# PHYSICALLY-GUIDED OPTICAL INVERSION ENABLE NON-CONTACT SIDE-CHANNEL ATTACK ON ISOLATED SCREENS

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

#### **ABSTRACT**

Noncontact exfiltration of electronic screen content poses a security challenge, with side-channel incursions as the principal vector. We introduce an optical projection side-channel paradigm that confronts two core instabilities: (i) the near-singular Jacobian spectrum of projection mapping breaches Hadamard stability, rendering inversion hypersensitive to perturbations; (ii) irreversible compression in light transport obliterates global semantic cues, magnifying reconstruction ambiguity. Exploiting passive speckle patterns formed by diffuse reflection, our Irradiance Robust Radiometric Inversion Network (IR<sup>4</sup>Net) fuses a Physically Regularized Irradiance Approximation (PRIrr-Approximation), which embeds the radiative transfer equation in a learnable optimizer, with a contour-to-detail cross-scale reconstruction mechanism that arrests noise propagation. Moreover, an Irreversibility Constrained Semantic Reprojection (ICSR) module reinstates lost global structure through context-driven semantic mapping. Evaluated across four scene categories, IR<sup>4</sup>Net achieves fidelity beyond competing neural approaches while retaining resilience to illumination perturbations.

#### 1 Introduction

Non-contact exfiltration of electronic screen information under unauthorized conditions represents a formidable challenge in information security. Long regarded as the ultimate safeguard, physical isolation may yet succumb to the merest reflection wall-scattered luminescence alone can betray sensitive content. This paper proposes a novel optical projection side-channel attack paradigm. Leveraging intrinsic optical characteristics, self-emissive targets enable imaging solely via their environmental projections. The resulting surveillance modality is passive and non-contact, with limited susceptibility to interception. An attacker can remotely capture the scattered light patterns projected onto nearby surfaces (e.g., walls) and use them to reconstruct the original screen content. As illustrated in Fig. 5, the attacker and the target remain physically separated, with no line-of-sight, no RF monitoring, and no communication link needed. This approach is highly stealthy and non-invasive, and it exposes new avenues of information leakage even in systems previously considered secure, such as laser protective glazing, electromagnetically shielded, or physically isolated environments.

Compare to traditional side-channel attacks, this optical approach leverages environmental media as a covert communication path. Microwave/electromagnetic techniques for tracking are vulnerable to attenuation and shielding; network-based channels are constrained by connectivity and congestion; hardware requirements are substantial; and active probing is readily detected, thereby revealing the operator's location. Electromagnetic-based attacks, for instance, rely on stray field emissions and are limited by distance, shielding, and ambient noise; Network-based attacks require connectivity and software vulnerabilities, making them inapplicable to air-gapped systems and often leaving traceable audit logs. In contrast, the optical projection side-channel attack proposed in this study circumvents these limitations, significantly enhancing attack feasibility and stealth, and prompting a fundamental reassessment of current defensive boundaries and strategies.

Despite its potential, this attack model presents substantial technical challenges. In everyday settings, self-luminous sources typically emit over a continuous spectrum, implying a continuously varying wavevector k. The corresponding Helmholtz solutions are therefore highly oscillatory,

which makes it impossible to construct an accurate spatial propagation model. Furthermore, non-linearity in the camera response undermines output stability and repeatability. The mapping from screen content to scattered speckles is ill-conditioned; the Jacobian matrix of the transformation has singular values that collapse in multiple directions, violating Hadamard's stability criterion. As a result, even minor irradiance fluctuations can be magnified into major structural distortions in the reconstructed image such as unpredictable edge displacement, false textures, or semantic drift. In addition, the inherently irreversible compression, along with occlusion, diffraction, and other optical effects, causes significant loss of global semantic structure and contextual cues. Without strong regularization, these losses manifest as blurry edges, disordered textures, and semantic discontinuities, leading to highly uncertain reconstructions.

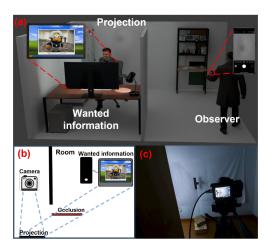


Figure 1: In the figure, (a), (b), and (c) correspond to the rendered scene, schematic diagram, and real-world scene respectively. An observer infers screen content via passive light projection. A light projection from the screen ("Wanted information") is cast onto a wall. By recording the wall's projection without viewing the screen, hacking, or capturing signals, the observer attempts to reconstruct the original content non-invasively.

To overcome these challenges, we propose IR<sup>4</sup>Net, a radiometric-inversion neural architecture that integrates physical modeling with deep learning priors, substantially improving the fidelity and stability of screen image reconstruction in optical side-channel scenarios. IR<sup>4</sup>Net comprises Physically-Regularized Irradiance Approximation (PRIrr-Approximation) and Irreversibility-Constrained Semantic Re-Projection (ICSR). PRIrr-Approximation recasts nonlinear optical-field inversion as a learnable iterative path, embedding forward/reverse propagation physics through neural modules to yield an estimate consistent with irradiance constraints; by constraining the solution's trajectory, amplification of minute perturbations is curtailed. Residual noise sensitivity and detail loss from multi-scale diffraction persist, so a frequency-selective upsampling network decouples perturbations via a multi-scale frequency separation module, enabling hierarchical reconstruction from low-frequency contours to high-frequency details while damping inconsistent components. Finally, to mitigate irreversible semantic loss, ICSR builds a stable mapping in deep semantic space that aligns global structure with visual context, re-embedding abstract features under perceptual consistency constraints to infer and complete missing information.

The contribution of this paper are summarized as follow:

- To the best of our knowledge, this work is the first to demonstrate that diffuse wall reflections can serve as a viable optical side channel for reconstructing on-screen content, and to propose the optical projection attack paradigm. This reveals a novel and previously overlooked avenue of information leakage in physically isolated environments.
- We introduce the PRIrr-Approximation module, reformulating optical field inversion as a physics-guided, learnable iterative trajectory to yield a stable initial estimate. A frequency-selective upsampling mechanism then drives progressive reconstruction from low to high frequencies, mitigating perturbation amplification and preserving structural integrity.

• We propose the ICSR module, which constructs a global-structure-aware semantic response within a deep semantic space. By embedding semantic features into a perceptual-consistency-constrained domain and applying context-driven completion rules to occluded and diffraction-corrupted regions, ICSR enhances edge continuity and semantic fidelity.

#### 2 RELATED WORK

Side-Channel Attacks (SCAs) using non-functional emissions such as cache traces, power signatures, electromagnetic emanations, acoustic signals, and timing discrepancies are expanding across domains. Cache-based SCAs include DeepCache targeting DNN traces Liu et al. (2024), asymmetric contention modeling Mahmud et al. (2023), and VM based isolation strategies Dhinakar et al. (2023). Power analysis covers M1/M2 vulnerabilities, ADC quantization leaks, and NCFET defenses Chawla et al. (2024); Chen & Savidis (2023); Sayedi & Kassiri (2022). EM and acoustic channels exploit transfer learned EM profiling and pressure or ultrasonic attacks on robotics and mobile systems Chen et al. (2024); Duan et al. (2024); Fang et al. (2022). Timing vectors leverage AVX based ASLR bypass, thermal interrupt hijacking, and DVFS abuse Choi et al. (2023); Liu et al. (2022); Zhang et al. (2024b). Deep learning amplifies extraction through DenseNet based feature recovery, architecture inference, adversarial migration, and cache based DNN interior probing Zhang et al. (2024a); Gupta et al. (2023); Meng et al. (2023); Wang et al. (2022a). Defenses include HPC anomaly detection, logic obfuscation, adversarial hardening, and FPGA specific classifiers Kapotoglu Koc & Altilar (2023); Kolhe et al. (2022); Cao et al. (2022); Zhao & Suh (2024). Emerging paradigms target bio leakage informed models, cloud profiling, post quantum (Kyber) masking SCAs, and multi tenant FPGA gradient extraction Zhu et al. (2024); Ji & Dubrova (2023); Johnson & Ward (2022); Albalawi et al. (2022). Yet, most approaches depend on network access or proximity, limiting covert recovery of isolated screen emissions.

Coherent Image generation from structured priors integrates realism with domain constraints. Super-resolution He et al. (2022); Hong & Lee (2024); Chen et al. (2025), denoising Ye et al. (2025); Yang et al. (2025), and dehazing Ma et al. (2025); Fu et al. (2025); Wang et al. (2025) models reflect evolving efficacy Ryou et al. (2024). Transformer encoders like Styleformer modulate diversity via attention-weighted embeddings Park & Kim (2022), and latent diffusion with implicit decoders enables scale-agnostic synthesis through multiresolution cascades Kim & Kim (2024). Patch-tokenization fused with global context boosts dehazing Jiuchen Chen & Li (2025), and inter-channel consistency drives unsupervised deraining Dong et al. (2025). However, reliance on optical acquisitions limits these methods, which struggle with multiscale diffraction and wavefront interference, failing to reconstruct concealed emissive patterns in diffuse projections.

#### 3 Method

Radiometric inversion under optical projection constitutes a severely ill-conditioned problem wherein nonlinear image-formation dynamics, perturbation amplification, and irreversible semantic degradation impede stable recovery. To address these challenges, we introduce the IR<sup>4</sup>Net (Fig 2) to integrate physical priors with learned optimization. First, PRIrr-Approximation formulates inversion as a physics-guided iterative trajectory embedding optical propagation operators with momentum-based updates to maintain consistency and mitigate cumulative error. A dual-path perturbation dissipation module concurrently performs spatial diffusion and semantic attenuation, while a frequency-selective multi-scale upsampling scheme constrains cross-scale energy propagation to reduce high-frequency amplification. Subsequently,ICSR establishes semantic completion and structural alignment within a perceptual space, enabling coherent reconstruction characterized by structurally preserved contours and contextually consistent textures.

#### 3.1 PHYSICALLY-REGULARIZED IRRADIANCE APPROXIMATION

Optical-projection side-channel attacks confront a fundamental challenge in the intricate physics of image formation: the observed image arises from a highly nonlinear mapping of the original irradiance through successive diffraction, scattering and reflection. This process imposes extreme information compression and yields an operator whose singular values tend toward zero, so that infinitesimal irradiance perturbations at the input become dramatically amplified in inversion, inducing severe

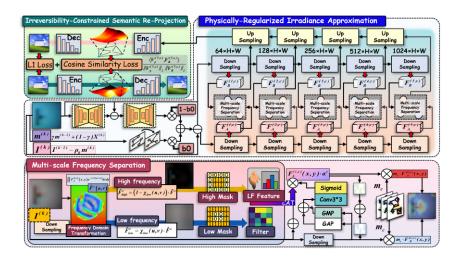


Figure 2: Overall architecture of IR<sup>4</sup>Net, comprising the PRIrr-Approximation and ICSR modules. The multi-scale frequency separation module, a key component of ICSR, is implemented via concatenation.

distortion and instability. To mitigate this, we introduce a physics-constrained module: guided inversion trajectory embeds optical-propagation modeling to guarantee physical consistency; in parallel, a frequency-selective upsampling network decouples perturbations and reconstructs multi-scale spectral components, structurally suppressing their amplification.

Our scheme models optical effects via a physics-consistent transfer operator  $\Phi(\cdot)$ , and derives its inverse approximation  $\Psi(\cdot)$  to harvest feedback. A momentum initialization melds local priors with multi-scale global feedback, steering each iteration along tenable, coherent directions. Momentum-guided gradient updates suppress noise and error accumulation, yielding feature estimates  $\hat{I}^{(k)}$  that converge toward an accurate inversion of the source radiance; derivations reside in the Appendix.

In the dual-path feature-dissipation stage, we deploy a frequency-selective upsampling network in parallel with spatial diffusion and semantic attenuation pathways to capture the rapid amplification of minute screen-to-wall perturbations. This decoupled architecture structurally restrains perturbation growth and disperses its energy, to maintain robustness against projection-induced distortions.

The input  $I^{(k)}$  flows through two paths: the spatial diffusion path applies a second-order differential kernel to the local gradient:

$$F_A^{(i,c)}(x,y) = \phi \left( \iint_{B_r} \kappa_A^{(i,c)}(\xi,\eta) \, \frac{\partial^2 I^{(k)}}{\partial x \partial y}(x-\xi,y-\eta) \, \mathrm{d}\xi \, \mathrm{d}\eta + b_A^{(i,c)} \right). \tag{1}$$

Here,  $I^{(k)}$  is the feature map at iteration k; i and c denote decoder and channel indices;  $\frac{\partial^2 I^{(k)}}{\partial x \partial y}$  is the mixed second-order derivative, capturing local curvature;  $B_r$  the neighborhood centered at (x,y) with radius r;  $\kappa_A^{(i,c)}(\xi,\eta)$  the second-order differential kernel for decoder i, channel c;  $b_A^{(i,c)}$  the bias term;  $\phi(\cdot)$  the activation; and  $F_A^{(i,c)}(x,y)$  the spatial diffusion output.

The semantic attenuation path, through an attention mechanism, mitigates disturbance components in the semantic dimension, where for the *i*-th attention head, the linear projection is given by

$$(Q^{(i)}, K^{(i)}, V^{(i)})(x) = I^{(k)}(x) (W_O^{(i)}, W_K^{(i)}, W_V^{(i)})$$
(2)

$$A^{(i)}(x,x') = \frac{\exp\langle Q^{(i)}(x), K^{(i)}(x')\rangle}{\int_{\Omega_s} \exp\langle Q^{(i)}(x), K^{(i)}(x')\rangle dx'}$$
(3)

$$F_B^{(i,c)}(x,y) = \phi \left( \int_{\Omega_s} A^{(i)}(x,x') V^{(i,c)}(x') dx' + b_B^{(i,c)} \right)$$
 (4)

In this context,  $W_Q^{(i)}, W_K^{(i)}, W_V^{(i)} \in \mathbb{R}^{C \times d}$  represent the projection matrices for query, key, and value, with C being the original number of feature channels and d the projected dimension. The attention mechanism disperses disturbance components in the semantic space to ensure that the disturbance does not concentrate spatially. Consequently,  $F_B^{(i,c)}\left(x,y\right)$  represents the output feature of this semantic attenuation path.

Subsequently, the multi-scale frequency separation module concatenates the outputs of both paths in the spatial domain and performs gating in the frequency domain. This step guarantees that only lowfrequency components with cross-scale consistency and structural robustness are amplified layer by layer, while high-frequency components that lack scale consistency attenuate during propagation.

$$\widehat{F}^{c}\left(u,v\right) = \iint F_{A}^{(i,c)}\left(x,y\right) e^{-j2\pi(ux+vy)} dxdy,\tag{5}$$

where,  $\hat{F}^{c}(u,v)$  represents the frequency-domain transformation of the concatenated features, and u and v are the spatial frequency coordinates along the x- and y-axes.

$$(F_{\text{low}}^c, F_{\text{high}}^c) = (\mathcal{F}^{-1}[\chi_{\text{low}} \widehat{F}^c], \, \mathcal{F}^{-1}[(1 - \chi_{\text{low}}) \widehat{F}^c]).$$
 (6)

The channel response  $\alpha^c$  passes through adaptive gating to fuse low/high-frequency features, enabling cross-scale propagation; final fusion is

$$\alpha^{c} = \sigma \Big( W_{2} \, \phi \big( W_{1} \, \int_{\Omega_{c}} \big| \nabla F_{\text{low}}^{c} + \nabla F_{\text{high}}^{c} \big| \, dx \, dy \Big) \Big) \tag{7}$$

$$F_A^{(i,c)}(x,y) = F_A^{(i,c)}(x,y) (1 + \alpha^c)$$
(8)

This attention-based fusion ensures a balanced contribution from both spatial diffusion and semantic attenuation, with each component adapting based on the gradient magnitude of low- and highfrequency features. Thus,  $F_A^{\prime(i,c)}(x,y)$  represents the feature after weighted fusion.

The channel attention weight  $m_c$  is generated through global average pooling and differential operations, with the calculation formula given as:

$$m_c = \sigma \left( \frac{d}{d\hat{g}_c} \left( \int_{x=0}^H \int_{y=0}^W (\hat{g}_c \cdot g_c(x, y)) \, dx \, dy \right) \right). \tag{9}$$

Here,  $g_c(x,y)$  is the value at position (x,y) of the c-th channel in the input feature map, and  $\hat{g}_c$  is its global average.  $m_c$  denotes the attention weight for channel c, and  $\sigma$  is the Sigmoid function ensuring  $m_c \in [0,1]$ . Using  $m_c$ , the feature maps  $F_A^{\prime(i,c)}(x,y)$  and  $F_B^{(i,c)}(x,y)$  are fused at each (x,y) to produce the output map  $\widetilde{F}^c(x,y)$ :

$$\widetilde{F}^{c}(x,y) = m_{c} \cdot \left( F_{A}^{\prime(i,c)}(x,y) + F_{B}^{(i,c)}(x,y) \right).$$
 (10)

In this equation,  $F_A^{\prime(i,c)}(x,y)$  and  $F_B^{(i,c)}(x,y)$  represent the values of the c-th channel of the input feature maps  $F_A^\prime$  and  $F_B$  at position (x,y). Through this weighted fusion process, the final output feature map  $F^c(x, y)$  incorporates the fused channel information.

Following this, multi-head attention mechanisms are employed to capture and suppress any remaining perturbation structures within the fused features  $F^c$ :

$$A_h(x, x') = \exp \langle \partial_x Q_h(x), K_h(x') \rangle + \langle Q_h(x), \partial_{x'} K_h(x') \rangle , \qquad (11)$$

$$O_h(x) = \int_{\Omega_s} A_h(x, x') V_h(x') dx',$$
 (12)

$$Z^{(i)} = \operatorname{Concat}_{h}(O_{h}(x)) + \widetilde{F}^{(c)}(x, y). \tag{13}$$

In this context, the space-derivative mappings of each attention head allow for precise quantification of perturbation effects on attention distribution. Residual connections preserve stable structural information throughout the process.

During the multi-scale frequency-selective upsampling and output synthesis stage, perturbation growth along successive upsampling layers is mitigated through hierarchical decomposition and reconstruction of enhanced features  $Z^{(i)}$  with the preceding layer  $Z^{(i-1)}$ . Specifically, the intermediate interpolation  $U^{(i)}(x)$  is computed as:

$$U^{(i)}(x) = \iint Z^{(i)}(x') \prod_{j=1}^{2} \max(0, 1 - |x_j - x'_j|) dx', \tag{14}$$

and the upsampled representation  $F_{\rm up}^{(i)}(x)$  is expressed as:

$$F_{\rm up}^{(i)}(x) = \phi \left( \iint \kappa_{\rm up}^{(i)}(x, x') \left[ U^{(i)}(x') + Z^{(i-1)}(x') \right] dx' \right), \tag{15}$$

where the bilinear interpolation kernel  $\prod_{j=1}^2 \max(0, 1 - |x_j - x_j'|)$  operates in concert with the learned upsampling kernel  $\kappa_{\rm up}^{(i)}$ , enabling progressive reconstruction. This hierarchical scheme introduces information from low to high frequencies in a controlled manner to permit expansion only of cross-scale-consistent structural features when attenuating perturbations lack multi-scale support.

The final stage maps the first-level upsampled feature into the pixel domain via an output convolution with kernel  $\kappa_{\rm out}$  and bias  $b_{\rm out}$ :

$$\widehat{\mathbf{J}}^{(k)}(x,y) = \iint \kappa_{\text{out}}(x,x') F_{\text{up}}^{(1)}(x') \, dx' + b_{\text{out}}.$$
 (16)

Through the integration of physical constraints with frequency-selective hierarchical fusion, this mechanism is designed to suppress propagation of fine-scale irradiance perturbations and maintain structural consistency during reconstruction of the projected image.

#### 3.2 IRREVERSIBILITY-CONSTRAINED SEMANTIC RE-PROJECTION

The inversion of optical projection requires irreversible, high-compression mapping original imagery. However, it suffers severe loss of global semantic structure and visual context, manifesting as edge blur, texture artifacts and semantic misalignment due to occlusion, diffraction and reflection. To address this challenge, we introduce the ICSR, comprising two parallel modules: a primary mapping network, driven by a prior-guided map, devoted to restoration of low-level structural detail; and a collaborative completion network, which extracts stable abstract semantic embeddings from the projected observation to capture global semantics and contextual cues. Building upon these, a stable mapping from semantic to structural space is established to enable high-dimensional semantic features to be dynamically fed back into the primary network's representation domain, thereby enforcing constrained completion and inference over missing regions. Here, the primary network's structural-space mapping features  $\operatorname{are} V_P^{(5,c)}(x,y) \in \mathbb{R}^d$  and the abstract semantic-space features  $\operatorname{are} V_R^{(5,c)}(x,y) \in \mathbb{R}^d$  where c denotes input channels, 5 denotes the encoder stage, (x,y) spatial coordinates and d the feature dimension; derivation appears in the Appendix.

$$\mathbf{v}_{P}^{(5,c)}(x,y) = \left(v_{P,1}^{(5,c)}(x,y), v_{P,2}^{(5,c)}(x,y), \dots, v_{P,d}^{(5,c)}(x,y)\right),\tag{17}$$

$$\mathbf{v}_{R}^{(5,c)}(x,y) = \left(v_{R,1}^{(5,c)}(x,y), v_{R,2}^{(5,c)}(x,y), \dots, v_{R,d}^{(5,c)}(x,y)\right). \tag{18}$$

In order to preserve the consistency between the semantic and structural feature spaces, to prevent semantic drift, and to enhance the stability of the completion inference process, we compute the cosine similarity between the projected features as follows:

$$\operatorname{CosSim}(x,y) = \frac{\sum_{i=1}^{d} v_{P,i}^{(5,c)}(x,y) v_{R,i}^{(5,c)}(x,y)}{\sqrt{\sum_{i=1}^{d} \left(v_{P,i}^{(5,c)}(x,y)\right)^{2} + \epsilon} \sqrt{\sum_{i=1}^{d} \left(v_{R,i}^{(5,c)}(x,y)\right)^{2} + \epsilon}}$$
(19)

where  $\epsilon > 0$  is introduced to prevent division by zero.

Subsequently, for each batch  $\mathcal{B} = \{(x_j, y_j)\}_{j=1}^N$ , the batch loss function is defined as:

$$s_{j} = \frac{\sum_{i=1}^{d} v_{P,i}^{(5,c)}(x_{j}, y_{j}) v_{R,i}^{(5,c)}(x_{j}, y_{j})}{\sqrt{\sum_{i=1}^{d} (v_{P,i}^{(5,c)}(x_{j}, y_{j}))^{2} + \epsilon} \sqrt{\sum_{i=1}^{d} (v_{R,i}^{(5,c)}(x_{j}, y_{j}))^{2} + \epsilon}} .$$
 (20)

$$\mathcal{L}_{\text{batch}} = \frac{1}{N} \sum_{j=1}^{N} (1 - s_j)^{\alpha} + \lambda \|\Theta\|_2^2.$$
 (21)

where  $\lambda \parallel \Theta \parallel_2^2$  represents the L2 regularization term.

This loss leverages multi-scale semantic alignment to improve missing-region completion, yielding sharp, realistic, and coherent reconstructions.

#### **EXPERIMENTS**

Four datasets: ReSh-WebSight, ReSh-Password, ReSh-Chart, and ReSh-Screen were employed to emulate user-interface layouts, password-entry interfaces, chart renderings, and desktop scenarios, randomized into training, validation, and test subsets in an 8:1:1 ratio. All experiments were implemented in PyTorch on an NVIDIA RTX 3090 GPU cluster, using Adam optimizer with a fixed learning rate of  $1 \times 10^{-4}$  and a batch size of 16; other hyperparameters were set to their default values unless stated otherwise.

Methods	Source	ReS	h-WebS	ight	ReS	Sh-Passw	ord	Re	eSh-Scre	en	R	eSh-Cha	ır
Methods	Source	PSNR↑	RMSE↓	SSIM↑	PSNR↑	$RMSE\!\!\downarrow$	SSIM↑	PSNR↑	$RMSE \!\!\downarrow$	SSIM↑	PSNR↑	$RMSE \!\!\downarrow$	SSIM↑
HVI-CIDNet	CVPR,25	18.940	33.837	0.792	13.024	57.269	0.858	21.027	26.686	0.708	15.720	44.843	0.692
DarkIR	CVPR,25	19.234	32.587	0.779	13.580	53.883	0.855	21.609	25.215	0.706	16.861	39.011	0.709
AST	CVPR,24	19.502	31.026	0.787	14.022	51.199	0.832	21.574	24.823	0.673	16.909	38.515	0.709
ConvIR	CVPR,24	19.573	30.678	0.799	14.779	47.077	0.867	22.010	23.718	0.731	16.678	39.569	0.707
C2PNet	CVPR,23	15.641	52.514	0.769	11.209	70.458	0.813	16.428	44.883	0.552	15.278	46.885	0.666
Uformer	CVPR,22	19.698	30.262	0.798	14.142	50.578	0.874	22.299	22.885	0.725	16.909	38.515	0.709
UNet	MICCAI,15	17.735	38.744	0.764	11.891	65.055	0.827	20.195	28.114	0.664	16.133	42.120	0.682
BicycleGAN	NIPS,17	18.680	35.453	0.775	9.784	82.939	0.781	18.305	46.888	0.600	15.289	48.376	0.632
DivCo	CVPR,21	13.353	66.098	0.721	9.266	87.803	0.730	12.280	72.146	0.369	11.091	76.426	0.523
pix2pix	CVPR,17	13.582	62.361	0.651	8.146	99.907	0.684	8.043	103.377	0.232	12.475	63.885	0.452
CycleGAN	ICCV,17	13.068	66.529	0.680	6.206	124.912	0.601	10.358	89.134	0.316	12.348	67.200	0.494
IR <sup>4</sup> Net	Ours	20.708	26.719	0.820	15.030	45.911	0.887	25.812	16.531	0.817	17.363	36.748	0.731

Table 1: Quantitative comparison of IR<sup>4</sup>Net against reconstruction-centric methods (HVI-CIDNetYan et al. (2025), DarkIRFeijoo et al. (2025), ASTZhou et al. (2024), ConVIRCui et al. (2024), C2PNetZheng et al. (2023), UformerWang et al. (2022b), UNetRonneberger et al. (2015)) and generation-centric methods (BicycleGANZhu et al. (2017b),DivcoLiu et al. (2021),pix2pixIsola et al. (2017), CycleGANZhu et al. (2017a)) on four benchmarks (ReSh-WebSight, ReSh-Password, ReSh-Screen, ReSh-Chart) under identical data splits and optimization.

#### 4.1 Comparative Evaluation

Deployed across four canonical benchmarks ReSh-WebSight, ReSh-Password, ReSh-Screen, and ReSh-Chart for assessment under disparate projection scenarios, IR4Net is juxtaposed with reconstruction-centric (Uformer, ConvIR, UNet) and generation-centric (pix2pix, CycleGAN, BicycleGAN) counterparts, each trained and tested under identical data partitions and optimisation regimes. Table 1 reports results on PSNR, RMSE, and SSIM: Specifically, PSNR on ReSh-Screen increases by 15.7% relative to Uformer, and RMSE on ReSh-WebSight falls by 27.9% compared to AST. Qualitative illustrations (Fig 3) reveal more consistent restoration of edges, textures, and occluded regions. Such behaviour likely reflects the interplay of two structural modules: PRIrr-Approximation, embedding physics-consistent, momentum-guided inversion trajectories with

frequency-selective perturbation dissipation, and ICSR, enacting a stable semantic-space mapping to align and replenish irreversibly lost information. In both perceptual and physical domains, these mechanisms operate in concert to sustain reconstruction fidelity and robustness.

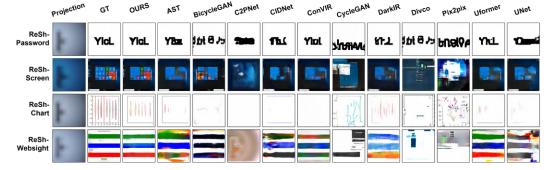


Figure 3: Visual comparison of IR<sup>4</sup>Net and baseline methods on four datasets. Our model yields visually more faithful restorations across various scenes.

#### 

#### ABLATION EXPERIMENT 4.2

 Inversion behaviour was evaluated across three datasets using four iterative schemes: classical momentum formulations including ADMM, NAG, and Heavy-Ball, and the proposed update strategy. Table 2 reports the performance under PSNR, SSIM, RMSE, and LPIPS; the mean relative improvement ranges from 8% to 15%. This variation may derive from a dual coupling design: structureaware momentum initialization, achieved through a learnable convolutional operator over local receptive fields, yielding priors aligned with intrinsic structural patterns; and a physics-feedback pathway, where inverse approximations are constructed from encoded residuals to capture projectioninduced perturbations to constrain the update direction in physically admissible regimes. Residualgated dynamic weighting integrates these cues to mitigate error amplification introduced by nearsingular transmission operators while accumulated momentum smooths the update trajectory. Stability observed under diverse conditions suggests adaptability in high-compression, nonlinear inversion scenarios. Additional ablation studies are provided in the Appendix.

Metric		C	Chart				Screen				WebSight			
11101110	OURS	ADMM	NAG	Heavyball	OURS	ADMM	NAG	Heavyball	OURS	ADMM	NAG	Heavyball		
PSNR↑	17.363	17.180	17.214	17.192	25.812	25.155	25.090	25.077	20.708	20.707	20.621	20.533		
RMSE↓	36.748	37.367	37.308	37.447	16.531	17.680	17.672	17.754	26.719	27.024	27.349	27.629		
SSIM↑	0.731	0.725	0.724	0.724	0.817	0.806	0.808	0.808	0.820	0.808	0.808	0.807		
LPIPS↓	0.431	0.468	0.465	0.462	0.216	0.232	0.235	0.231	0.282	0.299	0.299	0.300		

Table 2: Performance comparison of four iterative schemes across three datasets.

#### 4.3 LUMINANCE ROBUSTNESS EVALUATION

To assess stability under low illumination, an experimental setup was devised where display luminance was progressively attenuated to emulate irradiance decay in real projection scenarios. Experiments were conducted on the four previously mentioned datasets, with screen brightness reduced by 0–300 nits. PSNR values were recorded for each method at incremental luminance levels.

As summarized in Table 3, pronounced performance degradation emerged for several baselines under reduced brightness. For instance, UNet exhibited a PSNR decline of approximately 68% on ReSh-Screen, whereas the proposed architecture registered a reduction of 25.9% under identical conditions. Visual evidence Fig 4 indicates that when luminance decreased, competing methods produced outputs with structural misalignment and blurred contours, while the proposed approach maintained coherent edge geometry and stable texture patterns. Results for other datasets, together with qualitative exemplars, appear in the Appendix.

Model	0	25	50	75	100	125	150	175	200	225	250	275	300
OURS	25.812	25.726	25.702	25.634	25.533	25.288	24.990	24.537	24.016	23.220	22.306	20.983	19.136
UNet	20.195	6.302	6.461	6.121	6.546	6.104	6.301	7.015	6.089	6.686	6.257	6.144	6.412
C2PNet	16.428	15.913	14.951	13.452	12.311	11.520	11.195	10.948	10.542	10.039	9.666	9.370	9.144
DarkIR	21.609	21.360	20.660	19.288	17.423	15.355	13.672	12.481	11.491	10.614	10.077	9.686	9.424
CIDNet	21.027	20.808	20.337	19.366	18.118	16.576	15.131	13.853	12.791	11.739	10.913	10.192	9.745
ConvIR	22.010	21.824	21.480	20.601	19.223	17.698	16.215	14.885	13.662	12.537	11.707	10.991	10.435

Table 3: PSNR comparison of different models under screen brightness reductions (in nits).

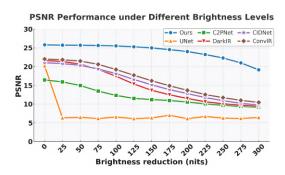


Figure 4: As screen brightness decreases on the ReSh-Screen dataset, our model's PSNR degrades significantly less than that of other methods.

These observations indicate that robustness to luminance attenuation arises from three architectural constraints: (i) a physics-regularized propagation path limiting perturbation diffusion; (ii) a frequency-selective hierarchical upsampling scheme ensuring cross-scale consistency; and (iii) a semantic-stability module restoring global context via feature-space completion. Without these constraints, conventional models suffer error amplification, causing structural collapse under low-intensity conditions. In contrast, the proposed design suppresses perturbations through physics-guided modeling, applies frequency-domain gating to limit non-structural energy propagation, and employs context-consistent semantic inference to recover projection-induced information loss. Together, these mechanisms preserve texture fidelity and ensure controlled, monotonic degradation across the luminance continuum.

#### 5 CONCLUSION AND DISCUSSION

Non-contact exfiltration of screen content in physically isolated or shielded environments is achieved via an optical-projection side channel, realized by IR<sup>4</sup>Net, a physics-constrained reconstruction framework embedding irradiance-consistent modeling and spectral regulation. Addressing two core challenges, namely nonlinear mapping ill-conditioning and semantic attrition, this architecture invalidates the notion that an air gap guarantees security. In the inversion-path stage, PRIrr-Approximation reformulates optical-field inversion as a learnable iterative trajectory that integrates forward and reverse propagation physics, mitigating perturbation amplification. In the spectral domain, a multi-scale frequency separation module decouples and hierarchically restores spectral components to reinforce cross-scale structural coherence and suppress noise. Furthermore, ICSR's abstract semantic-space mapping drives global semantic completion to bridge gaps induced by strong projection compressions. Experimental results demonstrate stable content restoration under attenuated irradiance, with SSIM and related metrics exceeding those of existing end-to-end models, to confirm the effectiveness and robustness of the physics-prior and deep-model fusion.

#### REFERENCES

Abdullah Albalawi, Vassilios Vassilakis, and Radu Calinescu. Side-channel attacks and countermeasures in cloud services and infrastructures. In *NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium*, pp. 1–4. IEEE, 2022.

- Pei Cao, Hongyi Zhang, Dawu Gu, Yan Lu, and Yidong Yuan. Al-pa: cross-device profiled side-channel attack using adversarial learning. In *Proceedings of the 59th ACM/IEEE Design Automation Conference*, pp. 691–696, 2022.
  - Nikhil Chawla, Chen Liu, Abhishek Chakraborty, Igor Chervatyuk, Thais Moreira Hamasaki, Ke Sun, and Henrique Kawakami. Uncovering software-based power side-channel attacks on apple m1/m2 systems. In *Proceedings of the 61st ACM/IEEE Design Automation Conference*, pp. 1–6, 2024.
  - Bin Chen, Gehui Li, Rongyuan Wu, Xindong Zhang, Jie Chen, Jian Zhang, and Lei Zhang. Adversarial diffusion compression for real-world image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025.
  - Peter Chen, Guannan Liu, and Haining Wang. Poster: Acoustic side-channel attack on robot vacuums. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*, pp. 5027–5029, 2024.
  - Ziyi Chen and Ioannis Savidis. A power side-channel attack on flash adc. In 2023 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1–5. IEEE, 2023.
  - Hyunwoo Choi, Suryeon Kim, and Seungwon Shin. Avx timing side-channel attacks against address space layout randomization. In 2023 60th ACM/IEEE Design Automation Conference (DAC), pp. 1–6. IEEE, 2023.
  - Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Revitalizing convolutional network for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.
  - Narayanasetti Mehar Dhinakar, Kasineni Kishan Rao, Neradi Jayanath, Repalle Devi Vara Prasad, Vijaya Chandra Jadala, and Radhika Rani Chintala. Defending against cache-based side-channel attack using virtual machine migration in cloud. In 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), pp. 239–242. IEEE, 2023.
  - Guanglu Dong, Tianheng Zheng, Yuanzhouhan Cao, Linbo Qing, and Chao Ren. Channel consistency prior and self-reconstruction strategy based unsupervised image deraining. *arXiv* preprint *arXiv*:2503.18703, 2025.
  - Yingli Duan, Weizhi Meng, Wei-Yang Chiu, and Yu Wang. Towards a novel ultrasonic side-channel attack on mobile devices. In *Proceedings of the ACM SIGCOMM 2024 Conference: Posters and Demos*, pp. 101–103, 2024.
  - Mingzhu Fang, Baolei Mao, and Wei Hu. A transfer learning approach for electromagnetic sidechannel attack and evaluation. In 2022 7th International Conference on Integrated Circuits and Microsystems (ICICM), pp. 636–640. IEEE, 2022.
  - Daniel Feijoo, Juan C. Benito, Alvaro Garcia, and Marcos V. Conde. Darkir: Robust low-light image restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pp. 10879–10889, June 2025.
  - Jiayi Fu, Siyu Liu, Zikun Liu, Chun-Le Guo, Hyunhee Park, Ruiqi Wu, Guoqing Wang, and Chongyi Li. Iterative predictor-critic code decoding for real-world image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.
  - Naina Gupta, Arpan Jati, and Anupam Chattopadhyay. Ai attacks ai: Recovering neural network architecture from nvdla using ai-assisted side channel attack. *ACM Transactions on Embedded Computing Systems*, 2023.
  - Jingwen He, Wu Shi, Kai Chen, Lean Fu, and Chao Dong. Gcfsr: a generative and controllable face super resolution method without facial and gan priors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1889–1898, 2022.
  - Cheeun Hong and Kyoung Mu Lee. Adabm: on-the-fly adaptive bit mapping for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2641–2650, 2024.

- Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Computer Vision and Pattern Recognition (CVPR)*, 2017 IEEE Conference on, 2017.
  - Yanning Ji and Elena Dubrova. A side-channel attack on a masked hardware implementation of crystals-kyber. In *Proceedings of the 2023 Workshop on Attacks and Solutions in Hardware Security*, pp. 27–37, 2023.
  - Qizhi Xu Jiuchen Chen, Xinyu Yan and Kaiqi Li. Tokenize image patches: Global context fusion for effective haze removal in large images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2025.
  - Andrew Johnson and Richard Ward. 'unified side-channel attack-model' (usca-m): An extension with biometrics side-channel type. In 2022 10th International Symposium on Digital Forensics and Security (ISDFS), pp. 1–5. IEEE, 2022.
  - Melis Kapotoglu Koc and Deniz Turgay Altilar. Selection of best fit hardware performance counters to detect cache side-channel attacks. In *Proceedings of the 2023 ACM Workshop on Secure and Trustworthy Cyber-Physical Systems*, pp. 17–22, 2023.
  - Jinseok Kim and Tae-Kyun Kim. Arbitrary-scale image generation and upsampling using latent diffusion model and implicit neural decoder. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9202–9211, 2024.
  - Gaurav Kolhe, Tyler Sheaves, Kevin Immanuel Gubbi, Soheil Salehi, Setareh Rafatirad, Sai Manoj PD, Avesta Sasan, and Houman Homayoun. Lock&roll: Deep-learning power side-channel attack mitigation using emerging reconfigurable devices and logic locking. In *Proceedings of the 59th ACM/IEEE Design Automation Conference*, pp. 85–90, 2022.
  - Chen Liu, Abhishek Chakraborty, Nikhil Chawla, and Neer Roggel. Frequency throttling sidechannel attack. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1977–1991, 2022.
  - Rui Liu, Yixiao Ge, Ching Lam Choi, Xiaogang Wang, and Hongsheng Li. Divco: Diverse conditional image synthesis via contrastive generative adversarial network. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2021.
  - Zhibo Liu, Yuanyuan Yuan, Yanzuo Chen, Sihang Hu, Tianxiang Li, and Shuai Wang. Deepcache: Revisiting cache side-channel attacks in deep neural networks executables. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*, pp. 4495–4508, 2024.
  - Long Ma, Yuxin Feng, Yan Zhang, Jinyuan Liu, Weimin Wang, Guang-Yong Chen, Chengpei Xu, and Zhuo Su. Coa: Towards real image dehazing via compression-and-adaptation. arXiv preprint arXiv:2504.05590, 2025.
  - Farabi Mahmud, Sungkeun Kim, Harpreet Singh Chawla, Eun Jung Kim, Chia-Che Tsai, and Abdullah Muzahid. Attack of the knights: Non uniform cache side channel attack. In *Proceedings of the 39th Annual Computer Security Applications Conference*, pp. 691–703, 2023.
  - Fanfei Meng, Zhanbo Li, Baolei Mao, Wei Hu, Maoyuan Qin, and Qiang Zhou. Adversarial profiled side-channel attack with unsupervised domain adaptation. In 2023 9th International Conference on Computer and Communications (ICCC), pp. 974–979. IEEE, 2023.
  - Jeeseung Park and Younggeun Kim. Styleformer: Transformer based generative adversarial networks with style vector. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8983–8992, 2022.
  - Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pp. 234–241. Springer, 2015.

- Donghun Ryou, Inju Ha, Hyewon Yoo, Dongwan Kim, and Bohyung Han. Robust image denoising through adversarial frequency mixup. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2723–2732, 2024.
- Mina Sayedi and Hossein Kassiri. Activity-adaptive architectures for energy-efficient scalable neural recording microsystems: A review of current and future directions. In 2022 20th IEEE Interregional NEWCAS Conference (NEWCAS), pp. 393–396. IEEE, 2022.
- Han Wang, Syed Mahbub Hafiz, Kartik Patwari, Chen-Nee Chuah, Zubair Shafiq, and Houman Homayoun. Stealthy inference attack on dnn via cache-based side-channel attacks. In 2022 Design, Automation & Test in Europe Conference & Exhibition (DATE), pp. 1515–1520. IEEE, 2022a.
- Ruiyi Wang, Yushuo Zheng, Zicheng Zhang, Chunyi Li, Shuaicheng Liu, Guangtao Zhai, and Xiaohong Liu. Learning hazing to dehazing: Towards realistic haze generation for real-world image dehazing. *arXiv preprint arXiv:2503.19262*, 2025.
- 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 (CVPR)*, pp. 17683–17693, June 2022b.
- Qingsen Yan, Yixu Feng, Cheng Zhang, Guansong Pang, Kangbiao Shi, Peng Wu, Wei Dong, Jinqiu Sun, and Yanning Zhang. Hvi: A new color space for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5678–5687, June 2025.
- Sidi Yang, Binxiao Huang, Yulun Zhang, Dahai Yu, Yujiu Yang, and Ngai Wong. Dnlut: Ultra-efficient color image denoising via channel-aware lookup tables. *arXiv* preprint arXiv:2503.15931, 2025.
- Xin Ye, Burhaneddin Yaman, Sheng Cheng, Feng Tao, Abhirup Mallik, and Liu Ren. Bevdiffuser: Plug-and-play diffusion model for bev denoising with ground-truth guidance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2025.
- Runlian Zhang, Minghui Hou, Wentao Cheng, and Xiaonian Wu. Side channel attacks based on densely connected convolutional networks with attention mechanism. In *Proceedings of the 2024 International Conference on Generative Artificial Intelligence and Information Security*, pp. 229–234, 2024a.
- Xin Zhang, Zhi Zhang, Qingni Shen, Wenhao Wang, Yansong Gao, Zhuoxi Yang, and Zhonghai Wu. Thermalscope: A practical interrupt side channel attack based on thermal event interrupts. In *Proceedings of the 61st ACM/IEEE Design Automation Conference*, pp. 1–6, 2024b.
- Mark Zhao and G Edward Suh. Remote power side-channel attacks on fpgas. *IEEE Design & Test*, 2024.
- Yu Zheng, Jiahui Zhan, Shengfeng He, Junyu Dong, and Yong Du. Curricular contrastive regularization for physics-aware single image dehazing. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.
- 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.
- Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networkss. In *Computer Vision (ICCV)*, 2017 IEEE International Conference on, 2017a.
- Jun-Yan Zhu, Richard Zhang, Deepak Pathak, Trevor Darrell, Alexei A Efros, Oliver Wang, and Eli Shechtman. Toward multimodal image-to-image translation. In *Advances in Neural Information Processing Systems*, 2017b.
- Yankun Zhu, Siting Liu, Liyu Yang, and Pingqiang Zhou. Ldl-sca: Linearized deep learning side-channel attack targeting multi-tenant fpgas. In *Proceedings of the Great Lakes Symposium on VLSI 2024*, pp. 583–587, 2024.

### A APPENDIX

#### A.1 DESCRIPTION OF THE EXPERIMENTAL SETUP

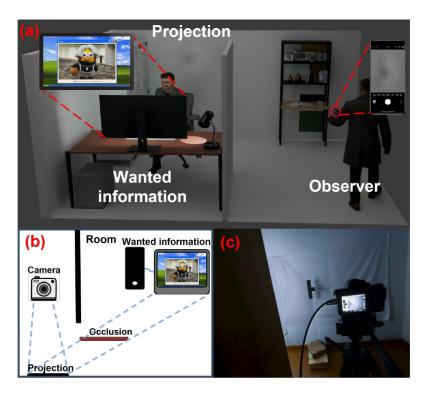


Figure 5: In the figure, (a), (b), and (c) correspond to the rendered scene, schematic diagram, and real-world scene respectively. An observer infers screen content via passive light projection. A light projection from the screen ("Wanted information") is cast onto a wall. By recording the wall's projection without viewing the screen, hacking, or capturing signals, the observer attempts to reconstruct the original content non-invasively.

As shown in Figure 5, the experimental setup comprises three sub-figures: (a) a rendered scene, (b) a schematic diagram, and (c) a real-world scene. In this setup, an observer attempts to infer the content displayed on a target screen through passive light projection. Specifically, the screen emits light containing the "Wanted information", which is indirectly projected onto a wall. The observer, without any direct visual access to the screen or active intrusion (e.g., hacking, signal tapping), records this projection in an attempt to non-invasively reconstruct the original screen content.

Figure (a) presents a 3D rendering of the experimental layout. The target screen, located in an enclosed space, displays sensitive content and is physically shielded from direct view. The light it emits is partially occluded before reaching a wall surface, where it undergoes diffuse reflection and forms a low-contrast, spatially degraded light patch. Blue dashed lines denote the boundaries of light propagation, while the green region marks the area visible to the camera. The inset on the right illustrates how multiple scattering and non-ideal reflections introduce severe nonlinear compression and information loss, eliminating most high-frequency textures and fine details.

Figure (b) presents a two-dimensional schematic of the optical path and imaging logic. It emphasizes the indirect transmission of information, from the screen to the wall and then to the camera, under conditions of severe degradation and highlights the significant compression effects within this high loss optical channel.

Figure (c) depicts the actual experimental environment. A standard computer monitor displays critical information to simulate a practical side-channel attack scenario. The screen is placed approximately 0.9m away from the wall behind it. To evaluate the method's robustness under varying reflective properties, the wall material in the observation area is designed to be interchangeable.

This simulates differences in wall reflectance commonly found in offices or server rooms and allows assessment of the system's sensitivity to environmental perturbations. The camera is located in a separate room, about 2m from the wall, with solid partitions ensuring complete physical isolation. This guarantees there is no direct line of sight or light path between the camera and the screen, thereby excluding traditional attack vectors such as network intrusion, infrared sensing, or electromagnetic eavesdropping.

#### A.2 DERIVATION OF THE OPTIMIZATION OBJECTIVE FOR INVERSE PROBLEM

In the context of wall-based indirect imaging, the screen acts as a radiation source, with each of its surface elements emitting luminous energy into space. The camera records the re-emission of these rays after they are reflected by the wall. To model this energy transfer, we begin with the radiative transfer equation and consider the radiance emitted from a point on the wall. According to classical photometry, the radiance emitted from surface point  $\mathbf{p}_w$  in direction  $\omega_o$  is expressed as an integral over all incident directions, as given by the rendering equation:

$$L_o(\mathbf{p}_w, \omega_o) = \int_{\Omega^+} f_r(\mathbf{p}_w, \omega_i \to \omega_o) L_i(\mathbf{p}_w, \omega_i) \cos \theta_i d\omega_i.$$
 (22)

Here,  $L_o\left(\mathbf{p}_w,\omega_o\right)$  denotes the radiance from the wall point  $\mathbf{p}_w$  in the outgoing direction  $\omega_o$ ,  $\Omega^+$  represents the set of all incident directions in the hemisphere, and  $f_r\left(\cdot\right)$  is the bidirectional reflectance distribution function (BRDF) at the point, describing the energy mapping between the incident direction  $\omega_i$  and outgoing direction  $\omega_o$ .  $L_i\left(\mathbf{p}_w,\omega_i\right)$  is the incident radiance, while  $\cos\theta_i$  reflects the angle between the incident ray and the surface normal, thus accounting for the energy projection effect.

To obtain the wall radiance, the incident radiance term must be further expanded. The light energy received by a point  $\mathbf{p}_w$  on the wall originates from various locations on the screen. Let  $\mathbf{p}_s$  be a point on the screen emitting radiance  $L_s(\mathbf{p}_s, \omega_s)$ , which, according to optical propagation principles, contributes to the incident radiance at  $\mathbf{p}_w$  as follows:

$$L_{i}\left(\mathbf{p}_{w}, \omega_{i}\right) = L_{s}\left(\mathbf{p}_{s}, \omega_{s}\right) \cdot V\left(\mathbf{p}_{s}, \mathbf{p}_{w}\right) \cdot \frac{\cos \theta_{s}}{\left\|\mathbf{p}_{s} - \mathbf{p}_{w}\right\|^{2}}.$$
(23)

Here,  $V(\mathbf{p}_s, \mathbf{p}_w)$  is the visibility function (taking value 1 if the propagation path is unobstructed and 0 otherwise),  $\cos \theta_s$  is the cosine of the angle between the screen normal and the light propagation direction, and  $\|\mathbf{p}_s - \mathbf{p}_w\|$  represents the distance between the two points, indicating the inverse square attenuation of energy during free-space propagation.

Substituting this expression for the incident radiance into the rendering equation, and transforming the integral domain from direction space to the screen parameter space  $\Omega_s$ , we obtain the integral expression for the wall radiance:

$$L_{o}\left(\mathbf{p}_{w},\omega_{o}\right) = \iint_{\Omega_{s}} f_{r}\left(\mathbf{p}_{w},\omega_{i}\to\omega_{o}\right) L_{s}\left(\mathbf{p}_{s},\omega_{s}\right) V\left(\mathbf{p}_{s},\mathbf{p}_{w}\right) \frac{\cos\theta_{s}\cos\theta_{i}}{\left\|\mathbf{p}_{s}-\mathbf{p}_{w}\right\|^{2}} dA_{s}.$$
 (24)

In this equation,  $\cos \theta_i$  represents the angle cosine between the wall normal and the incident direction, and  $dA_s$  is the area element of the screen surface. This equation indicates that the wall radiance is a weighted integral of the screen's pixel radiance, where the weight is determined by the reflection properties, geometric factors, and visibility.

To render the model computable, we introduce assumptions regarding the reflective properties of the wall material. If the wall is considered an ideal Lambertian diffuse reflector, the BRDF simplifies to a constant  $f_r = \rho/\pi$ , where  $\rho \in [0,1]$  is the reflectance of the surface. Substituting this into the equation, the wall radiance formula reduces to:

$$L_w\left(\mathbf{p}_w\right) = \iint_{\Omega_s} L_s\left(\mathbf{p}_s\right) \frac{\rho}{\pi} \frac{\cos \theta_s \cos \theta_i}{\left\|\mathbf{p}_s - \mathbf{p}_w\right\|^2} dA_s. \tag{25}$$

Thus, we derive the precise physical equation that describes how the screen radiance is mapped to the wall radiance through optical propagation and diffuse reflection.

Consider the imaging process of the camera. The optical system of the camera projects the luminance field of the wall onto the sensor plane, performing spatial sampling and digitization while introducing noise interference. If lens distortion is neglected and the camera response is assumed linear, the value of each pixel can be expressed as the wall's luminance plus noise:

$$y(x', y') = L_w(x', y') + n(x', y'), \quad n(x', y') \sim \mathcal{N}(0, \sigma^2),$$
 (26)

where  $n\left(x',y'\right)$  represents Gaussian noise, accounting for sensor quantization errors and environmental disturbances. To facilitate numerical treatment, both the screen image and the camera observations are discretized. Let the screen pixels be expanded into a vector of length N,  $\mathbf{x} = \left[x_1, x_2, \ldots, x_N\right]^T$ , and the wall observations into a vector of length M,  $\mathbf{y} = \left[y_1, y_2, \ldots, y_M\right]^T$ . The integral equation can then be discretized as follows:

$$y_i = \sum_{j=1}^{N} H_{ij} x_j + n_i, (27)$$

where  $H_{ij}$  represents the elements of the light transmission matrix, describing the contribution of screen pixel j to wall pixel i. It can be further written as:

$$H_{ij} = \frac{\rho}{\pi} \frac{\cos \theta_{i,j}^{(s)} \cos \theta_{i,j}^{(i)}}{\left\| \mathbf{p}_{i} - \mathbf{p}_{i} \right\|^{2}} \Delta A_{j}. \tag{28}$$

Here,  $\theta_{i,j}^{(s)}$  and  $\theta_{i,j}^{(i)}$  are the angles between the screen normal and wall normal, respectively, and  $\Delta A_j$  is the pixel area of the screen. Expressing all pixel relationships in matrix form yields:

$$y = Hx + n. (29)$$

Further abstraction of this expression into operator form gives the final model:

$$\mathbf{y} = \Phi(\mathbf{x}) + \mathbf{n}, \quad \mathbf{n} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}).$$
 (30)

This expression indicates that the observed signal y is the result of the original image x mapped by the optical operator  $\Phi$ , with added noise. Since the operator  $\Phi(\cdot)$  inherently causes information loss and noise interference, the problem is a typical ill-posed inverse problem, where direct inversion leads to instability and potential non-solvability.

To recover x, we apply statistical modeling techniques and first derive the likelihood function. According to the noise model, the observation vector y follows a Gaussian distribution conditioned on x, with the probability density function given by:

$$p(\mathbf{y} \mid \mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \parallel \mathbf{y} - \Phi(\mathbf{x}) \parallel_2^2\right), \tag{31}$$

where  $\|\cdot\|_2$  denotes the Euclidean norm. Thus, the negative log-likelihood (NLL) can be written

$$-\log p(\mathbf{y} \mid \mathbf{x}) = \frac{1}{2\sigma^2} \parallel \mathbf{y} - \Phi(\mathbf{x}) \parallel_2^2 + \text{const}, \tag{32}$$

where const is a constant independent of x and can be ignored. Therefore, employing maximum likelihood estimation (MLE), the optimization problem becomes:

$$\hat{\mathbf{x}}_{\text{MLE}} = \arg\min_{\mathbf{x}} \parallel \mathbf{y} - \Phi(\mathbf{x}) \parallel_2^2.$$
 (33)

However, due to the non-invertibility of  $\Phi(\cdot)$  and the noise amplification effects, such reconstruction relying solely on observation consistency leads to severe degradation. Therefore, it is necessary to introduce prior information to constrain the solution space, rendering the problem well-posed.

Within a Bayesian framework, a prior distribution  $p(\mathbf{x})$  is introduced to describe the statistical regularities of the image. According to Bayes' theorem, the posterior distribution satisfies:

$$p(\mathbf{x} \mid \mathbf{y}) \propto p(\mathbf{y} \mid \mathbf{x})p(\mathbf{x}).$$
 (34)

The goal of maximum a posteriori (MAP) estimation is to maximize the posterior probability:

$$\hat{\mathbf{x}}_{\text{MAP}} = \arg \max_{\mathbf{y}} p(\mathbf{x} \mid \mathbf{y}) = \arg \max_{\mathbf{y}} \left[ \log p(\mathbf{y} \mid \mathbf{x}) + \log p(\mathbf{x}) \right]. \tag{35}$$

Equivalently, taking the negative log and ignoring the constant term, this transforms into a minimization problem:

$$\hat{\mathbf{x}}_{\text{MAP}} = \arg\min_{\mathbf{x}} \left[ -\log p(\mathbf{y} \mid \mathbf{x}) - \log p(\mathbf{x}) \right]. \tag{36}$$

Substituting the likelihood and prior terms, we know that  $-\log p(\mathbf{y} \mid \mathbf{x}) = \frac{1}{2\sigma^2} \parallel \mathbf{y} - \Phi(\mathbf{x}) \parallel_2^2$ . Assuming the prior distribution has the form:

$$p(\mathbf{x}) \propto \exp\left(-\lambda R(\mathbf{x})\right),$$
 (37)

where  $R(\mathbf{x})$  represents a regularization function (such as total variation or sparsity constraints), and  $\lambda$  is a weight parameter, the negative log prior becomes:

$$-\log p(\mathbf{x}) = \lambda R(\mathbf{x}). \tag{38}$$

Thus, the MAP optimization problem can be written as:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left[ \frac{1}{2\sigma^2} \parallel \mathbf{y} - \Phi(\mathbf{x}) \parallel_2^2 + \lambda R(\mathbf{x}) \right]. \tag{39}$$

To simplify notation, the factor  $\frac{1}{2\sigma^2}$  can be absorbed into the data term or normalized directly, yielding the final form:

$$\hat{\mathbf{x}}^* = \arg\min_{\mathbf{x}} \| \Phi(\mathbf{x}) - \mathbf{y} \|_2^2 + \lambda R(\mathbf{x}). \tag{40}$$

In this objective function,  $\|\Phi(\mathbf{x}) - \mathbf{y}\|_2^2$  is the data fidelity term, ensuring the reconstructed image is consistent with the observed data;  $R(\mathbf{x})$  is the prior regularization term, incorporating statistical properties or deep learning priors of the image, thereby suppressing noise and recovering missing details; and  $\lambda$  is a balancing parameter that adjusts the relative weighting of these two terms.

# A.3 DERIVATION OF THE INVERSION PATH EMBEDDED IN OPTICAL PROPAGATION MODELING GUIDED BY THE PHYSICAL MODEL

To construct an inversion trajectory with momentum guidance and physical consistency, we introduce a structural-aware momentum initialization mechanism. The input image  $\mathbf{I}^{(0)}$  is first mapped to the initial momentum tensor  $\mathbf{m}^{(0)}$  through a learnable  $1 \times 1$  convolution operator  $\mathcal{C}_{1\times 1}(\cdot)$ , as given by:

given by.

$$\mathbf{m}^{(0)} = \mathcal{C}_{1\times 1}(\mathbf{I}^{(0)}) = \int_{\Omega_m} \mathcal{K}_{\text{init}}(\mathbf{u})\mathbf{I}^{(0)}(\mathbf{x} - \mathbf{u}) d\mathbf{u}_{\text{jl}}, \qquad (41)$$

where the process within the local receptive field  $\Omega_m$  extracts structural prior features using the convolution kernel  $\mathcal{K}_{\text{init}}$  as weights, providing the initial direction for the subsequent physical consistency optimization. Here,  $\mathbf{x} = (x_1, x_2) = (x, y)$  represents the two-dimensional spatial coordinates.

To cohesively model the coupled effects of reflection, diffraction, and scattering in light propagation, the forward light transport operator  $\Phi(\cdot)$  is constructed as a triple integral over space, depth, and channel:

$$\Phi(\mathbf{I}^{(k-1)}) = \iiint_{\Omega_m \times \mathcal{S}_d \times \mathcal{L}_e} K_{\phi}(i, j, \mathbf{u}) \cdot \phi(\mathbf{I}^{(k-1)}(\mathbf{x} - \mathbf{u})) \cdot \mathbf{1}_{\phi}(i, j) \, d\mathbf{u} \, di \, dj, \tag{42}$$

where  $\mathbf{I}^{(k-1)}$  denotes the response at the target position  $\mathbf{x}$ , and k indicates the iteration count. This is achieved by integrating over the local receptive field  $\Omega_m$ , the network depth layer  $i \in \mathcal{S}_d$ , and the feature channel  $j \in \mathcal{L}_e$ . The kernel  $K_\phi$  is channel-layer dependent, while  $\phi(\cdot)$  provides nonlinear modulation, and  $\mathbf{1}_\phi(i,j)$  dynamically activates the dominant physical mechanisms. The key advantage of this operator lies in its physical consistency: the spatial integral  $\int_{\Omega_m}$  models the specular/diffuse light spots and diffraction fringes around the neighborhood of  $\mathbf{x}$  via  $K_\phi$ ; the depth integral  $\int_{\mathcal{S}_d}$  accumulates the path superposition effects of multiple reflections and scattering, capturing indirect illumination; and the channel integration  $J_{\mathcal{L}_e}$  jointly accounts for multi-physical attribute responses, dynamically switching between reflection, diffraction, or scattering dominance using  $\mathbf{1}_\phi$  at specific locations. This cross-domain collaboration facilitates unified modeling of complex light transport.

To obtain the physical feedback direction from the current estimate  $I^{(k-1)}$ , the encoding residual is calculated as:

$$\Delta \Phi^{(k-1)} = \Phi(\mathbf{I}^{(k-1)}) - \mathbf{z}.\tag{43}$$

where z represents the deepest layer feature response along the encoding path of  $\Phi$ , encapsulating the current estimate's structural compressed representation in the projection path. Based on this residual, the inverse mapping approximation  $\Psi(\cdot)$  of the forward light transport operator  $\Phi(\cdot)$  is constructed. This operation extracts reconstruction information from  $\Delta\Phi$  using a multi-scale, multi-channel fusion approach:

$$\Psi(\Delta\Phi^{(k-1)}) = \iiint_{\Omega_m \times \mathcal{S}_u \times \mathcal{L}_d} K_{\psi}(i, j, \mathbf{v}) \cdot \psi(\Delta\Phi^{(k-1)}(\mathbf{x} + \mathbf{v})) \cdot \mathbf{1}_{\psi}(i, j) \, d\mathbf{v} \, di \, dj, \tag{44}$$

where  $S_u$  denotes the upsampling layer set,  $\mathcal{L}_d$  is the channel domain, and  $K_{\psi}(i,j,\mathbf{v})$  represents the deconvolution/upsampling kernel.  $\psi(\cdot)$  is the activation function, and  $\mathbf{1}_{\psi}(i,j)$  controls the information pathway. In physical terms,  $\Psi(\cdot)$  is equivalent to a backpropagation process that reconstructs the pre-projected image structure in the spatial and semantic domains.

Next, the model integrates the current structural estimate with physical feedback, constructing a structure-physical residual fusion term to guide optimization along physically plausible directions:

$$\widetilde{\mathbf{X}}^{(k)} = \beta_0 C_{1 \times 1} (\mathbf{I}^{(k-1)}) + (1 - \beta_0) \Psi(\Delta \Phi^{(k-1)}), \tag{45}$$

where the fusion coefficient  $\beta_0$  balances the local structural prior  $\mathcal{C}_{1\times 1}$  and the global physical feedback  $\Psi$ , maintaining a trade-off between perceptual consistency and physical interpretability.

Finally, a momentum mechanism is introduced to smooth the inversion trajectory, suppressing the propagation of unstable errors:

$$\mathbf{m}^{(k)} = \gamma \mathbf{m}^{(k-1)} + (1 - \gamma)\widetilde{\mathbf{X}}^{(k)},\tag{46}$$

 $\mathbf{I}^{(k)} = \mathbf{I}^{(k-1)} - \rho_k \mathbf{m}^{(k)}. \tag{47}$ 

where  $\gamma$  controls the degree of historical momentum retention, and  $\rho_k$  is the learning rate at the k-th step. This optimization trajectory explicitly constructs a dynamic inversion framework capable of guiding the process with structural awareness and physical consistency.

#### A.4 Mapping Function Derivation Process

To prevent redundancy in notation, we define the source set  $S = \{\text{prev}, \text{scr}\}$ , where prev and scr represent the input features for the primary mapping network and the collaborative completion network, respectively. Define:

$$s(p) = \begin{cases} \text{prev}, & p = P, \\ \text{scr}, & p = R. \end{cases}$$
 (48)

Consequently, both input paths are unified under the notation  $J_{s(p),c}^{(0)}(x,y)$ . In the dual-path perturbation dissipation feature extraction, the spatial diffusion path applies a second-order partial derivative convolution diffusion to the input:

$$F_{A,p}^{(5,c)}(x,y) = \phi \left( \iint_{B_r} \kappa_{A,p}^{(5,c)}(\xi,\eta) \frac{\partial^2 J_{s(p),c}^{(0)}}{\partial x \partial y} (x - \xi, y - \eta) \, d\xi d\eta + b_{A,p}^{(5,c)} \right). \tag{49}$$

This process simulates the local intensity gradient response of light waves encountering minute structural variations. Concurrently, the semantic attenuation path constructs a stable global mapping within the abstract space via a query-key-value mechanism:

$$A_p^{(5)}(x,x') = \frac{\exp\langle Q_p^{(5)}(x), K_p^{(5)}(x')\rangle}{\int_{\Omega} \exp\langle Q_p^{(5)}(x), K_p^{(5)}(u)\rangle du},\tag{50}$$

$$F_{B,p}^{(5,c)}(x,y) = \phi\left(\int_{\Omega_a} A_p^{(5)}(x,x') V_p^{(5,c)}(x') dx' + b_{B,p}^{(5,c)}\right). \tag{51}$$

This mapping establishes a collaborative response mechanism within the feature space for global semantic structures. Stable activation of features and participation in subsequent computations occurs only when regions satisfy contextual consistency and semantic coherence, thereby imposing constrained restoration for irreversible information loss.

Subsequently, the outputs from both paths are concatenated along the channel dimension, and Fourier transformed with low/high-frequency gated masks  $\chi_{\rm low}$  and  $1-\chi_{\rm low}$  for separation:

$$\widehat{F}^{c}(u,v) = \iint F_{A}^{(5,c)}(x,y) e^{-j2\pi(ux+vy)} dxdy,$$
 (52)

$$\hat{F}_p^{(5,c)}(u,v) = \iint F_{\text{cat},p}^{(5,c)} e^{-j2\pi(ux+vy)} \, dx \, dy, \tag{53}$$

$$\hat{F}_{\text{low},p}^{(5,c)} = \chi_{\text{low}} \hat{F}_p^{(5,c)}, \quad \hat{F}_{\text{high},p}^{(5,c)} = (1 - \chi_{\text{low}}) \hat{F}_p^{(5,c)}, \tag{54}$$

$$F_{\text{low},p}^{(5,c)} = \mathcal{F}^{-1}[\hat{F}_{\text{low},p}^{(5,c)}], \quad F_{\text{high},p}^{(5,c)} = \mathcal{F}^{-1}[\hat{F}_{\text{high},p}^{(5,c)}]. \tag{55}$$

The gradient integrals of low/high-frequency features yield the channel response  $s_p^c$ . This is then processed through two fully connected layers with activation functions to obtain the attention weights

  $\alpha_p^c$ , facilitating adaptive fusion of physical and semantic path features under attention guidance, producing the final perturbation feature  $\tilde{F}_p^{(5,c)}$ . Based on this, the STM module computes the partial derivative attention for each head h:

$$A_{h,p}(x,x') = \exp\langle \partial_x Q_{h,p}^{(5)}(x), K_{h,p}^{(5)}(x') \rangle + \langle Q_{h,p}^{(5)}(x), \partial_{x'} K_{h,p}^{(5)}(x') \rangle, \tag{56}$$

$$O_{h,p}^{(5,c)}(x,y) = \int_{\Omega_s} A_{h,p}(x,x') V_{h,p}^{(5,c)}(x') dx'.$$
 (57)

The outputs of all heads are concatenated and added to  $\tilde{F}_p^{(5,c)}$  to obtain the fused feature at layer 5:

$$Z_p^{(5,c)}(x,y) = \operatorname{Concat}_{h=1}^H(O_{h,p}^{(5,c)}(x,y)) + \tilde{F}_p^{(5,c)}(x,y), \quad V_p^{(5,c)}(x,y) = Z_p^{(5,c)}(x,y). \tag{58}$$

This mechanism not only enhances the expressiveness of multi-scale perturbations but also ensures edge clarity and semantic coherence.

#### A.5 Mapping Function Derivation Process

To prevent redundancy in notation, we define the source set  $S = \{\text{prev}, \text{scr}\}$ , where prev and scr represent the input features for the primary mapping network and the collaborative completion network, respectively. Define:

$$s(p) = \begin{cases} \text{prev}, & p = P, \\ \text{scr}, & p = R. \end{cases}$$
 (59)

Consequently, both input paths are unified under the notation  $J_{s(p),c}^{(0)}(x,y)$ . In the dual-path perturbation dissipation feature extraction, the spatial diffusion path applies a second-order partial derivative convolution diffusion to the input:

$$F_{A,p}^{(5,c)}(x,y) = \phi \left( \iint_{B_r} \kappa_{A,p}^{(5,c)}(\xi,\eta) \frac{\partial^2 J_{s(p),c}^{(0)}}{\partial x \partial y} (x - \xi, y - \eta) \, d\xi d\eta + b_{A,p}^{(5,c)} \right). \tag{60}$$

This process simulates the local intensity gradient response of light waves encountering minute structural variations. Concurrently, the semantic attenuation path constructs a stable global mapping within the abstract space via a query-key-value mechanism:

$$A_p^{(5)}(x,x') = \frac{\exp\langle Q_p^{(5)}(x), K_p^{(5)}(x')\rangle}{\int_{\Omega} \exp\langle Q_p^{(5)}(x), K_p^{(5)}(u)\rangle du},\tag{61}$$

$$F_{B,p}^{(5,c)}(x,y) = \phi\left(\int_{\Omega_c} A_p^{(5)}(x,x') V_p^{(5,c)}(x') dx' + b_{B,p}^{(5,c)}\right). \tag{62}$$

This mapping establishes a collaborative response mechanism within the feature space for global semantic structures. Stable activation of features and participation in subsequent computations occurs only when regions satisfy contextual consistency and semantic coherence, thereby imposing constrained restoration for irreversible information loss.

Subsequently, the outputs from both paths are concatenated along the channel dimension, and Fourier transformed with low/high-frequency gated masks  $\chi_{\text{low}}$  and  $1 - \chi_{\text{low}}$  for separation:

$$\widehat{F}^{c}(u,v) = \iint F_{A}^{(5,c)}(x,y) e^{-j2\pi(ux+vy)} dxdy,$$
(63)

$$\hat{F}_p^{(5,c)}(u,v) = \iint F_{\text{cat},p}^{(5,c)} e^{-j2\pi(ux+vy)} \, dx \, dy, \tag{64}$$

1028 1029

1032

1034 1035 1036

1037 1038

1039 1040

1041

1045

1046 1047

1048

1049 1050

1051 1052

1053 1054

1055

1056

1057 1058

1061 1062

1064 1065

1067

1068

1069 1070

1071 1072

1074 1075

1076

1078 1079

$$\hat{F}_{\text{low},p}^{(5,c)} = \chi_{\text{low}} \hat{F}_p^{(5,c)}, \quad \hat{F}_{\text{high},p}^{(5,c)} = (1 - \chi_{\text{low}}) \hat{F}_p^{(5,c)}.$$
 (65)

 $F_{\text{low},p}^{(5,c)} = \mathcal{F}^{-1}[\hat{F}_{\text{low},p}^{(5,c)}], \quad F_{\text{high},p}^{(5,c)} = \mathcal{F}^{-1}[\hat{F}_{\text{high},p}^{(5,c)}],$ (66)

The gradient integrals of low/high-frequency features yield the channel response  $s_p^c$ . This is then processed through two fully connected layers with activation functions to obtain the attention weights  $\alpha_p^c$ , facilitating adaptive fusion of physical and semantic path features under attention guidance, producing the final perturbation feature  $\tilde{F}_p^{(5,c)}$ . Based on this, the STM module computes the partial derivative attention for each head h:

$$A_{h,p}(x,x') = \exp\langle \partial_x Q_{h,p}^{(5)}(x), K_{h,p}^{(5)}(x') \rangle + \langle Q_{h,p}^{(5)}(x), \partial_{x'} K_{h,p}^{(5)}(x') \rangle, \tag{67}$$

$$O_{h,p}^{(5,c)}(x,y) = \int_{\Omega_s} A_{h,p}(x,x') V_{h,p}^{(5,c)}(x') dx'.$$
(68)

The outputs of all heads are concatenated and added to  $\tilde{F}_p^{(5,c)}$  to obtain the fused feature at layer 5:

$$Z_p^{(5,c)}(x,y) = \operatorname{Concat}_{h=1}^H(O_{h,p}^{(5,c)}(x,y)) + \tilde{F}_p^{(5,c)}(x,y), \quad V_p^{(5,c)}(x,y) = Z_p^{(5,c)}(x,y). \tag{69}$$

This mechanism not only enhances the expressiveness of multi-scale perturbations but also ensures edge clarity and semantic coherence.

## A.6 DATASET DESCRIPTION

In order to assess the efficacy of the proposed method, four datasets were utilized: ReSh-WebSight, ReSh-Password, ReSh-Chart, and ReSh-Screen.

ReSh-WebSight: As depicted in Figure 6, this is a publicly accessible large-scale synthetic English webpage dataset. Each sample comprises HTML/CSS (v0.2 using Tailwind CSS) along with its corresponding screenshot.

ReSh-Password: As shown in Figure 7, this dataset consists of garbled characters and is designed to simulate screen password entry scenarios, containing a total of 6800 images.

ReSh-Chart: As illustrated in Figure 8, this dataset includes various types of charts (e.g., line graphs, box plots, heatmaps), totaling 7000 images.

ReSh-Screen: As shown in Figure 9, this dataset consists of computer interface screenshots, comprising 1272 images. All datasets were partitioned using a random strategy, dividing the samples into training, validation, and test sets at an 8:1:1 ratio.

#### A.7 EVALUATION METRICS

#### (1) Mean Squared Error (MSE)

Mean Squared Error (MSE) serves as one of the fundamental metrics for image reconstruction and compression quality evaluation, directly quantifying the average squared difference between the reconstructed image I and the reference image I in pixel space. It is defined as:

$$MSE = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} (I_{ij} - \widehat{I}_{ij})^{2}.$$
 (70)

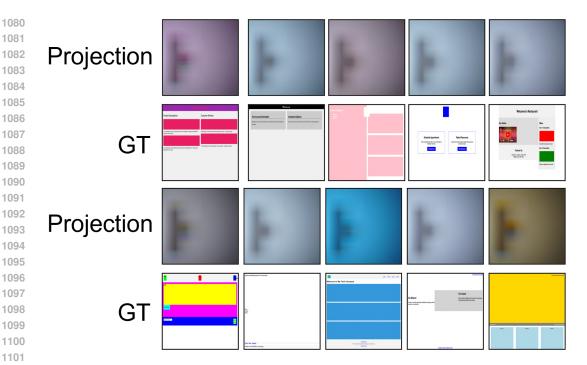


Figure 6: The ReSh-WebSight dataset is displayed: the first row shows the projection, and the second row shows the ground truth (GT).

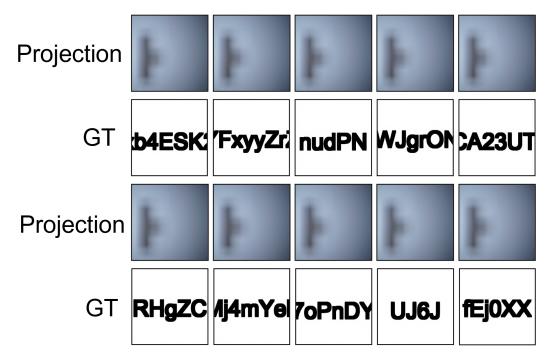


Figure 7: The ReSh-Password dataset is displayed: the first row shows the projection, and the second row shows the ground truth (GT).

where H and W represent the image height and width, and  $I_{ij}$ ,  $\hat{I}_{ij}$  denote pixel values. A smaller MSE indicates a closer match between the reconstructed and original image. Although MSE is straightforward to compute and interpretable, it lacks sensitivity to human visual perception, partic-

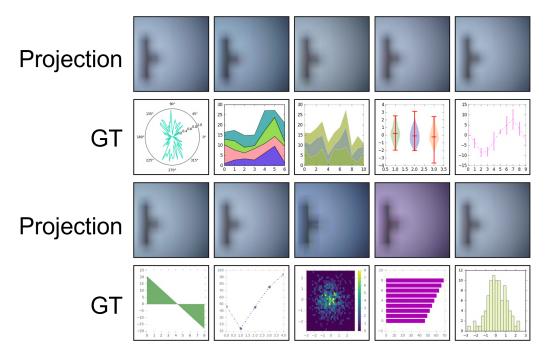


Figure 8: The ReSh-Chart dataset is displayed: the first row shows the projection, and the second row shows the ground truth (GT).



Figure 9: The ReSh-Screen dataset is displayed: the first row shows the projection, and the second row shows the ground truth (GT).

ularly in terms of structural and textural discrepancies, thereby potentially failing to capture perceptual image quality accurately.

(2) Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE), the square root of MSE, eliminates the dimensional change introduced by squaring and keeps the error measure consistent with pixel values, making it more intuitive for error interpretation. It is expressed as:

RMSE = 
$$\sqrt{\text{MSE}} = \sqrt{\frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} (I_{ij} - \hat{I}_{ij})^2}$$
. (71)

RMSE approximates the average pixel deviation, offering better interpretability than MSE in many cases. However, like MSE, it fails to account for perceptual differences in image structure or texture.

#### (3) Peak Signal-to-Noise Ratio (PSNR)

PSNR is a classical image quality metric that measures the ratio of signal strength to noise strength in decibels (dB). It is defined as:

$$PSNR = 10 \cdot \log_{10} \left( \frac{MAX^2}{MSE} \right). \tag{72}$$

Where MAX denotes the maximum possible pixel value (255 for 8-bit images). Higher PSNR values imply less distortion, with values above 30 dB typically indicating high-quality images. Although PSNR is computationally simple and widely used in signal processing, it is based solely on pixel differences and does not fully reflect perceptual quality, particularly in terms of structural or textural changes.

#### (4) Structural Similarity Index (SSIM)

The Structural Similarity Index (SSIM) is specifically designed to assess image similarity based on human visual system characteristics, evaluating luminance, contrast, and structure. It is expressed as:

$$SSIM(x,y) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}.$$
 (73)

where  $\mu_x$ ,  $\mu_y$  are the local means,  $\sigma_x^2$ ,  $\sigma_y^2$  are the local variances, and  $\sigma_{xy}$  is the covariance.  $C_1$  and  $C_2$  are constants to avoid division by zero. SSIM values range from 0 to 1, with higher values indicating better image quality. Unlike error-based metrics, SSIM aligns more closely with human visual perception and is widely used in image enhancement, super-resolution, and compression tasks.

#### (5) Multi-Scale Structural Similarity Index (MS-SSIM)

MS-SSIM is an enhancement of SSIM, calculated across multiple scales to capture structural information at various resolutions. It is defined as:

$$MS-SSIM(x,y) = \prod_{j=1}^{M} [l_j(x,y)]^{\alpha_j} [c_j(x,y)]^{\beta_j} [s_j(x,y)]^{\gamma_j}.$$
 (74)

where  $l_j, c_j, s_j$  denote the luminance, contrast, and structural components at scale j, and  $\alpha_j, \beta_j, \gamma_j$  are the corresponding weighting coefficients. MS-SSIM improves upon single-scale SSIM by incorporating structural fidelity across different scales, making it more suitable for tasks such as super-resolution and image compression quality assessment.

#### (6) Learned Perceptual Image Patch Similarity (LPIPS)

LPIPS is a deep feature-based perceptual quality metric, which compares multi-layer features extracted from pre-trained convolutional networks (e.g., AlexNet, VGG). It is defined as:

$$LPIPS(x,y) = \sum_{l} \frac{1}{H_l W_l} \sum_{h,w} \left\| w_l \odot \left( \widehat{f}_l^x(h,w) - \widehat{f}_l^y(h,w) \right) \right\|_2^2.$$
 (75)

where  $\hat{f}_l^x$ ,  $\hat{f}_l^y$  are the normalized feature maps at layer l,  $w_l$  is the learned channel weight, and  $\odot$  denotes element-wise multiplication. LPIPS captures perceptual differences at higher semantic levels, making it more aligned with human visual judgment than pixel-based metrics. However, it incurs higher computational costs and is more suitable for tasks involving image generation, style transfer, and super-resolution.

#### A.8 SUPPLEMENTARY ABLATION STUDY

#### (1) Neural Substitution of PRIrr-Approximation

To assess the influence of momentum-driven iterative design on radiometric inversion, comparative experiments were conducted across three projection scenarios, namely ReSh Chart, ReSh Screen, and ReSh WebSight. In each case, the proposed PRIrr Approximation was replaced by three canonical neural constructs: an attention-based transformer (AST), a multi-layer convolutional variant (ConvIR), and a residual network without physical modeling (DarkIR). The objective was to examine the effect of momentum-based updates under varying physical perturbation conditions.

Table 4 reports metric-wise outcomes. PRIrr-Approximation exhibits consistently favorable stability across evaluation criteria: in ReSh-Chart, PSNR exceeds ConvIR by approximately 15.1%, SSIM by 14.5%; under ReSh-Screen, LPIPS falls by 10.2% and MSE by 7.8% relative to DarkIR; within ReSh-WebSight, PSNR improves by 3.2% and SSIM by 3.3% compared with AST. These patterns indicate that momentum-embedded structures yield a broadly consistent impact on reconstruction quality across heterogeneous conditions.

This behavior may derive from momentum acting as a smoothing regulator along the iterative trajectory. Structure-aware initialization extracts stable directional cues through localized convolution, while cumulative momentum integrates historical gradients across iterations, constraining updates toward coherent evolution and mitigating oscillations induced by near-singular mappings. Consequently, this history-guided scheme forms an inversion path that remains stable and physically admissible, preserving convergence quality and robustness under multi-scale perturbations.

Dataset	Method	PSNR↑	MSE↓	RMSE↓	SSIM↑	MS-SSIM↑	LPIPS↓
	Ours	17.363	1513.986	36.748	0.731	0.641	0.431
Chart	DarkIR	16.960	1620.416	38.288	0.709	0.602	0.499
Chart	ConvIR	15.093	2521.525	47.734	0.639	0.463	0.603
	AST	16.958	1630.693	38.285	0.709	0.599	0.499
	Ours	25.812	451.633	16.531	0.817	0.845	0.216
Screen	DarkIR	24.871	490.223	17.799	0.802	0.829	0.241
Screen	ConvIR	24.803	510.053	17.989	0.800	0.826	0.242
	AST	24.697	507.025	18.050	0.792	0.824	0.237
	Ours	20.708	909.099	26.719	0.820	0.776	0.282
Wah Ciah 4	DarkIR	20.162	1178.313	29.365	0.790	0.745	0.321
WebSight	ConvIR	18.259	2041.559	37.487	0.732	0.688	0.402
	AST	20.067	1147.702	29.316	0.794	0.752	0.311

Table 4: Comparison of PRIrr-Approximation with AST, ConvIR, and DarkIR across three ReSh projection scenarios. PRIrr-Approximation consistently achieves higher reconstruction stability and quality, attributed to momentum-guided updates that enhance convergence and suppress perturbation-induced oscillations.

#### A.9 RECONSTRUCTION PERFORMANCE ACROSS DIFFERENT MATERIALS

This experiment aims to assess the adaptability and robustness of the proposed model under varying wall surface materials. Four surface types were selected for evaluation: a Typical Matte White Wall Surface, a Diffuse Scattering Wallpaper Surface, a Contaminated Diffuse Scattering Wallpaper Surface, and a Rough Textured Wallpaper Surface. Reconstructions were performed under consistent projection and imaging conditions.

As illustrated in Figure 10, the model successfully reconstructs images containing complete object contours and key semantic structures, even under conditions of surface contamination and high roughness. However, some degradation in texture fidelity is observed in detail areas. These results indicate that the proposed method maintains high structural consistency across different reflection and scattering patterns.

This performance may be attributed to the physical consistency constraints embedded within the inversion network, coupled with a frequency-selective upsampling mechanism. The former confines unreasonable light transmission paths during the iterative process, thereby reducing instability induced by surface scattering discrepancies. The latter, through cross-scale filtering, ensures the prioritization of low-frequency structural recovery, thereby mitigating the impact of surface feature variations on the overall reconstruction accuracy.

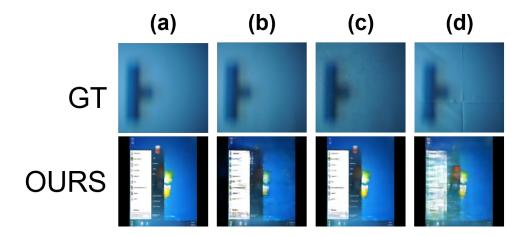


Figure 10: Figure 4. Reconstruction results under different wall surface materials: (a) Typical Matte White Wall; (b) Diffuse Scattering Wallpaper; (c) Contaminated Diffuse Scattering Wallpaper; (d) Rough Textured Wallpaper. All experiments were conducted under consistent projection and imaging conditions.

#### A.10 IMAGE CROPPING EXPERIMENT

The purpose of this experiment is to evaluate the model's reconstruction performance on images with varying cropped regions, further assessing the efficacy of the frequency-selective upsampling mechanism. Specifically, the projection image is divided into four subregions—top-left, top-right, bottom-left, and bottom-right—using a sliding window technique, while the remaining portion is filled with a gray tone. This approach generates different occlusion configurations to measure the impact of spatial occlusions on model performance.

As indicated in Table 5, the metric comparison demonstrates that the top-left and bottom-left regions consistently exhibit better performance than the top-right and bottom-right regions across different image categories, with average improvements of approximately 29% in Structural Similarity Index (SSIM) and 45% in Multi-Scale SSIM (MS-SSIM). Figure 11 further reveals that reconstructions in the top-left and bottom-left regions preserve richer structural details and texture information. The model is more responsive to occlusion edges in these areas, resulting in fewer blurring and misalignment artifacts.

This observation may be attributed to the multi-scale diffraction interference caused by the occlusion edges, which enhances local frequency activation and triggers the model's internal frequency upsampling mechanism. Feature heatmaps in Figure 11 show that the network more frequently captures fine-grained diffraction patterns in the top-left and bottom-left regions. This suggests that the frequency domain separation mechanism and physical regularization paths embedded in the design are effectively activated, aiding in the information recovery process.

Dataset	Region	SSIM↑	PSNR↑	MSE↓	RMSE↓	LPIPS↓	MS-SSIM↑
	Top-Left	0.621	15.044	2362.987	46.877	0.502	0.472
Chart	Top-Right	0.389	10.054	6604.782	80.704	0.629	0.124
Chart	Bottom-Left	0.637	15.026	2285.840	46.487	0.501	0.484
	Bottom-Right	0.406	10.515	5959.803	76.586	0.619	0.135
	Top-Left	0.801	11.305	4877.204	69.613	0.235	0.743
Password	Top-Right	0.749	9.896	6702.749	81.742	0.302	0.630
Passworu	Bottom-Left	0.847	13.179	3219.007	56.338	0.215	0.825
	Bottom-Right	0.769	10.110	6369.445	79.715	0.255	0.677
	Top-Left	0.614	17.732	1793.367	38.514	0.379	0.606
Screen	Top-Right	0.365	11.597	7093.265	77.646	0.600	0.278
Screen	Bottom-Left	0.684	19.727	1075.403	29.778	0.320	0.688
	Bottom-Right	0.374	12.037	6352.646	73.519	0.591	0.295
	Top-Left	0.624	14.975	2631.137	48.225	0.496	0.452
WahCiaht	Top-Right	0.523	11.101	5765.094	73.468	0.582	0.309
WebSight	Bottom-Left	0.667	13.757	3504.527	55.736	0.439	0.531
	Bottom-Right	0.524	12.221	5039.852	66.680	0.566	0.349

Table 5: Ablation results of quadrant-wise occlusion across four datasets. The top-right and bottomright occlusions consistently lead to better reconstruction quality, as reflected in higher SSIM and lower LPIPS, indicating that the left-side regions are more critical for structure-preserving inversion. Occluding left-side regions, especially top-left, causes more severe degradation in perceptual and structural metrics.



Figure 11: This figure presents the results of cropping experiments conducted on four different datasets. Each image is cropped into four regions: top-left, top-right, bottom-left, and bottom-right, with other areas filled in gray. The purpose of the experiment is to analyze, by combining model inference with heatmaps, whether the model can attend to multi-scale diffraction fringes caused by shadow edges, thereby achieving effective reconstruction.

#### A.11 SUPPLEMENTARY LUMINANCE EXPERIMENT

As shown in Table 6 - 11, with decreasing luminance, certain methods suffer from abrupt performance degradation across multiple datasets. For instance, UNet on the ReSh-Screen dataset experiences a drop of 69%, while **CIDNet** on *ReSh-WebSight* decreases by 58.6%. In contrast, the proposed method only exhibits reductions of 25.9% and 31.9% under the same conditions, respec-

tively. Furthermore, on the *ReSh-Password* dataset, **C2PNet** shows a substantial decline from 11.2 to 2.82, amounting to a decrease of over 74%, whereas our method only drops by 19.3%.

As depicted in Figure 12 - 15, when luminance falls, most models produce images with misaligned structures and blurred contours. In contrast, the proposed method maintains stable textures and consistent edge definition under the same low-luminance conditions.

The results indicate that the robustness of this architecture arises from three key components: (i) physical constraints that limit the propagation of disturbances, (ii) frequency-selective upsampling that enhances cross-scale consistency, and (iii) the semantic stability module that replenishes lost information. In contrast to traditional methods that tend to accumulate errors and experience structural degradation under low luminance, the proposed method suppresses perturbations through physical modeling, controls non-structural amplification in the frequency domain, and utilizes semantic consistency to restore missing regions, thereby achieving texture preservation and feature stability, significantly mitigating performance decline.

#### A.12 SUPPLEMENTARY NOISE EXPERIMENTS

The objective of this experiment is to evaluate the robustness of the proposed method in the context of image inversion under diverse noise conditions. The experiments are conducted across four datasets, with the application of five types of Gaussian noise and five types of salt-and-pepper noise. Comparisons are made against four baseline models as well as the proposed Physically-Regularized Inversion Network. The evaluation metrics include PSNR, SSIM, LPIPS, and MS-SSIM, which comprehensively assess structural fidelity, perceptual quality, and noise suppression performance. As depicted in Figures 16- 19 it is evident that the proposed method consistently outperforms the alternatives across various noise levels in most scenarios.

As shown in Table 12 - 16, the proposed method maintains a leading performance in both PSNR and SSIM. For instance, in the Gaussian noise scenario at 20 dB on the *ReSh-Screen* dataset, the PSNR improves by approximately 20%–30% compared to the second-best baseline, while SSIM increases by over 10%, accompanied by a significant reduction in LPIPS. This trend is similarly observed in the salt-and-pepper noise tests, indicating the method's stability in recovering structures even under destructive noise conditions. Overall, the results suggest that the proposed approach achieves superior reconstruction quality across various noise types and intensities.

The observed performance gains are likely attributable to the integration of physical consistency constraints and a frequency-selective feature fusion mechanism within the network architecture. On the one hand, the inversion path incorporating the optical propagation model effectively mitigates noise amplification, steering the estimation process toward physically plausible directions. On the other hand, the dual-path disturbance decoupling and frequency-domain gating strategy attenuate high-frequency noise components while maintaining the cross-scale consistency of low-frequency semantic features. These combined design elements likely explain the method's ability to maintain stable recovery under multiple noise scenarios.

#### A.13 VIDEO-BASED DYNAMIC IRRADIANCE RECONSTRUCTION EVALUATION

To assess the model's capacity to preserve temporal irradiance consistency and detail integrity under realistic human–computer interaction patterns, we construct a test sequence comprising common window operations within the Windows OS interface—namely, window switching, interface scrolling, and dialog box invocation. These user-driven events naturally induce dynamic irradiance modulations, encompassing localized brightness fluctuations, shadow transitions, and specular variations, thereby emulating realistic projection-induced perturbations in temporal irradiance fields.

As illustrated in Figure 20, the proposed method preserves inter-frame structural sharpness and exhibits smooth, artifact-free light transitions across temporally adjacent frames. Notably, during abrupt events such as control emergence or interface swapping—where luminance discontinuities become pronounced—the reconstruction maintains temporal coherence and structural integrity, avoiding edge fragmentation or texture drift. Highlight reflectance and shading continuity are preserved without explicit supervision, suggesting implicit stability under photometric discontinuities.

Dataset	Brightness reduced (nits)	SSIM↑	PSNR↑	MSE↓	RMSE↓	LPIPS↓	MS-SSIM
	25	0.727	17.259	1544.880	37.155	0.435	0.634
	50	0.725	17.201	1562.124	37.378	0.436	0.630
	75	0.725	17.193	1570.321	37.451	0.436	0.631
	100	0.724	17.187	1572.670	37.482	0.436	0.630
	125	0.722	17.134	1588.369	37.687	0.437	0.630
ReSh-Chart	150	0.718	17.066	1615.228	37.994	0.440	0.626
Resil-Chart	175	0.712	16.958	1650.673	38.438	0.445	0.619
	200	0.704	16.804	1709.271	39.124	0.450	0.610
	225	0.686	16.480	1830.240	40.551	0.464	0.586
	250	0.661	15.991	2023.473	42.763	0.481	0.550
	275	0.606	14.954	2472.827	47.676	0.516	0.461
	300	0.542	13.712	3162.928	54.383	0.554	0.365
	25	0.880	14.586	2388.873	48.242	0.128	0.885
	50	0.864	13.680	2909.675	53.385	0.150	0.857
	75	0.857	13.345	3145.904	55.500	0.158	0.846
	100	0.853	13.159	3272.432	56.650	0.164	0.839
	125	0.850	12.969	3400.880	57.820	0.168	0.833
DoCh Doggwood	150	0.846	12.796	3534.532	58.966	0.174	0.827
ReSh-Password	175	0.842	12.622	3671.877	60.129	0.179	0.820
	200	0.841	12.567	3716.191	60.499	0.179	0.819
	225	0.837	12.371	3877.600	61.836	0.185	0.811
	250	0.836	12.335	3909.543	62.090	0.187	0.809
	275	0.833	12.212	4011.323	62.932	0.192	0.802
	300	0.830	12.127	4091.547	63.556	0.198	0.797
	25	0.816	25.726	453.701	16.653	0.219	0.842
	50	0.814	25.702	451.119	16.658	0.219	0.841
	75	0.812	25.634	450.689	16.727	0.220	0.840
	100	0.810	25.533	453.817	16.841	0.222	0.837
	125	0.806	25.288	463.534	17.181	0.226	0.832
D-Cl- C	150	0.800	24.990	478.884	17.628	0.231	0.825
ReSh-Screen	175	0.790	24.537	509.553	18.388	0.238	0.814
	200	0.777	24.016	553.675	19.352	0.247	0.800
	225	0.757	23.220	635.812	21.025	0.262	0.779
	250	0.731	22.306	767.505	23.326	0.282	0.752
	275	0.691	20.983	1043.217	27.348	0.310	0.713
	300	0.641	19.136	1804.186	35.266	0.350	0.667
	25	0.815	20.050	1065.026	28.849	0.302	0.764
	50	0.813	20.142	1051.119	28.615	0.304	0.764
	75	0.812	20.134	1048.181	28.600	0.305	0.763
	100	0.810	20.062	1065.397	28.839	0.306	0.762
	125	0.806	19.925	1094.066	29.250	0.308	0.758
D-Cl-W/1 ' 1	150	0.798	19.595	1155.733	30.211	0.313	0.751
ReSh-Websight	175	0.794	19.483	1186.217	30.604	0.317	0.746
	200	0.773	18.747	1343.429	32.911	0.334	0.724
	225	0.746	18.029	1504.957	35.211	0.366	0.691
	250	0.714	17.049	1747.104	38.628	0.404	0.655
	275	0.690	15.950	2090.561	43.037	0.430	0.629
	300	0.663	14.108	2949.844	52.241	0.449	0.600

Table 6: Performance of our model(IR<sup>4</sup>Net) under varying brightness reduction levels across different datasets. The values on the left indicate the amount of brightness reduced (in nits). Higher SSIM, PSNR, and MS-SSIM and lower MSE, RMSE, and LPIPS represent better quality.

Dataset	Brightness Reduction (nits)	SSIM↑	<b>PSNR</b> ↑	MSE↓	RMSE↓	LPIPS↓	MS-SSIM
	25	0.648	15.196	2538.976	47.548	0.594	0.456
	50	0.611	14.752	2901.607	50.482	0.606	0.406
	75	0.553	13.816	3512.275	55.756	0.629	0.314
	100	0.476	12.196	4622.488	65.367	0.662	0.201
	125	0.410	10.152	6564.463	80.124	0.691	0.133
chart	150	0.392	9.495	7565.657	86.229	0.700	0.127
Chart	175	0.389	9.105	8272.198	90.182	0.708	0.131
	200	0.384	8.756	8966.084	93.889	0.712	0.133
	225	0.381	8.369	9808.027	98.194	0.716	0.134
	250	0.375	7.897	10939.179	103.712	0.716	0.141
	275	0.370	7.283	12652.558	111.464	0.723	0.153
	300	0.359	6.805	14241.234	118.054	0.730	0.157
	25	0.751	10.118	6352.994	79.630	0.364	0.614
	50	0.684	8.814	8576.730	92.524	0.436	0.461
	75	0.589	7.520	11563.799	107.413	0.528	0.295
	100	0.535	6.832	13561.278	116.295	0.569	0.246
	125	0.491	6.060	16180.022	127.066	0.610	0.165
naceword	150	0.423	5.449	18581.314	136.241	0.642	0.100
password	175	0.372	5.032	20449.292	142.938	0.666	0.081
	200	0.312	4.837	21369.060	146.149	0.709	0.073
	225	0.239	4.426	23509.117	153.256	0.734	0.048
	250	0.186	3.632	28249.950	167.967	0.731	0.030
	275	0.165	2.871	33594.730	183.263	0.729	0.035
	300	0.151	2.823	33954.049	184.254	0.742	0.061
	25	0.536	15.913	2726.712	47.457	0.521	0.474
	50	0.502	14.951	3661.368	54.182	0.546	0.430
	75	0.463	13.452	5886.223	66.939	0.575	0.374
	100	0.433	12.311	7786.670	76.863	0.598	0.340
	125	0.417	11.520	8912.763	83.150	0.616	0.320
corean	150	0.414	11.195	9137.827	85.224	0.625	0.318
screen	175	0.410	10.948	9255.518	86.737	0.630	0.320
	200	0.403	10.542	9749.742	90.004	0.640	0.317
	225	0.397	10.039	10738.375	95.089	0.652	0.321
	250	0.389	9.666	11653.274	99.311	0.660	0.319
	275	0.385	9.370	12583.531	103.077	0.667	0.324
	300	0.384	9.144	13365.562	106.060	0.671	0.336
	25	0.762	15.500	3897.519	52.999	0.436	0.591
	50	0.751	15.058	4173.101	55.312	0.445	0.583
	75	0.736	14.356	4806.018	59.771	0.457	0.573
	100	0.713	13.342	5888.361	66.785	0.474	0.557
	125	0.682	11.925	7562.463	77.095	0.499	0.536
websight	150	0.643	10.126	9957.350	91.080	0.528	0.515
websigiit	175	0.597	8.252	12982.224	107.414	0.560	0.499
	200	0.550	6.699	16394.340	123.773	0.591	0.488
	225	0.503	5.569	20038.779	138.662	0.616	0.488
	250	0.463	4.868	23097.679	149.561	0.631	0.492
	275	0.424	4.321	25990.787	158.970	0.642	0.499
	300	0.391	3.937	28323.736	166.064	0.647	0.504

Table 7: Performance of C2PNet model under varying brightness reduction levels across different datasets. The values in the left column represent the amount of brightness reduced (in nits). Higher SSIM, PSNR, and MS-SSIM, and lower MSE, RMSE, and LPIPS indicate better visual quality.

Dataset	Brightness Reduction (nits)	SSIM↑	<b>PSNR</b> ↑	$\mathbf{MSE}{\downarrow}$	$RMSE\!\!\downarrow$	$\textbf{LPIPS}\!\!\downarrow$	MS-SSIM
	25	0.688	15.662	2298.720	45.111	0.536	0.525
	50	0.686	15.601	2329.272	45.422	0.535	0.522
	75	0.682	15.516	2375.274	45.875	0.536	0.517
	100	0.675	15.360	2450.391	46.656	0.539	0.506
	125	0.662	15.084	2589.863	48.073	0.544	0.485
	150	0.644	14.649	2814.312	50.305	0.554	0.453
chart	175	0.621	14.060	3121.791	53.340	0.567	0.412
	200	0.591	13.338	3542.436	57.307	0.581	0.357
	225	0.552	12.462	4137.767	62.577	0.600	0.287
	250	0.517	11.675	4774.064	67.808	0.615	0.232
	275	0.490	10.984	5440.248	72.874	0.629	0.194
	300	0.473	10.543	5948.814	76.437	0.640	0.174
		0.473	10.545	3740.014	70.437	0.040	0.174
	25	0.850	12.650	3611.269	59.772	0.193	0.813
	50	0.846	12.433	3796.162	61.283	0.197	0.804
	75	0.843	12.329	3887.131	62.017	0.199	0.800
	100	0.844	12.383	3836.463	61.623	0.203	0.800
	125	0.841	12.227	3976.559	62.738	0.209	0.793
,	150	0.832	11.877	4307.986	65.311	0.224	0.776
password	175	0.823	11.622	4576.196	67.286	0.237	0.761
	200	0.808	11.160	5080.836	70.927	0.266	0.732
	225	0.786	10.529	5864.582	76.236	0.296	0.695
	250	0.764	9.952	6671.487	81.388	0.325	0.658
	275	0.730	9.181	7950.280	88.894	0.361	0.604
	300	0.696	8.466	9338.889	96.429	0.400	0.549
	25	0.701	20.808			0.380	0.683
	1	I		1052.193	27.296		
	50	0.684	20.337	1149.735	28.830	0.392	0.664
	75	0.654	19.366	1400.857	32.230	0.411	0.627
	100	0.616	18.118	1823.141	37.282	0.437	0.580
	125	0.567	16.576	2582.688	44.834	0.473	0.522
screen	150	0.521	15.131	3590.042	53.192	0.505	0.467
	175	0.480	13.853	4775.349	61.686	0.534	0.420
	200	0.445	12.791	6108.109	69.960	0.560	0.381
	225	0.415	11.739	7860.197	79.504	0.587	0.344
	250	0.394	10.913	9647.824	88.081	0.608	0.322
	275	0.381	10.192	11570.630	96.381	0.627	0.311
	300	0.379	9.745	12831.590	101.587	0.641	0.320
	25	0.661	8.574	9515.409	96.525	0.570	0.427
	50	0.664	8.903	8814.484	92.867	0.573	0.421
	75	0.653	8.314	10054.489	99.327	0.582	0.435
	100	0.657	8.296	10093.491	99.532	0.577	0.436
	125	0.640	7.691	11643.784	106.882	0.582	0.447
wahaiaht	150	0.629	7.301	12772.741	111.904	0.583	0.457
websight	175	0.615	7.001	13744.349	116.006	0.584	0.455
	200	0.579	6.196	16627.105	127.521	0.596	0.472
	225	0.625	7.253	12916.477	112.528	0.588	0.465
	250	0.611	6.874	14157.671	117.735	0.586	0.464
	275	0.571	6.067	17127.085	129.420	0.599	0.473
	300	0.645	7.924	11049.247	104.073	0.578	0.442

Table 8: Performance of HVI-CIDNet model under varying brightness reduction levels across different datasets. The values in the left column indicate the amount of brightness reduced (in nits). Higher SSIM, PSNR, and MS-SSIM and lower MSE, RMSE, and LPIPS indicate better visual quality.

Dataset	Brightness Reduction (nits)	SSIM↑	<b>PSNR</b> ↑	MSE↓	$RMSE\!\!\downarrow$	LPIPS↓	MS-SSIM↑
	25	0.705	16.601	1762.022	39.926	0.526	0.574
	50	0.699	16.421	1831.477	40.729	0.525	0.570
	75	0.691	16.183	1943.446	41.917	0.536	0.550
	100	0.675	15.809	2126.285	43.813	0.552	0.522
	125	0.635	15.130	2462.232	47.286	0.567	0.472
.1	150	0.549	13.046	3493.387	57.960	0.604	0.344
chart	175	0.492	11.733	4550.248	66.785	0.633	0.264
	200	0.471	11.253	5047.724	70.460	0.645	0.236
	225	0.462	10.877	5497.519	73.554	0.654	0.223
	250	0.451	10.324	6248.731	78.411	0.664	0.212
	275	0.443	9.537	7505.830	85.916	0.675	0.203
	300	0.436	8.843	8841.484	93.186	0.689	0.198
	25	0.849	13.422	3038.232	54.759	0.187	0.834
	50	0.845	13.422	3239.782	56.568	0.194	0.825
	75	0.842	12.952	3380.108	57.781	0.194	0.823
	100		12.734				
		0.839		3526.131	59.126	0.205	0.816
	125	0.833	12.432	3764.814	61.154	0.216	0.807
password	150	0.822	12.049	4106.705	63.892	0.230	0.786
	175	0.803	11.590	4544.990	67.285	0.253	0.758
	200	0.779	11.037	5145.557	71.648	0.283	0.719
	225	0.756	10.474	5852.138	76.426	0.315	0.688
	250	0.739	10.026	6488.705	80.474	0.336	0.673
	275	0.728	9.728	6953.603	83.298	0.348	0.669
	300	0.722	9.501	7325.570	85.498	0.358	0.670
	25	0.728	21.824	786.760	24.137	0.354	0.718
	50	0.721	21.480	826.035	24.909	0.359	0.711
	75	0.699	20.601	989.336	27.535	0.374	0.689
	100	0.671	19.223	1364.980	32.575	0.401	0.651
	125	0.636	17.698	2050.895	39.810	0.436	0.608
coroon	150	0.601	16.215	3052.722	48.381	0.468	0.569
screen	175	0.570	14.885	4357.889	57.576	0.500	0.536
	200	0.543	13.662	6016.988	67.405	0.528	0.508
	225	0.516	12.537	7943.358	77.409	0.553	0.486
	250	0.492	11.707	9622.782	85.364	0.571	0.472
	275	0.469	10.991	11226.232	92.501	0.585	0.462
	300	0.449	10.435	12495.862	97.994	0.595	0.455
	25	0.796	19.667	1219.075	30.491	0.329	0.748
	50	0.791	18.957	1297.736	32.081	0.336	0.741
	75	0.785	17.889	1486.076	35.249	0.345	0.733
	100	0.775	16.347	1890.953	40.976	0.357	0.721
	125	0.637	17.698	2050.895	39.810	0.436	0.608
	150	0.740	12.632	3966.833	61.330	0.395	0.687
websight	175	0.570	14.885	4357.889	57.576	0.500	0.536
	200	0.682	9.405	8163.928	88.666	0.442	0.656
	225	0.645	8.104	11025.751	103.115	0.442	0.646
	250	0.609	7.151	13782.521	115.249	0.488	0.638
	275	0.469	10.991	11226.232	92.501	0.585	0.462
	300	0.541	5.842	18674.280	134.159	0.530	0.621

Table 9: Performance of our ConvIR under varying brightness reduction levels across different datasets. The values in the left column represent the amount of brightness reduced (in nits). Higher SSIM, PSNR, and MS-SSIM and lower MSE, RMSE, and LPIPS indicate better visual quality.

Dataset	Brightness Reduction (nits)	SSIM↑	<b>PSNR</b> ↑	MSE↓	RMSE↓	LPIPS↓	MS-SSIM
	25	0.697	16.550	1821.971	40.350	0.489	0.576
	50	0.678	16.044	2010.286	42.574	0.503	0.546
	75	0.653	15.441	2286.395	45.526	0.521	0.506
	100	0.622	14.692	2692.348	49.479	0.542	0.456
	125	0.569	13.497	3448.897	56.325	0.568	0.386
.1	150	0.483	11.571	5069.740	69.262	0.612	0.275
chart	175	0.392	9.406	7882.010	87.622	0.659	0.166
	200	0.362	8.329	10083.420	99.197	0.685	0.138
	225	0.347	7.867	11192.537	104.568	0.697	0.123
	250	0.337	7.628	11772.978	107.367	0.704	0.116
	275	0.329	7.374	12472.405	110.540	0.706	0.111
	300	0.325	7.077	13415.085	114.532	0.705	0.109
	25	0.842	13.053	3314.844	57.168	0.179	0.824
	50	0.834	12.651	3627.064	59.838	0.187	0.806
	75	0.834	12.730	3564.317	59.306	0.186	0.811
	100	0.829	12.514	3729.618	60.730	0.192	0.802
	125	0.820	12.165	4032.859	63.182	0.199	0.788
	150	0.809	11.818	4351.762	65.695	0.207	0.773
password	175	0.792	11.244	4961.238	70.163	0.222	0.741
	200	0.776	10.747	5554.808	74.268	0.237	0.715
	225	0.759	10.281	6165.694	78.301	0.252	0.686
	250	0.747	10.027	6527.133	80.593	0.261	0.670
	275	0.693	9.283	7738.470	87.774	0.341	0.585
	300	0.648	8.635	8932.544	94.439	0.385	0.527
	25	0.698	21.360	903.737	25.729	0.352	0.703
	50	0.678	20.660	997.602	27.556	0.361	0.685
	75	0.637	19.288	1322.155	32.283	0.384	0.642
	100	0.577	17.423	2026.182	40.377	0.421	0.573
	125	0.511	15.355	3299.462	51.885	0.469	0.492
	150	0.467	13.672	4911.837	63.491	0.511	0.429
screen	175	0.438	12.481	6502.774	73.155	0.549	0.387
	200	0.412	11.491	8205.150	82.303	0.577	0.354
	225	0.393	10.614	9949.519	90.878	0.599	0.332
	250	0.383	10.077	11180.983	96.505	0.614	0.319
	275	0.379	9.686	12156.798	100.737	0.626	0.315
	300	0.378	9.424	12810.155	103.526	0.637	0.325
	25	0.154	5.414	19568.754	138.729	0.818	0.382
	50	0.102	6.789	14199.959	118.251	0.739	0.390
	75	0.085	5.868	17809.940	132.014	0.825	0.410
	100	0.117	4.735	23413.799	151.120	0.751	0.432
	125	0.165	6.456	15524.627	123.371	0.727	0.398
websicht	150	0.246	7.249	12871.406	112.405	0.675	0.397
websight	175	0.255	6.030	16908.888	128.927	0.662	0.460
	200	0.094	7.147	13087.513	113.473	0.676	0.368
	225	0.182	6.794	14243.713	118.289	0.702	0.422
	250	0.252	6.665	14682.275	120.069	0.718	0.433
	275	0.107	6.336	15799.116	124.619	0.656	0.415
	300	0.232	7.306	12615.531	111.383	0.646	0.437

Table 10: Performance of DarkIR model under varying brightness reduction levels across different datasets. The values in the left column represent the amount of brightness reduced (in nits). Higher SSIM, PSNR, and MS-SSIM and lower MSE, RMSE, and LPIPS indicate better visual quality.

Dataset	$\Big  \ Brightness \ Reduction \ (nits)$	SSIM↑	<b>PSNR</b> ↑	MSE↓	RMSE↓	LPIPS↓	MS-SSIM↑
	25	0.130	1.682	44923.767	211.063	0.840	0.287
	50	0.056	1.200	50139.276	223.041	0.852	0.253
	75	0.034	1.035	52072.622	227.307	0.867	0.225
	100	0.041	1.129	50963.553	224.872	0.864	0.244
	125	0.069	1.262	49451.318	221.486	0.870	0.259
	150	0.022	0.988	52629.401	228.528	0.865	0.212
chart	175	0.058	1.222	49884.200	222.478	0.856	0.256
	200	0.085	1.377	48153.784	218.558	0.889	0.269
	225	0.011	0.913	53539.982	230.503	0.876	0.110
	250	0.071	1.263	49445.035	221.468	0.876	0.259
	275	0.029	1.007	52412.011	228.047	0.872	0.217
	300	0.011	0.913	53539.982	230.503	0.876	0.110
	25	0.097	1.054	51025.639	225.871	0.758	0.349
	50	0.127	1.226	49046.130		0.745	0.366
	75	0.107	0.979	51915.575		0.786	0.341
	100	0.096	0.923	52587.651		0.774	0.333
	125	0.065	0.669	55766.004		0.758	0.234
	150	0.125	1.344	47734.110		0.714	0.374
password	175	0.065	0.667	55789.352		0.755	0.230
	200	0.070	0.788	54248.641	232.895	0.767	0.304
	225	0.128	1.228	49026.231	221.401	0.757	0.366
	250	0.088	0.877	53147.653		0.798	0.325
	275	0.092	0.903	52829.031		0.771	0.329
	300	0.085	0.864	53311.254		0.733	0.322
	25	0.176	6.302	23948.491	141.973	0.750	0.343
	50	0.200	6.461	23712.630	140.711	0.716	0.347
	75	0.147	6.121	24933.079	145.014	0.755	0.329
	100	0.197	6.546	23431.301	139.664	0.702	0.349
	125	0.140	6.104	25550.478	146.626	0.737	0.321
	150	0.180	6.301	25076.352		0.689	0.329
screen	175	0.272	7.015	21270.457	132.532	0.684	0.364
	200	0.120	6.089	26911.471	150.267	0.707	0.242
	225	0.227	6.686	22740.821	137.512	0.697	0.355
	250	0.146	6.257	25667.588	146.308	0.694	0.320
	275	0.124	6.144	25850.045	147.294	0.719	0.319
	300	0.182	6.412	24671.415	143.170	0.693	0.337
	25	0.087	1.741	47075.345	213.935	0.707	0.439
	50	0.050	1.458	14199.959	118.251	0.703	0.214
	75	0.107	1.788	46488.952	212.612	0.672	0.450
	100	0.094	1.721	47246.318	214.342	0.752	0.437
	125	0.112	1.880	45541.721	210.416	0.728	0.462
عملت ثم ملم	150	0.094	1.792	46522.294	212.662	0.679	0.448
websight	175	0.120	1.926	45103.866	209.376	0.715	0.467
	200	0.100	1.755	46883.500	213.497	0.726	0.443
	225	0.106	1.774	46641.908	212.966	0.762	0.448
	250	0.078	1.666	47902.625		0.681	0.424
	275	0.069	1.659	48007.867		0.700	0.421
	300	0.108	1.792	46447.455		0.672	0.450

Table 11: Performance of UNet model under varying brightness reduction levels across different datasets. The values in the left column indicate the amount of brightness reduced (in nits). Higher SSIM, PSNR, and MS-SSIM and lower MSE, RMSE, and LPIPS indicate better visual quality.

Dataset	Noise Type	SNR (dB)	SSIM↑	<b>PSNR</b> ↑	MSE↓	RMSE↓	LPIPS↓	MS-SSIM↑
	<u>'</u>	15	0.688	15.674	2293.165	45.056	0.533	8.311
		20	0.691	15.705	2280.971	44.916	0.533	8.703
	gaussian	25	0.691	15.715	2276.276	44.866	0.533	8.834
		30	0.692	15.716	2276.042	44.863	0.533	8.869
chart		35	0.692	15.715	2275.895	44.865	0.533	8.876
		25	0.678	15.437	2401.394	46.204	0.539	8.289
		30	0.681	15.499	2370.929	45.888	0.537	8.397
	salt pepper	35	0.681	15.496	2373.288	45.913	0.537	8.391
		40	0.681	15.481	2376.261	45.963	0.537	8.391
		45	0.681	15.495	2370.964	45.899	0.538	8.408
		15	0.801	10.953	5293.539	72.512	0.270	0.359
		20	0.836	12.071	4111.432	63.832	0.208	0.148
	gaussian	25	0.848	12.551	3689.944	60.436	0.196	0.155
		30	0.850	12.674	3589.991	59.599	0.194	0.160
password		35	0.850	12.676	3591.947	59.604	0.194	0.161
•		25	0.659	8.036	10325.171	101.359	0.517	4.148
	salt pepper	30	0.683	8.414	9482.130	97.089	0.486	3.153
		35	0.682	8.407	9497.168	97.163	0.486	3.215
		40	0.681	8.384	9538.681	97.397	0.488	3.216
		45	0.683	8.414	9472.439	97.061	0.486	3.151
		15	0.701	21.088	1009.091	26.535	0.373	3.428
		20	0.707	21.133	1007.925	26.476	0.374	3.910
	gaussian	25	0.709	21.145	1007.086	26.465	0.374	4.135
		30	0.709	21.149	1006.137	26.452	0.374	4.193
screen		35	0.709	21.149	1007.045	26.460	0.374	4.195
		25	0.702	21.298	1007.783	26.532	0.366	3.568
	_	30	0.704	21.258	1010.903	26.542	0.366	3.699
	salt pepper	35	0.701	21.095	1009.556	26.535	0.366	3.703
		40	0.703	21.095	1010.971	26.548	0.366	3.700
		45	0.704	21.299	1009.449	26.529	0.366	3.715
		15	0.638	9.416	7887.189	87.655	0.572	13.517
		20	0.620	7.447	12325.767	109.951	0.594	19.986
	gaussian	25	0.671	9.467	7793.617	87.134	0.568	13.071
		30	0.661	8.777	9083.029	94.265	0.569	14.724
websight	<u> </u>	35	0.628	7.462	12277.075	109.751	0.588	19.541
		25	0.626	8.376	9943.443	98.710	0.652	15.744
		30	0.602	7.292	12786.478	111.977	0.664	19.803
	salt pepper	35	0.636	8.723	9183.856	94.819	0.649	14.966
		40	0.656	9.946	7050.120	82.602	0.633	12.060
		45	0.619	8.060	10686.007	102.371	0.655	17.279

Table 12: Performance of our model(IR<sup>4</sup>Net) under Gaussian and Salt & Pepper noise with various SNR levels (dB). Values are reported as mean values rounded to three decimal places. Horizontal rules separate noise types for clarity.

Dataset	Noise Type	SNR (dB)	SSIM↑	<b>PSNR</b> ↑	MSE↓	RMSE↓	LPIPS↓	MS-SSIM↑
chart	gaussian	15	0.688	15.674	2293.165	45.056	0.533	8.311
		20	0.691	15.705	2280.971	44.916	0.533	8.703
		25	0.691	15.715	2276.276	44.866	0.533	8.834
		30	0.692	15.716	2276.042	44.863	0.533	8.869
		35	0.692	15.715	2275.895	44.865	0.533	8.876
	salt pepper	25	0.678	15.437	2401.394	46.204	0.539	8.289
		30	0.681	15.499	2370.929	45.888	0.537	8.397
		35	0.681	15.496	2373.288	45.913	0.537	8.391
		40	0.681	15.481	2376.261	45.963	0.537	8.391
		45	0.681	15.495	2370.964	45.899	0.538	8.408
password		15	0.801	10.953	5293.539	72.512	0.270	0.359
		20	0.836	12.071	4111.432	63.832	0.208	0.148
	gaussian	25	0.848	12.551	3689.944	60.436	0.196	0.155
		30	0.850	12.674	3589.991	59.599	0.194	0.160
		35	0.850	12.676	3591.947	59.604	0.194	0.161
	salt pepper	25	0.659	8.036	10325.171	101.359	0.517	4.148
		30	0.683	8.414	9482.130	97.089	0.486	3.153
		35	0.682	8.407	9497.168	97.163	0.486	3.215
		40	0.681	8.384	9538.681	97.397	0.488	3.216
		45	0.683	8.414	9472.439	97.061	0.486	3.151
screen	gaussian	15	0.701	21.088	1009.091	26.535	0.373	3.428
		20	0.707	21.133	1007.925	26.476	0.374	3.910
		25	0.709	21.145	1007.086	26.465	0.374	4.135
		30	0.709	21.149	1006.137	26.452	0.374	4.193
		35	0.709	21.149	1007.045	26.460	0.374	4.195
	salt pepper	25	0.702	21.298	1007.783	26.532	0.366	3.568
		30	0.704	21.258	1010.903	26.542	0.366	3.699
		35	0.701	21.095	1009.556	26.535	0.366	3.703
		40	0.703	21.095	1010.971	26.548	0.366	3.700
		45	0.704	21.299	1009.449	26.529	0.366	3.715
websight	gaussian	15	0.638	9.416	7887.189	87.655	0.572	13.517
		20	0.620	7.447	12325.767	109.951	0.594	19.986
		25	0.671	9.467	7793.617	87.134	0.568	13.071
		30	0.661	8.777	9083.029	94.265	0.569	14.724
		35	0.628	7.462	12277.075	109.751	0.588	19.541
	salt pepper	25	0.626	8.376	9943.443	98.710	0.652	15.744
		30	0.602	7.292	12786.478	111.977	0.664	19.803
		35	0.636	8.723	9183.856	94.819	0.649	14.966
		40	0.656	9.946	7050.120	82.602	0.633	12.060
		45	0.619	8.060	10686.007	102.371	0.655	17.279

Table 13: Performance of CIDNet model under Gaussian and Salt & Pepper noise with various SNR levels (dB). Horizontal rules are added to separate different noise types for each dataset.

Dataset	Noise Type	SNR (dB)	SSIM↑	<b>PSNR</b> ↑	MSE↓	RMSE↓	LPIPS↓	MS-SSIM↑
chart	gaussian	15	0.662	15.288	2442.810	46.837	0.588	0.469
		20	0.664	15.289	2442.064	46.828	0.588	0.470
		25	0.665	15.289	2442.062	46.828	0.589	0.470
		30	0.665	15.289	2441.872	46.826	0.589	0.470
		35	0.665	15.289	2441.750	46.825	0.589	0.470
	salt pepper	25	0.629	14.642	2941.085	51.041	0.603	0.414
		30	0.633	14.725	2865.547	50.455	0.601	0.421
		35	0.633	14.727	2866.524	50.451	0.601	0.421
		40	0.633	14.725	2869.631	50.470	0.601	0.421
		45	0.634	14.731	2860.082	50.410	0.601	0.422
password	gaussian	15	0.792	10.917	5327.250	72.778	0.268	0.702
		20	0.802	11.012	5213.150	71.992	0.265	0.713
		25	0.806	11.050	5168.650	71.681	0.264	0.717
		30	0.807	11.056	5161.045	71.628	0.264	0.717
		35	0.807	11.060	5156.797	71.599	0.264	0.717
		25	0.761	10.449	5944.950	76.848	0.346	0.626
	salt pepper	30	0.770	10.575	5778.882	75.755	0.329	0.645
		35	0.770	10.583	5766.338	75.675	0.328	0.645
		40	0.770	10.579	5772.561	75.715	0.330	0.644
		45	0.770	10.562	5795.700	75.867	0.330	0.643
screen	gaussian	15	0.531	16.155	2607.021	46.273	0.513	0.489
		20	0.543	16.164	2605.241	46.247	0.512	0.491
		25	0.547	16.168	2604.352	46.235	0.512	0.492
		30	0.548	16.170	2603.673	46.228	0.513	0.492
		35	0.548	16.169	2603.948	46.231	0.513	0.492
	salt pepper	25	0.527	15.908	2638.782	46.674	0.548	0.483
		30	0.531	15.925	2636.481	46.633	0.540	0.484
		35	0.531	15.927	2635.730	46.625	0.541	0.484
		40	0.531	15.931	2633.924	46.613	0.541	0.484
		45	0.531	15.932	2635.638	46.617	0.540	0.485
websight	gaussian	15	0.750	15.678	3850.090	52.363	0.430	0.595
		20	0.762	15.684	3849.236	52.349	0.430	0.596
		25	0.766	15.686	3848.946	52.345	0.430	0.597
		30	0.767	15.687	3848.896	52.344	0.430	0.597
		35	0.767	15.687	3848.868	52.344	0.430	0.597
	salt pepper	25	0.753	15.676	3859.689	52.405	0.426	0.594
		30	0.756	15.679	3858.871	52.395	0.424	0.595
		35	0.756	15.679	3859.115	52.396	0.424	0.595
		40	0.756	15.679	3858.920	52.396	0.425	0.595
		45	0.756	15.679	3858.924	52.396	0.424	0.595

Table 14: Performance of C2PNet model under Gaussian and Salt & Pepper noise with various SNR levels (dB). Horizontal rules are added to separate different noise types for each dataset.

Dataset	Noise Type	SNR (dB)	SSIM↑	<b>PSNR</b> ↑	MSE↓	RMSE↓	LPIPS↓	MS-SSIM↑
chart	gaussian	15	0.696	16.588	1806.016	40.172	0.490	0.573
		20	0.704	16.759	1747.515	39.443	0.487	0.585
		25	0.707	16.818	1727.139	39.193	0.485	0.590
		30	0.708	16.832	1722.569	39.136	0.485	0.591
		35	0.708	16.828	1721.878	39.138	0.485	0.591
	salt pepper	25	0.667	16.045	2000.447	42.522	0.496	0.532
		30	0.674	16.163	1955.533	41.991	0.494	0.542
		35	0.673	16.173	1954.646	41.965	0.495	0.541
		40	0.674	16.179	1945.937	41.900	0.494	0.542
		45	0.674	16.184	1949.674	41.916	0.494	0.543
password	gaussian	15	0.804	11.657	4526.573	66.966	0.209	0.755
		20	0.828	12.446	3790.971	61.215	0.190	0.798
		25	0.840	12.925	3412.466	58.008	0.181	0.819
		30	0.843	13.076	3299.814	57.029	0.178	0.826
		35	0.843	13.073	3304.336	57.059	0.178	0.826
•	salt pepper	25	0.772	10.956	5309.416	72.555	0.271	0.703
		30	0.783	11.195	5032.207	70.616	0.256	0.721
		35	0.783	11.207	5019.571	70.519	0.256	0.721
		40	0.782	11.163	5060.958	70.843	0.257	0.719
		45	0.783	11.191	5035.419	70.642	0.257	0.721
screen	gaussian	15	0.691	21.577	881.163	25.247	0.349	0.710
		20	0.701	21.598	879.713	25.213	0.348	0.710
		25	0.704	21.606	879.065	25.199	0.348	0.711
		30	0.705	21.608	878.769	25.194	0.348	0.711
		35	0.705	21.608	879.055	25.197	0.348	0.711
	salt pepper	25	0.681	21.280	892.819	25.564	0.382	0.704
		30	0.686	21.308	890.678	25.516	0.374	0.705
		35	0.685	21.309	891.003	25.518	0.374	0.705
		40	0.685	21.316	890.951	25.512	0.375	0.705
		45	0.686	21.320	889.921	25.495	0.374	0.705
websight	gaussian	15	0.289	6.797	13928.000	117.393	0.689	0.442
		20	0.262	6.688	14502.166	119.533	0.754	0.428
		25	0.125	8.492	9677.961	97.341	0.819	0.353
		30	0.219	10.417	6542.657	78.783	0.642	0.369
		35	0.360	12.754	5151.496	64.950	0.582	0.419
	salt pepper	25	0.147	6.049	16884.073	128.802	0.808	0.416
		30	0.300	9.792	7373.893	84.169	0.693	0.408
		35	0.071	5.599	18737.056	135.688	0.808	0.405
		40	0.100	5.560	19054.560	136.686	0.767	0.402
		45	0.063	5.156	21122.362	143.680	0.767	0.393

Table 15: Performance of DarkIR method under Gaussian and Salt & Pepper noise with various SNR levels (dB). Values are reported as mean values rounded to three decimal places. Horizontal rules separate noise types for clarity.

Dataset	Noise Type	SNR (dB)	SSIM↑	<b>PSNR</b> ↑	MSE↓	RMSE↓	LPIPS↓	MS-SSIM↑
chart	gaussian	15	0.039	1.061	51774.245	226.646	0.866	0.231
		20	0.011	0.913	53539.983	230.503	0.876	0.110
		25	0.094	1.448	47371.925	216.784	0.843	0.274
		30	0.038	1.059	51790.868	226.687	0.866	0.231
		35	0.011	0.913	53539.873	230.503	0.876	0.111
	salt pepper	25	0.131	1.696	44769.063	210.715	0.838	0.288
		30	0.089	1.411	47794.117	217.728	0.859	0.272
		35	0.043	1.085	51494.270	226.029	0.867	0.236
		40	0.011	0.913	53539.983	230.503	0.876	0.110
		45	0.011	0.913	53539.983	230.503	0.876	0.110
password		15	0.105	0.968	52044.985	228.116	0.784	0.340
		20	0.091	1.133	50111.171	223.838	0.748	0.357
	gaussian	25	0.054	0.734	54926.827	234.346	0.749	0.284
		30	0.091	1.056	51003.151	225.821	0.782	0.350
		35	0.100	0.943	52350.750	228.785	0.778	0.336
1		25	0.060	0.690	55489.232	235.543	0.761	0.258
	salt pepper	30	0.076	0.813	53939.530	232.230	0.757	0.311
		35	0.106	0.974	51975.131	227.963	0.785	0.341
		40	0.090	0.892	52971.265	230.137	0.802	0.327
		45	0.081	0.839	53614.884	231.531	0.786	0.317
screen	gaussian	15	0.196	6.393	23446.930	140.414	0.729	0.349
		20	0.114	6.098	26527.193	149.210	0.715	0.299
		25	0.163	6.182	25439.589	145.921	0.731	0.325
		30	0.147	6.095	25191.974	145.751	0.752	0.324
		35	0.131	6.142	26396.819	148.658	0.704	0.296
	salt pepper	25	0.220	6.596	23569.123	139.666	0.687	0.350
		30	0.197	6.596	23637.809	139.995	0.689	0.351
		35	0.151	6.282	25545.719	145.907	0.695	0.323
		40	0.196	6.497	23871.159	140.838	0.693	0.349
		45	0.167	6.323	24592.989	143.382	0.724	0.341
websight	gaussian	15	0.190	2.377	40709.124	198.837	0.669	0.503
		20	0.092	1.789	46578.809	212.777	0.701	0.447
		25	0.096	1.726	47189.535	214.213	0.754	0.438
		30	0.107	1.790		212.608	0.733	0.449
		35	0.095	1.733	47113.800	214.046	0.750	0.440
	salt pepper	25	0.132	2.061	43767.778	206.224	0.669	0.479
		30	0.158	2.115	43144.857	204.777	0.756	0.486
		35	0.107	1.784	46539.378	212.725	0.672	0.449
		40	0.081	1.702	47488.963	214.890	0.680	0.432
		45	0.098	1.738	47052.677	213.903	0.756	0.441

Table 16: Performance of UNet method under Gaussian and Salt & Pepper noise with various SNR levels (dB). Values are reported as mean values rounded to three decimal places. Horizontal rules separate noise types for clarity.

Such behavior may be attributed to the architectural coupling of structural constraint enforcement and perturbation-resilient feature processing. The model dynamically balances global irradiance trends and localized textural fidelity, suppressing illumination-induced bias without sacrificing high-frequency detail. Multi-scale feature decomposition facilitates low-frequency irradiance smoothing while concurrently preserving high-frequency structural features, enabling temporally consistent recovery under visually non-stationary conditions.

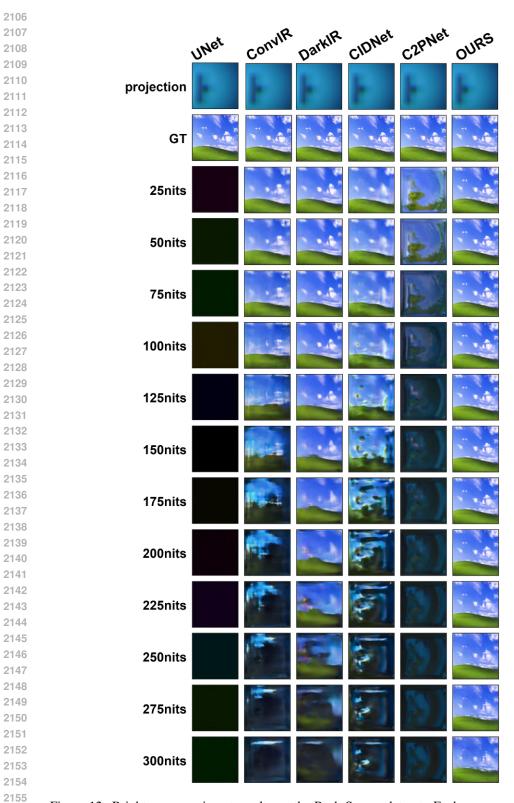


Figure 12: Brightness experiment results on the Resh-Screen dataset. Each row corresponds to a model. The first column shows the projected images, and the second column shows the ground truth (GT). From left to right, the brightness of the subsequent columns gradually decreases, illustrating the performance of the models under different lighting conditions. This figure reveals the models' sensitivity to changes in brightness and demonstrates how the preservation of image details and prediction quality vary as brightness decreases.

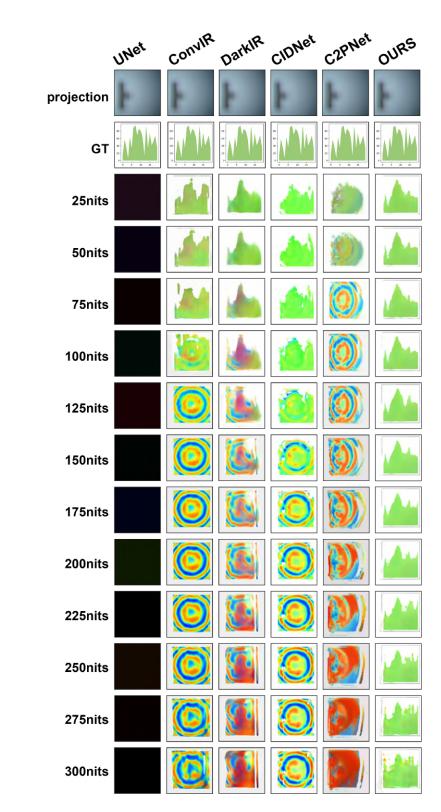


Figure 13: Brightness experiment results on the Resh-Chart dataset. Each row corresponds to a model. The first column shows the projected images, and the second column shows the ground truth (GT). From left to right, the brightness of the subsequent columns gradually decreases, illustrating the performance of the models under different lighting conditions. This figure reveals the models' sensitivity to changes in brightness and demonstrates how the preservation of image details and prediction quality vary as brightness decreases.

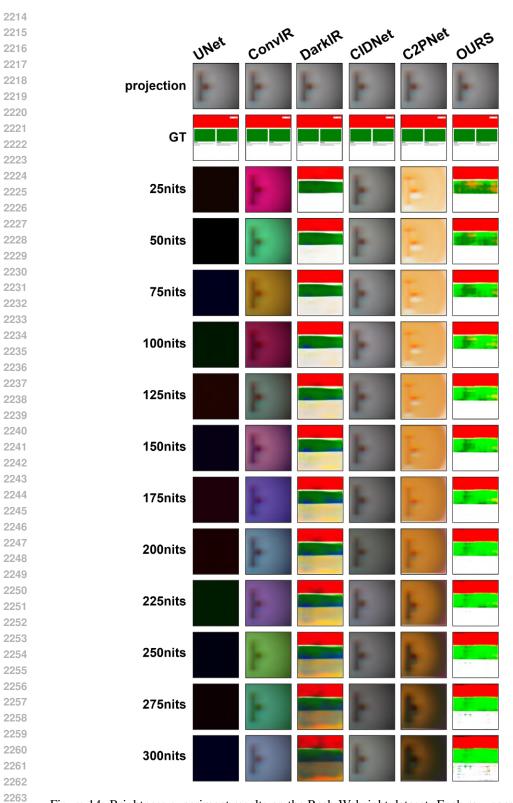


Figure 14: Brightness experiment results on the Resh-Websight dataset. Each row corresponds to a model. The first column shows the projected images, and the second column shows the ground truth (GT). From left to right, the brightness of the subsequent columns gradually decreases, illustrating the performance of the models under different lighting conditions. This figure reveals the models' sensitivity to changes in brightness and demonstrates how the preservation of image details and prediction quality vary as brightness decreases.



Figure 15: Brightness experiment results on the Resh-Password dataset. Each row corresponds to a model. The first column shows the projected images, and the second column shows the ground truth (GT). From left to right, the brightness of the subsequent columns gradually decreases, illustrating the performance of the models under different lighting conditions. This figure reveals the models' sensitivity to changes in brightness and demonstrates how the preservation of image details and prediction quality vary as brightness decreases.

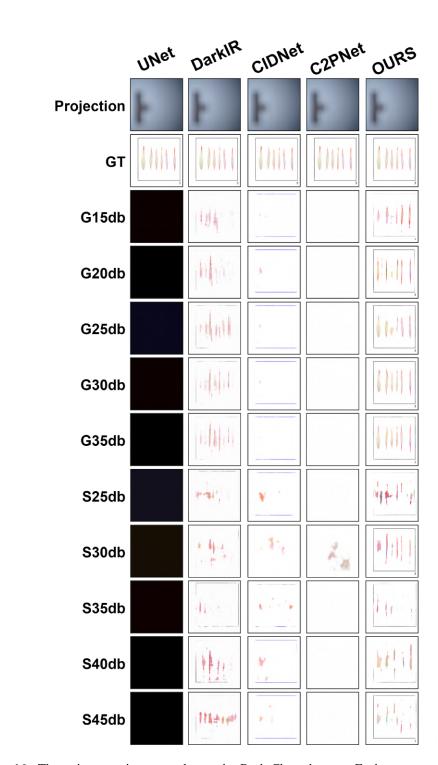


Figure 16: The noise experiment results on the Resh-Chart dataset. Each row corresponds to a model; the first column shows the projected image, the second column shows the ground truth (GT), and the subsequent columns represent the model's performance under different types and levels of injected noise, where G denotes Gaussian noise and S denotes salt-and-pepper noise.

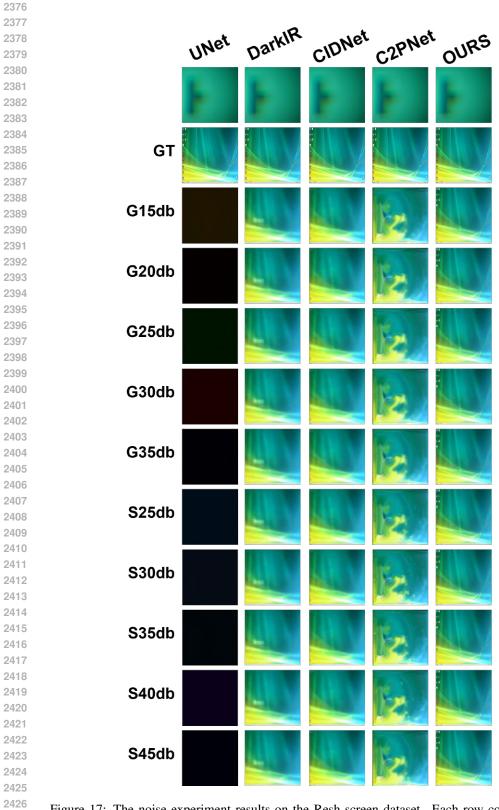


Figure 17: The noise experiment results on the Resh-screen dataset. Each row corresponds to a model; the first column shows the projected image, the second column shows the ground truth (GT), and the subsequent columns represent the model's performance under different types and levels of injected noise, where G denotes Gaussian noise and S denotes salt-and-pepper noise.



Figure 18: The noise experiment results on the Resh-screen dataset. Each row corresponds to a model; the first column shows the projected image, the second column shows the ground truth (GT), and the subsequent columns represent the model's performance under different types and levels of injected noise, where G denotes Gaussian noise and S denotes salt-and-pepper noise.

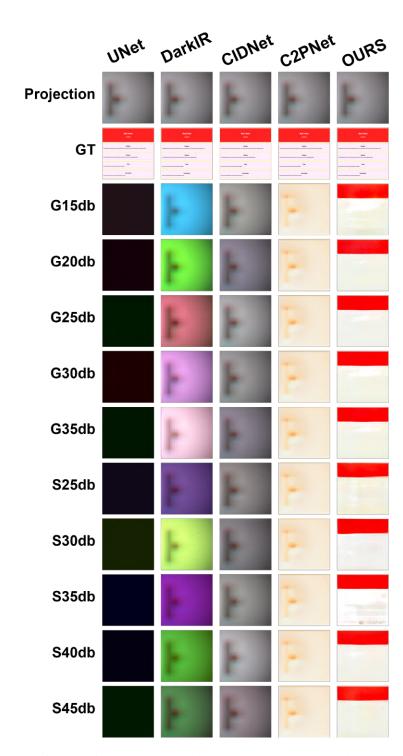


Figure 19: The noise experiment results on the Resh-screen dataset. Each row corresponds to a model; the first column shows the projected image, the second column shows the ground truth (GT), and the subsequent columns represent the model's performance under different types and levels of injected noise, where G denotes Gaussian noise and S denotes salt-and-pepper noise.

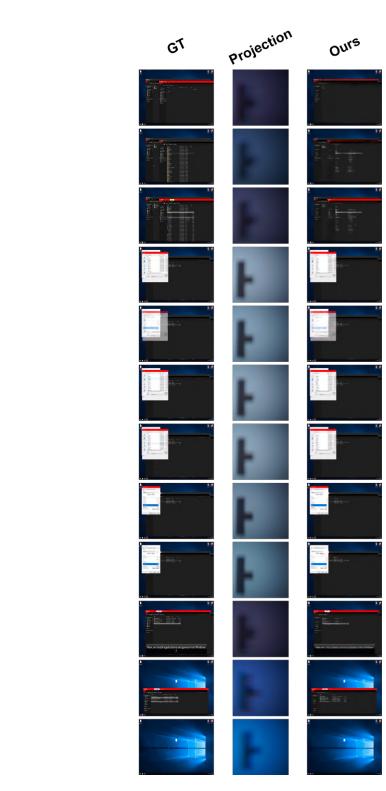


Figure 20: Our model's reconstruction results on video: the first column shows the video frames after sampling, the second column presents the corresponding projected frames, and the third column displays the results reconstructed by our model.