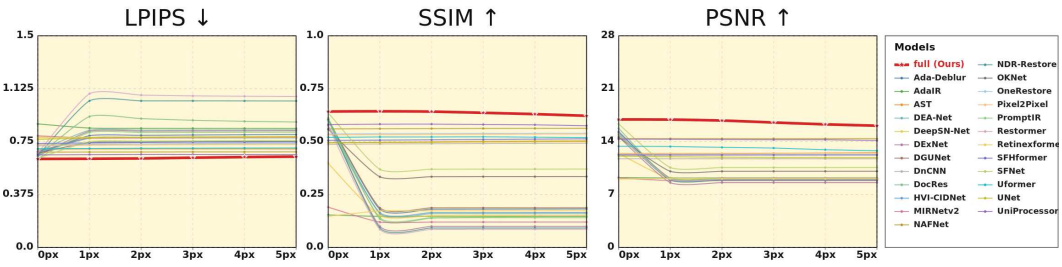
On Game, OARL preserves the dark blue board region and the bright control area near the lower right. At the smaller blur kernels, its reconstruction retains more of the internal board pattern than several competing outputs, which remain diffuse or emphasize unrelated boundaries. The OARL metric profiles vary only slightly as the kernel grows and preserve a clear lead in LPIPS, SSIM, and PSNR through the strongest blur setting. The favorable aggregate response supports the stronger reconstruction of the main game organization under reduced local evidence. These results demonstrate robust preservation of the screen's regional structure as motion blur increasingly smooths the observed wall glow. The bright control area remains located beside the darker board region, preserving a useful positional reference when local patterns are smoothed. Several competing outputs lose this distinction in large uniform fields, whereas OARL retains the main spatial division, reinforcing its structural advantage across the blur conditions.

C.19 Robustness to camera-shake intensity

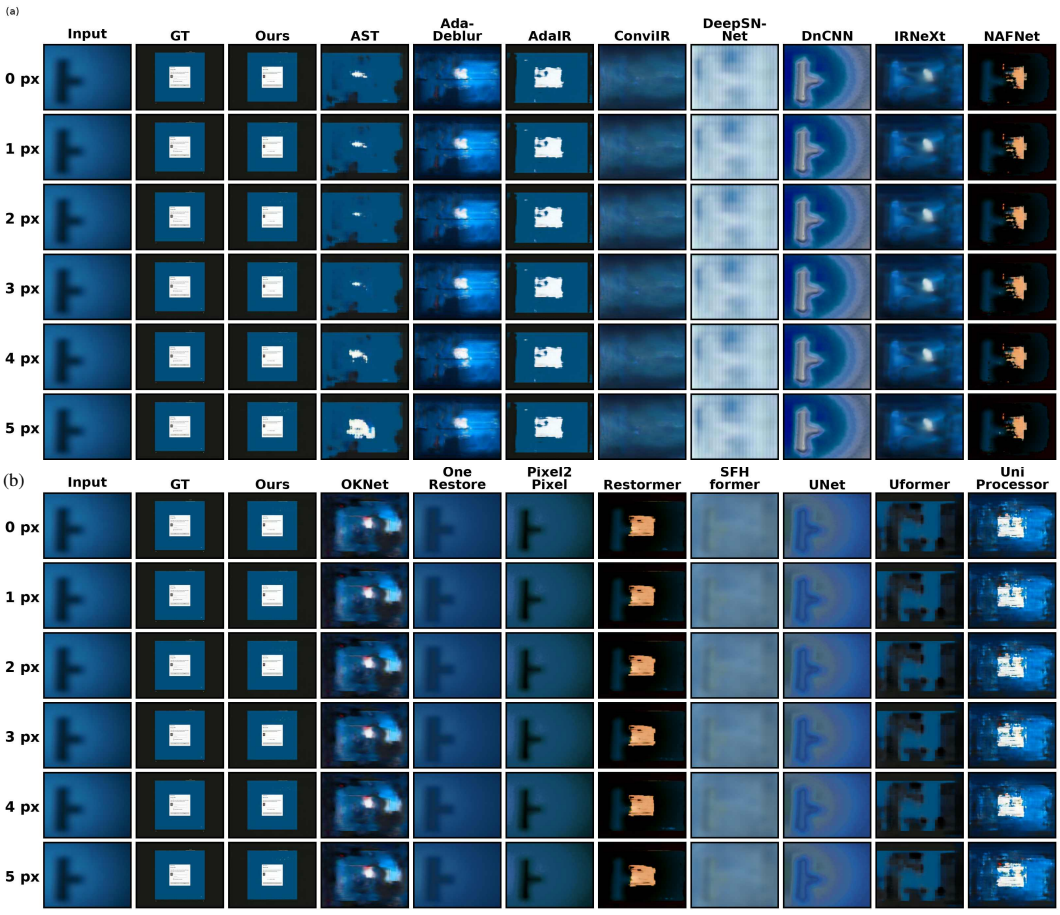
Simulated camera shake perturbs alignment by averaging seven copies of the recorded wall image, including one unchanged copy. Horizontal and vertical translations are sampled independently with maximum displacements of 1, 2, 3, 4, or 5 pixels, paired with maximum rotations of $0.10°$, $0.20°$, $0.35°$, $0.50°$, and $0.70°$, respectively. The 0 px reference is unchanged. Screen content, reconstruction, and evaluation protocols remain fixed.



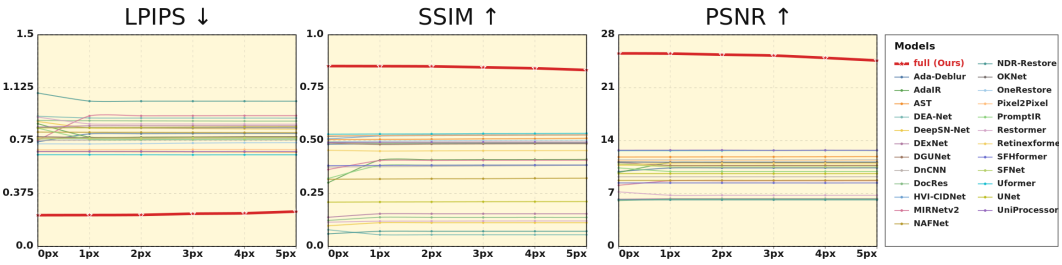
Supplementary Figure 68: OSLock under camera-shake intensity changes. Panels (a) and (b) show the two baseline groups. Rows correspond to jitter amplitudes of 0, 1, 2, 3, 4, and 5 px; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



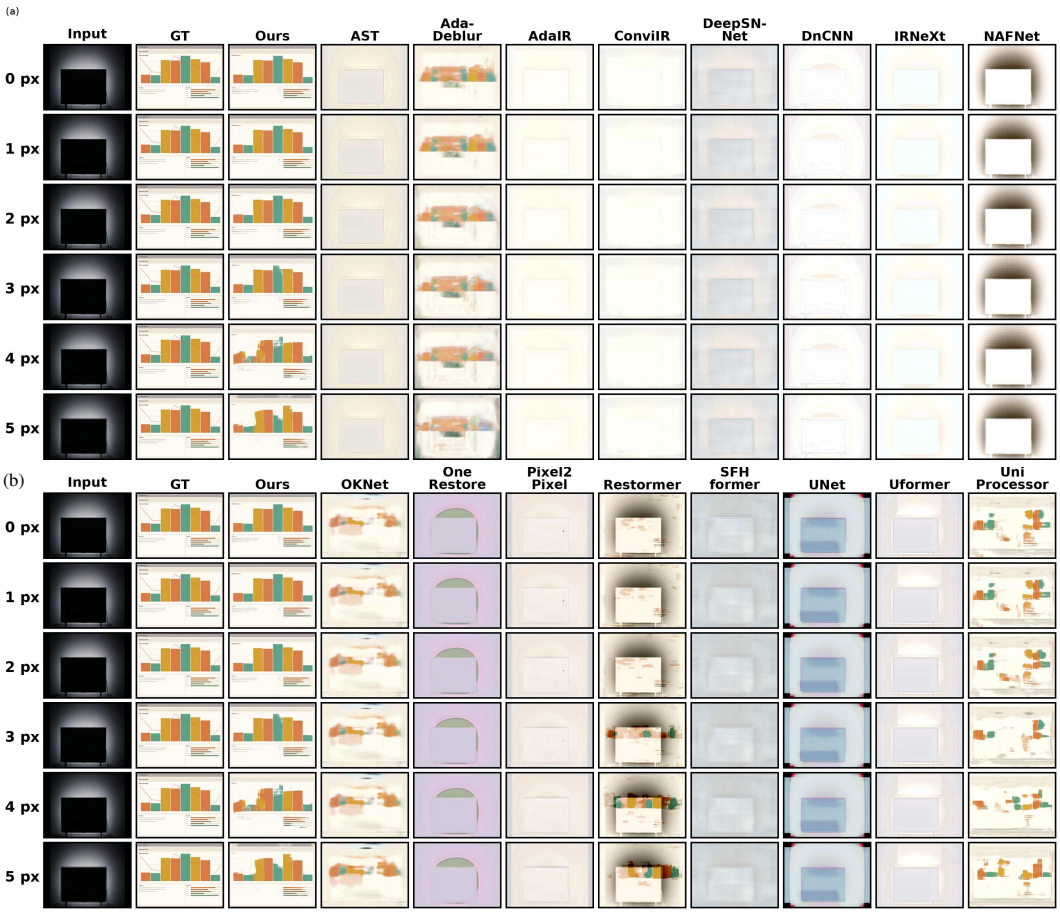
On OSLock, OARL preserves the blue upper background, bright curved lower region, and their relative placement across all six shake amplitudes. Several baselines produce broad horizontal bands or dark blocks with less faithful regional organization. Nearly constant LPIPS and gradually varying SSIM and PSNR remain favorable through 5 px, supporting recovery of the lock-screen's dominant appearance under simulated local displacement.



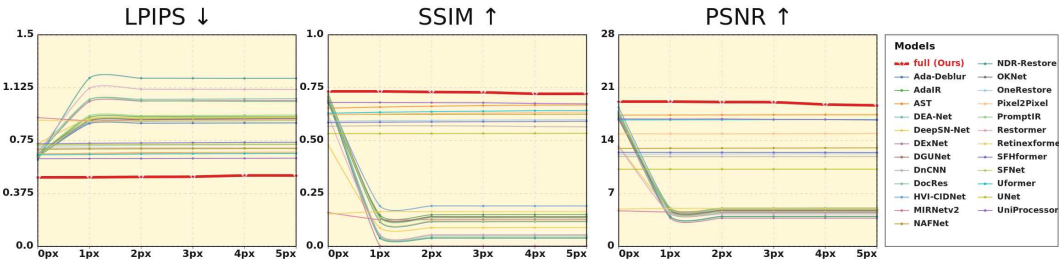
Supplementary Figure 69: Screen under camera-shake intensity changes. Panels (a) and (b) show the two baseline groups. Rows correspond to jitter amplitudes of 0, 1, 2, 3, 4, and 5 px; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



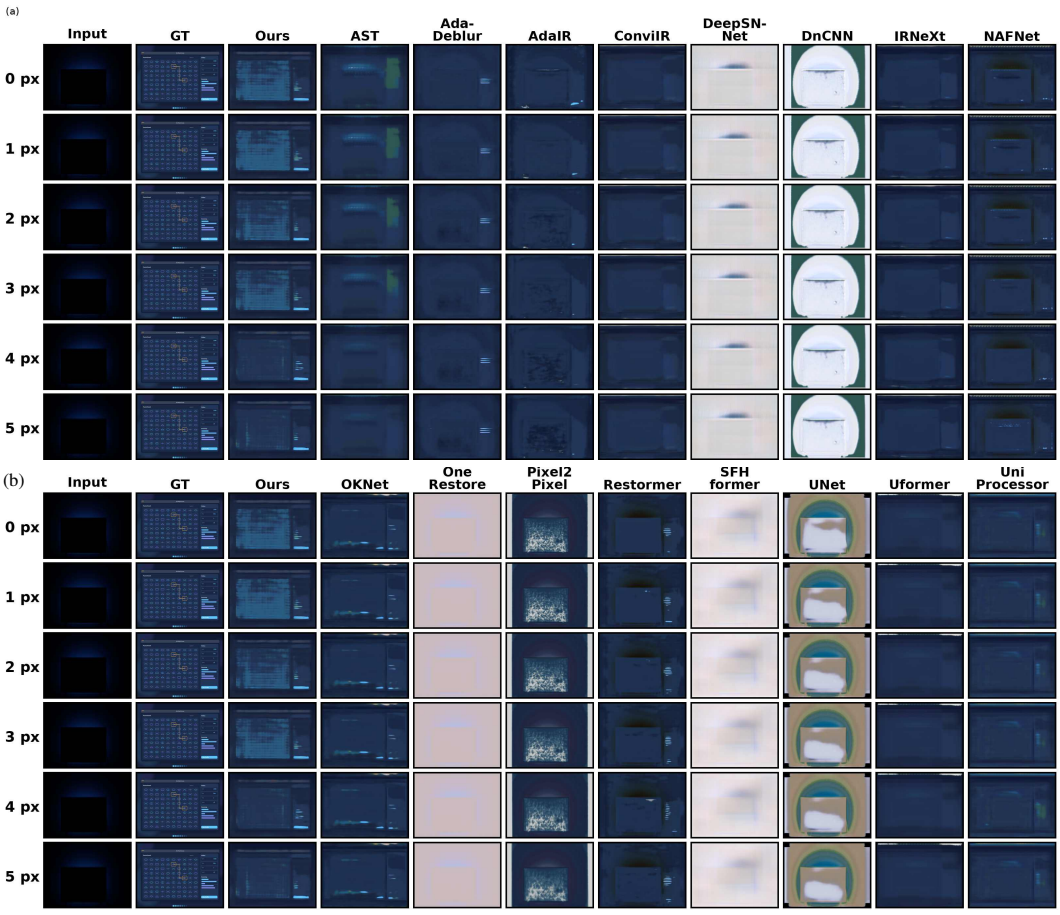
On Screen, OARL keeps the small white dialog centered within the blue desktop region as camera shake increases from 0 to 5 px. Several competing outputs disperse the foreground response or reproduce a tinted input silhouette. OARL preserves the dialog's rectangular outline and its contrast with the surrounding screen throughout the qualitative comparison. Its LPIPS and SSIM curves show little variation, and PSNR retains a substantial lead at the larger jitter amplitudes. The stable foreground location agrees with these metric profiles, demonstrating robust recovery of compact interface content under local registration changes in the observed wall glow. This persistent advantage extends across the full shake sweep, including the strongest tested displacement. The outer dark background, blue desktop block, and bright dialog form three nested regions in the reference. OARL preserves their placement under jitter, giving a direct visual counterpart to its low perceptual error and high structural similarity even when the observation undergoes the strongest tested positional displacement.



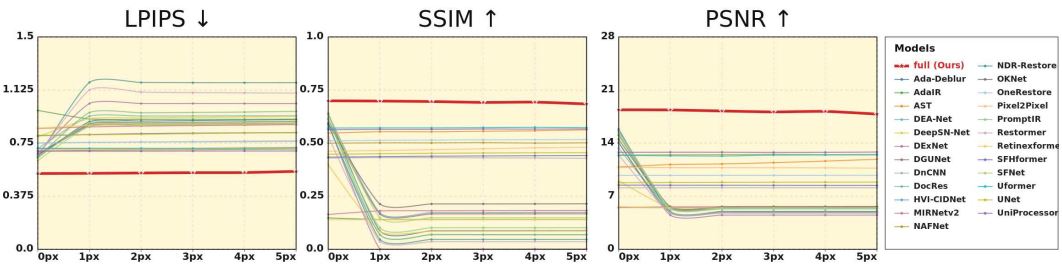
Supplementary Figure 70: Figure under camera-shake intensity changes. Panels (a) and (b) show the two baseline groups. Rows correspond to jitter amplitudes of 0, 1, 2, 3, 4, and 5 px; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



On Figure, OARL preserves the ordered green and orange bar sequence against the light chart background. The bars remain distinct over the lower shake amplitudes and retain their overall arrangement at the stronger settings, whereas several baselines produce nearly uniform outputs or mixed patches. OARL's metric curves change gradually and maintain the leading ranking in LPIPS, SSIM, and PSNR through 5 px. The persistent quantitative separation supports the stronger recovery of the chart's colored structures despite local positional errors. These results demonstrate robust reconstruction of organized graphical content under camera shake, with coherent bar placement and a clear aggregate advantage across the evaluated jitter amplitudes. The light canvas and colored bar groups also remain spatially separated in OARL's outputs. This relationship distinguishes the recovered chart from the diffuse fields returned by several baselines and connects the favorable perceptual response to a concrete organization of target colors and graphical regions under local camera displacement.



Supplementary Figure 71: Game under camera-shake intensity changes. Panels (a) and (b) show the two baseline groups. Rows correspond to jitter amplitudes of 0, 1, 2, 3, 4, and 5 px; columns show the input, ground truth, and compared reconstructions. The curves below report LPIPS (↓), SSIM and PSNR (↑) for this dataset.



On Game, OARL preserves the dark blue board and the bright control region near the lower right under camera-position jitter. Its outputs retain more of the board pattern at the smaller amplitudes than several competing reconstructions, which remain diffuse or emphasize unrelated contours. As displacement increases, the OARL curves retain low LPIPS together with leading SSIM and PSNR, showing only gradual changes over the sweep. The persistent advantage at 3–5 px demonstrates stronger recovery under local registration errors. Together with the more faithful organization of the board and its controls, these results establish OARL's robustness to camera shake in reconstructing the principal regions of a dark game interface. The lower-right control area supplies a stable reference for the extent of the board, while the dark background preserves the target's overall appearance. OARL maintains these main relationships more faithfully than outputs dominated by smooth blocks or curved contours, supporting the aggregate structural advantage at larger shake amplitudes.

C.20 Summary of the robustness protocol

The preceding sweeps provide complementary tests of the observation-to-screen mapping. Photometric and sensor perturbations change evidence strength; geometric perturbations change its spatial arrangement; surface and ambient-light perturbations change the recorded background; and motion perturbations remove or displace local structure. Across these studies, OARL preserves the dominant screen layout over the evaluated conditions, with the metric responses documenting sensitivity to each perturbation. The next tests connect these reconstruction results to target-relevant evidence, agreement with the observed wall glow, learning efficiency, and state inheritance.

Mechanism stress tests

The following five experiments examine evidence selection and evidence accumulation under controlled conditions. REP-PK1 and REP-PK3 test the enhanced representation under misleading and weak target contours. LRN-PK1 evaluates wall consistency after refinement, LRN-PK2 tests learning efficiency under a matched one-epoch budget, and LRN-PK4 examines strong-to-weak evidence transfer with validation-selected stage gains. Together, these comparisons show how the evaluated representation, learning, refinement, and stage-transfer configurations support screen recovery.

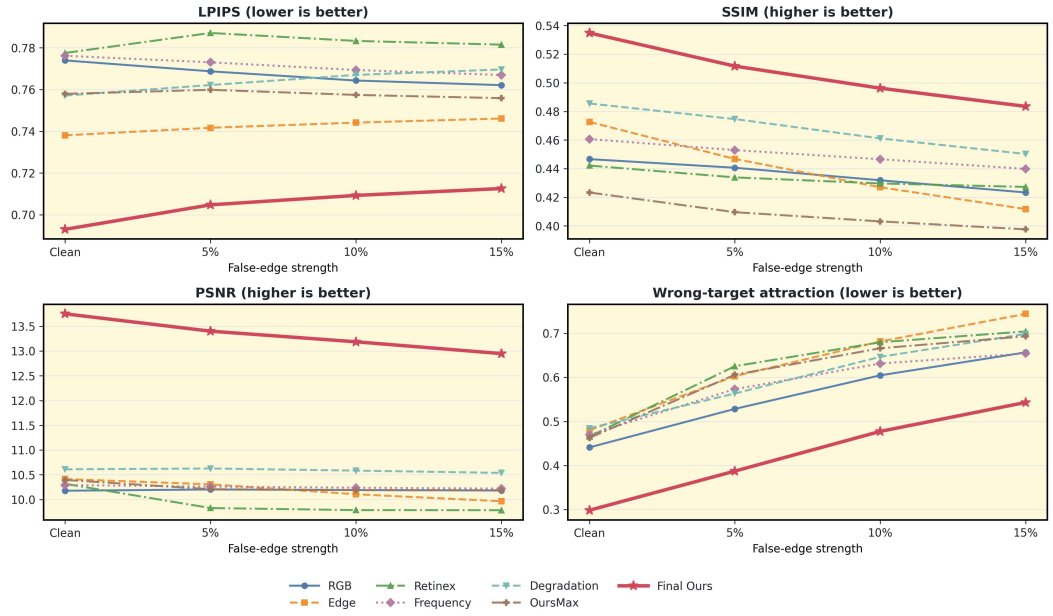
C.21 REP-PK1: Strong false-cue competition

The enhanced OARL representation used here and in the following visibility analysis uses a guarded multi-scale configuration for Figure, an evidence-fusion head for OSLock, and a mixture-of-experts fusion configuration for Screen, Game, and Browser. The diagnostic subset named Screen2000 belongs to the Screen content category; its sample scope is specific to these diagnostics. These category-specific settings define the enhanced OARL representation evaluated in the two representation stress tests; Table 3 evaluates the distinct main representation configuration. The diagnostic label "Final Ours" denotes this enhanced representation. This experiment tests whether the representation follows a conspicuous but incorrect screen contour when that contour competes with the weaker, correct wall reflection. We keep the original wall observation and target screen fixed, then add a contour from a nonmatching donor screen at 5%, 10%, and 15% strength; the clean condition is the reference. The model should reconstruct the paired target rather than the donor. We report LPIPS (lower is better), SSIM and PSNR (higher is better), and wrong-target attraction rate (lower is better). For each sample, attraction is recorded when the reconstruction has a strictly smaller LPIPS distance to the nonmatching donor than to the paired target; the rate is the fraction of such samples. Each table entry averages the five category-level rates equally. The formal test set contains five content categories, four conditions, and four metrics (80 comparison units); Final Ours is first in all 80 units.

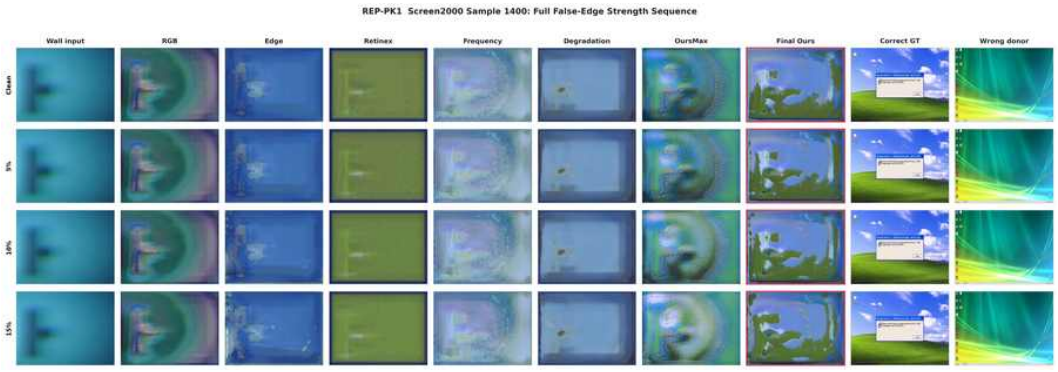
Table S1: REP-PK1 five-dataset macro averages under increasing false-contour strength. The table reports macro-averaged PSNR, SSIM, LPIPS, and wrong-target attraction for all seven representations under four false-contour conditions.

Representation	Clean				5% decoy				10% decoy				15% decoy			
	PSNR↑	SSIM↑	LPIPS↓	Wrong↓	PSNR↑	SSIM↑	LPIPS↓	Wrong↓	PSNR↑	SSIM↑	LPIPS↓	Wrong↓	PSNR↑	SSIM↑	LPIPS↓	Wrong↓
RGB	10.1771	0.4467	0.7740	0.4411	10.2020	0.4406	0.7688	0.5285	10.1907	0.4319	0.7643	0.6046	10.1815	0.4234	0.7621	0.6567
Edge	10.4125	0.4726	0.7381	0.4800	10.3058	0.4468	0.7417	0.6028	10.1054	0.4270	0.7442	0.6814	9.9643	0.4118	0.7461	0.7442
Retinex	10.3161	0.4422	0.7775	0.4659	9.8268	0.4339	0.7871	0.6253	9.7847	0.4297	0.7833	0.6797	9.7831	0.4272	0.7815	0.7043
Frequency	10.2918	0.4607	0.7762	0.4697	10.2596	0.4530	0.7731	0.5733	10.2387	0.4466	0.7694	0.6315	10.2198	0.4398	0.7670	0.6544
Degradation	10.6095	0.4855	0.7571	0.4846	10.6266	0.4747	0.7622	0.5634	10.5834	0.4611	0.7670	0.6465	10.5378	0.4503	0.7696	0.6987
OursMax	10.3951	0.4233	0.7580	0.4627	10.2068	0.4096	0.7599	0.6060	10.1892	0.4032	0.7574	0.6660	10.1851	0.3976	0.7559	0.6933
Final Ours	**13.7530**	**0.5348**	**0.6930**	**0.2985**	**13.4025**	**0.5117**	**0.7048**	**0.3873**	**13.1878**	**0.4962**	**0.7093**	**0.4774**	**12.9492**	**0.4835**	**0.7126**	**0.5431**

As false-contour interference increases, the enhanced representation's macro-averaged PSNR and SSIM decrease and its LPIPS and wrong-target attraction increase, while it retains the leading value for each metric at every tested level. The lower wrong-target attraction indicates a stronger preference for paired content relative to the compared representations as false-contour interference increases.



Supplementary Figure 72: REP-PK1 macro-averaged LPIPS, SSIM, PSNR, and wrong-target attraction as the false-contour strength increases from clean to 15%. Curves are shown for the same seven fixed representations at every condition; the reported macro averages cover five datasets.



Supplementary Figure 73: REP-PK1 qualitative outputs for a Screen2000 sample under clean, 5%, 10%, and 15% donor-contour conditions. Final Ours is compared with the fixed representation baselines, the correct target, and the wrong donor.

C.22 REP-PK3: Contour-visibility stratification

We next examine recoverability-guided reconstruction as target-contour visibility varies from low to high. Original test samples are grouped by target-contour visibility into Low, Mid, and High bins. These labels stratify evaluation only; model inputs remain the original wall-glow observations. Within each category, visibility is the mean product of standardized 3×3 Sobel gradient-magnitude maps of the grayscale wall observation and target screen over their valid crops. Each map is centered and divided by its standard deviation with a 10^{-6} floor. Samples are sorted by this score, with sample identifier breaking ties, and divided into three approximately equal-sized groups in increasing order. Each bin is evaluated on five datasets using LPIPS, SSIM, and PSNR. Across the 45 dataset–bin–metric comparisons, Final Ours ranks first in every unit.

Table S2: REP-PK3 contour-visibility stratification. The table reports all seven fixed representations for each dataset and visibility bin. The strongest non-Ours value is shown in blue and the Final Ours value in red.

Representation	Figure			Screen2000			Game			OSLock			Browser		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Low visibility															
RGB-Rep	12.6162	0.5921	0.7133	8.4099	0.3260	0.8016	10.8065	0.5198	0.7654	11.9847	0.5101	0.7364	9.1708	0.4567	0.7641
Edge-Rep	12.6835	0.5766	0.6866	8.9347	0.4534	0.7910	9.8069	0.4931	0.7216	12.5656	0.5375	0.7260	9.8483	0.4735	0.7098
Retinex-Rep	11.9762	0.5002	0.7217	9.6430	0.4529	0.8070	9.7404	0.4721	0.7600	12.7566	0.5218	0.7605	9.0223	0.4182	0.7770
Frequency-Rep	12.3957	0.5828	0.7255	9.4349	0.4256	0.8211	10.0269	0.4946	0.7749	12.2870	0.4942	0.7618	9.6858	0.4573	0.7346
Degradation-Rep	12.3283	0.5752	0.7219	9.5034	0.4691	0.7897	9.8091	0.5115	0.7853	12.2966	0.5306	0.7348	11.9923	0.5354	0.6677
Ours-Rep-Max	12.7062	0.5410	0.7007	8.7507	0.3786	0.8275	10.5165	0.4730	0.7205	12.1934	0.4588	0.7751	10.2286	0.4140	0.7068
Final Ours	**14.4559**	**0.5960**	**0.6621**	**13.3250**	**0.4970**	**0.7110**	**14.4239**	**0.5812**	**0.6914**	**14.5148**	**0.6168**	**0.6829**	**15.0104**	**0.5682**	**0.6250**
Mid visibility															
RGB-Rep	12.3019	0.5705	0.7075	7.8495	0.2748	0.8474	10.0707	0.4686	0.7809	10.8875	0.4383	0.7791	9.1685	0.3978	0.7910
Edge-Rep	12.5643	0.5732	0.6700	7.9794	0.3570	0.8274	10.5044	0.4767	0.7190	11.3838	0.4806	0.7624	8.5538	0.3794	0.7543
Retinex-Rep	11.7121	0.4961	0.7118	9.0155	0.3649	0.8480	10.4857	0.4588	0.7520	11.4945	0.4600	0.8019	8.6294	0.3428	0.8120
Frequency-Rep	12.0937	0.5676	0.7184	8.3321	0.3390	0.8902	9.6494	0.4525	0.7727	11.1687	0.4489	0.7807	8.9509	0.3928	0.7791
Degradation-Rep	12.2702	0.5694	0.7132	8.5297	0.3675	0.8315	10.2398	0.4825	0.7807	11.1634	0.4722	0.7703	9.4308	0.4210	0.7522
Ours-Rep-Max	12.4630	0.5241	0.6969	8.1712	0.3151	0.8714	9.8155	0.4355	0.7282	11.3169	0.4145	0.7918	8.9125	0.3329	0.7562
Final Ours	**14.6389**	**0.5929**	**0.6450**	**12.9575**	**0.4242**	**0.7498**	**15.0816**	**0.5511**	**0.6870**	**13.1564**	**0.5553**	**0.7466**	**11.4069**	**0.4549**	**0.7073**
High visibility															
RGB-Rep	12.0761	0.5685	0.6968	7.6460	0.3174	0.8605	9.7693	0.4395	0.7894	10.9378	0.4134	0.7882	8.9282	0.4046	0.7884
Edge-Rep	12.3349	0.5706	0.6566	8.0958	0.3985	0.8053	10.5324	0.4581	0.7271	11.2438	0.4430	0.7770	9.1628	0.4177	0.7365
Retinex-Rep	11.4748	0.4964	0.7061	8.5983	0.4196	0.8293	10.0748	0.4351	0.7706	11.4766	0.4307	0.8050	8.6371	0.3622	0.8003
Frequency-Rep	11.8671	0.5678	0.7057	8.6637	0.4145	0.8445	9.9915	0.4444	0.7752	10.9848	0.4176	0.7906	8.8261	0.4093	0.7675
Degradation-Rep	12.0235	0.5710	0.7079	8.4727	0.4200	0.8089	10.3847	0.4685	0.7839	11.0465	0.4355	0.7804	9.6427	0.4522	0.7286
Ours-Rep-Max	12.3915	0.5236	0.6879	8.5649	0.3767	0.8408	9.3397	0.4127	0.7379	10.9705	0.3811	0.7957	9.5612	0.3675	0.7318
Final Ours	**14.8139**	**0.5931**	**0.6302**	**12.9429**	**0.4848**	**0.7127**	**15.4149**	**0.5269**	**0.6919**	**12.7027**	**0.5069**	**0.7664**	**11.4520**	**0.4720**	**0.6856**

The Low bin tests reconstruction under weak target contours. Final Ours ranks first in every metric for all five datasets and all three visibility bins. Its advantage in the low-visibility group shows that the enhanced representation remains effective when target contours are weak. This behavior is consistent with CHRRL's joint use of chromatic, contextual, and gradient cues to represent recoverable screen information.

The stratification separates this result from performance averaged across all visibility levels. Within each category and bin, every representation is evaluated on the same samples, so the comparison reflects recovery of the same target content from the same observations. Across bins, however, the sample membership changes: Low and High are not progressively modified versions of an identical input. Differences between their scores therefore describe the evaluated subsets rather than a controlled response to increasing contour strength. The consistent within-bin ranking indicates that the overall advantage is not confined to samples with stronger contour correspondence. Because the visibility score uses the paired target only to organize evaluation, this analysis does not provide additional target information to the reconstruction model at inference.

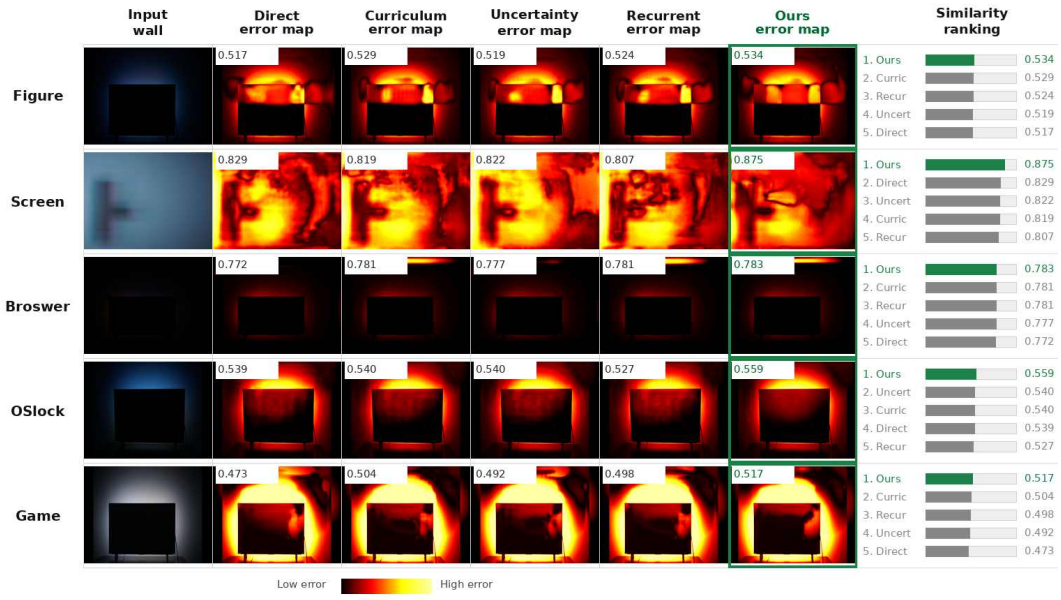
C.23 LRN-PK1: Wall-reprojection consistency

Wall-reprojection consistency directly tests whether a predicted screen explains the wall observation that generated the input. We keep the U-Net, representation, data split, and screen-to-wall forward operator fixed. Each Learning method produces a screen from the same wall image; the frozen operator reprojects that screen to the wall, and we measure wall-reprojection SSIM. OARL learning + refinement applies eight wall-consistency correction steps, whereas the alternatives use their trained single-output predictions. The correction optimizes the predicted screen with Adam (learning rate 0.02), using wall-domain L1, MSE, and SSIM losses with weights 1, 0.75, and 0.20, plus an L1 penalty of weight 0.01 relative to the initial prediction. Predictions are clamped to $[0, 1]$; the forward operator is frozen and correction uses no target screen. This comparison evaluates the configuration including refinement, rather than isolating the effect of the learning rule under equal inference budgets.

Table S3: Wall-reprojection consistency across five screen-content categories. Entries are category-mean wall-reprojection SSIM (↑); the macro average weights categories equally. Only OARL learning + refinement uses eight test-time correction steps. Blue marks the best alternative in each column.

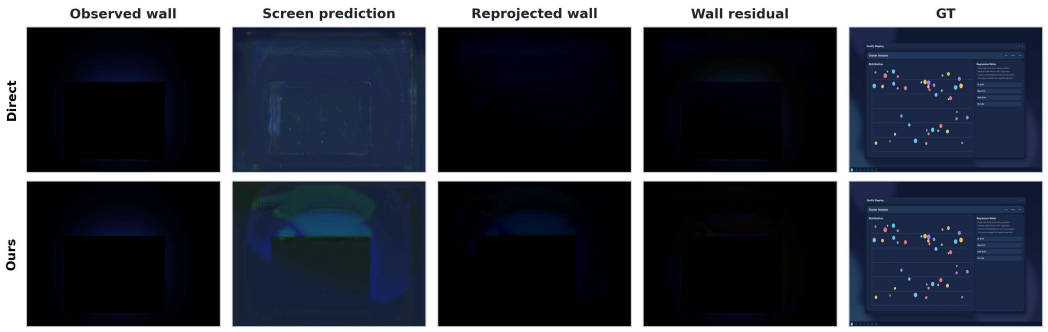
Learning configuration	Figure	Screen	Game	OSLock	Browser	Macro
Test samples	100	251	100	279	200	—
Direct learning	0.5378	0.7916	0.5117	0.6487	0.5472	0.6074
Perceptual learning	0.5355	0.7657	0.5159	0.6512	0.5508	0.6038
Multi-scale supervision	0.5324	0.7794	0.5147	0.6520	0.5530	0.6063
Curriculum learning	0.5370	0.7757	0.5149	0.6537	0.5552	0.6073
Contrastive learning	0.5359	0.7806	0.5127	0.6292	0.5465	0.6010
Distillation	0.5368	0.7773	0.5133	0.6346	0.5507	0.6025
Uncertainty weighting	0.5372	0.7734	0.5148	0.6520	0.5591	0.6073
Recurrent learning	0.5340	0.7712	0.5137	0.6501	0.5478	0.6034
Task-guided learning	0.5329	0.7699	0.5149	0.6476	0.5621	0.6055
OARL learning + refinement	**0.5891**	**0.8374**	**0.5861**	**0.7142**	**0.5894**	**0.6632**

OARL learning with eight-step refinement ranks first on every dataset in Table S3, with a macro-average SSIM of 0.6632 versus 0.6074 for Direct learning, the strongest alternative by macro average. The refined screen predictions yield wall reprojections with closer structural agreement to the observed glow under the frozen forward operator.



Supplementary Figure 74: LRN-PK1 wall-reprojection similarity maps and rankings for representative samples from the five content categories.

Figure 74 complements the category averages with spatial similarity maps and rankings for representative samples. The maps locate agreement in the wall domain, while the accompanying rankings summarize the selected examples; these sample-level values are distinct from the category means in Table S3. Agreement after reprojection provides a consistency check on the recovered screen under the specified forward operator. It does not by itself establish a unique screen solution, since spatial mixing can suppress differences between screen predictions. The wall-domain result is therefore considered alongside the screen-domain reconstruction metrics. In this comparison, the improved consistency belongs to the evaluated learning-and-refinement configuration, including its additional correction steps, rather than to a matched-cost comparison of training objectives alone.



Supplementary Figure 75: LRN-PK1 comparison of observed walls, predicted screens, reprojected walls, and wall residuals for Direct Learning and Ours.

C.24 LRN-PK2: One-epoch takeover

To compare unit-budget learning efficiency, all Learning variants take over from the same checkpoint with the same U-Net, representation, decoder, optimizer, batch order, random seed, and number of updates. Only one training epoch is allowed. We report LPIPS, SSIM, and PSNR on five complete test sets. Across the 15 dataset–metric comparisons, Ours ranks first in every unit.

Table S4: LRN-PK2 one-epoch takeover results using the same U-Net recovery architecture. All learning variants are trained for one epoch under the same initialization, data order, optimizer, and update budget.

Learning	Origin	Figure			Screen2000			Game			OSLock			Browser		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Supervision-Based Learning																
Direct Learning	This study	16.4946	0.6049	0.6601	9.5877	0.3480	0.7996	13.5464	0.4980	0.7114	12.7481	0.4088	0.7851	13.3542	0.5064	0.6484
Perceptual Learning	This study	16.0943	0.5870	0.6370	9.5178	0.3420	0.7897	13.7324	0.4851	0.6842	12.7426	0.4086	0.7759	13.7783	0.5195	0.6156
Deep Supervision Learning	This study	15.6377	0.6109	0.6484	9.5377	0.3618	0.7875	13.5781	0.5000	0.6938	12.7217	0.4070	0.7826	13.2940	0.5255	0.6301
Contrastive Learning	This study	16.3837	0.5996	0.6554	9.4277	0.3489	0.7874	13.6577	0.4922	0.6939	12.8183	0.4044	0.7812	13.3713	0.5389	0.6139
Distillation Learning	This study	15.7385	0.5964	0.6481	9.5708	0.3455	0.7911	13.3246	0.4929	0.6998	12.7863	0.4098	0.7815	14.0478	0.5252	0.6344
Adaptive Learning Strategies																
Curriculum Learning	This study	16.5069	0.6157	0.6435	9.3475	0.3489	0.7838	13.6582	0.4982	0.6936	12.8059	0.4173	0.7782	13.1282	0.5228	0.6227
Uncertainty Learning	This study	16.3028	0.6115	0.6465	9.4066	0.3479	0.7836	13.4636	0.4960	0.6970	12.7265	0.4065	0.7816	13.4961	0.5380	0.6182
Recurrent Learning	This study	16.5669	0.6126	0.6471	9.4075	0.3475	0.7947	13.1827	0.5005	0.6981	12.7506	0.4110	0.7824	13.1159	0.5247	0.6283
Task-Guided Learning	This study	16.6318	0.6288	0.6420	9.6705	0.3749	0.7820	13.7144	0.5141	0.6953	12.8387	0.4320	0.7788	13.7786	0.5392	0.6277
Ours Learning	–	**16.7903**	**0.6430**	**0.6235**	**10.5428**	**0.4622**	**0.7548**	**14.5237**	**0.5853**	**0.6816**	**13.0903**	**0.5193**	**0.7613**	**14.2124**	**0.5753**	**0.6115**

Under the identical one-epoch optimization budget, Final Ours ranks first for every dataset and metric. These consistent gains demonstrate superior learning efficiency under matched initialization, data order, optimizer, and update count.

C.25 LRN-PK4: Strong-to-weak cue relay

This experiment examines strong-to-weak evidence transfer by keeping the checkpoint, recoverability map, representation, and stage heads fixed while changing only the stage weighting, order, or inheritance rule. This stage-selection diagnostic applies category-specific, fixed nonnegative stage gains to the OARL updates; a zero gain suppresses the corresponding update. This extra selection operation is distinct from the continuous spatial gates in the main model. Flat weighting, reverse ordering, and removal of inheritance compare alternative combinations of stage weighting, order, and state reuse with the proposed validation-selected strategy. The mean of the four recoverability maps defines a scalar support map. Within the valid image region, values at or above its 70th percentile define the strong-cue mask, and values at or below its 30th percentile define the weak-cue mask; tied values are included. Stage gains are selected on validation data and frozen before testing. Relay Score is

$$2\,\Delta_{\text{strong,first}} + \Delta_{\text{strong,hold}} + 2\,\Delta_{\text{weak,late}},$$

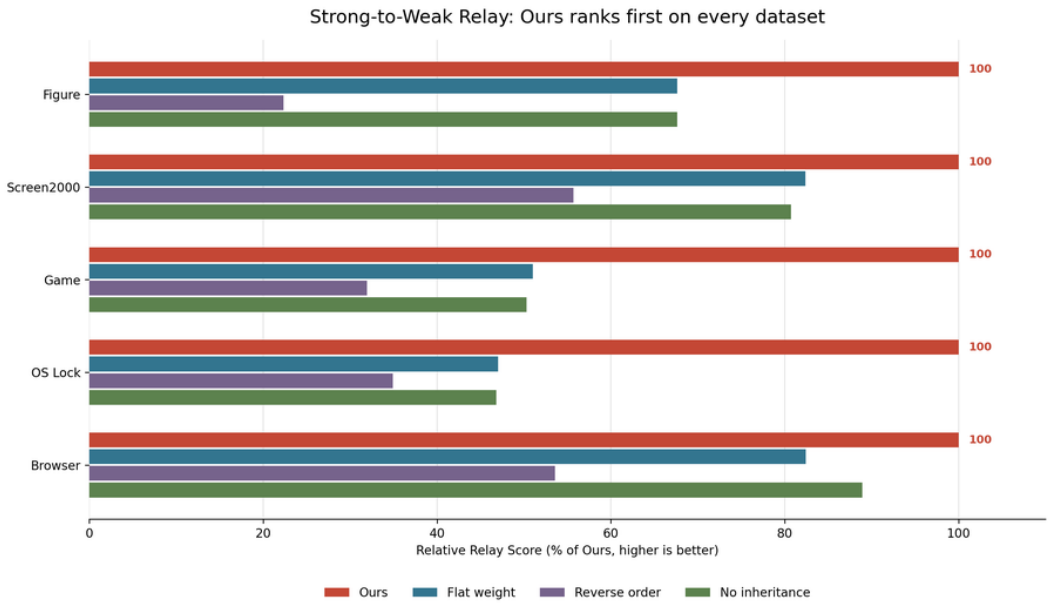
where larger values are better and each dataset is ranked independently. Let $E_R(x)$ be the mean absolute RGB error to the target over mask R, x_0 the initial estimate, x_a the state after the first

positive-gain stage in the evaluated order, and x_f the final state. The terms are $\Delta_{\text{strong,first}} = E_S(x_0) - E_S(x_a)$, $\Delta_{\text{strong,hold}} = E_S(x_a) - E_S(x_f)$, and $\Delta_{\text{weak,late}} = E_W(x_a) - E_W(x_f)$, with S and W denoting the strong and weak masks. Positive values indicate error reduction. Scores are averaged over samples. The color/layout/structure/detail gains used for OARL are $(8, 0, 0.25, 1.5)$ for Figure, $(8, 0, 3, 1)$ for Screen, $(8, 0, 0.75, 0.75)$ for Game, $(0, 8, 8, 0.75)$ for OSLock, and $(4, 0, 0, 1)$ for Browser; control gains are all one. The first active stage is therefore Layout for OSLock and Color for the other categories under the proposed rule.

Table S5: LRN-PK4 strong-to-weak cue relay. Relay Score is multiplied by 10^3 for readability; each dataset is ranked separately.

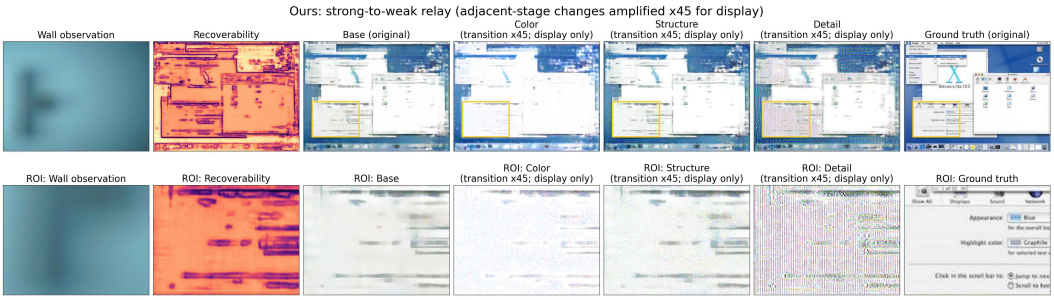
Learning strategy	**Relay Score $\times 10^3$ (higher is better)**				
	Figure ($n = 100$)	**Screen2000** ($n = 251$)	**Game** ($n = 100$)	**OSLock** ($n = 279$)	**Browser** ($n = 200$)
Ours	**0.0940**	**0.9357**	**0.3490**	**2.3844**	**7.6695**
Flat weight	0.0635	0.7710	0.1781	1.1212	6.3201
Reverse order	0.0210	0.5211	0.1115	0.8324	4.1103
No inheritance	0.0635	0.7553	0.1755	1.1162	6.8197

Ours attains the highest Relay Score on all five complete test sets, consistently outperforming flat weighting, reverse ordering, and removal of inheritance. The higher Relay Scores support the evaluated accretive configuration, which combines validation-selected stage gains with inherited reconstruction states to guide updates in weak-cue regions. These comparisons evaluate the configuration as a whole. The $\times 45$ residual amplification in the stage-error visualization displays the per-stage changes, while quantitative scores use the original outputs.

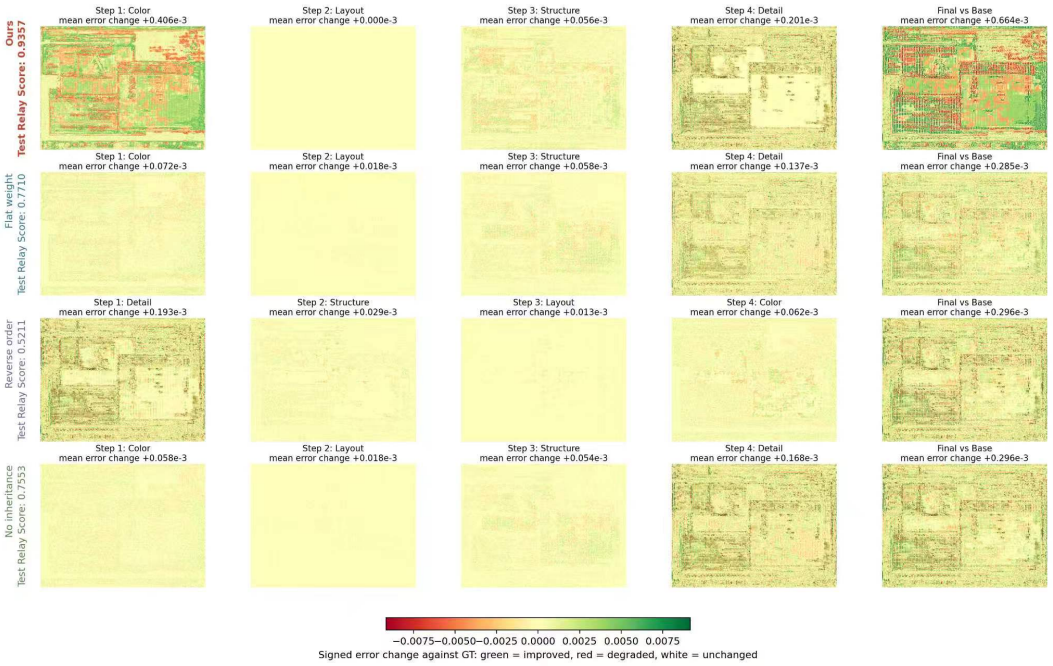


Supplementary Figure 76: LRN-PK4 relative Relay Score across the five datasets, with each dataset normalized independently to Ours = 100.

The normalization in Figure 76 makes the within-category comparison explicit: each proposed score sets its own reference, and the alternative scores are expressed relative to that value. Equal reference bars therefore do not imply equal absolute error reductions across categories; the unnormalized values remain available in Table S5. The lower scores under reversed order and removed inheritance are consistent with the role of stage sequence and state reuse in the evaluated configuration. Since that configuration also uses validation-selected gains, the comparison does not isolate either factor from weighting.



Supplementary Figure 77: LRN-PK4 strong-to-weak relay process for a Screen2000 sample. Stage panels use $\times 45$ display-only residual amplification; quantitative scores use the unmodified outputs.



Supplementary Figure 78: LRN-PK4 per-stage error changes for four Learning strategies on Screen2000. Green indicates reduced error, red increased error, and white negligible change.

D Reproducibility and Evidence Summary

The experiments evaluate OARL under a closed-room, single-frame wall-glow protocol across five content categories. Robustness sweeps cover photometric, sensor, geometric, visibility, surface, illumination, blur, and camera-shake perturbations within the stated ranges; the mechanism comparisons use their specified formal test sets.

Exact numerical reproduction uses the corresponding data partitions, checkpoints, baseline implementations, perturbation generators, and metric scripts. Wall-reprojection evaluation (LRN-PK1) uses the frozen screen-to-wall operator and eight-step refinement; the strong-to-weak diagnostic (LRN-PK4) uses validation-selected coefficients fixed for testing. The C.21–C.25 tables and visualizations document the evaluated configurations and formal comparison units.

Together, these findings provide initial experimental evidence for recoverability-guided reconstruction of the main content of a hidden display from wall-reflected light recorded by an ordinary camera under the tested indoor conditions.