Lumen: Latent Uncertainty-Modulated Evidence Network for Seeing People Around Corners with an Ordinary Camera and Steady Flashlight Illumination

Anonymous authors
Paper under double-blind review

ABSTRACT

Can an ordinary camera viewing a wall under steady flashlight illumination reveal a person beyond an L-shaped corner, providing a view before direct entry for inspection and rescue? We study reconstruction of the hidden person from the RGB light pattern recorded on the visible wall. The flashlight–wall–person–wall–camera path contains three diffuse reflections that attenuate and spatially mix body cues with wall background and noise; different body arrangements can also produce similar steady patterns, making coherent reconstruction ambiguous. We propose LUMEN, whose Path-Calibrated Evidence Routing (PACER) estimates relative cue support and routes multiscale evidence, while Evidence-Conditioned Hypothesis Optimization (ECHO) conditions global body-structure selection and progressive refinement on the routed wall evidence. LUMEN achieves the highest PSNR, SSIM, and MS-SSIM and the lowest LPIPS in all 16 dataset–metric comparisons. In the extreme L-shaped-corner scene where the wall and right-angle turn block the direct view of a person and the camera sees only a wall light pattern from the observation side, using only an ordinary camera and steady flashlight illumination, LUMEN estimates the relative reliability of wall clues from the attenuation and mixing caused by three diffuse reflections, routes evidence across scales, and conditions global selection and progressive refinement of structural hypotheses on this evidence to reconstruct and observe a clear person from extremely weak wall-light information mixed with substantial noise. Across four datasets, LUMEN achieves this goal with higher reconstruction quality than all compared benchmarks and baselines. LUMEN initially enables observation of a target behind a wall using only flashlight illumination. This demonstrates the feasibility of observing a target behind a wall using only flashlight illumination.

1 INTRODUCTION

Can we see a person around a corner by illuminating a wall with a flashlight? We study this question in an L-shaped corridor, where the walls and right-angle turn block the person from direct view and an ordinary camera records only the illuminated wall patch (Figure 1). Recovering the person's location, body outline, and pose from this indirect return would support inspection and rescue before direct entry.

The visible wall is approximately 1.5 m wide, the camera-to-wall distance is approximately 2 m, the person-to-corner distance is approximately 3 m, and the corridor width is approximately 2 m.

The flashlight–wall–person–wall–camera path creates two successive difficulties. Three diffuse reflections weaken and spatially mix the human-dependent return (Chen et al., 2019) against a stronger wall background. Which unevenly supported wall cues can guide body reconstruction? Once weak cues are identified, structural ambiguity remains: different body configurations can produce similar wall patterns. This motivates identifying reliable cues and combining them into one coherent body configuration.

We evaluate Reflect-Corridor (R-C), Reflect-Room (R-R), Reflect-NIR880 (R-N), and Reflect-Garage (R-G), covering corridor, room, near-infrared, and low-light settings.

We propose LUMEN. Path-Calibrated Evidence Routing (PACER) estimates relative cue reliability, aligns information across scales, and separates local residual evidence from wall context. Evidence-Conditioned Hypothesis Optimization (ECHO) uses this evidence to connect distant body cues, screen compatible structural hypotheses, and refine the reconstruction from overall layout to local detail. Czajkowski & Murray-Bruce (2024) reconstructed three-dimensional hidden scenes using two orthogonal occluding edges and a visible matte surface illuminated by the hidden scene. LUMEN places both the flashlight and camera on the observation side of the L-shaped corner.

In the extreme L-shaped-corner scene where the wall and right-angle turn block the direct view of a person and the camera sees only a wall light pattern from the observation side, using only an ordinary camera and steady flashlight illumination, LUMEN estimates the relative reliability of wall clues from the attenuation and mixing caused by three diffuse reflections, routes evidence across scales,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

and conditions global selection and progressive refinement of structural hypotheses on this evidence to reconstruct and observe a clear person from extremely weak wall-light information mixed with substantial noise. Across four datasets, LUMEN achieves this goal with higher reconstruction quality than all compared benchmarks and baselines. LUMEN initially enables observation of a target behind a wall using only flashlight illumination. This demonstrates the feasibility of observing a target behind a wall using only flashlight illumination.

Our main contributions are:

1. We establish the initial feasibility of observing and reconstructing a person beyond an L-shaped corner with an ordinary camera and steady flashlight illumination, recovering connected body extent and pose with leading reconstruction results and metrics over four datasets against current well-known methods.
2. We design Path-Calibrated Evidence Routing (PACER) to estimate relative cue reliability, align multiscale wall evidence, and route local residuals around wall context.
3. We develop Evidence-Conditioned Hypothesis Optimization (ECHO) to couple distant cues globally, screen compatible structural hypotheses, and progressively refine coherent human reconstructions.

2 Related Work

2.1 Transient Measurements

Active methods resolve photon travel time after controlled illumination. Ultrafast measurements, confocal inversion, wave migration, and learned transient representations recover hidden structure (Velten et al., 2012; O'Toole et al., 2018; Lindell et al., 2019; Shen et al., 2021; Li et al., 2023). Isogawa et al. (2020) recover 3D human pose from transient optical measurements. LUMEN reconstructs people from ordinary-camera steady wall observations under flashlight illumination.

2.2 Steady-State Wall Observations

Under ambient or controlled illumination, intensity variations reveal hidden content. Examples include floor penumbras, occluders, steady-state wall observations, and transport modulation (Bouman et al., 2017; Saunders et al., 2019; Yedidia et al., 2019; Chen et al., 2019; Zhang et al., 2025). Ordinary

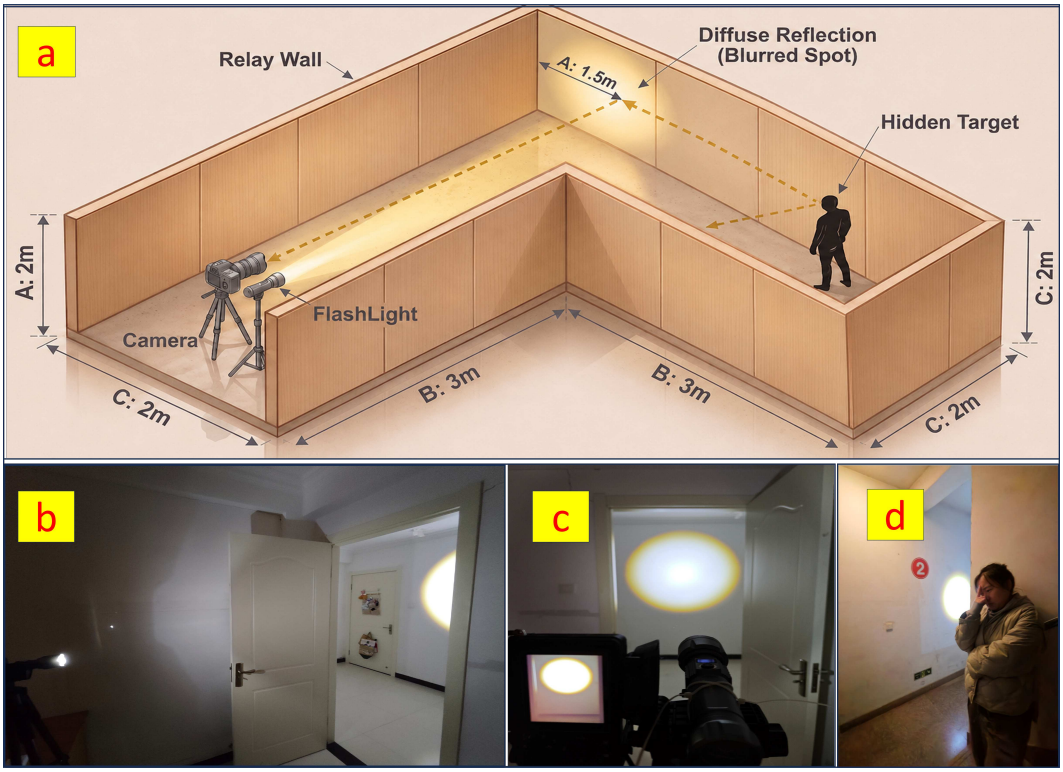

Figure 1: Reconstructing a person beyond an L-shaped corner. (a) The flashlight illuminates the visible wall and the camera records its light pattern while the person stands in the other corridor. (b–d) Real views of corner illumination, the capture apparatus, and the hidden person's position.

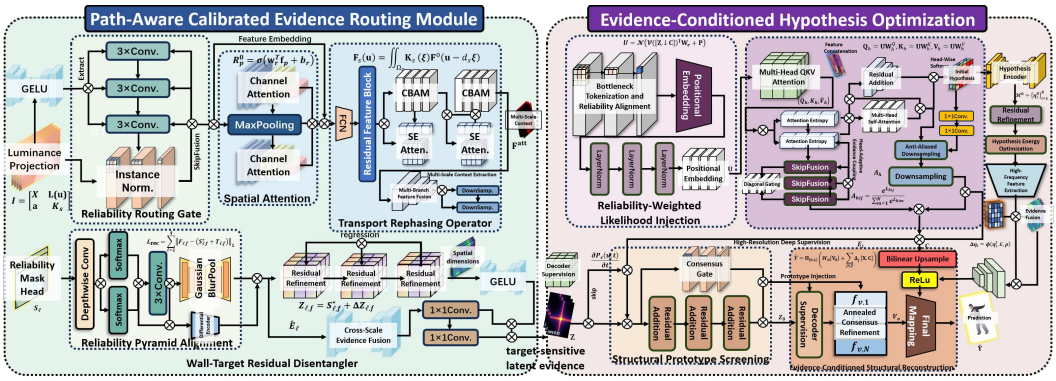

Figure 2: LUMEN architecture. PACER uses the Reliability Routing Gate (RRG), Transport Rephasing Operator (TRO), and Wall-Target Residual Disentangler (WRD) to produce relative reliability, multiscale evidence, and a bottleneck. ECHO applies Reliability-Weighted Likelihood Injection (RLI), reliability-biased global attention, Structural Prototype Screening (SPS), and Annealed Consensus Refinement (ACR) for global coupling and progressive structural readout. Reconstruction supervision is used only during training.

cameras recover hidden geometry with structured edges (Czajkowski & Murray-Bruce, 2024) or localize and reconstruct hidden targets using consumer illumination (Tancik et al., 2018); other work uses multimodal, optimal-transport, propagation-calibration, or light-transport representations (Beckus et al., 2019; Geng et al., 2022; Wang et al., 2023). Blank-wall videos classify count and activity through temporal illumination changes (Sharma et al., 2021); controlled spotlighting supports human localization and posture classification (Chandran et al., 2024). Our inverse-problem baselines span optimal transport, total-variation regularization, and steady-state learned imaging (Geng et al., 2022; Rudin et al., 1992; Chen et al., 2019).

3 Method

3.1 Overall Architecture

Let $\mathbf{X}$ and $\mathbf{Y}$ denote paired wall observations and hidden-person references of size $H \times W$. Training uses both; inference uses only $\mathbf{X}$. Under steady illumination, the wall pattern is

$$\mathbf{X_u} = \mathbf{B_u} + \mathbf{M}_c \mathfrak{F}^{-1}\left[\widehat{\mathbf{H}}_{\mathbf{\Gamma}} \odot \mathfrak{F}\left(\iint_{\Omega_w \times \Omega_h} \mathbf{K_\Gamma L_s} \boldsymbol{\rho_q}\, \mathrm{dq ds}\right)\right] + \boldsymbol{\varepsilon}_s + \boldsymbol{\varepsilon}_q. \tag{1}$$

$\mathbf{M}_c$, $\mathfrak{F}$, and $\widehat{\mathbf{H}}_{\mathbf{\Gamma}}$ denote camera response, Fourier transform, and geometry/scattering response. $\mathbf{K_\Gamma}$ encodes path, angle, reflectance, and visibility; $\boldsymbol{\varepsilon}_s, \boldsymbol{\varepsilon}_q$ capture variation. $\mathbf{B_u}$ is wall-dominated background; Ω_w and Ω_h are wall and hidden-person domains; $\mathbf{L}_s$ and $\boldsymbol{\rho}_q$ describe source illumination and target reflectance. Repeated diffuse reflections attenuate returns, while spatial integration mixes hidden-surface contributions (Chen et al., 2019), leaving weak, ambiguous human evidence in our model (Equation 1). Aittala et al. (2019) recovered hidden videos by factoring visible clutter's shading and mutual shadows into hidden content and light transport.

In LUMEN (Figure 2), PACER, $\mathcal{P}_{\boldsymbol{\theta}}$, maps $\mathbf{X}$ to relative reliability $\mathbf{C}$, multiscale evidence $\{\mathbf{E}_\ell\}$, and bottleneck $\mathbf{Z}$; ECHO, $\mathcal{H}_{\boldsymbol{\phi}}$, maps them to $\widehat{\mathbf{Y}}$. Shared reliability guides routing, global body-structure selection, and refinement. Related designs use multiscale fusion (Dong et al., 2020; Zamir et al., 2022) and iterative refinement (Chun et al., 2023).

3.2 PACER: Reliability Routing and Transport Evidence

Under attenuation, spatial mixing, and scale loss (Equation 1), PACER estimates relative cue support and routes multiscale wall evidence. The Reliability Routing Gate (RRG) combines RGB and fixed luminance $\mathbf{W}_L\mathbf{X}$, defining reliability through embedding, normalization, a skip path, channel response, and neighborhood agreement:

$$\begin{aligned} \mathbf{F}_r &= \Phi_{\boldsymbol{\theta}_r}[\mathbf{X}, \mathbf{W}_L\mathbf{X}] + \mathcal{N}_r(\mathbf{W}_s\mathbf{X}) + \mathcal{S}_r[\mathbf{X}], \\ C_i &= \sigma\left(p_{\boldsymbol{\theta}}(1 \mid \mathbf{f}_i) + \lambda_c[\mathbf{W}_c\mathcal{A}_c(\mathbf{F}_r)]_i + \lambda_s \int_{\Omega_i} \kappa_i(\mathbf{v})\langle \mathbf{W}_q\mathbf{f}_i, \mathbf{W}_k\mathbf{f}_\mathbf{v}\rangle \mathrm{d}\mathbf{v}\right). \end{aligned} \tag{2}$$

Here $\mathbf{f}_i$ is the feature at i in $\mathbf{F}_r$. The posterior assigns features to observation-specific centers; $\mathcal{A}_c$ and κ_i encode channel and neighborhood agreement. $\mathbf{C}$ expresses relative wall-cue support for evidence routing and structural reconstruction.

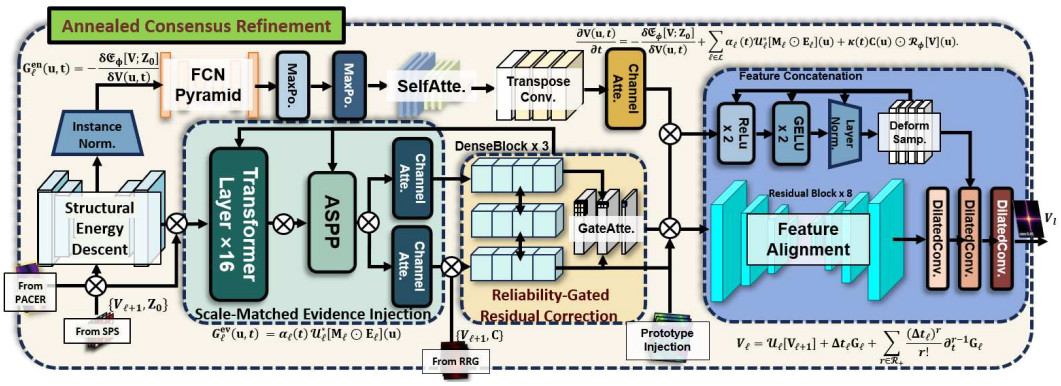

Figure 3: Mechanism of Annealed Consensus Refinement (ACR). The hypothesis state $\mathbf{Z}_0$ from SPS and multiscale evidence $\{\mathbf{E}_\ell\}$ from PACER enter the refinement path. Structural-energy descent sets the update direction; scale-matched evidence injection, consensus-gated residual correction, feature alignment, and the Euler-series realization progressively update $\mathbf{V}_{\ell+1}$ to $\mathbf{V}_\ell$ before final mapping. The reliability field $\mathbf{C}$ gates residual correction, while reconstruction supervision is used only during training. The displayed flow corresponds to Equation 7.

The Transport Rephasing Operator (TRO) carries routed content across antialiased scales. The Wall-Target Residual Disentangler (WRD) extracts local residual cues against wall context and fuses aligned scales. For scale ℓ, $\widehat{\mathbf{h}}_\ell$ is the fixed BlurPool response, $\mathcal{D}_\ell$ learned sampling, and $\mathcal{C}_\ell, \mathcal{S}_\ell$ the branches in Figure 2:

$$\begin{aligned}\mathbf{T}_{\ell+1} &= \mathbf{B}_{\ell+1}\mathcal{D}_\ell\mathfrak{F}^{-1}[\widehat{\mathbf{h}}_\ell \odot \mathfrak{F}(\mathbf{T}_\ell)] + \mathcal{C}_\ell[\mathbf{T}_\ell] + \mathcal{S}_\ell[\mathbf{T}_\ell],\\ \mathbf{E}_\ell &= \mathbf{T}_\ell + \tanh(\alpha_\ell)(\mathbf{I} - \mathbf{A}_\ell)\mathbf{T}_\ell \odot \mathcal{I}_\ell[\mathbf{C}] + \sum_{r\in\mathcal{N}_\ell} \mathbf{W}_{\ell r}\mathcal{U}_{\ell r}[\mathbf{E}_r].\end{aligned} \tag{3}$$

The fused antialiased, reliability-gated features preserve body layout and local structure, forming the PACER–ECHO interface $(\mathbf{C}, \{\mathbf{E}_\ell\}, \mathbf{Z})$.

3.3 ECHO: Global Coupling and Structured Readout

ECHO globally couples routed cues, screens hypotheses, and progressively refines coherent body structure. Reliability-Weighted Likelihood Injection (RLI) tokenizes $[\mathbf{Z}, \downarrow \mathbf{C}]$, adds position, and produces $\mathbf{U}$:

$$\mathbf{U} = \mathcal{N}\Big(\mathcal{V}([\mathbf{Z}, \downarrow \mathbf{C}])^\mathsf{T}\,\mathbf{W}_e + \mathbf{P}\Big). \tag{4}$$

Here $\mathcal{V}$ tokenizes, $\mathbf{W}_e$ embeds, and $\mathbf{P}$ adds position. $\mathbf{c}$ pools reliability $\mathbf{C}$ over tokens of $\mathbf{U}$. Per head h, $\mathcal{S}$ is row-wise normalized exponentiation, $\mathfrak{H}_h$ an attention-entropy functional, $\mathfrak{D}$ diagonal embedding, and d the head dimension:

$$\begin{aligned}\mathbf{A}_h &= \mathcal{S}\left(\frac{(\mathbf{U}\mathbf{W}_h^Q)(\mathbf{U}\mathbf{W}_h^K)^\mathsf{T}}{\sqrt{d}} + \lambda_h \log(\mathbf{c}\mathbf{c}^\mathsf{T} + \epsilon_h) - \beta_h\mathfrak{D}(\mathfrak{H}_h[\mathbf{U}])\right),\\ \widetilde{\mathbf{U}} &= \mathbf{U} + \sum_{h\in\mathcal{H}} \mathbf{A}_h(\mathbf{U}\mathbf{W}_h^V)\mathbf{W}_h^O + \mathcal{R}[\mathcal{N}(\mathbf{U})].\end{aligned} \tag{5}$$

Attention aggregates correlations (Li et al., 2023); LUMEN's reliability bias favors supported wall cues, while learned compatibility combines distant body features.

Structural Prototype Screening (SPS) reshapes $\widetilde{\mathbf{U}}$ as $\widetilde{\mathbf{Z}}$ for comparison with learned prototypes. Let $\mathbf{q}_n$ be a normalized pooled query, $\mathbf{P}_{\text{proto}}$ the normalized prototype matrix, and $\mathcal{I}_n$ its top-K set. The simplex and indicator enforce nonnegative weights on this set:

$$\begin{aligned}\boldsymbol{\pi}_n &= \operatorname*{arg\,min}_{\boldsymbol{\pi}\in\Delta_M}\left\{-\boldsymbol{\pi}^\mathsf{T}\mathbf{P}_{\text{proto}}^\mathsf{T}\mathbf{q}_n + \tau_p\sum_{m\in\mathcal{I}_n}\pi_m\log\pi_m + \iota_{\Delta(\mathcal{I}_n)}[\boldsymbol{\pi}]\right\},\\ \mathbf{Z}_0 &= \widetilde{\mathbf{Z}} + \tanh(\gamma_p)\mathbf{W}_p\mathcal{R}_p\left[\left\{\sum_{m\in\mathcal{I}_n}\pi_{nm}\mathbf{p}_m\right\}_n\right] + \sum_{r\in\mathcal{R}}\beta_r\mathcal{K}_r[\widetilde{\mathbf{Z}}].\end{aligned} \tag{6}$$

Codebooks supply reusable visual parts (Zhou et al., 2022); LUMEN's SPS selects observation-dependent prototype mixtures, guiding hypotheses while its residual connection retains wall evidence.

Annealed Consensus Refinement (ACR) refines body extent and pose with scale-matched wall evidence and reliability-gated residual corrections, progressing toward local detail (Figure 3). Here $\mathfrak{E}_\phi$ is the hypothesis energy, $\mathcal{U}_\ell^*$ the aligned evidence operator, $\mathbf{M}_\ell$ the consensus gate, and $\alpha_\ell(t)$ the coarse-to-fine schedule:

$$\frac{\partial \mathbf{V}(\mathbf{u},t)}{\partial t} = -\frac{\delta \mathfrak{E}_{\boldsymbol{\phi}}[\mathbf{V};\mathbf{Z}_0]}{\delta \mathbf{V}(\mathbf{u},t)} + \sum_{\ell\in\mathcal{L}} \alpha_\ell(t)\mathcal{U}_\ell^*[\mathbf{M}_\ell \odot \mathbf{E}_\ell](\mathbf{u}) + \kappa(t)\mathbf{C}(\mathbf{u}) \odot \mathcal{R}_{\boldsymbol{\phi}}[\mathbf{V}](\mathbf{u}),$$
$$\mathbf{V}_\ell = \mathcal{U}_\ell[\mathbf{V}_{\ell+1}] + \Delta t_\ell \mathbf{G}_\ell[\mathbf{V}_{\ell+1}, \mathbf{E}_\ell, \mathbf{C}] + \sum_{r\in\mathcal{R}_+} \frac{(\Delta t_\ell)^r}{r!}\, \partial_t^{r-1}\mathbf{G}_\ell. \tag{7}$$

The truncated Euler realization uses learned $\mathbf{G}_\ell$ and retained orders $\mathcal{R}_+$; Final Mapping produces the human image.

The RGB head applies bounded residual and high-frequency corrections to $\mathbf{V}_0$. With $\mathbf{e} = \widehat{\mathbf{Y}} - \mathbf{Y}$, $\mathfrak{S}$ the structural functional, and $\mathcal{T}_{\boldsymbol{\Gamma}}$ the transport operator from Equation 1, the output and training objective are

$$\widehat{\mathbf{Y}} = \Pi_{[0,1]}\left\{ \mathcal{H}_0[\mathbf{V}_0] + \sum_{j\in\mathcal{J}} \Delta_j[\mathbf{X}, \mathbf{O}^{(j)}, \mathbf{C}] + \eta_f \mathfrak{F}^{-1}[\widehat{\mathbf{H}}_f \odot \mathfrak{F}(\mathcal{H}_0[\mathbf{V}_0])] \right\},$$
$$\boldsymbol{\Theta}^\star = \arg\min_{\boldsymbol{\Theta}} \mathbb{E}_{(\mathbf{X},\mathbf{Y})\sim\mathcal{D}}\left[\|\mathbf{e}\|_1 + \lambda_s(1 - \mathfrak{S}[\widehat{\mathbf{Y}}, \mathbf{Y}]) + \lambda_g \int_\Omega \|\nabla \mathbf{e}\|_1 \mathrm{d}\mathbf{u} + \lambda_r \|\mathbf{M}_c \mathcal{T}_{\boldsymbol{\Gamma}}[\widehat{\mathbf{Y}}] - \mathbf{X}\|_1 \right]. \tag{8}$$

Equation 8 combines reconstruction, structural, gradient, and transport consistency. Paired references and decoder-side supervision guide training only; inference follows $\mathbf{X} \to (\mathbf{C}, \{\mathbf{E}_\ell\}, \mathbf{Z}) \to \widetilde{\mathbf{U}} \to \mathbf{Z}_0 \to \mathbf{V}_0 \to \widehat{\mathbf{Y}}$.

4 Experiment

4.1 Experimental Setup

R-C, R-R, and R-G use visible light; R-N uses 880-nm acquisition. The camera and steady source are co-located on the visible side of the L-shaped corner; a second camera records direct-view human references. R-C and R-R cover corridor- and room-scale multipath; R-G covers a low-light, long-path garage.

Before frame extraction, we partition intact subject, sequence, and recording-session groups.

Cameras share trial triggers and timestamps; individual intrinsic calibration, a common target, and fixed landmarks establish extrinsics and a shared canvas. Before resizing or normalization, raw wall/reference pairs retain subject, sequence, session, and frame identifiers. Camera/flashlight placement, corner geometry, wall, distances, and angles are fixed except for specified physical factors. PSNR, SSIM, MS-SSIM, and LPIPS score complete aligned images. Comparisons cover convolutional, Transformer, state-space, generative, and unrolled models, plus NLOS-OT (Geng et al., 2022), ROF-TV (Rudin et al., 1992), and SteadyStateNLOS (Chen et al., 2019). Here, NLOS-RescueNet denotes the model of Mu et al. (2025). Pointwise baseline envelopes select the best competitor.

4.2 Comparison with Baseline Methods

Using the same wall observations as eight representative baselines, LUMEN reconstructs the person's location and body configuration (Figure 4). Leaning torsos and arms remain separated on R-C, upright and crouched structures remain distinguishable on R-R, arm span and clothing layout survive on R-N, and coarse bodies remain localized on R-G. Competing outputs more often duplicate contours, merge limbs, over-smooth bodies, or collapse toward low-frequency blobs. Sharma et al. (2021) classify hidden count and activity from blank-wall videos; we reconstruct location, connected body outline, and pose from wall light.

Pixel, structural, and perceptual agreement also reflect recovered body structure (Table 1). LUMEN strictly leads all 16 displayed protocol–metric cells. Its PSNR gains over the strongest baseline are 4.88, 1.83, 1.15, and 12.03 dB on R-C, R-R, R-N, and R-G; SSIM gains are 0.15, 0.06, 0.04, and 0.15. LPIPS is lower by 0.03, 0.01, 0.01, and 0.17, while MS-SSIM improves by 0.20, 0.05, 0.01, and 0.41, respectively. On R-G, the strongest PSNR baselines are ConvIR and PGDNet (15.69 dB), while NLOS-OT reaches the strongest baseline MS-SSIM (0.54 versus LUMEN's 0.95). R-G gains are largest under weak, mixed wall observations; R-N extends improvement to 880-nm acquisition with smaller margins.



Figure 4: Human reconstructions for two examples each from R-C, R-R, R-N, and R-G. Each row presents the wall input, eight baseline predictions, LUMEN, and the direct-view reference. Red frames mark LUMEN; black frames mark the other cells.

Type	Method	Venue	R-C				R-R				R-N				R-G			
			P	S	L	M	P	S	L	M	P	S	L	M	P	S	L	M
Transformer	Restormer [37]	CVPR'22	14.18	0.68	0.22	0.61	14.31	0.77	0.15	0.73	11.07	0.81	**0.16**	0.80	13.55	0.62	0.53	0.40
	NLOST [20]	CVPR'23	12.23	**0.71**	0.33	0.37	12.17	0.71	0.26	0.60	23.30	0.80	0.21	0.75	13.34	0.59	0.53	0.36
	SwinIR [21]	ICCVW'21	10.92	0.70	0.34	0.34	9.59	0.68	0.53	0.39	20.22	0.65	0.46	0.57	14.91	0.75	0.48	0.46
U-Net	MSBDN-DFF [12]	CVPR'20	11.68	0.71	0.33	0.37	13.62	0.77	0.17	0.70	19.76	0.82	0.17	0.80	13.63	0.64	0.44	0.39
	ConvIR [9]	TPAMI'24	14.20	0.68	0.22	0.62	13.44	0.75	0.19	0.69	24.16	**0.85**	0.16	**0.81**	**15.69**	**0.78**	**0.29**	0.45
	U-Net [26]	MICCAI'15	14.35	0.68	0.25	0.61	15.26	0.80	0.15	0.76	24.91	0.84	0.17	0.80	13.42	0.62	0.50	0.41
	DarkIR [13]	CVPR'25	13.27	0.48	0.39	0.47	13.12	0.70	0.22	0.66	23.72	0.84	0.17	0.80	13.34	0.58	0.54	0.38
	DnCNN [39]	TIP'17	12.51	0.69	0.31	0.43	11.62	0.71	0.23	0.58	25.40	0.66	0.37	0.55	14.22	0.71	0.37	0.30
	AST [42]	CVPR'24	13.80	0.65	0.25	0.54	13.97	0.77	0.16	0.71	25.01	0.84	0.16	0.80	13.61	0.61	0.55	0.39
	C2PNet [40]	CVPR'23	12.24	0.69	0.33	0.37	11.32	0.70	0.25	0.56	20.30	0.84	0.18	0.78	13.52	0.62	0.50	0.38
Diffusion	DiffIR [35]	ICCV'23	12.80	0.63	0.38	0.48	13.17	0.73	0.21	0.67	22.64	0.84	0.18	0.80	13.75	0.63	0.50	0.40
Unrolled	DGUNet [23]	CVPR'22	14.25	0.69	0.23	0.60	13.80	0.77	0.15	0.71	24.93	0.80	0.16	0.80	14.29	0.64	0.52	0.41
VAE	IntroVAE [17]	NeurIPS'18	14.21	0.68	**0.21**	**0.64**	**16.29**	**0.82**	**0.13**	**0.80**	22.02	0.84	0.16	0.81	13.77	0.64	0.42	0.40
Codebook	CodeFormer [41]	NeurIPS'22	11.93	0.61	0.50	0.43	11.32	0.58	0.51	0.49	23.66	0.80	0.39	0.80	15.03	0.69	0.45	0.46
Unrolled	PGDNet [5]	MICCAI'20	14.00	0.69	0.23	0.58	12.93	0.73	0.21	0.65	24.93	0.82	0.16	0.81	15.69	0.78	0.29	0.45
Newton	DeepSN-Net [11]	TPAMI'25	13.87	0.56	0.34	0.58	13.63	0.59	0.33	0.65	21.15	0.83	0.18	0.80	15.12	0.68	0.49	0.44
Unrolled	Momentum-Net [7]	TPAMI'23	14.13	0.69	0.22	0.60	14.18	0.77	0.16	0.72	22.15	0.81	0.16	0.81	14.05	0.64	0.42	0.40
NLOS	VS-NLOS [8]	NeurIPS'24	9.00	0.63	0.37	0.49	12.80	0.73	0.21	0.65	25.65	0.84	0.17	0.79	13.71	0.62	0.54	0.41
NLOS	NLOS-RescueNet [24]	TPAMI'25	12.90	0.58	0.46	0.41	11.14	0.69	0.26	0.55	23.43	0.84	0.17	0.80	13.27	0.61	0.49	0.38
Inverse	NLOS-OT [14]	TIP'22	**14.52**	0.71	0.27	0.62	14.51	0.79	0.19	0.73	26.49	0.82	0.25	0.78	15.03	0.73	0.29	**0.54**
	ROF-TV [27]	Physica D'92	12.63	0.67	0.51	0.40	10.05	0.67	0.55	0.32	22.99	0.76	0.34	0.71	14.93	0.75	0.49	0.47
	SteadyStateNLOS [6]	CVPR'19	14.17	0.68	0.45	0.56	14.45	0.74	0.31	0.70	23.76	0.69	0.57	0.78	14.85	0.69	0.60	0.50
Mamba	MambaIR [16]	ECCV'24	12.97	0.68	0.30	0.49	11.19	0.63	0.47	0.41	26.96	0.84	0.21	0.80	13.82	0.65	0.44	0.39
	VMambaIR [31]	TCSVT'25	13.92	0.63	0.44	0.42	12.61	0.75	0.19	0.65	**27.10**	0.84	0.17	0.78	13.32	0.63	0.48	0.35
	MambaIRv2 [15]	CVPR'25	11.68	0.71	0.33	0.37	10.39	0.49	0.59	0.34	26.51	0.82	0.35	0.78	13.30	0.63	0.44	0.39
	FreqMamba [43]	ACM MM'24	10.70	0.67	0.58	0.37	10.19	0.61	0.39	0.47	26.33	0.80	0.37	0.74	13.95	0.66	0.44	0.36
	MaIR [19]	CVPR'25	11.36	0.56	0.56	0.32	10.26	0.53	0.62	0.36	17.27	0.60	0.54	0.54	13.85	0.64	0.51	0.41
Ours	LUMEN	–	**19.40**	**0.86**	**0.18**	**0.84**	**18.12**	**0.88**	**0.12**	**0.85**	**28.25**	**0.89**	**0.15**	**0.82**	**27.72**	**0.93**	**0.12**	**0.95**

Table 1: Comparison with 27 baseline methods on R-C, R-R, R-N, and R-G. Red shading marks LUMEN; blue marks a representative strongest baseline (IntroVAE for R-C/R-R; ConvIR for R-N/R-G). P/S/L/M denote PSNR, SSIM, LPIPS, and MS-SSIM. Each metric marks one red and one blue value; ties favor LUMEN, then table order. P/S/M are higher-is-better; L is lower-is-better.

4.3 Ablation Study

Ablations test which wall evidence to retain and how to assemble a coherent body. Table 2 compares the complete model and seven component-removal variants through pixel, structural, perceptual, and gradient agreement.

Protocol	Stage	Configuration	Mechanism	PSNR	SSIM	LPIPS	MS-SSIM	GMSD	Endpoint change
R-C	PACER	w/o PACER	PACER routing	12.61	**0.78**	0.30	0.46	**0.29**	58.56
		w/o TRO	TRO transport	12.70	0.77	0.30	0.45	0.29	60.04
		w/o WRD	WRD residuals	14.38	0.77	0.24	**0.63**	0.29	41.77
		w/o RRG	RRG reliability	14.37	0.77	**0.23**	0.63	0.29	**41.21**
	ECHO	w/o RLI	RLI coupling	14.33	0.77	0.23	0.63	0.29	41.34
		w/o SPS	SPS support	14.18	0.78	0.25	0.61	0.29	42.79
		w/o ACR	ACR refinement	**14.41**	0.77	0.24	0.63	0.29	41.68
	Joint	Full	Complete	**19.40**	**0.86**	**0.18**	**0.84**	**0.23**	**0.00**
R-R	PACER	w/o PACER	PACER routing	10.94	0.79	0.24	0.55	0.26	53.89
		w/o TRO	TRO transport	10.97	0.79	0.23	0.55	0.26	53.24
		w/o WRD	WRD residuals	16.73	0.86	0.15	0.81	0.23	10.25
		w/o RRG	RRG reliability	**17.42**	**0.87**	**0.12**	**0.83**	**0.22**	**4.29**
	ECHO	w/o RLI	RLI coupling	16.62	0.86	0.13	0.81	0.23	9.43
		w/o SPS	SPS support	13.73	0.83	0.19	0.71	0.25	29.93
		w/o ACR	ACR refinement	16.72	0.86	0.15	0.81	0.23	10.28
	Joint	Full	Complete	**18.12**	**0.88**	**0.12**	**0.85**	**0.22**	**0.00**
R-N	PACER	w/o PACER	PACER routing	26.80	0.82	0.24	0.79	0.16	20.35
		w/o TRO	TRO transport	26.79	0.82	0.24	0.79	0.16	19.97
		w/o WRD	WRD residuals	26.92	**0.84**	0.23	0.77	0.16	17.98
		w/o RRG	RRG reliability	26.72	0.80	**0.20**	0.79	**0.15**	17.99
	ECHO	w/o RLI	RLI coupling	**27.99**	0.79	0.20	0.78	0.15	17.71
		w/o SPS	SPS support	27.53	0.82	0.22	**0.80**	0.15	**16.05**
		w/o ACR	ACR refinement	26.41	0.81	0.23	0.74	0.15	24.14
	Joint	Full	Complete	**28.25**	**0.89**	**0.15**	**0.82**	**0.15**	**0.00**
R-G	PACER	w/o PACER	PACER routing	15.23	**0.85**	0.22	0.44	0.26	56.13
		w/o TRO	TRO transport	15.26	0.85	0.22	0.44	0.26	56.08
		w/o WRD	WRD residuals	15.27	0.85	0.22	0.44	0.26	56.07
		w/o RRG	RRG reliability	**17.80**	0.85	**0.20**	**0.71**	**0.24**	**38.04**
	ECHO	w/o RLI	RLI coupling	15.76	0.84	0.23	0.59	0.24	49.49
		w/o SPS	SPS support	15.54	0.85	0.23	0.48	0.26	54.27
		w/o ACR	ACR refinement	15.23	0.85	0.22	0.44	0.26	56.13
	Joint	Full	Complete	**27.72**	**0.93**	**0.12**	**0.95**	**0.17**	**0.00**

Table 2: Component ablation across four protocols using five endpoint metrics. Higher PSNR, SSIM, and MS-SSIM are better; lower LPIPS and GMSD are better. Values use two decimals.

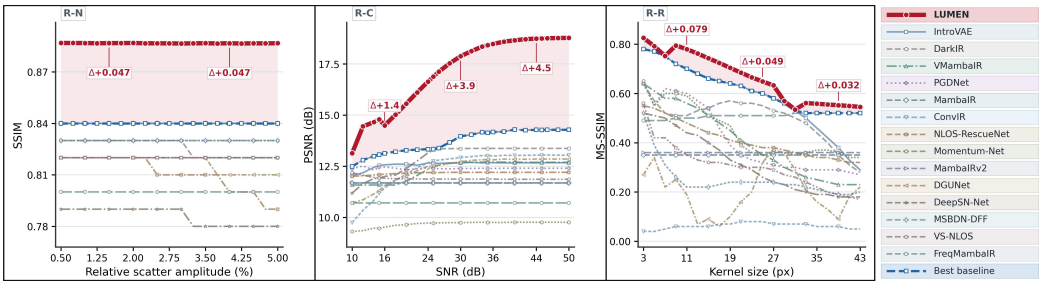

Figure 5: Reconstruction responses to single-variable observation perturbations. From left to right: R-N SSIM versus scatter amplitude, R-C PSNR versus SNR, and R-R MS-SSIM versus motion-blur kernel size. Higher SNR indicates weaker Gaussian noise. Red denotes LUMEN and dashed blue the strongest evaluated baseline at each condition.

PACER removals test evidence formation and transport alignment; ECHO removals test global coupling, screening, and refinement. The normalized endpoint-change scale is protocol-specific.

Removing PACER reduces R-C PSNR from 19.40 to 12.61 dB and R-G MS-SSIM from 0.95 to 0.44, supporting wall-evidence organization before human decoding. Routing or transport removal causes the largest degradation in R-C, R-R, and R-G. WRD and RRG alter residual and reliability paths, with regime-dependent effects.

ECHO's reliability coupling, structural screening, and multiscale refinement have protocol-dependent effects. ACR is most costly in R-N and ties the largest change in R-G. Removing SPS reduces R-R PSNR from 18.12 to 13.73 dB; removing ACR lowers R-N MS-SSIM from 0.82 to 0.74. These effects support complementary structural selection and refinement after evidence formation.

The complete model leads PSNR, SSIM, and MS-SSIM in every dataset block, supporting faithful, coherent body reconstruction.

4.4 Robustness under Noise and Physical Perturbations

Robustness tests use zero-shot evaluation with frozen source-trained checkpoints. Single-variable noise and physical sweeps use about 25 measured nodes each. Each paired study directly evaluates all methods on $25 \times 25 = 625$ joint conditions, and Figures 7 and 8 display the complete measured grids. Perturbed observations, preprocessing, samples, and evaluation are shared across methods.

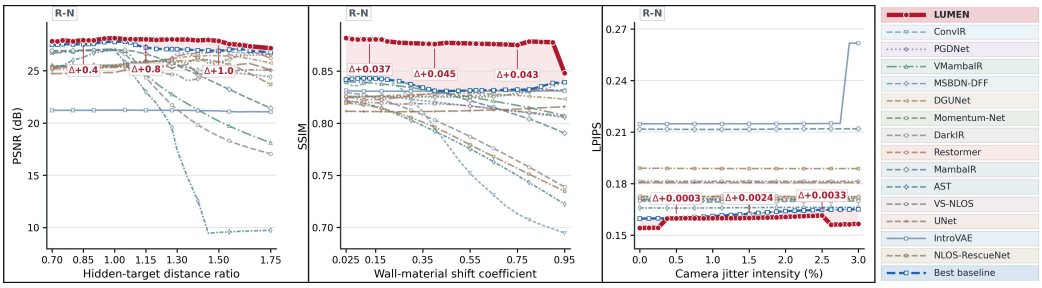

Figure 6: R-N reconstruction under changes in hidden-target distance ratio, wall-material shift coefficient, and camera-jitter intensity, from left to right. The panels report PSNR, MS-SSIM, and LPIPS, respectively. Red denotes LUMEN and dashed blue the strongest evaluated baseline at each condition.

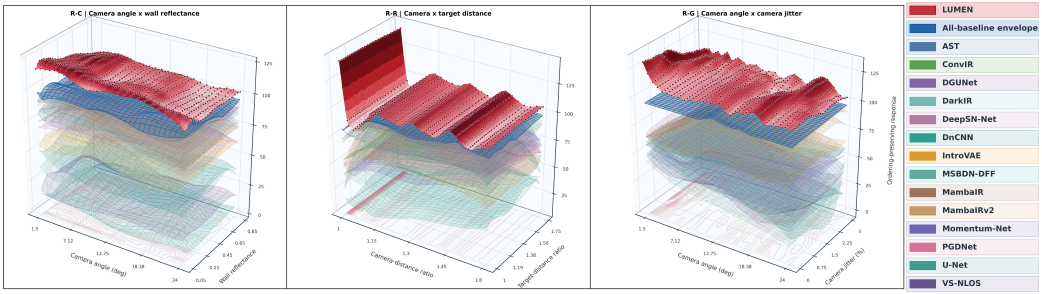

Figure 7: Paired physical perturbations. From left to right: camera angle with wall material on R-C, camera distance with hidden-target distance on R-R, and camera angle with wall material on R-G. Height is the weighted score S; red denotes LUMEN.

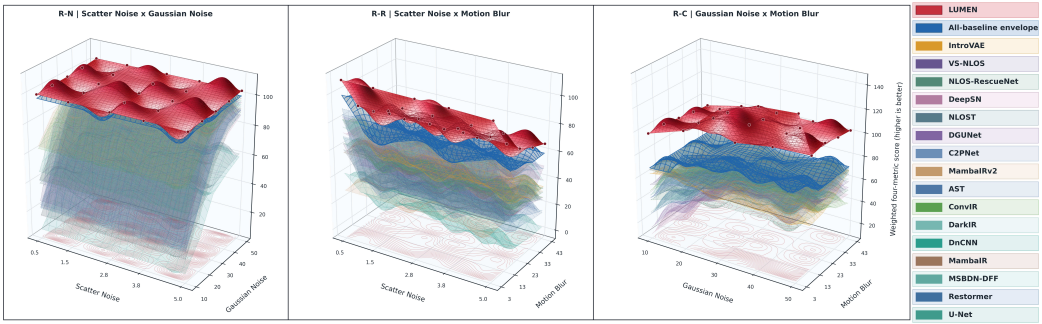

Figure 8: Paired noise perturbations. From left to right: scatter noise with Gaussian noise on R-N, scatter noise with motion blur on R-R, and Gaussian noise with motion blur on R-C. Height is the weighted score S; red denotes LUMEN.

LUMEN preserves structure as separate scatter noise, Gaussian noise, and motion blur weaken or redistribute wall cues (Figure 5).

Across the three R-N sweeps in Figure 6 and camera distance/angle, LUMEN maintains mean SSIM 0.87–0.88 and MS-SSIM 0.80–0.81 over all five factors.

Each of the three physical pairs in Figure 7 uses the full $25 \times 25 = 625$-condition grid, which is displayed in full.

The noise pairs in Figure 8 share measured coordinates and evaluation, and the figure displays the complete $25 \times 25 = 625$-condition grid for each pair. Height is the weighted score $S = 100\Big(0.30\widehat{\text{PSNR}} + 0.25\widehat{\text{SSIM}} + 0.25\widehat{\text{MS-SSIM}} + 0.20(1 - \widehat{\text{LPIPS}})\Big)$; larger S is better. LUMEN leads mean SSIM and MS-SSIM across all six full 625-node grids (appendix), extending structural recovery to simultaneous geometry, wall-response, or noise changes.

4.5 Cross-Dataset Generalization

Table 3 compares LUMEN and eight baselines on four zero-shot routes with frozen source-trained checkpoints; target references serve only for scoring.

LUMEN leads all four metrics on all four transfer routes. Zhang et al. (2025) study hidden-display recovery with light-transport modulation. This transfer advantage is consistent with relative wall-cue support conditioning structural selection and refinement.

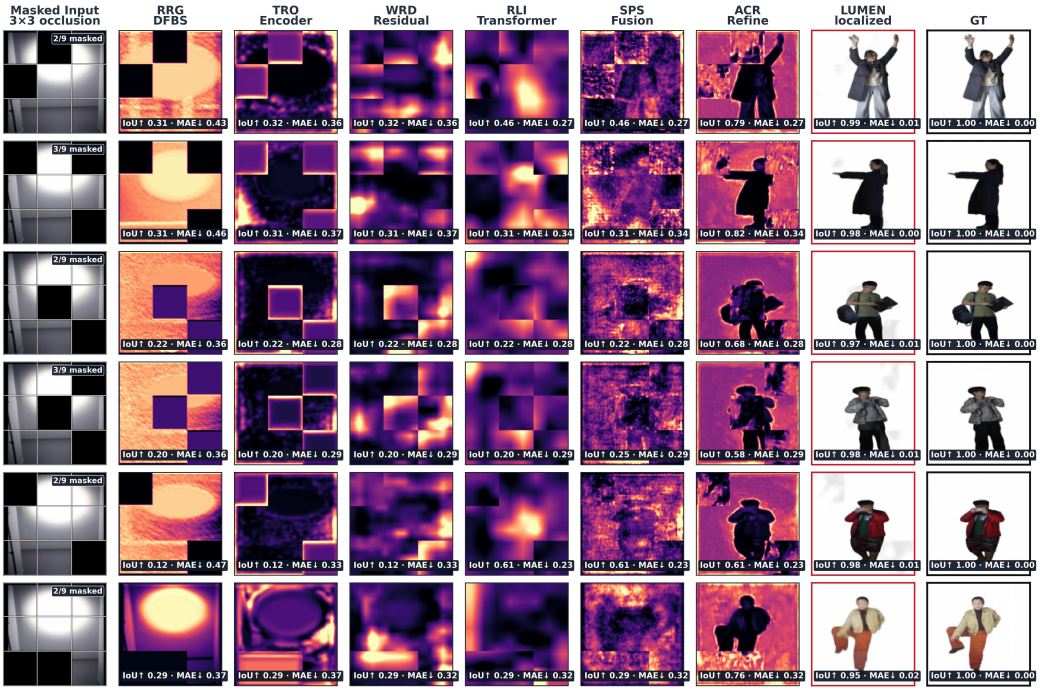

Figure 9: Structural responses to random masking of 2–3 cells in a 3×3 wall-light grid. Rows show the masked input and scoring reference; columns trace RRG, TRO, WRD, RLI, SPS, ACR, localized LUMEN, and GT, visualizing PACER-to-ECHO propagation of remaining cues. Localized visualization annotations report IoU (higher is better) and MAE (lower is better): 0.95–0.99/0.00–0.02 for LUMEN and $1.00/0.00$ for GT.

Method	R-R → R-C				R-R → R-N				R-G → R-R				R-C → R-G			
	P	S	L	M	P	S	L	M	P	S	L	M	P	S	L	M
ConvIR [9]	10.15	0.61	0.37	0.26	2.32	0.30	0.66	**0.42**	9.32	**0.74**	**0.30**	0.35	12.47	0.70	0.33	0.19
DGUNet [23]	9.58	0.61	0.34	0.30	**3.63**	0.30	0.65	0.26	9.43	0.38	0.65	0.16	13.40	0.63	0.51	0.37
IntroVAE [17]	9.66	0.61	0.37	0.28	3.03	0.29	0.66	0.30	**10.07**	0.58	0.47	0.19	11.93	0.59	0.41	0.29
MambaIR [16]	**11.68**	**0.71**	0.33	0.37	2.32	0.30	0.66	0.42	7.81	0.60	0.72	**0.36**	**15.13**	**0.78**	**0.29**	**0.45**
Momentum-Net [7]	10.94	0.65	**0.32**	**0.39**	3.10	0.28	0.68	0.23	9.34	0.51	0.54	0.08	10.36	0.63	0.35	0.20
MSBDN-DFF [12]	9.54	0.60	0.39	0.16	3.43	**0.31**	**0.64**	0.26	8.99	0.50	0.68	0.23	14.41	0.77	0.38	0.45
PGDNet [5]	10.70	0.64	0.32	0.36	3.52	0.30	0.65	0.26	9.32	0.74	0.30	0.35	7.26	0.51	0.46	0.07
U-Net [26]	10.70	0.59	0.43	0.30	2.78	0.27	0.74	0.38	8.83	0.38	0.71	0.21	13.09	0.52	0.57	0.33
LUMEN	**14.16**	**0.76**	**0.29**	**0.56**	**26.91**	**0.88**	**0.29**	**0.81**	**13.76**	**0.79**	**0.24**	**0.66**	**16.63**	**0.83**	**0.27**	**0.63**

Table 3: Zero-shot cross-dataset transfer over four source-to-target routes. Each block reports PSNR, SSIM, LPIPS, and MS-SSIM for eight baselines and LUMEN; colored entries mark best and second-best values.

4.6 Localized Responses under Structured Occlusion

A 3×3 wall-light grid has 2–3 randomly masked cells; methods share inputs, and GT serves only scoring. Figure 9 traces wall-evidence removal through PACER's RRG, TRO, and WRD, then ECHO's RLI, SPS, and ACR, to the localized LUMEN output and GT reference. Early maps show gaps and mixed wall structure; later maps show more connected body support. Localized visualization annotations report IoU 0.95–0.99 and MAE 0.00–0.02 (GT: $1.00/0.00$).

5 Conclusion

LUMEN links relative-reliability estimation and multiscale routing to global structural selection and progressive refinement. Visible-light R-C, R-R, and R-G results, the 880-nm R-N extension, and perturbation and transfer tests support this design. Recovering location, connected body outline, and pose after three diffuse reflections establishes the initial feasibility of observing a person beyond an L-shaped corner from a wall-light pattern recorded by an ordinary camera under steady flashlight illumination.

REFERENCES

Miika Aittala, Prafull Sharma, Lukas Murmann, Adam B. Yedidia, Gregory W. Wornell, William T. Freeman, and Frédo Durand. Computational mirrors: Blind inverse light transport by deep matrix factorization. In *Advances in Neural Information Processing Systems*, volume 32, pp. 14311–14321, 2019.

Andre Beckus, Alexandru Tamasan, and George K. Atia. Multi-modal non-line-of-sight passive imaging. *IEEE Transactions on Image Processing*, 28(7):3372–3382, 2019. doi: 10.1109/TIP.2019.2896517.

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

Sreenithy Chandran, Tatsuya Yatagawa, Hiroyuki Kubo, and Suren Jayasuriya. Learning-based spotlight position optimization for non-line-of-sight human localization and posture classification. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 4206–4215, 2024. doi: 10.1109/WACV57701.2024.00417.

Dongdong Chen, Mike E. Davies, and Mohammad Golbabaee. Compressive MR fingerprinting reconstruction with neural proximal gradient iterations. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, volume 12262 of *Lecture Notes in Computer Science*, pp. 13–22, 2020. doi: 10.1007/978-3-030-59713-9_2.

Wenzheng Chen, Simon Daneau, Fahim Mannan, and Felix Heide. Steady-state non-line-of-sight imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6790–6799, 2019. doi: 10.1109/CVPR.2019.00695.

Il Yong Chun, Zhengyu Huang, Hongki Lim, and Jeffrey A. Fessler. Momentum-Net: Fast and convergent iterative neural network for inverse problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4915–4931, 2023. doi: 10.1109/TPAMI.2020.3012955. Early access 2020.

Xingyu Cui, Huanjing Yue, Song Li, Xiangjun Yin, Yusen Hou, Yun Meng, Kai Zou, Xiaolong Hu, and Jingyu Yang. Virtual scanning: Unsupervised non-line-of-sight imaging from irregularly undersampled transients. In *Advances in Neural Information Processing Systems*, volume 37, pp. 109381–109406, 2024a. doi: 10.52202/079017-3472.

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

Robinson Czajkowski and John Murray-Bruce. Two-edge-resolved three-dimensional non-line-of-sight imaging with an ordinary camera. *Nature Communications*, 15(1):1162, 2024. doi: 10.1038/s41467-024-45397-7.

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

Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2154–2164, 2020. doi: 10.1109/CVPR42600.2020.00223.

Daniel Feijoo, Juan C. Benito, Alvaro Garcia, and Marcos V. Conde. DarkIR: Robust low-light image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10879–10889, 2025. doi: 10.1109/CVPR52734.2025.01016.

Ruixu Geng, Yang Hu, Zhi Lu, Cong Yu, Houqiang Li, Hengyu Zhang, and Yan Chen. Passive non-line-of-sight imaging using optimal transport. *IEEE Transactions on Image Processing*, 31: 110–124, 2022. doi: 10.1109/TIP.2021.3128312.

Hang Guo, Yong Guo, Yaohua Zha, Yulun Zhang, Wenbo Li, Tao Dai, Shu-Tao Xia, and Yawei Li. MambaIRv2: Attentive state space restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 28124–28133, 2025a. doi: 10.1109/CVPR52734.2025.02619.

Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. MambaIR: A simple baseline for image restoration with state-space model. In *Computer Vision – ECCV 2024*, volume 15076 of *Lecture Notes in Computer Science*, pp. 222–241, 2025b. doi: 10.1007/978-3-031-72649-1_13.

Huaibo Huang, Zhihang Li, Ran He, Zhenan Sun, and Tieniu Tan. IntroVAE: Introspective variational autoencoders for photographic image synthesis. In *Advances in Neural Information Processing Systems*, volume 31, pp. 52–63, 2018.

Mariko Isogawa, Ye Yuan, Matthew O'Toole, and Kris M. Kitani. Optical non-line-of-sight physics-based 3D human pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7011–7020, 2020. doi: 10.1109/CVPR42600.2020.00704.

Boyun Li, Haiyu Zhao, Wenxin Wang, Peng Hu, Yuanbiao Gou, and Xi Peng. MaIR: A locality- and continuity-preserving mamba for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7491–7501, 2025. doi: 10.1109/CVPR52734.2025.00702.

Yue Li, Jiayong Peng, Juntian Ye, Yueyi Zhang, Feihu Xu, and Zhiwei Xiong. NLOST: Non-line-of-sight imaging with transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13313–13322, 2023. doi: 10.1109/CVPR52729.2023.01279.

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

David B. Lindell, Gordon Wetzstein, and Matthew O'Toole. Wave-based non-line-of-sight imaging using fast f-k migration. *ACM Transactions on Graphics*, 38(4):1–13, 2019. doi: 10.1145/3306346.3322937.

Chong Mou, Qian Wang, and Jian Zhang. Deep generalized unfolding networks for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17378–17389, 2022. doi: 10.1109/CVPR52688.2022.01688.

Fangzhou Mu, Sicheng Mo, Jiayong Peng, Xiaochun Liu, Ji Hyun Nam, Siddeshwar Raghavan, Andreas Velten, and Yin Li. Physics to the rescue: Deep non-line-of-sight reconstruction for high-speed imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(8):6146–6158, 2025. doi: 10.1109/TPAMI.2022.3203383.

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

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*, volume 9351 of *Lecture Notes in Computer Science*, pp. 234–241, 2015. doi: 10.1007/978-3-319-24574-4_28.

Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1–4):259–268, 1992. doi: 10.1016/0167-2789(92)90242-F.

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

Prafull Sharma, Miika Aittala, Yoav Y. Schechner, Antonio Torralba, Gregory W. Wornell, William T. Freeman, and Frédo Durand. What you can learn by staring at a blank wall. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2310–2319, 2021. doi: 10.1109/ICCV48922.2021.00233.

Siyuan Shen, Zi Wang, Ping Liu, Zhengqing Pan, Ruiqian Li, Tian Gao, Shiying Li, and Jingyi Yu. Non-line-of-sight imaging via neural transient fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(7):2257–2268, 2021. doi: 10.1109/TPAMI.2021.3076062.

Yuan Shi, Bin Xia, Xiaoyu Jin, Xing Wang, Tianyu Zhao, Xin Xia, Xuefeng Xiao, and Wenming Yang. VmambaIR: Visual state space model for image restoration. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(6):5560–5574, 2025. doi: 10.1109/TCSVT.2025.3530090.

Matthew Tancik, Guy Satat, and Ramesh Raskar. Flash photography for data-driven hidden scene recovery. *arXiv preprint arXiv:1810.11710*, 2018. doi: 10.48550/arXiv.1810.11710.

Andreas Velten, Thomas Willwacher, Otkrist Gupta, Ashok Veeraraghavan, Moungi G. Bawendi, and Ramesh Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Communications*, 3(1):745, 2012. doi: 10.1038/ncomms1747.

Yihao Wang, Zhigang Wang, Bin Zhao, Dong Wang, Mulin Chen, and Xuelong Li. Propagate and calibrate: Real-time passive non-line-of-sight tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 972–981, 2023. doi: 10.1109/CVPR 52729.2023.00100.

Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, and Luc Van Gool. DiffIR: Efficient diffusion model for image restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 13049–13059, 2023. doi: 10.1109/ICCV51070.2023.01204.

Adam B. Yedidia, Manel Baradad, Christos Thrampoulidis, William T. Freeman, and Gregory W. Wornell. Using unknown occluders to recover hidden scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12223–12231, 2019. doi: 10.1109/ CVPR.2019.01251.

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

Jiarui Zhang, Ruixu Geng, Xiaolong Du, Yan Chen, Houqiang Li, and Yang Hu. Passive non-line-of-sight imaging with light transport modulation. *IEEE Transactions on Image Processing*, 34: 410–424, 2025. doi: 10.1109/TIP.2024.3518097.

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

Yu Zheng, Jiahui Zhan, Shengfeng He, Junyu Dong, and Yong Du. Curricular contrastive regularization for physics-aware single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5785–5794, 2023. doi: 10.1109/CVPR52729.2023 .00560.

Shangchen Zhou, Kelvin C. K. Chan, Chongyi Li, and Chen Change Loy. Towards robust blind face restoration with codebook lookup transformer. In *Advances in Neural Information Processing Systems*, volume 35, pp. 30599–30611, 2022. doi: 10.52202/068431-2218.

Shihao Zhou, Duosheng Chen, Jinshan Pan, Jinglei Shi, and Jufeng Yang. Adapt or perish: Adaptive sparse transformer with attentive feature refinement for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2952–2963, 2024. doi: 10.1109/CVPR52733.2024.00285.

Zhen Zou, Hu Yu, Jie Huang, and Feng Zhao. FreqMamba: Viewing mamba from a frequency perspective for image deraining. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 1905–1914, 2024. doi: 10.1145/3664647.3680862.

SUPPLEMENTARY EVIDENCE FOR LUMEN

Anonymous authors
Paper under double-blind review

Table notation. The tables use the following abbreviations. Mom.-Net denotes Momentum-Net, MSBDN denotes MSBDN-DFF, and DeepSN denotes DeepSN-Net. The metric initials are P for peak signal-to-noise ratio (PSNR), S for structural similarity (SSIM), L for learned perceptual image patch similarity (LPIPS), and M for multi-scale structural similarity (MS-SSIM). In Table S1, ER is the compact label for ERGAS (relative global dimensional synthesis error); the table additionally reports GMSD, ER, DCT, and FSIM as complementary metrics, with their directions given in its caption. Prefixes m, w, min, and sd denote mean, worst, minimum, and standard deviation; d denotes a direction-aligned endpoint change, Q denotes the rank-based quality percentile, S denotes the weighted quality score, and rank denotes the mean rank within the table's comparison pool.

A EXPERIMENTAL SCENE AND OBSERVATION

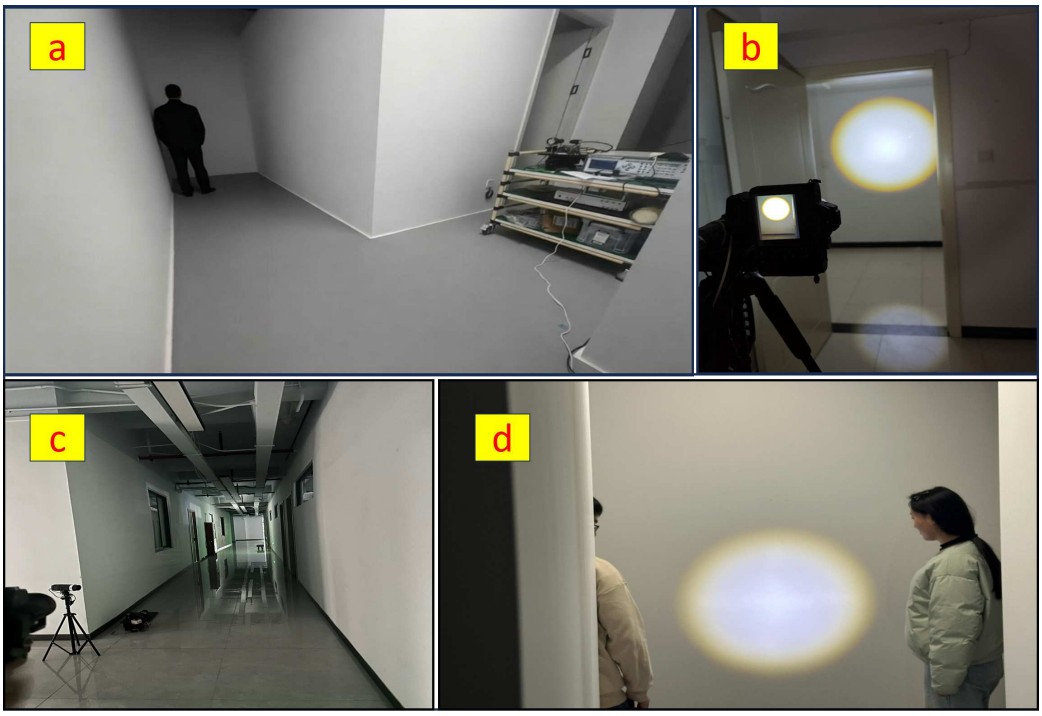

Figure S1: Photographs of the supplementary experimental scene. The L-shaped indoor geometry places the camera and steady flashlight on the observation side and the person in the other branch, while the corner blocks direct sight. The photographs document the observation-side apparatus and hidden-target region.

Figure S1 shows the acquisition setup used throughout the appendix. An ordinary RGB camera and a continuous flashlight-class source are placed on the visible side of an L-shaped indoor corner. They face a relay wall with an effective field width of approximately 1.5 m, the camera-to-wall distance is approximately 2 m, and the hidden human is approximately 3 m beyond the corner in the other branch. The corner geometry removes the direct line of sight, so the camera does not receive an ordinary photograph of the target. This setting follows established diffuse-corner computational imaging protocols while retaining a single steady-state observation. The four protocols retain the same observation logic while emphasizing different transport conditions, including directional mixing in a corridor-like branch, multipath in a room-like branch, an 880-nm spectral shift, and a large low-illumination rough-wall setting.

The measured signal follows the source to relay wall to hidden target to relay wall to camera path. The first wall reflection redirects illumination around the corner and acts as an extended secondary source. The hidden target changes the returning radiance through its coarse shape and reflectance, and

the relay wall scatters this target-dependent component toward the sensor. The camera integrates the three diffuse surface reflections together with wall background, ambient light, and sensor effects. The inverse output is consequently a two-dimensional coarse reconstruction evaluated by PSNR, SSIM, LPIPS, and MS-SSIM. The reconstruction depicts the dominant outline, connected body extent, and pose trend supported by the observation.

$$
\begin{aligned}
I(\mathbf{u}) &= B_w(\mathbf{u}) + \int_{\Omega_h} \rho_h(\mathbf{x}) K_\Gamma(\mathbf{u}, \mathbf{x}) \, \mathrm{d}\mathbf{x} + \varepsilon(\mathbf{u}), \\
K_\Gamma(\mathbf{u}, \mathbf{x}) &= \int_{\Omega_w} L_s(\mathbf{w}) \rho_w(\mathbf{w})^2 V_\Gamma(\mathbf{w}, \mathbf{x}) G_\Gamma(\mathbf{u}, \mathbf{w}, \mathbf{x}) \, \mathrm{d}\mathbf{w}.
\end{aligned}
\tag{S1}
$$

Here, Ω_w and Ω_h are the visible-wall and hidden-target domains, B_w is the wall-dominated background, L_s is source illumination on the wall, and V_Γ and G_Γ collect visibility, distance attenuation, angular projection, and aperture terms. Equation (S1) is an explanatory steady-state transport model. Its factors describe the physical contributions to the mixed wall observation. The nested integration mixes hidden-surface contributions at the wall and again through the wall-to-camera aperture before sampling, explaining how spatial body information is lost in the observation.

A.1 Acquisition, Registration, and Evaluation Units

The observation and reference cameras serve different roles. The observation camera records the illuminated wall patch used as the only model input, whereas the reference camera views the person directly and supplies the paired target used for supervision and scoring. A shared trial trigger and timestamp associate the two streams before any frame is admitted to a pair. Intrinsic parameters are estimated for each camera separately, and a common calibration target together with fixed scene landmarks establishes the transformation into a shared image canvas. Capture geometry and retained metadata define registration before inference.

The raw synchronized pair is stored before resizing, normalization, or foreground processing. Subject, sequence, session, and frame identifiers remain attached to both members of the pair, allowing each processed sample to be traced to one acquisition event. The camera, flashlight, corner, and wall remain fixed inside a protocol unless a named physical-perturbation experiment explicitly varies one factor. This ordering separates acquisition variation from preprocessing variation: a change in camera angle, distance, jitter, target distance, or wall material is recorded as an experimental coordinate, while ordinary benchmark pairs follow their protocol's fixed geometry.

Dataset partitioning is performed before frame extraction. All frames belonging to one subject, sequence, and recording session remain in one partition, and temporally adjacent frames cannot be divided between training and test sets. This subject-and-capture separation prevents nearby motion states or repeated wall patterns from entering both sides of the evaluation through a random frame-level split. The four protocol labels identify different acquisition populations, but they do not replace the split identifiers. Consequently, within-protocol evaluation and cross-protocol transfer can use the same pairing rules while asking different questions about reconstruction and transport change.

Primary reconstruction metrics are computed on the complete aligned image. PSNR, SSIM, MS-SSIM, and LPIPS therefore evaluate the same full spatial support for LUMEN and every baseline. Foreground-specific measures are reserved for the structured-occlusion study, where overlap and reconstruction error answer a separate spatial question. This distinction is important because full-image fidelity and foreground recovery need not rank methods identically. Full-image metrics and localized visualization annotations retain separate evaluation units.

The four acquisition protocols also provide different stress conditions for the same task. R-C emphasizes directional transport around a corridor corner, R-R includes room-scale multipath, R-N changes the spectral response near 880 nm, and R-G combines a longer low-light path with a rougher garage environment. These descriptions identify measured scene regimes. Unless a protocol indicator is explicitly part of a reported baseline, the input remains the wall observation itself. The comparisons below therefore retain protocol boundaries in every table block and avoid pooling raw metric values across physically different ranges.

A.2 Module Derivation Overview

Figure S2 summarizes the quantities exchanged by the two model stages. Given the wall observation $\mathbf{X}$, PACER first forms a luminance projection and a learned embedding, then routes transport evidence across scales. With $\mathbf{C}$ denoting relative reliability, $\mathbf{E}_\ell$ multiscale evidence, and $\mathbf{Z}$ the bottleneck, the

routing chain is

$$(\mathbf{C}, \{\mathbf{E}_\ell\}, \mathbf{Z}) = \mathcal{P}_{\boldsymbol{\theta}}(\mathbf{X}), \qquad \mathbf{E}_\ell = \mathcal{D}_\ell(\mathcal{T}_\ell(\mathbf{X}, \mathbf{C})), \quad \mathbf{Z} = \mathcal{B}(\{\mathbf{E}_\ell\}, \mathbf{C}). \tag{S2}$$

PACER passes multiscale wall evidence and its relative support through this shared interface; ECHO combines them with the bottleneck hypothesis. RLI injects reliability-weighted evidence, SPS removes inconsistent structural candidates, and ACR refines the retained state through scale-matched corrections. The complete readout can therefore be written as

$$\widehat{\mathbf{Y}} = \mathcal{M}(\mathcal{A}(\mathcal{S}_{\mathrm{SPS}}(\mathcal{I}(\mathbf{Z}, \{\mathbf{E}_\ell\}, \mathbf{C})))) = \mathcal{H}_{\boldsymbol{\phi}}(\mathbf{Z}, \{\mathbf{E}_\ell\}, \mathbf{C}). \tag{S3}$$

Here $\mathcal{I}$, $\mathcal{S}_{\mathrm{SPS}}$, $\mathcal{A}$, and $\mathcal{M}$ denote reliability injection, structural screening, annealed consensus refinement, and final mapping, respectively. The diagram and equations make the dependency explicit: reliability accompanies multiscale evidence through structural screening and refinement, and decoder-side supervision is used only during training.

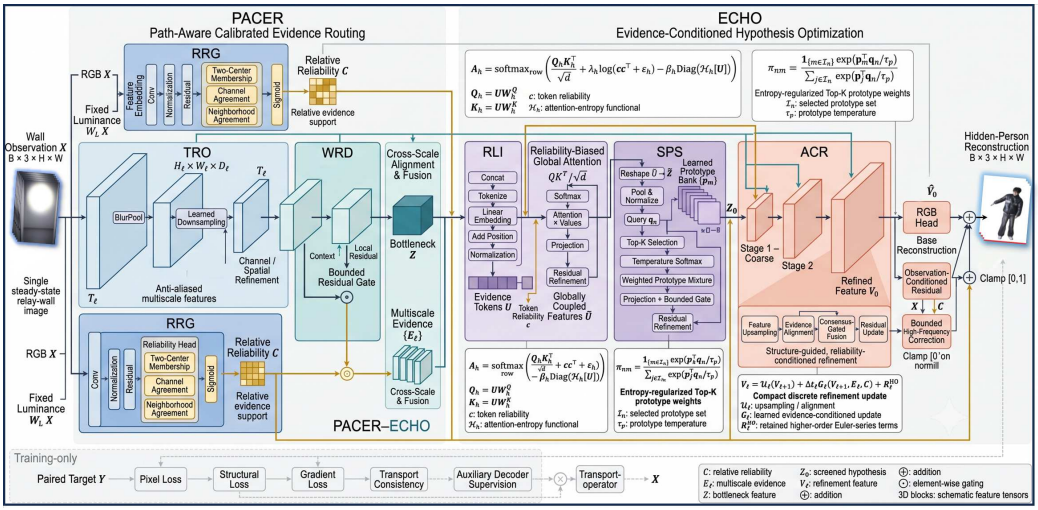

Figure S2: Module derivation overview. The appendix diagram follows the dependency from the wall observation through PACER evidence routing, ECHO hypothesis formation, and ACR refinement to the final reconstruction. Eqs. S2–S3 summarize the two-stage interface.

B Reconstruction Quality and Metric-Space Structure

We test whether reconstructed people retain gradient, radiometric, frequency, and feature agreement using complementary metrics. Figure S3 gives the broader measured-result map, while Table S1 reports the complementary-metric reference used to compare all 25 methods on the same four datasets. Benchmark values describe unperturbed reconstruction; later sections examine architectural and input changes.

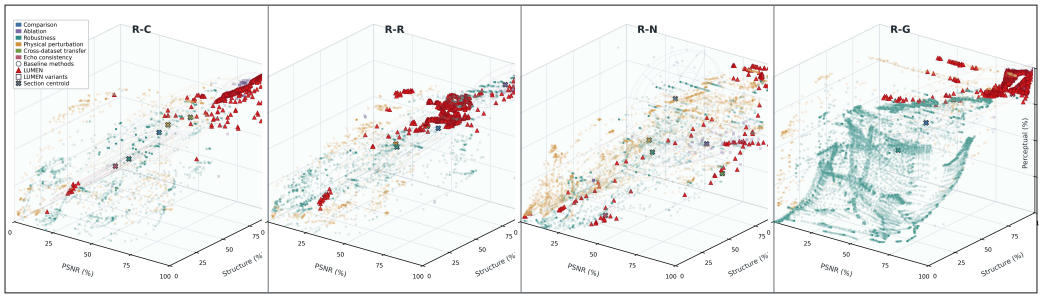

Figure S3: Complete measured results in a common direction-aligned metric space. The four panels correspond to R-C, R-R, R-N, and R-G. Coordinates are empirical within-protocol percentiles of PSNR, the mean of SSIM and MS-SSIM, and reverse LPIPS. Circles denote baselines, red triangles denote LUMEN records, and squares denote LUMEN variants. Crosses and covariance wireframes summarize the displayed experiment groups.

Table S1 reports the 25-method complementary-metric benchmark in the same grouped layout as the main comparison: each dataset block contains GMSD, ER, DCT, and FSIM, including the Mamba-family results. Tables S2 and S3 display eight methods using the four primary reconstruction metrics, with cross-metric placements computed against the complete 28-model main-text benchmark.

Type	Method	Venue	R-C				R-R				R-N				R-G			
			GMSD	ER	DCT	FSIM	GMSD	ER	DCT	FSIM	GMSD	ER	DCT	FSIM	GMSD	ER	DCT	FSIM
Transformer	Restormer	CVPR'22	**0.30**	5.85	0.18	0.78	0.28	5.99	0.25	0.84	**0.20**	4.55	2.28	**0.94**	0.34	5.77	0.76	0.70
	NLOST	CVPR'23	0.37	7.32	1.04	0.76	0.32	7.42	0.55	0.79	0.21	6.30	3.24	0.90	0.34	5.90	1.06	0.67
	SwinIR	ICCVW'21	0.37	8.71	0.77	0.72	0.30	9.88	0.97	0.78	0.24	10.44	3.03	0.75	**0.30**	5.66	0.64	**0.76**
U-Net	MSBDN-DFF	CVPR'20	0.36	7.80	0.78	0.72	0.28	6.40	0.25	0.84	0.20	4.67	1.53	**0.94**	0.36	5.73	0.54	0.72
	ConvIR	TPAMI'24	0.31	5.63	0.26	0.79	0.29	6.55	0.30	0.83	0.20	4.48	1.05	0.94	0.30	6.04	0.67	0.69
	U-Net	MICCAI'15	0.32	5.84	0.32	0.75	0.27	5.52	0.26	0.85	0.20	4.65	1.71	0.94	0.34	5.93	0.89	0.70
	DarkIR	CVPR'25	0.31	6.33	0.41	0.65	0.30	6.72	0.41	0.81	0.20	4.69	1.48	0.94	0.33	5.91	0.62	0.70
	DnCNN	TIP'17	0.35	7.16	0.42	0.77	0.31	7.92	0.27	0.80	0.23	8.92	1.71	0.83	0.33	5.93	0.41	0.75
	AST	CVPR'24	0.30	6.06	0.32	0.76	0.28	6.20	0.29	0.84	0.20	4.56	1.68	0.94	0.33	5.73	1.02	0.70
	C2PNet	CVPR'23	0.36	7.32	1.05	0.74	0.31	8.18	0.47	0.79	0.21	4.86	1.64	0.93	0.34	5.79	1.00	0.69
Diffusion	DiffIR	ICCV'23	0.35	6.86	0.32	0.75	0.30	6.68	0.39	0.82	0.20	4.74	1.57	0.93	0.33	5.64	0.85	0.69
Unrolled	DGUNet	CVPR'22	0.31	5.85	0.32	0.79	0.29	6.31	0.27	0.84	0.20	4.57	1.54	0.94	0.32	5.32	0.79	0.71
VAE	IntroVAE	NeurIPS'18	0.31	**5.46**	0.28	**0.80**	**0.25**	**5.02**	**0.19**	**0.87**	0.21	**4.46**	1.08	0.94	0.33	5.68	0.71	0.72
Codebook	CodeFormer	NeurIPS'22	0.38	7.81	0.59	0.69	0.38	8.30	0.81	0.65	0.25	7.31	2.67	0.84	0.32	5.11	0.38	0.73
Unrolled	PGDNet	MICCAI'20	0.32	6.26	0.21	0.79	0.30	6.89	0.33	0.82	0.20	4.59	1.24	0.94	0.30	6.04	0.67	0.69
Newton	DeepSN	TPAMI'25	0.32	6.05	**0.09**	0.73	0.29	6.39	0.31	0.76	0.21	4.89	**0.93**	0.93	0.31	**4.98**	0.71	0.73
Unrolled	Momentum-Net	TPAMI'23	0.31	6.03	0.20	0.79	0.28	6.08	0.27	0.84	0.20	4.50	1.86	0.94	0.34	5.48	0.87	0.71
NLOS	VS-NLOS	NeurIPS'24	0.32	16.15	0.35	0.74	0.30	6.96	0.42	0.81	0.21	4.72	1.51	0.94	0.34	5.68	0.82	0.70
NLOS	NLOS-RescueNet	TPAMI'25	0.35	6.68	0.68	0.69	0.32	8.38	0.49	0.78	0.20	4.79	2.07	0.93	0.33	5.95	0.97	0.70
Mamba	MambaIR	ECCV'24	0.34	6.58	0.37	0.77	0.30	10.11	0.91	0.77	0.21	5.18	2.41	0.93	0.34	5.60	0.90	0.72
	VMambaIR	TCSVT'25	0.31	5.89	1.08	0.74	0.29	7.15	0.20	0.83	0.20	5.20	2.37	0.93	0.34	5.93	0.96	0.72
	MambaIRv2	CVPR'25	0.36	7.80	0.78	0.71	0.33	8.91	1.38	0.62	0.20	5.41	2.16	0.92	0.35	5.93	0.84	0.71
	FreqMamba	ACM MM'24	0.38	8.34	2.46	0.67	0.35	9.19	1.24	0.72	0.22	5.38	1.63	0.90	0.32	5.56	0.78	0.74
	MaIR	CVPR'25	0.37	8.27	0.54	0.62	0.33	9.05	1.34	0.63	0.29	14.51	6.78	0.61	0.37	5.95	**0.25**	0.68
Ours	LUMEN	–	**0.23**	**3.25**	**0.08**	**0.91**	**0.22**	**4.12**	**0.09**	**0.93**	**0.15**	**4.37**	**0.70**	0.93	**0.17**	**1.24**	**0.06**	**0.93**

Table S1: Complementary-metric comparison of 25 methods on R-C, R-R, R-N, and R-G using the synchronized canonical benchmark records. The four displayed complementary metrics are GMSD (gradient magnitude similarity deviation), ER (ERGAS, relative global dimensional synthesis error), DCT (absolute discrete-cosine-transform entropy difference), and FSIM (feature similarity index); values are shown to two decimals. GMSD, ER, and DCT are lower-is-better, whereas FSIM is higher-is-better. Red data bands identify LUMEN and blue data bands identify the representative strongest baseline for each dataset block (Restormer for R-C, IntroVAE for R-R, ConvIR for R-N, and CodeFormer for R-G). Each dataset–metric column marks one best value in red and one second-best value in blue; ties favor LUMEN, then table order.

On R-C, LUMEN reduces GMSD from the strongest baseline value of 0.30 to 0.23 and raises FSIM from 0.80 to 0.91. LUMEN reaches FSIM 0.91, 0.93, 0.93, and 0.93 on R-C, R-R, R-N, and R-G; on R-N, its lowest GMSD, ER, and DCT accompany FSIM 0.93, compared with 0.94 for several baselines.

These endpoints also define the reference for subsequent perturbations. The 25-method complementary comparison retains each architecture family. A raw endpoint measurement can therefore be distinguished from a statistic computed over perturbation nodes. This separation is useful when a method changes relative position across protocols, because the displayed complementary metrics retain their dataset-specific values.

The direction-aligned quantities in this section are used only to compare heterogeneous metrics on a common reading scale. For the metric-space figure, PSNR, SSIM, and MS-SSIM retain their native higher-is-better orientation, whereas LPIPS is reversed before aggregation. Table S1 keeps a separate raw record of complementary gradient, radiometric, transform-domain, and feature-similarity measurements; GMSD, ER, and DCT are lower-is-better, while FSIM is higher-is-better. Taken together, the figure and Table S1 show the distribution of reconstruction quality; Tables S2 and S3 provide the primary-metric view needed to interpret any apparent separation in that map.

The benchmark, ranking, and projection views use explicitly defined comparison units. The selected-method comparison and derived placements use the metric directions stated above. The subsequent projection relates these values in a common metric space.

The complementary metrics in Table S1 add views that are not identical to the four headline measures. GMSD responds to gradient-magnitude disagreement, ER records a relative global synthesis error, DCT summarizes transform-domain discrepancy, and FSIM measures feature similarity. Their joint presentation is useful because diffuse transport can preserve broad intensity while removing edges, or preserve a body-shaped support while shifting its local contrast. Agreement across these measures strengthens the interpretation of a reconstruction row; disagreement identifies the type of residual error.

Metric directions distinguish error from similarity. Lower GMSD, ER, and DCT indicate smaller error, whereas higher FSIM indicates greater feature agreement. The red and blue type therefore follows the direction of each column instead of the largest printed numeral. Row shading is applied only to the numerical block, leaving method and family text uncolored. Numeric emphasis follows these metric directions.

Protocol-wise comparison is essential for these secondary measures. Their scales depend on wall appearance, illumination range, spectral response, and the dynamic range of the aligned target. A raw ER or DCT value from R-G is not treated as interchangeable with the same numeral from R-N. Ranking is performed inside each protocol block, with separate values for each acquisition regime.

The complementary and primary metrics provide distinct comparisons. The complementary-metric table exposes its 25-method population, while the main-text primary-metric table provides the complete 28-model pool for the compact ranking. The compact tables make four-metric patterns readable at normal page width. The closest competitor is determined separately for each protocol and metric.

The method families are retained because they fail in different ways under the wall-observation task. Transformer and state-space restorers can capture long-range context but may retain wall texture while losing body structure. Convolutional encoders preserve local continuity but can blur distant limbs when evidence is weak. Diffusion, unfolding, and inverse-imaging baselines introduce different reconstruction biases and therefore provide complementary tests of numerical and structural fidelity. Reporting the complete family population makes a favorable LUMEN result less dependent on the weakness of one architecture class.

Method identity and measured performance are reported separately. A citation and venue identify the registered baseline, while the numerical cells report checkpoint outputs with shared preprocessing and metric computation. All listed methods are retained across protocols. The strongest competitors are determined independently for each metric.

Benchmark and ablation highlighting uses the displayed values. Each dataset–metric column marks one best numeral in red and one second-best numeral in blue; ties favor LUMEN or Full, then table order. The benchmark summary uses competition ranks over the 28-model main-text pool, so equal displayed values receive equal ranks before averaging and counting.

These benchmark records define the undisturbed endpoint to which the intervention and robustness sections return. An ablation asks how a changed architecture departs from its protocol-specific Full row. A noise or physical sweep asks how a frozen method departs from its clean operating condition. Establishing the benchmark population first therefore gives the appendix a consistent reference without mixing an unperturbed method comparison with statistics over altered inputs or altered modules.

Dataset	Type	Model	PSNR	SSIM	LPIPS	MS-SSIM
R-C	U-Net	ConvIR	14.20	0.68	0.22	0.62
	VAE	IntroVAE	**14.21**	0.68	**0.21**	**0.64**
	Mamba	MambaIR	12.97	0.68	0.30	0.49
	Mamba	VMambaIR	13.92	0.63	0.44	0.42
	Mamba	MambaIRv2	11.68	**0.71**	0.33	0.37
	Trans.	NLOST	12.23	0.71	0.33	0.37
	Trans.	Restormer	14.18	0.68	0.22	0.61
	Proposed	LUMEN	**19.40**	**0.86**	**0.18**	**0.84**
R-G	U-Net	ConvIR	**15.69**	**0.78**	**0.29**	**0.45**
	VAE	IntroVAE	13.77	0.64	0.42	0.40
	Mamba	MambaIR	13.82	0.65	0.44	0.39
	Mamba	VMambaIR	13.32	0.63	0.48	0.35
	Mamba	MambaIRv2	13.30	0.63	0.44	0.39
	Trans.	NLOST	13.34	0.59	0.53	0.36
	Trans.	Restormer	13.55	0.62	0.53	0.40
	Proposed	LUMEN	**27.72**	**0.93**	**0.12**	**0.95**
R-N	U-Net	ConvIR	24.16	**0.85**	**0.16**	**0.81**
	VAE	IntroVAE	22.02	0.84	0.16	0.81
	Mamba	MambaIR	26.96	0.84	0.21	0.80
	Mamba	VMambaIR	**27.10**	0.84	0.17	0.78
	Mamba	MambaIRv2	26.51	0.82	0.35	0.78
	Trans.	NLOST	23.30	0.80	0.21	0.75
	Trans.	Restormer	11.07	0.81	0.16	0.80
	Proposed	LUMEN	**28.25**	**0.89**	**0.15**	**0.82**
R-R	U-Net	ConvIR	13.44	0.75	0.19	0.69
	VAE	IntroVAE	**16.29**	**0.82**	**0.13**	**0.80**
	Mamba	MambaIR	11.19	0.63	0.47	0.41
	Mamba	VMambaIR	12.61	0.75	0.19	0.65
	Mamba	MambaIRv2	10.39	0.49	0.59	0.34
	Trans.	NLOST	12.17	0.71	0.26	0.60
	Trans.	Restormer	14.31	0.77	0.15	0.73
	Proposed	LUMEN	**18.12**	**0.88**	**0.12**	**0.85**

Table S2: Four-metric comparison of seven representative baselines and LUMEN, synchronized with the main-text comparison. Red and blue row bands identify the best and representative second-best aggregate rows within each dataset block; metric values are marked independently.

The eight-method table reports the four primary metrics at main-text precision. Its role is to preserve the raw comparison before the next table compresses those values into cross-metric placements.

Each benchmark row is kept at the method level so that a favorable aggregate cannot conceal a weak metric. The four columns should be read together: PSNR reflects numerical fidelity, SSIM and MS-SSIM retain structural agreement at different scales, and LPIPS records perceptual distance with its native lower-is-better interpretation. These columns retain the complete four-metric comparison.

The four primary columns should be interpreted jointly when comparing reconstruction behavior. PSNR captures numerical agreement, SSIM and MS-SSIM capture structure at different neighborhoods, and LPIPS captures perceptual distance with the opposite direction. Reporting them together prevents a gain in one dimension from being mistaken for a uniform improvement. The later figures show metric organization and response trends.

The paired benchmark summaries are useful for a second check. Repeated best or top-three placements indicate whether a method is competitive across metrics, while a single strong value does not receive the same interpretation. This distinction is important for the later perturbation sections, where the ordering can change with the transport condition. Ranks and counts summarize the same measured rows.

The benchmark pair also fixes the comparison vocabulary used below. A method can be numerically strong on one metric and structurally weaker on another, so the appendix keeps the metric columns separate even when a later analysis reports a joint quality score. This separation is particularly important for LPIPS, whose lower-is-better direction is opposite to the other three columns. The colored rows summarize overall placement, while the uncolored rows preserve the full competing set needed to reproduce the ordering.

Dataset	Type	Model	rank↓	Q (%)↑	Best↑	Top3↑
R-C	VAE	IntroVAE	**5.25**	**84.26**	**0**	**2**
	U-Net	ConvIR	6.00	81.48	0	2
	Trans.	Restormer	6.75	78.70	0	1
	Mamba	MambaIR	12.75	56.48	0	0
	Trans.	NLOST	14.50	50.00	0	1
	Mamba	MambaIRv2	15.00	48.15	0	1
	Mamba	VMambaIR	18.25	36.11	0	0
	Proposed	LUMEN	**1.00**	**100.00**	**4**	**4**
R-G	U-Net	ConvIR	**3.25**	**91.67**	**0**	**3**
	VAE	IntroVAE	12.25	58.33	0	0
	Mamba	MambaIR	13.25	54.63	0	0
	Mamba	MambaIRv2	17.75	37.96	0	0
	Trans.	Restormer	19.75	30.56	0	0
	Mamba	VMambaIR	21.00	25.93	0	0
	Trans.	NLOST	24.75	12.04	0	0
	Proposed	LUMEN	**1.00**	**100.00**	**4**	**4**
R-N	U-Net	ConvIR	**4.75**	**86.11**	**0**	**3**
	VAE	IntroVAE	7.25	76.85	0	3
	Mamba	MambaIR	7.50	75.93	0	2
	Mamba	VMambaIR	8.00	74.07	0	2
	Trans.	Restormer	13.50	53.70	0	1
	Mamba	MambaIRv2	14.50	50.00	0	0
	Trans.	NLOST	19.75	30.56	0	0
	Proposed	LUMEN	**1.00**	**100.00**	**4**	**4**
R-R	VAE	IntroVAE	**2.00**	**96.30**	**0**	**4**
	Trans.	Restormer	4.50	87.04	0	1
	U-Net	ConvIR	10.50	64.81	0	0
	Mamba	VMambaIR	12.50	57.41	0	0
	Trans.	NLOST	17.50	38.89	0	0
	Mamba	MambaIR	23.00	18.52	0	0
	Mamba	MambaIRv2	26.50	5.56	0	0
	Proposed	LUMEN	**1.00**	**100.00**	**4**	**4**

Table S3: Cross-metric placements for the eight displayed methods in Table S2, ranked within the full 28-model main-text benchmark pool. Each of the four primary metrics uses the main-text displayed values and competition rank (one plus the number of strictly better values). Rank is the mean of these four ranks; Q is 100 times (28 minus mean rank)/27. Best and Top3 count ranks of 1 and at most 3, respectively. Data bands mark the two lowest displayed mean ranks; each column marks one red and one blue value, with ties favoring LUMEN, then table order.

As the leading competitor changes across acquisition settings, LUMEN retains first place in all four primary metrics on each dataset within the 28-model benchmark. Best and Top-3 counts expose repeated placement across metrics while the underlying values remain available in the first table.

The summary complements the metric values. Its counts indicate repeated competitiveness across the four metrics while retaining ties. A method can enter the Top-3 set without being best on every column. This keeps the aggregate view concise while preserving the metric-level evidence needed for a careful comparison with the later intervention and robustness tables.

The count view uses competition ranks across the 28-model pool. Equal main-text displayed measurements receive the same rank before Best and Top3 counts are formed. For that reason, the summary should be interpreted together with the preceding metric table whenever two methods occupy the same printed value. Counts expose repeated placement across protocols.

Cross-protocol differences in the identity of the second-best method are also informative. The closest competitor depends on the acquisition regime. A restoration model can be strong in a room-scale protocol yet weaker under the spectral or low-light shift, while another family can provide the closest comparison there. Retaining this change of competitor makes the LUMEN comparison more demanding than selecting one fixed baseline whose weakness is concentrated in a particular setting.

The PCA analysis uses these benchmark rows as a descriptive reference frame, with the raw table values preserved. This sequencing is intentional: raw reconstruction evidence is established first, repeated placement is summarized second, and only then are standardized coordinates used to visualize how the methods occupy metric space.

B.1 Baseline-Reference PCA and Illumination Sensitivity

Figure S4 uses a baseline-only PCA fit to make the reference frame explicit. PSNR, SSIM, reverse LPIPS, and MS-SSIM are standardized after their quality directions are aligned. The fit uses only baseline rows; LUMEN rows are projected out of sample. The baseline population therefore determines the projection axes. Table S5 summarizes the projected responses and matched-condition margins. Separately, Table S4 compares LUMEN directly with seven baselines under changes in input brightness.

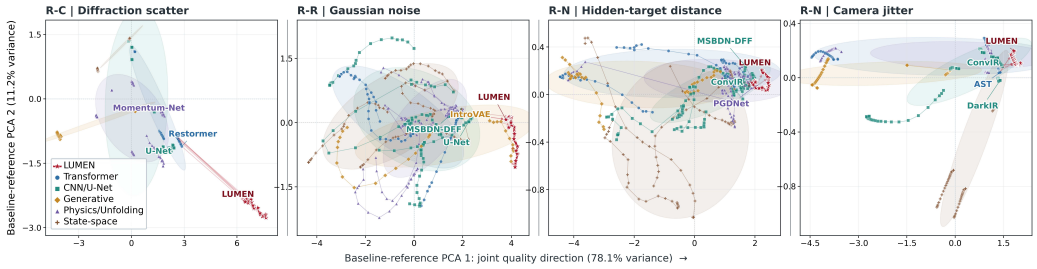

Figure S4: Baseline-reference PCA trajectories for four non-illumination conditions. PCA is fitted on baseline rows only, and the LUMEN trajectory is projected out of sample. Connections follow measured levels; pale family ellipses summarize displayed baseline clouds. Red arrows compare measured LUMEN nodes with the same-coordinate baseline PCA1 frontier. All plotted coordinates are retained measurements.

The PCA display describes joint metric organization. Its first component captures most of the baseline variance and is oriented toward larger direction-aligned quality. The four non-illumination trajectories include two physical R-N conditions; their coordinates describe the displayed observations.

Standardization is performed inside the displayed reference construction so that a metric with a large numerical range does not dominate only because of its units. PSNR, SSIM, reverse LPIPS, and MS-SSIM first receive a consistent quality direction and are then centered and scaled using the baseline population. The fitted axes are held fixed when LUMEN is projected. This makes the visual displacement interpretable relative to the competing methods with the same axes used for all projected methods.

The connected points show ordering over retained condition levels. A displacement along the first component summarizes correlated changes in the direction-aligned metrics, whereas displacement along the second component exposes differences that are not explained by that dominant pattern. The accompanying projection statistics summarize coordinate means, standard deviations, and matched-condition margins. The dimensionless components describe metric covariation; the original metrics quantify reconstruction quality.

Brightness scaling tests body reconstruction when wall-return strength changes while retaining its spatial organization. For the illumination experiment, Table S4 uses R-C brightness factors 0.80, 1.15, and 1.50, spanning the endpoints and midpoint of its retained range. These are measured nodes in the four-metric record. Each block contains the same seven baselines and LUMEN, with their recorded checkpoint outputs and native metric values. LUMEN leads all four metrics at each of these three levels, including comparisons against the complete 21-baseline envelope in the retained measurements. This supports reconstruction under the displayed illumination changes.

Factor	Type	Model	PSNR↑	SSIM↑	LPIPS↓	MS-SSIM↑
0.80	U-Net	ConvIR	13.07	0.66	0.26	0.50
	VAE	IntroVAE	12.67	0.63	0.33	0.45
	Mamba	MambaIR	11.68	**0.71**	0.33	0.37
	Mamba	VMambaIR	11.68	0.71	0.33	0.37
	Mamba	MambaIRv2	11.68	0.71	0.33	0.37
	Trans.	NLOST	12.23	0.71	0.33	0.37
	Trans.	Restormer	**14.06**	0.68	**0.23**	**0.60**
	Proposed	LUMEN	**18.99**	**0.85**	**0.20**	**0.83**
1.15	U-Net	ConvIR	13.03	0.66	0.26	0.51
	VAE	IntroVAE	12.67	0.63	0.33	0.45
	Mamba	MambaIR	11.68	**0.71**	0.33	0.37
	Mamba	VMambaIR	11.68	0.71	0.33	0.37
	Mamba	MambaIRv2	11.68	0.71	0.33	0.37
	Trans.	NLOST	12.23	0.71	0.33	0.37
	Trans.	Restormer	**14.17**	0.69	**0.24**	**0.61**
	Proposed	LUMEN	**19.26**	**0.86**	**0.19**	**0.83**
1.50	U-Net	ConvIR	12.65	0.65	**0.28**	0.48
	VAE	IntroVAE	12.68	0.63	0.33	0.45
	Mamba	MambaIR	11.68	**0.71**	0.33	0.37
	Mamba	VMambaIR	11.68	0.71	0.33	0.37
	Mamba	MambaIRv2	11.68	0.71	0.33	0.37
	Trans.	NLOST	12.23	0.71	0.33	0.37
	Trans.	Restormer	**13.74**	0.66	0.31	**0.56**
	Proposed	LUMEN	**18.61**	**0.85**	**0.20**	**0.81**

Table S4: R-C illumination sensitivity at three measured brightness factors. The seven baselines match Table S3; LUMEN is last in each block. Red/blue data bands mark the best/second-best mean four-metric ranks; numerals mark the best/second-best values per metric using displayed values, with ties favoring LUMEN, then table order.

The strongest baseline varies by metric: Restormer leads PSNR and MS-SSIM among the displayed baselines, MambaIR, VMambaIR, MambaIRv2, and NLOST tie at displayed SSIM 0.71, and ConvIR leads LPIPS at factor 1.50. LUMEN retains its advantage against these different competitors. Table S5 returns to the PCA analysis and gives the projection statistics and matched-condition margins for the plotted trajectories.

Brightness factors are applied within the same R-C acquisition setup, so the comparison isolates input-level illumination variation with the room and subject partition fixed. Every method receives the same adjusted observation at a given factor, and the reference target and metric implementation remain unchanged. The factor labels identify measured levels; no intermediate level is inferred from the three displayed blocks. LUMEN maintains higher four-metric quality at all three brightness settings.

The same methods are compared across all three factors. The repeated comparison shows how each method responds to brightness. It also exposes whether the closest competitor changes as brightness changes. LUMEN remains the final row in every block, while red and blue numeric formatting is recomputed from the values at that factor. Each factor has its own metric comparison.

Table S4 reports reconstruction measurements, whereas Table S5 describes a fitted reference frame. The projection and reconstruction metrics remain separate quantities.

The two views answer different questions. The PCA plot describes how baseline measurements are arranged in a reference coordinate system, whereas the illumination table compares recorded reconstruction metrics directly. The brightness table reports three selected measured levels, while PCA summarizes metric covariation over the plotted conditions.

Table S5 retains the mean and spread of the fitted coordinates together with the joint envelope margin. The mean locates the displayed response population, and the spread records variation over retained levels. The minimum joint margin is particularly important because a positive average can otherwise hide a local baseline crossing. The four-metric win-rate column reports the percentage of nodes satisfying the matched-node comparison and is kept separate from the mean coordinate so that repeated local placement is not conflated with geometric distance in PCA space.

Condition	Method	mPC1	sdPC1	mPC2	sdPC2	mJ	minJ	mGap	4W (%)
R-C — scatter noise	Restormer	2.63	0.11	-0.93	0.08	**-0.10**	**-0.13**	**0.00**	**0.00**
	U-Net	2.19	0.23	-1.14	0.04	-0.32	-0.67	-0.43	0.00
	Mom.-Net	1.72	0.06	-0.31	0.04	-0.55	-0.59	-0.91	0.00
	DGUNet	0.04	0.04	-1.25	0.03	-1.40	-1.45	-2.58	0.00
	PGDNet	0.03	0.04	0.38	0.03	-1.40	-1.47	-2.60	0.00
	MSBDN	-0.17	0.01	1.36	0.01	-1.49	-1.57	-2.79	0.00
	LUMEN	7.01	0.39	-2.53	0.14	**2.09**	**1.72**	**4.38**	**100.00**
R-N — jitter	MSBDN	1.12	0.02	0.17	0.01	**-0.21**	**-0.25**	**-0.28**	**0.00**
	PGDNet	1.06	0.03	0.20	0.01	-0.24	-0.28	-0.34	0.00
	Mom.-Net	1.04	0.03	0.22	0.01	-0.25	-0.29	-0.36	0.00
	DGUNet	1.02	0.03	0.21	0.01	-0.26	-0.30	-0.38	0.00
	Restormer	0.94	0.03	0.26	0.02	-0.30	-0.34	-0.46	0.00
	U-Net	0.81	0.04	0.20	0.01	-0.36	-0.41	-0.59	0.00
	LUMEN	1.86	0.08	0.13	0.04	**0.16**	**0.12**	**0.46**	**100.00**
R-N — target dist.	PGDNet	1.41	0.36	0.17	0.05	**-0.18**	-0.28	**-0.29**	**0.00**
	MSBDN	1.39	0.26	0.18	0.04	-0.19	**-0.25**	-0.31	0.00
	Mom.-Net	1.39	0.36	0.20	0.03	-0.19	-0.29	-0.31	0.00
	DGUNet	1.38	0.37	0.18	0.05	-0.20	-0.30	-0.32	0.00
	Restormer	1.35	0.46	0.21	0.09	-0.21	-0.34	-0.35	0.00
	U-Net	1.16	0.35	0.13	0.07	-0.31	-0.41	-0.54	0.00
	LUMEN	2.16	0.19	0.11	0.05	**0.19**	**0.09**	**0.46**	**100.00**
R-R — Gaussian	MSBDN	1.65	0.22	0.17	0.04	**-0.83**	**-1.18**	**-1.63**	**0.00**
	U-Net	1.59	1.09	-0.14	0.27	-0.86	-2.71	-1.69	0.00
	DGUNet	1.58	0.35	0.23	0.06	-0.86	-1.73	-1.70	0.00
	Restormer	1.40	0.50	-0.47	0.68	-0.95	-1.85	-1.88	0.00
	PGDNet	1.26	0.43	0.42	0.29	-1.02	-1.84	-2.01	0.00
	Mom.-Net	0.99	0.85	0.43	0.33	-1.15	-2.67	-2.28	0.00
	LUMEN	3.98	0.17	-0.18	0.36	**0.34**	**0.06**	**0.70**	**100.00**

Table S5: Compact PCA-response summary for LUMEN and six globally strongest baselines in the four-panel projection. The selected baselines have the highest pooled joint-quality z score over the displayed conditions; means and population standard deviations summarize retained nodes for each condition-method pair. mJ and minJ are direction-aligned joint envelope margins, mGap is the mean PCA1 gap to the matched frontier, and 4W is the four-metric win rate (%), the percentage of nodes above all four matched baseline envelopes. PCA is fitted on baseline rows and LUMEN is projected out of sample.

Taken together, Figure S4 and Tables S4–S5 provide three levels of evidence: raw four-metric reconstruction values, a baseline-defined descriptive projection, and matched-node margin summaries. The matched-condition comparisons in Table S5 report a 100.00% four-metric win rate for LUMEN in each of the four reported conditions, connecting the projection analysis to reconstruction measurements. All interpretations remain tied to the displayed conditions and the frozen evaluation records.

C Component and Pathway Interventions

We test two sources of reconstruction gains: each computation's contribution and the use and learning of evidence passed between computations. Table S6 compares Full LUMEN with removal of PACER, TRO, WRD, RRG, RLI, SPS, and ACR across all four datasets. Table S7 changes information or gradient pathways while retaining the surrounding architecture. Table S8 then expresses the same endpoint evidence relative to the matched Full result. These interventions measure how reconstruction depends on evidence routing and structural readout.

All variants use the same paired test samples, aligned targets, image support, and metric implementations as the corresponding Full checkpoint. Comparisons are made within a dataset block, because an endpoint change is meaningful only against the intact configuration evaluated under the same transport regime. Each variant is compared with its protocol-specific Full row. This matched construction lets the reader attribute the displayed output sensitivity to the registered intervention without introducing a change of data split.

The three tables retain raw endpoints before derived changes. Table S6 gives canonical removal values, Table S7 records pathway-specific variants, and Table S8 subtracts each canonical endpoint from its matched Full reference in a direction-aware manner. This order exposes the measurements before summarization and allows every delta to be checked against a displayed pair. It also keeps a metric-specific effect visible when the aggregate row ordering is unchanged.

Dataset	Module	Configuration	PSNR	SSIM	MS-SSIM	LPIPS
R-C	PACER	w/o PACER	12.61	**0.78**	0.46	0.30
	TRO	w/o TRO	12.70	0.77	0.45	0.30
	WRD	w/o WRD	14.38	0.77	**0.63**	0.24
	RRG	w/o RRG	14.37	0.77	0.63	**0.23**
	RLI	w/o RLI	14.33	0.77	0.63	0.23
	SPS	w/o SPS	14.18	0.78	0.61	0.25
	ACR	w/o ACR	**14.41**	0.77	0.63	0.24
	LUMEN	Full	**19.40**	**0.86**	**0.84**	**0.18**
R-G	PACER	w/o PACER	15.23	**0.85**	0.44	0.22
	TRO	w/o TRO	15.26	0.85	0.44	0.22
	WRD	w/o WRD	15.27	0.85	0.44	0.22
	RRG	w/o RRG	**17.80**	0.85	**0.71**	**0.20**
	RLI	w/o RLI	15.76	0.84	0.59	0.23
	SPS	w/o SPS	15.54	0.85	0.48	0.23
	ACR	w/o ACR	15.23	0.85	0.44	0.22
	LUMEN	Full	**27.72**	**0.93**	**0.95**	**0.12**
R-N	PACER	w/o PACER	26.80	0.82	0.79	0.24
	TRO	w/o TRO	26.79	0.82	0.79	0.24
	WRD	w/o WRD	26.92	**0.84**	0.77	0.23
	RRG	w/o RRG	26.72	0.80	0.79	**0.20**
	RLI	w/o RLI	**27.99**	0.79	0.78	0.20
	SPS	w/o SPS	27.53	0.82	**0.80**	0.22
	ACR	w/o ACR	26.41	0.81	0.74	0.23
	LUMEN	Full	**28.25**	**0.89**	**0.82**	**0.15**
R-R	PACER	w/o PACER	10.94	0.79	0.55	0.24
	TRO	w/o TRO	10.97	0.79	0.55	0.23
	WRD	w/o WRD	16.73	0.86	0.81	0.15
	RRG	w/o RRG	**17.42**	**0.87**	**0.83**	**0.12**
	RLI	w/o RLI	16.62	0.86	0.81	0.13
	SPS	w/o SPS	13.73	0.83	0.71	0.19
	ACR	w/o ACR	16.72	0.86	0.81	0.15
	LUMEN	Full	**18.12**	**0.88**	**0.85**	**0.12**

Table S6: Matched component removals evaluated with four reconstruction metrics. Full LUMEN is the final row in every dataset block; the metric columns retain their native directions.

The matched-removal table first reports endpoint values under shared evaluation conditions. Each component is compared with the same Full configuration before the more targeted pathway interventions are considered.

The grouping by dataset makes the intervention comparison local to each protocol. A change in one component is not averaged with a different optical condition before it is inspected. The final Full row in each block provides the reference endpoint, and the four metric columns expose whether a component affects numerical, structural, perceptual, or multiscale agreement. The resulting differences quantify reconstruction sensitivity to each intervention.

The canonical removals are grouped by stage. PACER, TRO, WRD, and RRG concern the formation and transport of evidence from the wall observation, whereas RLI, SPS, and ACR concern its use during structural readout. This grouping connects the endpoint study to the architecture diagram through the components' evidence-routing and structural-readout roles. The output of every variant is scored against the same aligned target, so differences among rows arise on the same complete aligned image support.

Metric-level changes need not have identical magnitude. Removing a routing operation may strongly reduce PSNR while preserving part of the connected support measured by SSIM, and changing structural refinement may alter perceptual distance more than average pixel fidelity. The table retains this disagreement because it helps distinguish broad degradation from a narrower endpoint effect. A single row-level color summarizes overall placement only; the red and blue numerals remain the authoritative best and second-best indicators for each metric.

Cross-protocol variation is expected because attenuation, multipath, spectrum, and wall response differ among R-C, R-R, R-N, and R-G. The same removal can therefore produce different endpoint changes without contradicting a shared architectural role. Each protocol block closes with its own Full result, allowing the reader to compare a variant with the intact method under the same scene regime before comparing patterns across regimes. Protocol-specific Full values retain these scene-dependent differences.

Dataset	Module	Category	Configuration	PSNR	SSIM	MS-SSIM	LPIPS
R-C	RRG	Grad.	RRG stop-grad.	14.34	0.77	0.63	0.23
	WRD	Uniform	WRD uniform	14.34	0.77	0.63	0.23
	RLI	Grad.	RLI no-grad.	14.80	**0.78**	0.65	0.23
	PACER–ECHO	Grad.	cross-stage grad cut	14.75	0.78	0.65	0.23
	SPS	Uniform	SPS uniform	14.38	0.77	0.63	0.23
	TRO	Grad.	TRO no-grad.	**14.91**	0.78	**0.66**	**0.22**
	LUMEN	Full	Full	**19.40**	**0.86**	**0.84**	**0.18**
R-G	RRG	Grad.	RRG stop-grad.	17.84	0.85	0.70	0.20
	WRD	Uniform	WRD uniform	17.99	0.86	0.71	0.20
	RLI	Grad.	RLI no-grad.	17.14	0.85	0.65	0.21
	PACER–ECHO	Grad.	cross-stage grad cut	16.62	0.85	0.59	0.21
	SPS	Uniform	SPS uniform	16.11	0.85	0.53	0.22
	TRO	Grad.	TRO no-grad.	**20.55**	**0.88**	**0.81**	**0.17**
	LUMEN	Full	Full	**27.72**	**0.93**	**0.95**	**0.12**
R-N	RRG	Grad.	RRG stop-grad.	**27.88**	**0.88**	**0.81**	**0.26**
	WRD	Uniform	WRD uniform	20.41	0.75	0.52	0.40
	RLI	Grad.	RLI no-grad.	20.43	0.75	0.52	0.40
	PACER–ECHO	Grad.	cross-stage grad cut	20.41	0.75	0.52	0.40
	SPS	Uniform	SPS uniform	20.43	0.75	0.52	0.40
	TRO	Grad.	TRO no-grad.	26.29	0.87	0.77	0.26
	LUMEN	Full	Full	**28.25**	**0.89**	**0.82**	**0.15**
R-R	RRG	Grad.	RRG stop-grad.	16.09	0.85	0.80	**0.16**
	WRD	Uniform	WRD uniform	**16.60**	**0.86**	**0.81**	0.16
	RLI	Grad.	RLI no-grad.	14.54	0.84	0.74	0.17
	PACER–ECHO	Grad.	cross-stage grad cut	14.65	0.84	0.75	0.17
	SPS	Uniform	SPS uniform	13.55	0.83	0.70	0.19
	TRO	Grad.	TRO no-grad.	9.25	0.80	0.35	0.23
	LUMEN	Full	Full	**18.12**	**0.88**	**0.85**	**0.12**

Table S7: Pathway interventions. Grad. changes gradient access while retaining forward computation; Uniform replaces observation-dependent selection; Full denotes intact LUMEN. Full LUMEN is the final row in every dataset block; the four metric columns retain their native directions.

The pathway table refines that comparison by changing information flow or gradient access while holding the surrounding configuration fixed. This isolates the intervention being tested with the surrounding computation retained.

The pathway records complement complete component removal because they preserve more of the surrounding computation. Their purpose is to test whether a targeted change in access or propagation is reflected at the output. The table should be read against the matched-removal table: consistent changes across both views support endpoint sensitivity to the selected component or path, whereas a metric-specific change identifies a narrower effect.

The pathway records are read as controlled endpoint perturbations. A row that changes a routing or gradient path alters the information available to the rest of the model, while evaluation conditions remain unchanged. Comparing that row with the matched Full row indicates how sensitive the final reconstruction is to the selected path. The measured output differences quantify sensitivity to the selected pathway.

Two interpretations are retained. A complete removal measures the endpoint effect of omitting a component, while a pathway intervention measures the effect of changing route or gradient access with the other modules present. The four metric columns expose changes in numerical fidelity, structural similarity, perceptual distance, and multiscale structural agreement.

Stop-gradient and no-gradient variants change optimization access while retaining the corresponding forward computation. Uniform replacements preserve the surrounding tensor interface but remove observation-dependent selection. A cross-stage cut interrupts gradient exchange between the two major stages while retaining their forward connection. Within the same surrounding architecture, observation-dependent selection improves the reconstructed body structure. The Category column identifies gradient restrictions, uniform replacements, and the full model. On R-G, SPS uniform reaches MS-SSIM 0.53 and the cross-stage gradient cut reaches 0.59, compared with Full at 0.95. These results support learning the PACER–ECHO connection and selecting structural candidates from current wall evidence, with the surrounding computations retained.

The intact LUMEN row is again placed last within each dataset block. This makes it possible to compare every pathway intervention with one common endpoint without searching across the table. Highlighting begins at the first numerical column, leaving Dataset, Module, Category, and Configuration as black descriptive text. The formatting separates what was changed from how the

output responded, which is essential when two interventions share a similar endpoint value but alter different parts of the information flow.

Dataset	Module	dP	dS	dM	dL
R-C	PACER	6.79	**0.08**	0.38	0.12
	TRO	6.70	0.09	0.39	0.12
	WRD	5.02	0.09	**0.21**	0.06
	RRG	5.03	0.09	0.21	**0.05**
	RLI	5.07	0.09	0.21	0.05
	SPS	5.22	0.08	0.23	0.07
	ACR	**4.99**	0.09	0.21	0.06
	LUMEN	**0.00**	**0.00**	**0.00**	**0.00**
R-G	PACER	12.49	**0.08**	0.51	0.10
	TRO	12.46	0.08	0.51	0.10
	WRD	12.45	0.08	0.51	0.10
	RRG	**9.92**	0.08	**0.24**	**0.08**
	RLI	11.96	0.09	0.36	0.11
	SPS	12.18	0.08	0.47	0.11
	ACR	12.49	0.08	0.51	0.10
	LUMEN	**0.00**	**0.00**	**0.00**	**0.00**
R-N	PACER	1.45	0.07	0.03	0.09
	TRO	1.46	0.07	0.03	0.09
	WRD	1.33	**0.05**	0.05	0.08
	RRG	1.53	0.09	0.03	**0.05**
	RLI	**0.26**	0.10	0.04	0.05
	SPS	0.72	0.07	**0.02**	0.07
	ACR	1.84	0.08	0.08	0.08
	LUMEN	**0.00**	**0.00**	**0.00**	**0.00**
R-R	PACER	7.18	0.09	0.30	0.12
	TRO	7.15	0.09	0.30	0.11
	WRD	1.39	0.02	0.04	0.03
	RRG	**0.70**	**0.01**	**0.02**	**0.00**
	RLI	1.50	0.02	0.04	0.01
	SPS	4.39	0.05	0.14	0.07
	ACR	1.40	0.02	0.04	0.03
	LUMEN	**0.00**	**0.00**	**0.00**	**0.00**

Table S8: Direction-aligned component-removal changes computed from the displayed values in Table S6. dP, dS, and dM are Full minus variant; dL is variant minus Full. Positive values indicate reduced reconstruction quality. dP is in dB; the other columns retain their metric scales. Full LUMEN is the final row in every dataset block.

Table S8 reports direction-aligned changes caused by component removal relative to Full LUMEN. Positive values indicate reduced reconstruction quality; PSNR changes are in dB and the remaining columns retain their metric scales.

Because the delta columns are direction aligned, positive or negative entries can be compared across the four metrics without changing the raw values in the preceding tables. The differences quantify the loss relative to the intact model. It helps locate the largest endpoint changes and check whether a small aggregate hides a larger metric-specific shift, but it remains bounded by the measured configurations.

C.1 Scene-Dependent Intervention Effects

The three ablation tables form a progression from architecture to pathway to direction-aligned endpoint change. Component removal asks whether the complete reconstruction chain still reaches the Full endpoint when one named computation is absent. Pathway intervention keeps more of the architecture intact but alters the information available along a selected route. The delta table then places the component-removal changes in one direction-aligned frame. Removal effects include the network's response to the missing computation.

PACER-side interventions are read against the first physical challenge. RRG changes the relative support assigned to features, TRO changes how evidence is transported across scale, and WRD changes the separation and fusion of local departures from wall context. Their endpoint losses support the need for evidence formation and routing under weak mixed observations. Together, these operations regulate which weak wall cues reach structural reconstruction.

ECHO-side interventions are read against the second challenge. RLI changes how routed evidence enters global interaction, SPS changes which structural candidates remain compatible with that evidence, and ACR changes the sequence of corrections used before final mapping. A reduction in PSNR or structural similarity, or an increase in LPIPS, indicates sensitivity of the final reconstruction

to that operation. These interventions connect the information path to measured structural and perceptual agreement.

The joint Full row closes each protocol block and provides the common complete-model reference. Placing it last makes two checks immediate: all ablated rows precede the intact configuration, and every direction-aligned change can be recomputed from a displayed Full endpoint. The red and blue metric typography is determined separately within each numerical column. Thus a row-level highlight summarizes overall placement, while colored numerals retain metric-level best and second-best information even when different ablations exchange position across metrics.

The intervention evidence is most persuasive when the tables agree at two levels. Repeated degradation after both removal and pathway restriction indicates that the final result depends on the named part of the computation. A larger response on one protocol than another can then be related to the different transport regime without pooling scales. Small and metric-specific changes identify different sensitivities across acquisition regimes. This reporting rule keeps the claims proportional to the measured endpoints and leaves room for interactions among PACER, ECHO, and the final mapping.

D Single-Variable Robustness

The robustness appendix separates absolute response from ranking and endpoint change. The paired figures show measured nodes as markers and the pointwise strongest-baseline envelope in blue. Lines connect the displayed measurements. Response curves trace how human reconstruction changes as disturbances alter wall evidence; Tables S9–S12 summarize measured levels, retained sweeps, endpoint changes, and matched-node ranks.

D.1 Cross-Protocol Response Traces

Figure S5 spans Gaussian noise, scatter noise, and motion blur across the four protocols. Tables S9, S10, and S11 retain the measured endpoints, while Table S12 summarizes response, rank, percentile, and top-three metric-placement counts.

The trace figure and endpoint tables are paired deliberately. A trace exposes whether separation is retained, narrows, or crosses as severity changes, while an endpoint table preserves exact values for a specified measured coordinate. The red curve represents LUMEN at the retained nodes, and the dashed blue curve is the strongest available baseline value recomputed at each of those nodes. Because the envelope can change method along a sweep, it is more demanding than comparing LUMEN with one fixed competitor selected at the clean endpoint.

All curves use frozen checkpoints and matched observations. The perturbation is generated before inference, and the same perturbed observation is supplied to every method associated with that dataset. Input size, test sample, node coordinate, perturbation realization, and metric implementation remain matched. Line segments are drawn only after the node values have been retained, keeping the visual progression connected to direct evaluations under fixed weights.

The three displayed higher-is-better metrics emphasize numerical and structural preservation as the wall pattern degrades. LPIPS remains available in the accompanying numerical tables with its native lower-is-better direction. The three plotted columns share a higher-is-better direction. A favorable visual trajectory is therefore checked against all four core metrics before the text describes the model as leading at a retained condition.

Protocol boundaries remain visible because equal perturbation parameters can interact differently with corridor multipath, room-scale mixing, the near-infrared response, or the low-light garage path. A severity coordinate identifies the corruption applied to the registered input, not a universal amount of reconstruction difficulty shared across datasets. The endpoint blocks and pooled summaries consequently retain their dataset labels, and comparisons are made locally before any cross-protocol pattern is discussed.

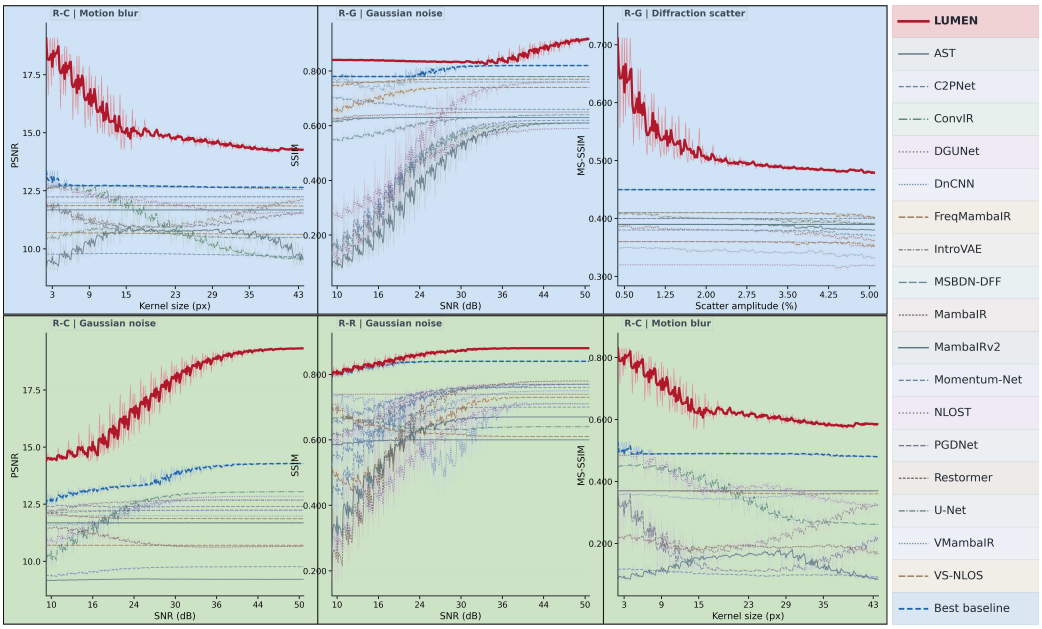

Figure S5: Measured single-variable perturbation traces for Gaussian noise, scatter noise, and motion blur. Columns are PSNR, SSIM, and MS-SSIM. Red denotes LUMEN and dashed blue the pointwise baseline envelope; markers are measured nodes.

Dataset	Model	PSNR	SSIM	MS-SSIM	LPIPS
R-C	ConvIR	13.04	0.66	0.51	**0.26**
	DeepSN	**14.28**	0.65	**0.60**	0.33
	IntroVAE	12.67	0.63	0.45	0.33
	MambaIR	11.68	**0.71**	0.37	0.33
	Restormer	10.65	0.66	0.23	0.36
	LUMEN	**18.77**	**0.85**	**0.82**	**0.17**
R-G	ConvIR	**15.68**	**0.78**	**0.45**	**0.29**
	DeepSN	13.20	0.58	0.38	0.57
	IntroVAE	13.70	0.76	0.40	0.40
	MambaIR	13.81	0.65	0.39	0.44
	Restormer	13.56	0.62	0.40	0.53
	LUMEN	**16.66**	**0.86**	**0.64**	**0.21**
R-N	ConvIR	19.12	0.70	0.45	0.39
	DeepSN	25.17	0.79	0.77	0.24
	IntroVAE	**27.58**	**0.87**	0.79	0.19
	MambaIR	26.96	0.84	**0.80**	0.21
	Restormer	25.44	0.82	0.77	**0.18**
	LUMEN	**27.94**	**0.89**	**0.81**	**0.16**
R-R	ConvIR	7.62	0.64	0.04	0.37
	DeepSN	13.68	0.70	0.68	0.31
	IntroVAE	**15.95**	**0.84**	**0.78**	0.16
	MambaIR	9.32	0.74	0.35	0.30
	Restormer	14.31	0.78	0.73	**0.15**
	LUMEN	**17.87**	**0.88**	**0.84**	**0.12**

Table S9: Gaussian-noise endpoint summaries at SNR 50 dB in all four datasets. Endpoints have the largest recorded SNR, corresponding to weaker Gaussian noise. All included nodes have four metrics; LPIPS is lower-is-better.

At the weaker-noise endpoint of 50 dB, the largest recorded SNR, LUMEN reaches SSIM 0.85, 0.88, 0.89, and 0.86 on R-C, R-R, R-N, and R-G, respectively (Table S9). The following scatter noise table reports R-C at relative amplitude 5%.

The endpoint tables and response traces describe complementary parts of each sweep. The trace shows how relative ordering evolves over retained nodes, while the table gives a precise multi-method view at the recorded endpoint. This distinction matters when curves cross: one endpoint documents a local comparison, but it cannot summarize every intermediate node in the measured sweep.

The matched layout makes differences between Gaussian noise and scatter noise readable across their measured severity nodes. The scatter-noise block is reported for the protocol with the retained structured-transport sweep.

Dataset	Model	PSNR	SSIM	MS-SSIM	LPIPS
R-C	ConvIR	13.03	0.66	0.51	0.26
	DeepSN	13.27	0.66	0.50	0.39
	IntroVAE	12.64	0.63	0.44	0.33
	MambaIR	11.68	**0.71**	0.37	0.33
	Restormer	**14.18**	0.69	**0.61**	**0.23**
	LUMEN	**17.31**	**0.82**	**0.75**	**0.22**

Table S10: Scatter noise endpoint summary on R-C at relative scatter amplitude 5%. The table summarizes the R-C sweep.

Gaussian noise adds stochastic intensity variation, while scatter noise changes the transport response in a structured way. At a 5% relative scatter amplitude on R-C, LUMEN achieves 17.31 dB PSNR, 0.82 SSIM, 0.22 LPIPS, and 0.75 MS-SSIM (Table S10), supporting structural recovery from the perturbed wall pattern.

Figure S6 emphasizes that the same four core metrics are evaluated under different corruption families and datasets. LPIPS retains its lower-is-better direction in the numerical summary, while the plotted columns show the three higher-is-better metrics. Table S12 retains the complete measured response, rank, percentile, and top-three metric-placement evidence.

Across these displays, the baseline envelope is a reference assembled at each retained node, selecting the strongest competing value at each coordinate. It therefore answers a strict local question: how does the LUMEN value compare with the strongest available competing value at this same severity? The endpoint tables answer the same question at selected endpoint nodes, while the pooled summary reports how often the comparison is favorable over the retained sweep. Keeping these views together avoids reading a smooth trace as evidence at levels that were never measured.

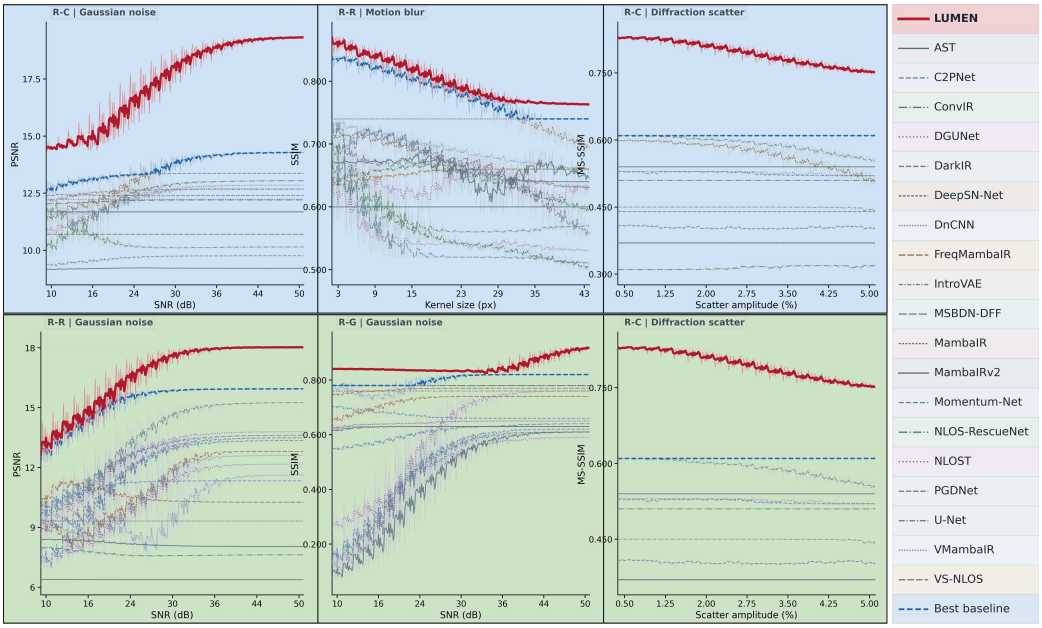

Figure S6: Cross-protocol robustness traces at measured nodes. The upper and lower rows retain distinct measured corruption and dataset combinations. Red and dashed blue encode LUMEN and the complete baseline envelope, respectively; no interpolated node is treated as a measurement.

At the largest evaluated motion-blur kernels—15 px on R-C, 43 px on R-G, and 7 px on R-R—LUMEN retains MS-SSIM of 0.59, 0.47, and 0.75, respectively (Table S11). The table is kept beside the pooled summary so endpoint behavior can be checked against the measured sweep.

Taken together, the three endpoint tables preserve the asymmetry of the available records. Each table covers the available protocol–perturbation combinations. Each row identifies a retained node, and each highlighted comparison is evaluated within its own dataset block and metric direction.

The pooled robustness summary aggregates the same retained nodes into mean, worst-case, rank, and top-three metric-placement views. Across 75, 42, 39, and 22 evaluated nodes on R-C, R-G, R-N, and R-R, LUMEN records 300, 168, 156, and 88 Top-3 metric placements, respectively, among the six methods in Table S12, covering four metric placements at each node.

Dataset	Model	PSNR	SSIM	MS-SSIM	LPIPS
R-C	ConvIR	11.97	0.64	0.42	0.32
	DeepSN	11.98	0.63	0.34	0.42
	IntroVAE	10.92	0.58	0.16	0.39
	MambaIR	**12.72**	**0.77**	**0.49**	**0.26**
	Restormer	9.92	0.62	0.17	0.40
	LUMEN	**14.62**	**0.78**	**0.59**	**0.24**
R-G	ConvIR	**15.68**	**0.78**	**0.45**	**0.29**
	DeepSN	13.52	0.58	0.40	0.63
	IntroVAE	13.73	0.64	0.40	0.42
	MambaIR	14.28	0.68	0.40	0.41
	Restormer	13.55	0.61	0.40	0.54
	LUMEN	**15.78**	**0.84**	**0.47**	**0.23**
R-R	ConvIR	7.66	0.65	0.05	0.36
	DeepSN	10.47	0.72	0.50	0.35
	IntroVAE	**15.03**	**0.83**	**0.75**	**0.17**
	MambaIR	9.32	0.74	0.35	0.30
	Restormer	12.23	0.72	0.62	0.23
	LUMEN	**15.14**	**0.84**	**0.75**	**0.16**

Table S11: Motion-blur endpoint summaries at kernel sizes 15 px on R-C, 43 px on R-G, and 7 px on R-R, the largest recorded kernel in each dataset.

Dataset	Model	mP	wP	mS	wS	mM	wM	mL	wL	rank	Q	Top3
R-C	ConvIR	12.90	11.82	0.66	0.62	0.50	**0.42**	**0.27**	**0.32**	**3.46**	**50.73**	135
	DeepSN	**13.82**	**11.98**	0.63	0.27	**0.55**	0.29	0.37	0.63	4.03	39.33	136
	IntroVAE	12.54	10.92	0.63	0.57	0.43	0.16	0.33	0.39	4.73	25.40	28
	MambaIR	11.78	11.68	**0.72**	**0.71**	0.38	0.37	0.32	0.33	4.30	33.93	95
	Restormer	12.85	9.92	0.67	0.45	0.46	0.16	0.30	0.56	3.47	50.60	**206**
	LUMEN	**18.28**	**14.62**	**0.84**	**0.78**	**0.79**	**0.59**	**0.20**	**0.24**	**1.00**	**100.00**	**300**
R-G	ConvIR	**15.68**	**15.68**	**0.78**	**0.78**	**0.45**	**0.45**	**0.29**	**0.29**	**2.02**	**79.64**	**168**
	DeepSN	13.41	13.11	0.58	0.50	0.38	0.37	0.59	0.67	5.70	5.95	5
	IntroVAE	13.75	13.70	0.66	0.64	0.40	0.40	0.41	0.42	3.65	47.02	66
	MambaIR	13.71	11.64	0.70	0.65	0.37	0.30	0.42	0.49	3.78	44.40	95
	Restormer	13.51	12.82	0.61	0.60	0.40	0.39	0.54	0.57	4.85	22.98	2
	LUMEN	**16.29**	**15.71**	**0.85**	**0.83**	**0.57**	**0.47**	**0.21**	**0.25**	**1.00**	**100.00**	**168**
R-N	ConvIR	19.07	18.88	0.70	0.69	0.34	0.19	0.39	0.40	5.22	15.64	0
	DeepSN	24.94	24.63	0.78	0.67	0.76	0.74	0.24	0.47	4.04	39.10	38
	IntroVAE	**27.15**	**26.58**	**0.87**	**0.87**	**0.78**	**0.77**	**0.19**	**0.19**	**2.29**	**74.10**	**150**
	MambaIR	17.15	9.21	0.62	0.40	0.24	0.01	0.57	0.73	5.21	15.77	35
	Restormer	25.19	24.96	0.81	0.74	0.76	0.75	0.19	0.37	3.23	55.38	89
	LUMEN	**27.46**	**26.85**	**0.88**	**0.88**	**0.80**	**0.79**	**0.16**	**0.16**	**1.00**	**100.00**	**156**
R-R	ConvIR	7.61	7.55	0.64	0.63	0.04	0.03	0.38	0.39	5.83	3.41	0
	DeepSN	12.78	10.47	0.64	0.35	0.62	0.43	0.38	0.57	4.66	26.82	0
	IntroVAE	**15.79**	**15.03**	**0.84**	**0.83**	**0.78**	**0.75**	**0.16**	**0.17**	**2.09**	**78.18**	**88**
	MambaIR	9.32	9.32	0.74	0.74	0.35	0.35	0.30	0.30	4.41	31.82	8
	Restormer	13.79	12.23	0.75	0.63	0.70	0.62	0.20	0.36	3.01	59.77	80
	LUMEN	**17.26**	**15.14**	**0.87**	**0.84**	**0.82**	**0.75**	**0.13**	**0.16**	**1.00**	**100.00**	**88**

Table S12: Single-variable robustness summary over the retained nodes. m and w denote mean and worst-case values in the native metric direction; rank, Q, and Top3 are direction-aligned summaries over the six displayed methods. Top3 counts top-three metric placements; each evaluated node contributes four metric placements. The retained node counts are 75, 42, 39, and 22 for R-C, R-G, R-N, and R-R.

The pooled summary is most informative when used as a cross-check on the endpoint measurements. Mean response describes the central tendency of retained nodes, worst response exposes the weakest measured case, and rank-based quantities describe relative placement against the local baseline envelope. None of these summaries creates a new noise level or replaces the node-level measurements shown in the supporting tables.

D.2 Noise Families and Matched-Node Comparisons

The three single-variable corruptions alter different aspects of the observed wall pattern. Gaussian noise adds spatially distributed intensity variation, scatter noise introduces structured transport-like disturbance, and motion blur spreads local contrast along a kernel trajectory. They are therefore retained as separate experiment families even when the same four reconstruction metrics are used. A common metric panel makes their effects comparable, but a severity coordinate from one family is never numerically equated with a coordinate from another.

Every reported comparison is matched at the node level. LUMEN and all baselines use the same clean input, corruption rule, strength coordinate, random realization, and preprocessing; outputs are scored against the same reference target. The strongest-baseline envelope is recomputed at each node because the identity of the leading baseline may change with severity or metric. This is stricter than comparing with the average of a fixed subset, and it prevents a method from appearing favorable merely because a competing trajectory deteriorates under an unmatched perturbation realization.

Endpoint, sweep, and percentile views answer three distinct questions. The endpoint tables expose an exact multi-method comparison at the largest recorded coordinate: higher SNR means weaker Gaussian noise, while larger scatter amplitude or blur kernel means stronger perturbation. Sweep summaries describe behavior across all retained coordinates, including a worst value and spread. Percentile maps describe relative position among methods at each coordinate before aggregation. Agreement among these views indicates that a favorable endpoint is not isolated; disagreement identifies where the ordering changes and directs the reader back to the node-level trace.

Every robustness node uses zero-shot evaluation. A source-trained checkpoint is frozen and applied directly to the perturbed test input. Parameters, thresholds, and normalization statistics remain fixed as corruption strength follows the sweep coordinates. Ground truth enters only after inference for metric computation. The sweep therefore tests one fixed reconstruction rule.

The response curves connect measured anchors only to make progression legible. They are not fitted dose-response laws, and the area between anchors is not counted as additional evidence. A curve crossing is interpreted using its adjacent measured coordinates. The pooled table retains the number of valid nodes so that datasets with different measured coverage are not silently given the same apparent sample count.

D.3 Within-Node Relative Performance

Figure S7 aggregates matched-node quality percentiles for R-C, R-R, R-N, and R-G. The percentiles summarize relative competitiveness across corruption families. Ties receive the same minimum rank and LPIPS is direction-reversed before percentile conversion. The heatmap complements the raw endpoint values.



Figure S7: Within-node robustness quality percentiles across R-C, R-R, R-N, and R-G. Each cell aggregates matched measured nodes for one method and core metric; darker cells indicate larger direction-aligned percentiles and the red outline identifies LUMEN.

E Single-Variable Physical Perturbations

Physical perturbations change the acquisition geometry or wall response while preserving the relay-wall task. Figure S8 shows PSNR, SSIM, and MS-SSIM in three columns across a 2 by 3 plate. The measurements cover hidden-target distance, camera distance, camera jitter, wall material, and camera angle. Tables S13 and S14 report matched-node ranks and pooled distribution summaries.

The two rows contain different sweeps; the lower row includes both wall-material and camera-angle changes. The metric columns remain PSNR, SSIM, and MS-SSIM. This makes the direction of a response change readable without combining variables that have different physical meanings. The red trace is compared with the blue pointwise envelope at retained nodes; the visual connection between nodes is only a rendering aid. The rank table below then summarizes the same measured coordinates, while the box plots pool the five sweeps to expose how much variation is hidden by an endpoint trace.

At the acquisition level, the sweeps separate changes in geometry from changes in the wall response. Hidden-target distance and camera distance alter path length and attenuation; jitter and camera

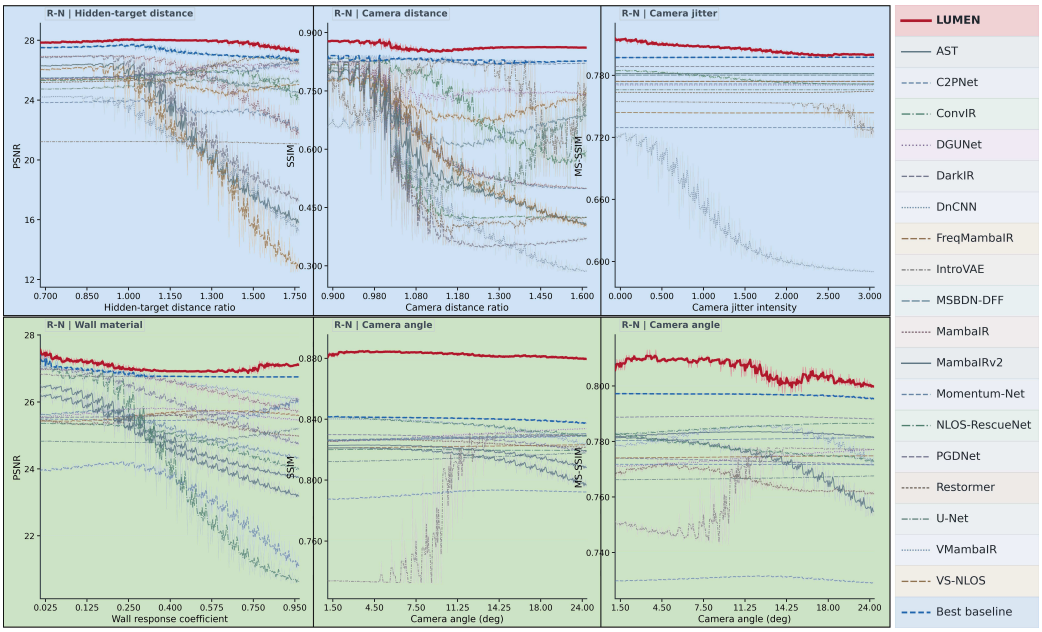

Figure S8: Single-variable physical response traces. The two rows retain separate measured sweeps and the three columns show PSNR, SSIM, and MS-SSIM. Red denotes LUMEN and dashed blue denotes the pointwise baseline envelope. All markers correspond to measured nodes; the local trace rendering does not introduce samples.

angle alter sampling geometry; wall material changes the diffuse response itself. The common metric columns make their endpoint effects comparable, but the five sweeps are not collapsed into one physical variable. This separation is why rank and distribution summaries accompany the factor-specific traces.

Perturbation	Method	rank	mP	mS	mL	mM
camera angle	ConvIR	**2.86**	**27.05**	**0.84**	**0.17**	0.78
	DeepSN	5.84	10.72	0.56	0.33	0.67
	IntroVAE	4.04	24.31	0.79	0.19	0.76
	MambaIR	2.89	26.92	0.84	0.21	**0.80**
	Restormer	4.37	25.51	0.82	0.17	0.77
	LUMEN	**1.00**	**27.50**	**0.88**	**0.16**	**0.81**
camera dist.	ConvIR	**2.88**	**23.34**	**0.70**	0.32	0.64
	DeepSN	5.83	8.75	0.46	0.45	0.53
	IntroVAE	3.79	13.78	0.62	**0.28**	**0.70**
	MambaIR	3.21	20.90	0.67	0.37	0.66
	Restormer	4.29	17.20	0.63	0.36	0.63
	LUMEN	**1.00**	**27.06**	**0.87**	**0.20**	**0.80**
jitter	ConvIR	**2.48**	**27.36**	**0.84**	**0.16**	0.78
	DeepSN	6.00	8.65	0.53	0.34	0.65
	IntroVAE	5.00	20.76	0.73	0.22	0.75
	MambaIR	2.77	26.97	0.84	0.21	**0.80**
	Restormer	3.75	25.33	0.82	0.17	0.76
	LUMEN	**1.00**	**27.84**	**0.88**	**0.16**	**0.81**
target dist.	ConvIR	**2.34**	**27.26**	**0.84**	**0.16**	**0.78**
	DeepSN	5.21	15.75	0.66	0.29	0.70
	IntroVAE	5.08	21.19	0.73	0.22	0.75
	MambaIR	4.03	25.62	0.80	0.25	0.74
	Restormer	3.34	25.82	0.83	0.17	0.77
	LUMEN	**1.00**	**27.87**	**0.88**	**0.16**	**0.81**
wall	ConvIR	4.53	24.21	0.79	0.31	0.71
	DeepSN	5.83	8.65	0.53	0.36	0.65
	IntroVAE	**2.70**	**26.76**	**0.83**	**0.17**	**0.78**
	MambaIR	3.14	26.55	0.83	0.26	0.77
	Restormer	3.80	25.38	0.83	0.20	0.76
	LUMEN	**1.00**	**27.05**	**0.88**	**0.16**	**0.80**

Table S13: Matched four-metric summaries for the five single-factor physical sweeps on R-N. Numeral colors use displayed values; ties favor LUMEN, then table order. Each row aggregates the directly measured nodes for one perturbation family; LUMEN is the final row in every block.

The rank table compares methods at matched physical nodes across the five single-factor sweeps. The distribution table below retains the within-metric spread that a rank alone cannot show.

Matched-node ranking and pooled spread are complementary because physical factors can change both the center and the variability of the response. The rank view asks which method is strongest at the same coordinate, while the distribution view asks how retained values are distributed after the five sweeps are combined. Reporting both prevents a favorable average from hiding a localized crossing or a broad response spread.

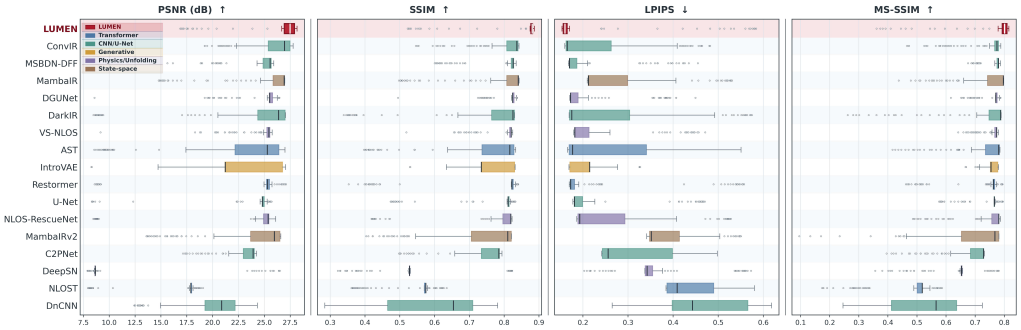

Figure S9: Pooled physical single-factor quality distributions on R-N. Boxes show measured interquartile ranges, center lines show medians, and individual points retain observations outside the whiskers. Panels report PSNR, SSIM, LPIPS, and MS-SSIM for the displayed methods; the boxes summarize empirical variability across the 105 evaluated physical conditions on R-N.

Figure S9 provides a distributional complement to the traces. Its pooled observations combine five directly measured physical sweeps, so a wide box can reflect variation between perturbation families as well as variation within one family. Tables S14 and S13 retain the quartiles, pointwise win rates, metric means, and matched four-metric rank used for cross-checking.

Metric	Method	rank	Mean	Q1	Median	Q3	Win
LPIPS	ConvIR	**2.73**	0.21	**0.16**	**0.16**	**0.17**	**0.00**
	DeepSN	5.75	0.34	0.34	0.34	0.35	0.00
	IntroVAE	3.86	0.20	0.17	0.21	0.22	0.00
	MambaIR	4.43	0.24	0.21	0.21	0.24	0.00
	Restormer	3.23	**0.19**	0.17	0.17	0.17	0.00
	LUMEN	**1.00**	**0.16**	**0.16**	**0.16**	**0.16**	**100.00**
MS-SSIM	ConvIR	3.56	0.75	0.77	0.78	0.78	**0.00**
	DeepSN	5.65	0.66	0.65	0.65	0.66	0.00
	IntroVAE	4.13	0.76	0.75	0.75	0.78	0.00
	MambaIR	**2.56**	**0.77**	**0.78**	**0.80**	**0.80**	0.00
	Restormer	4.10	0.76	0.76	0.76	0.77	0.00
	LUMEN	**1.00**	**0.80**	**0.80**	**0.80**	**0.81**	**100.00**
PSNR	ConvIR	**2.78**	**26.31**	26.52	**27.05**	**27.52**	**0.00**
	DeepSN	5.70	10.84	8.64	8.65	8.66	0.00
	IntroVAE	4.31	22.68	21.19	21.22	26.76	0.00
	MambaIR	3.05	26.20	**26.61**	26.94	26.97	0.00
	Restormer	4.15	25.03	25.31	25.35	25.57	0.00
	LUMEN	**1.00**	**27.54**	**27.21**	**27.58**	**27.93**	**100.00**
SSIM	ConvIR	3.04	**0.82**	**0.83**	**0.84**	**0.84**	**0.00**
	DeepSN	5.80	0.56	0.53	0.53	0.53	0.00
	IntroVAE	4.48	0.76	0.73	0.73	0.83	0.00
	MambaIR	**2.79**	0.82	0.83	0.84	0.84	0.00
	Restormer	3.90	0.81	0.82	0.82	0.83	0.00
	LUMEN	**1.00**	**0.88**	**0.88**	**0.88**	**0.88**	**100.00**

Table S14: Pooled distributions over the 105 physical nodes on R-N. Numeral colors use displayed values; ties favor LUMEN, then table order. Mean, quartiles, and rank retain native metric directions; Win is the pointwise best-or-tied percentage against the five displayed baselines.

The quartiles and pointwise win rates expose whether a method's physical response is concentrated and whether its advantage is retained across nodes. Together with the preceding ranks, they separate consistent placement from a favorable average.

The distribution table quantifies the box-plot summary. Its quartiles are computed from retained measurements, and its win rates compare methods at matched coordinates. This makes the table suitable for checking the visual spread and median while keeping the interpretation descriptive. On R-N, LUMEN's SSIM mean, median, and both quartiles are 0.88; MS-SSIM has mean 0.80 and quartiles 0.80–0.81, indicating concentrated structural agreement across the five sweeps.

The distributional view is especially useful for the physical study because a single geometry variable can produce very different radiometric responses at different nodes. A narrow interquartile range with

a favorable median indicates concentration of the measured response, whereas a wide range warns that the mean is mixing distinct operating regimes. The pooled construction is therefore descriptive: it compares the spread of the retained observations across the five sweeps, with quartiles describing the measured spread. The matched rank table and the distribution table answer complementary questions and should be consulted together.

E.1 Physical Factors and Controlled Variation

Each physical sweep changes one named factor while retaining the other registered settings. Camera distance modifies the wall-to-sensor leg and the sampled wall footprint. Camera angle changes projection and the spatial weighting of the visible patch. Camera jitter tests small departures from the nominal alignment. Hidden-target distance changes the source-wall-target and target-wall path lengths, while wall material changes the reflection and scattering response. Keeping these factors separate makes the associated degradation interpretable under the stated physical settings.

The single-factor protocol uses the same frozen-checkpoint rule as the noise study. The checkpoint, input transformation, and metric implementations are unchanged across the measured coordinates, and models and thresholds remain fixed throughout target-domain evaluation. The direct-view reference is paired to the observation at acquisition and is introduced only for scoring. Response changes therefore describe the frozen mapping at each capture coordinate.

Path-length variables and angular variables should not be expected to have identical response shapes. Distance primarily changes attenuation and spatial mixing, whereas angle and jitter also change correspondence between the observed wall support and the learned sampling pattern. Wall material can alter contrast and frequency content without moving the geometry. The trace plate therefore preserves the original horizontal coordinates and does not normalize all five factors into one generic severity axis. The pooled distribution is used only after these factor-specific traces have been displayed.

Matched-node rank complements the raw metric trajectory when different baselines respond differently to a physical change. At each coordinate, the comparison set receives the same observation and is ordered in the native metric direction. The pointwise envelope may consequently switch from one baseline family to another. This switching retains the strongest local comparison, making any reported LUMEN advantage conditional on the strongest available comparison at that measured coordinate.

The physical evidence covers the registered distances, angles, jitter levels, target positions, and wall materials. The dual-variable section extends this coverage to paired coordinates, where two named factors change together and all methods are evaluated on the same Cartesian grid. Reading the single-factor results first helps distinguish an interaction pattern from the marginal response of either axis alone.

F Dual-Variable Physical Response

For each physical pair, LUMEN and every compared baseline are directly evaluated on the same $25 \times 25 = 625$ grid. The figures simplify this grid to 25 measured display anchors; the accompanying statistics summarize all 625 nodes, except Best and Top3 counts over the display anchors. The combined tables retain means, spread, extrema, ranks, and node-placement counts. The weighted surface uses $S = 100(0.30\widehat{\mathrm{PSNR}} + 0.25\widehat{\mathrm{SSIM}} + 0.25\widehat{\mathrm{MS\text{-}SSIM}} + 0.20(1 - \widehat{\mathrm{LPIPS}}))$; metric-specific surfaces retain raw metric semantics.

F.1 Weighted Joint Response

Figure S10 shows the weighted physical manifolds for R-C, R-R, and R-G. Red surfaces and markers identify LUMEN, the blue surface is the pointwise connected baseline envelope, and subdued surfaces retain individual baseline variation. Tables S15 and S16 combine the physical and robustness dual-grid summaries, separating average response from matched-node ranks. Together they describe average quality and mean rank over all 625 nodes, with Best and Top3 counts over the 25 display anchors.

F.2 Metric-Specific Joint Response

Each weighted panel contains the same measured coordinate set for LUMEN and the baseline population. The score combines direction-aligned, normalized versions of PSNR, SSIM, MS-SSIM, and reverse LPIPS using the stated fixed weights. It summarizes joint metric behavior for visualization.

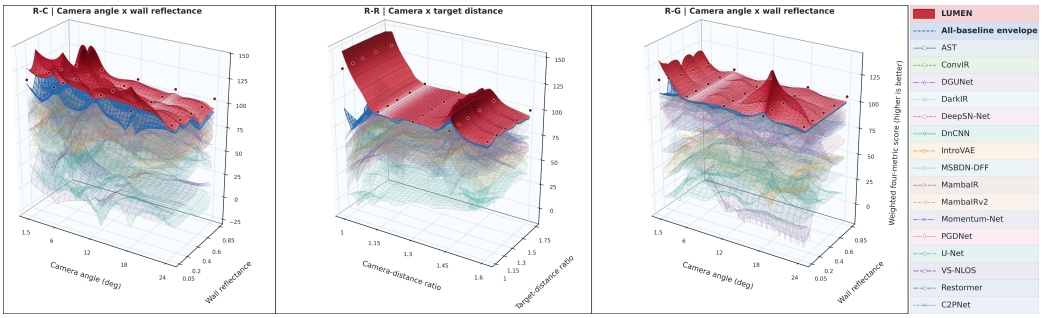

Figure S10: Weighted dual-variable physical response. Each panel displays 25 measured anchors from the full $25 \times 25 = 625$ grid evaluated for LUMEN and every frozen baseline. Height is $S = 100(0.30\widehat{\text{PSNR}} + 0.25\widehat{\text{SSIM}} + 0.25\widehat{\text{MS-SSIM}} + 0.20(1 - \widehat{\text{LPIPS}}))$; bounded connections make the measured grid legible. The right-hand legend lists LUMEN, the baseline envelope, and the individual baselines.

The raw metric surfaces below are therefore required to determine whether a high combined surface reflects broad agreement or compensation between individual measures.

The physical pairs represent complementary interactions. Camera angle with wall material combines projection change with reflection change on R-C and R-G. Camera distance with hidden-target distance combines two path-length terms on R-R. These pairings are kept in separate panels because their axes have different units and because a surface slope has meaning only relative to its named physical coordinates.

The metric-specific surfaces in Figures S11, S12, S13, and S14 prevent the weighted height from hiding metric disagreement. PSNR, SSIM, and MS-SSIM retain their native higher-is-better direction; the LPIPS panel displays one minus the raw perceptual distance so that the vertical direction is consistent, while the combined dual-response table preserves raw LPIPS together with its direction-aware summaries. The response and rank tables provide the matched-coordinate context for reading those surfaces.

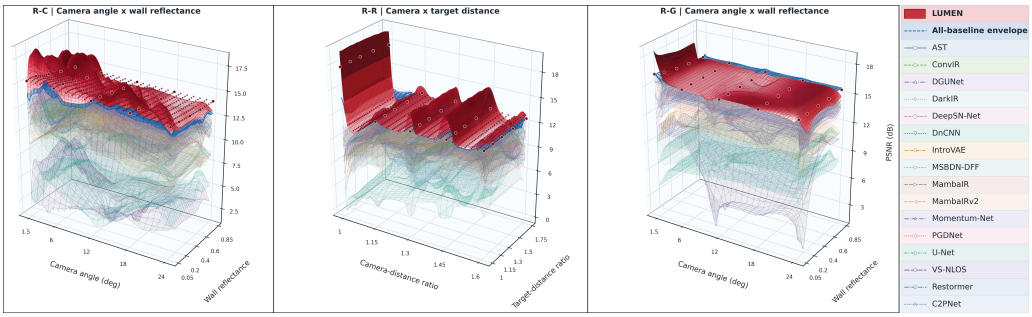

Figure S11: Metric-specific physical response for PSNR. The surfaces connect the 25 measured display anchors for each pair. Red marks LUMEN and blue marks the connected baseline envelope; the vertical axis is raw PSNR in dB.

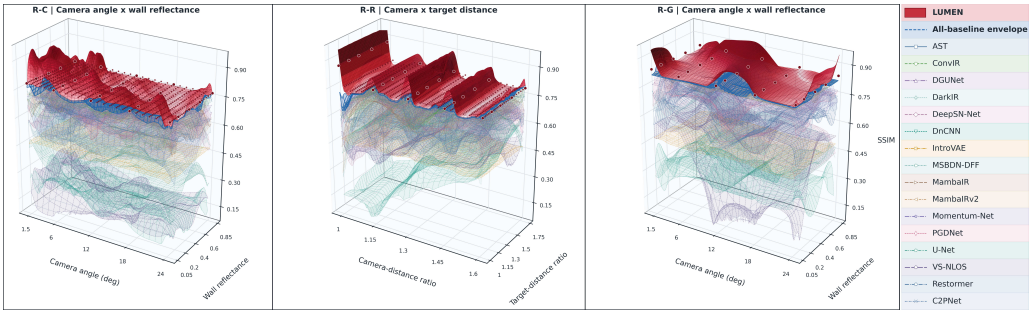

Figure S12: Metric-specific physical response for SSIM. The vertical axis is raw SSIM, higher is better. The common measured-grid and envelope conventions are unchanged from the weighted surface.

The four metric-specific views are read as a set: the weighted surface provides a common joint scale, whereas these panels keep each metric's native response visible. Agreement across surfaces indicates broad separation at the measured anchors; a crossing in one surface identifies a localized metric trade-off that the weighted score can conceal.

PSNR and SSIM provide complementary readings of the first two surfaces. A PSNR change can arise from distributed pixel error even when the body layout is preserved, whereas SSIM responds more directly to local structural agreement. Their joint inspection therefore distinguishes radiometric drift from a loss of coherent shape. The axes and anchor coordinates remain identical between panels, so a reader can compare the same physical node without estimating correspondence between different plots.

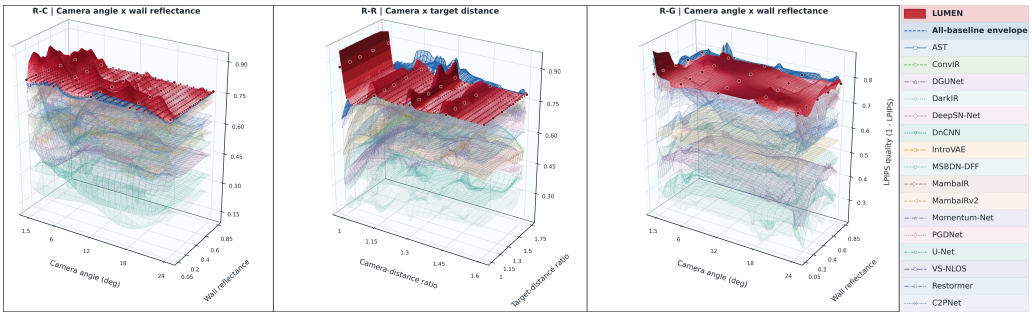

Figure S13: Metric-specific physical response for perceptual quality displayed as one minus raw LPIPS. Larger height therefore denotes lower perceptual distance; raw LPIPS values and direction-aware ranks remain available in the source tables.

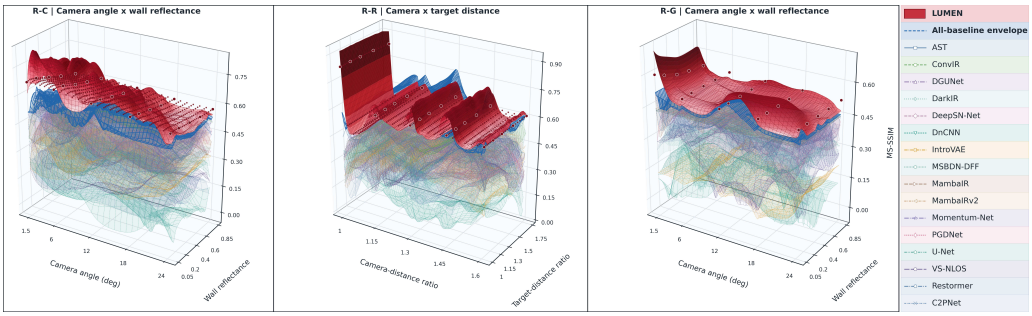

Figure S14: Metric-specific physical response for MS-SSIM. The vertical axis is raw MS-SSIM, higher is better, and all displayed nodes are retained measured coordinates.

G Dual-Variable Robustness

The dual-variable perturbation surface couples two observation noises at a time. Figure S15 uses the same fixed weighted quality definition as the physical surface, but the axes are scatter-noise amplitude, SNR, and motion-blur kernel. Table S15 summarizes the magnitude and variation of the paired response, while Table S16 reports the corresponding rankings and placement counts.

The response table is the direct numerical companion to the dual-variable surfaces. Its means, minima, and spread summarize all 625 nodes; the next table reports mean rank over the same full grid.

The paired-grid summary retains the mean, minimum, and spread over all 625 conditions. These quantities supply the numerical context for reading the red surface and connected baseline envelope.

The rank table reports mean rank and quality over all 625 nodes, with Best and Top3 counts over the 25 display anchors. Its full-grid summaries use the same matched coordinates as the response table.

The paired-grid response and rank tables are kept in this order so that absolute values precede relative placement. The response table shows the scale and spread of the full 625-node grid, and the rank table summarizes relative placement over that grid, with Best and Top3 counts over the display anchors. Reading them together is especially important near a crossing, where a similar mean can arise from different local patterns.

The matched-node view changes the question from absolute height to relative placement at the same physical coordinates. Its S and Q columns distinguish the weighted score from its ranking percentile,

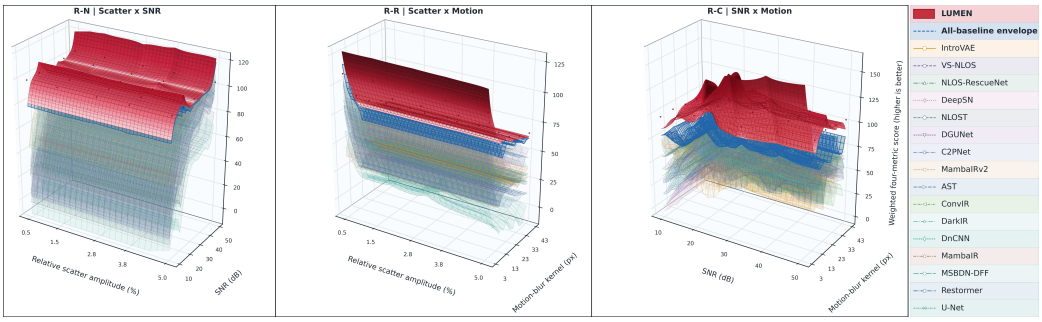

Figure S15: Weighted dual-variable noise-perturbation response. Panels show scatter-noise amplitude with SNR, scatter-noise amplitude with motion-blur kernel, and SNR with motion-blur kernel. Red denotes LUMEN, blue denotes the connected all-baseline envelope, and subdued wireframes retain individual baselines. The 25 measured anchors per panel form a display subset of each method's full $25 \times 25 = 625$ evaluation grid; surfaces connect those anchors.

with both summarized over all 625 nodes. Best and Top3 count nodes with rank 1 and rank at most 3, respectively, within the full 17-method pool. These summaries are useful when methods have different response scales, with both count columns restricted to the 25 display anchors. The response and rank summaries retain score minima, variation, and display-node counts alongside the full-grid average rank.

The three noise pairings cover scatter noise with Gaussian noise, scatter noise with motion blur, and Gaussian noise with motion blur. Each pair combines distinct disturbances on the same full $25 \times 25 = 625$ grid for every method; the figures use 25 display anchors, while table summaries cover all 625 nodes except the display-anchor Best and Top3 counts. The application order, coordinate values, random realization, and preprocessing are shared across methods at every joint node; each method retains its own frozen checkpoint.

Two-dimensional responses are measured with jointly perturbed inputs. A pair can amplify, suppress, or spatially redistribute an effect in a way that is not captured by adding two marginal changes. The surface anchors display 25 directly evaluated joint inputs; table summaries cover all 625, except Best and Top3 display-anchor counts. The single-variable section supplies context for each axis, while the paired section reports only what is observed on the Cartesian product of evaluated coordinates.

The weighted score gives a concise height for visualization, while the tables preserve means, minima, dispersion, ranks, and measured-node comparisons. The matched-node rank table reports display-anchor placement counts alongside full-grid mean rank. Across all six full 625-node grids, LUMEN leads mean SSIM and MS-SSIM. Under joint R-G angle/wall changes, LUMEN retains a multiscale structural advantage, with mean MS-SSIM of 0.53 versus 0.45 for the strongest displayed baseline. Mean PSNR is 15.52 dB for LUMEN and 15.57 dB for ConvIR and MSBDN-DFF. Under R-R scatter noise/motion blur, LUMEN reaches mean MS-SSIM 0.60 and LPIPS 0.22, compared with DarkIR's 0.50 and 0.34. DarkIR targets low-light restoration (Feijoo et al., 2025); this joint-corruption result is consistent with PACER routing usable multiscale wall evidence and ECHO assembling and refining it into coherent human structure.

As in the single-factor experiments, the dual-variable study uses a frozen source-trained checkpoint; references are used only after reconstruction. The surfaces therefore characterize one fixed mapping under compound input degradation.

H Fixed-Checkpoint Feature Responses

Endpoint ablations locate sensitivity at the output. We therefore use two fixed-checkpoint response atlases with held-out examples from R-C, R-R, R-N, and R-G. The two atlases probe complementary changes: removing local wall evidence and distorting it with structured noise, then tracing each change through the same reconstruction stages.

The feature-response protocol records the same named hooks shown in the architecture: RRG, TRO, WRD, RLI, SPS, and ACR. Clean and perturbed observations are forwarded independently through one frozen checkpoint, and activations are captured without changing the model computation. Hook order therefore follows the implemented PACER-to-ECHO dependency. The direct-view target is never supplied to a hook. It appears only in the final GT column, where it provides a spatial reference for interpreting the localized reconstruction.

Dataset	Pair	Method	mP	minP	mSSIM	wSSIM	mL	mM	Mean S	SD S	Min S
R-C	angle / wall	ConvIR	13.10	12.70	0.73	0.72	**0.25**	0.50	**95.95**	0.87	**93.09**
		DGUNet	**13.32**	**13.05**	0.72	0.71	0.28	**0.51**	95.36	1.21	92.76
		Mom.-Net	12.77	12.43	0.72	0.72	0.25	0.48	93.08	1.22	90.26
		PGDNet	12.41	11.78	0.71	0.71	0.25	0.49	92.17	1.96	88.60
		Restormer	12.61	11.99	0.69	0.67	0.33	0.47	87.32	4.86	81.38
		MSBDN	11.58	11.58	**0.76**	**0.76**	0.26	0.36	83.59	0.00	83.59
		AST	12.55	11.85	0.63	0.62	0.39	0.47	82.52	1.53	79.31
		LUMEN	**14.56**	**13.93**	**0.77**	**0.77**	**0.23**	**0.62**	**109.96**	4.18	**104.15**
R-G	angle / wall	ConvIR	**15.57**	**15.57**	**0.82**	**0.82**	**0.22**	0.44	**98.53**	0.00	**98.53**
		MSBDN	**15.57**	**15.57**	0.82	0.82	0.22	0.44	98.53	0.00	98.53
		MambaIRv2	15.25	14.88	0.82	0.82	0.22	**0.45**	98.32	0.65	97.01
		MambaIR	15.20	13.99	0.82	0.81	0.23	0.43	96.63	3.75	81.63
		Mom.-Net	15.39	14.19	0.82	0.80	0.22	0.42	96.49	4.85	82.87
		DGUNet	14.47	14.41	0.77	0.76	0.26	0.45	92.89	0.81	91.29
		Restormer	14.15	13.44	0.75	0.74	0.30	0.45	89.10	4.36	83.77
		LUMEN	15.52	15.36	**0.84**	**0.84**	**0.22**	**0.53**	**106.05**	2.37	**104.38**
R-R	camera / target	Restormer	11.63	10.25	**0.75**	0.69	**0.21**	**0.60**	**95.80**	3.99	**83.23**
		DGUNet	11.24	9.50	0.72	**0.70**	0.27	0.48	85.36	5.57	70.36
		ConvIR	9.90	8.30	0.73	0.67	0.24	0.50	83.57	7.00	70.43
		Mom.-Net	9.73	8.65	0.72	0.67	0.24	0.48	81.63	5.07	70.78
		PGDNet	9.60	8.26	0.70	0.64	0.26	0.46	79.14	6.02	63.97
		AST	**11.87**	**11.83**	0.64	0.61	0.41	0.44	76.87	0.82	74.66
		MSBDN	8.65	4.25	0.67	0.37	0.29	0.48	74.66	12.65	30.02
		LUMEN	**13.54**	**12.36**	**0.81**	**0.79**	**0.20**	**0.66**	**106.67**	13.66	**99.11**
R-C	noise / blur	NLOST	12.24	**12.23**	**0.71**	0.70	0.35	0.37	**70.45**	0.92	68.23
		DnCNN	11.98	11.98	0.71	**0.71**	0.34	0.37	69.49	0.09	**69.29**
		MambaIR	11.68	11.68	0.71	0.71	**0.33**	0.37	67.99	0.00	67.98
		MambaIRv2	11.68	11.68	0.71	0.71	0.33	0.37	67.97	0.00	67.97
		MSBDN	11.67	11.67	0.71	0.71	0.35	0.37	67.19	0.22	66.86
		U-Net	**12.57**	11.84	0.58	0.38	0.45	**0.39**	64.95	16.01	43.50
		C2PNet	12.21	12.09	0.60	0.30	0.41	0.36	63.49	10.39	43.23
		LUMEN	**15.06**	**13.77**	**0.78**	**0.75**	**0.24**	**0.63**	**107.15**	13.62	**93.61**
R-N	scatter / noise	IntroVAE	**27.01**	**26.71**	**0.83**	**0.83**	**0.17**	**0.78**	**98.15**	0.72	**97.44**
		ConvIR	24.37	19.21	0.79	0.67	0.31	0.72	86.65	13.71	64.01
		MambaIRv2	25.55	22.83	0.76	0.57	0.45	0.75	84.19	10.08	65.54
		U-Net	23.55	18.47	0.68	0.35	0.36	0.70	80.19	17.84	48.20
		MSBDN	23.98	19.46	0.62	0.20	0.41	0.71	77.68	20.30	42.43
		MambaIR	21.44	12.38	0.71	0.47	0.37	0.67	76.48	22.16	42.62
		VS-NLOS	22.16	10.98	0.65	0.20	0.39	0.69	75.94	25.90	27.73
		LUMEN	**27.30**	**26.93**	**0.88**	**0.87**	**0.16**	**0.80**	**101.09**	0.96	**100.00**
R-R	scatter / blur	DarkIR	10.52	7.65	0.65	0.55	0.34	**0.50**	**60.12**	13.59	35.86
		AST	10.44	8.94	0.66	0.61	0.33	0.48	60.01	14.08	46.99
		MambaIR	9.32	**9.32**	**0.74**	**0.74**	**0.30**	0.35	58.48	0.00	**58.48**
		IntroVAE	**10.66**	8.22	0.63	0.53	0.32	0.39	56.56	26.93	32.00
		DeepSN	10.41	8.54	0.64	0.58	0.41	0.44	54.02	11.79	37.45
		DGUNet	9.19	6.90	0.61	0.53	0.35	0.41	51.12	18.40	30.63
		U-Net	10.01	8.35	0.60	0.56	0.43	0.37	47.71	17.66	32.40
		LUMEN	**12.54**	**10.38**	**0.79**	**0.74**	**0.22**	**0.60**	**81.24**	16.02	**65.34**

Table S15: Response summaries over the full $25 \times 25 = 625$ grid for each physical or noise pair. All statistics here summarize all 625 evaluated nodes per method; the figures retain 25 display anchors. P, L, and M denote PSNR, LPIPS, and MS-SSIM; mSSIM and wSSIM are mean and minimum SSIM. Mean S, sample SD S (denominator $n - 1$), and Min S summarize the main-text weighted score S over all 625 nodes per block. Each performance column, excluding SD S, marks one red and one blue value; displayed ties favor LUMEN, then table order.

Response definitions are selected to match the tensor exposed by each stage. RRG uses the output of its reliability head, while feature tensors from later stages are reduced over channels to obtain spatial maps. Each stage has its own response scale. Independent percentile normalization makes each spatial pattern visible, and Table S17 reports the corresponding display percentile and alignment quantities so that saturation alone is never treated as a quantitative comparison between modules.

The intervention is defined before examining the target. Grid cells are selected by a deterministic target-blind rule, and the same selected support is used for the clean-to-perturbed comparison associated with a displayed case. This prevents the visualization from choosing only regions that overlap a convenient body part or produce a large reconstruction difference. It also makes unchanged regions a meaningful reference for evaluating how far a localized input change propagates through the global structural operations.

These response atlases complement the robustness and ablation experiments. Robustness tables score complete outputs over registered perturbation nodes. Component ablations change a module or pathway and measure the endpoint. The atlases keep the weights fixed and change only local input

Dataset	Pair	Method	Mean S	Min S	rank	Q (%)	Best	Top3
R-C	angle / wall	ConvIR	**95.95**	**93.09**	**2.40**	**91.25**	**0**	**24**
		DGUNet	95.36	92.76	2.80	88.75	0	23
		Mom.-Net	93.08	90.26	4.48	78.25	0	1
		PGDNet	92.17	88.60	4.72	76.75	0	2
		Restormer	87.32	81.38	6.08	68.25	0	0
		MSBDN	83.59	83.59	7.24	61.00	0	0
		AST	82.52	79.31	7.68	58.25	0	0
		LUMEN	**109.96**	**104.15**	**1.00**	**100.00**	**25**	**25**
R-G	angle / wall	ConvIR	**98.53**	**98.53**	**2.64**	**89.75**	**0**	**25**
		MSBDN	98.53	98.53	2.64	89.75	0	25
		MambaIRv2	98.32	97.01	3.04	87.25	0	16
		Mom.-Net	96.49	82.87	3.84	82.25	0	19
		MambaIR	96.63	81.63	4.44	78.50	0	12
		DGUNet	92.89	91.29	7.08	62.00	0	0
		PGDNet	88.87	87.10	8.32	54.25	0	0
		LUMEN	**106.05**	**104.38**	**1.00**	**100.00**	**25**	**25**
R-R	camera / target	Restormer	**95.80**	**83.23**	**2.00**	**93.75**	**0**	**25**
		DGUNet	85.36	70.36	4.12	80.50	0	16
		ConvIR	83.57	70.43	5.00	75.00	0	7
		Mom.-Net	81.63	70.78	5.56	71.50	0	0
		PGDNet	79.14	63.97	6.36	66.50	0	2
		AST	76.87	74.66	7.76	57.75	0	0
		MambaIRv2	73.63	32.67	8.08	55.75	0	0
		LUMEN	**106.67**	**99.11**	**1.00**	**100.00**	**25**	**25**
R-C	noise / blur	NLOST	**70.45**	68.23	**3.40**	**85.00**	**0**	**18**
		DnCNN	69.49	**69.29**	4.72	76.75	0	10
		MambaIR	67.99	67.98	6.08	68.25	0	0
		C2PNet	63.49	43.23	6.80	63.75	0	5
		MambaIRv2	67.97	67.97	7.08	62.00	0	0
		AST	63.05	30.87	7.60	58.75	0	7
		U-Net	64.95	43.50	8.00	56.25	0	5
		LUMEN	**107.15**	**93.61**	**1.00**	**100.00**	**25**	**25**
R-N	scatter / noise	IntroVAE	**98.15**	**97.44**	**2.16**	**92.75**	**0**	**25**
		ConvIR	86.65	64.01	3.24	86.00	0	15
		U-Net	80.19	48.20	6.12	68.00	0	0
		MambaIR	76.48	42.62	6.20	67.50	0	0
		MambaIRv2	84.19	65.54	6.20	67.50	0	10
		MSBDN	77.68	42.43	6.56	65.25	0	0
		VS-NLOS	75.94	27.73	6.80	63.75	0	0
		LUMEN	**101.09**	**100.00**	**1.00**	**100.00**	**25**	**25**
R-R	scatter / blur	AST	60.01	46.99	**4.40**	**78.75**	**0**	5
		DarkIR	**60.12**	35.86	4.40	78.75	0	**15**
		IntroVAE	56.56	32.00	5.20	73.75	0	10
		DeepSN	54.02	37.45	5.60	71.25	0	0
		MambaIR	58.48	**58.48**	5.60	71.25	0	10
		DGUNet	51.12	30.63	6.76	64.00	0	5
		U-Net	47.71	32.40	8.12	55.50	0	0
		LUMEN	**81.24**	**65.34**	**1.00**	**100.00**	**25**	**25**

Table S16: Matched-node ranks over the same full $25 \times 25 = 625$ grids as Table S15, with Best and Top3 counts restricted to the 25 display anchors. Mean S and Min S summarize the main-text weighted score S over all 625 nodes per method and pair. The eight displayed methods are ranked at each of the 625 nodes within the full 17-method pool by S, using competition rank (one plus the number of strictly better scores). Rank is the mean node rank; Q (%) is the mean of 100 times (17 minus node rank)/16, both over all 625 nodes. Best and Top3 count display anchors with rank 1 and rank at most 3. Each numerical column marks one red and one blue value; displayed ties favor LUMEN, then table order.

evidence while recording intermediate responses. Together they relate endpoint sensitivity to the spatial propagation of local perturbations.

H.1 Local Evidence Occlusion Responses

Figure S16 examines local evidence removal along the RRG, TRO, WRD, RLI, SPS, and ACR stages before showing the localized output and GT. Each input is partitioned into a 3×3 grid, and a target-blind deterministic rule selects one or two cells for RGB-zero masking. Clean and masked inputs are passed independently through the same dataset-matched fixed checkpoint. The response maps show absolute clean-to-masked changes, with RRG represented by its mask-head response and the other stages by channel-mean activation magnitude. Table S17 records the upper display percentile and descriptive alignment values.

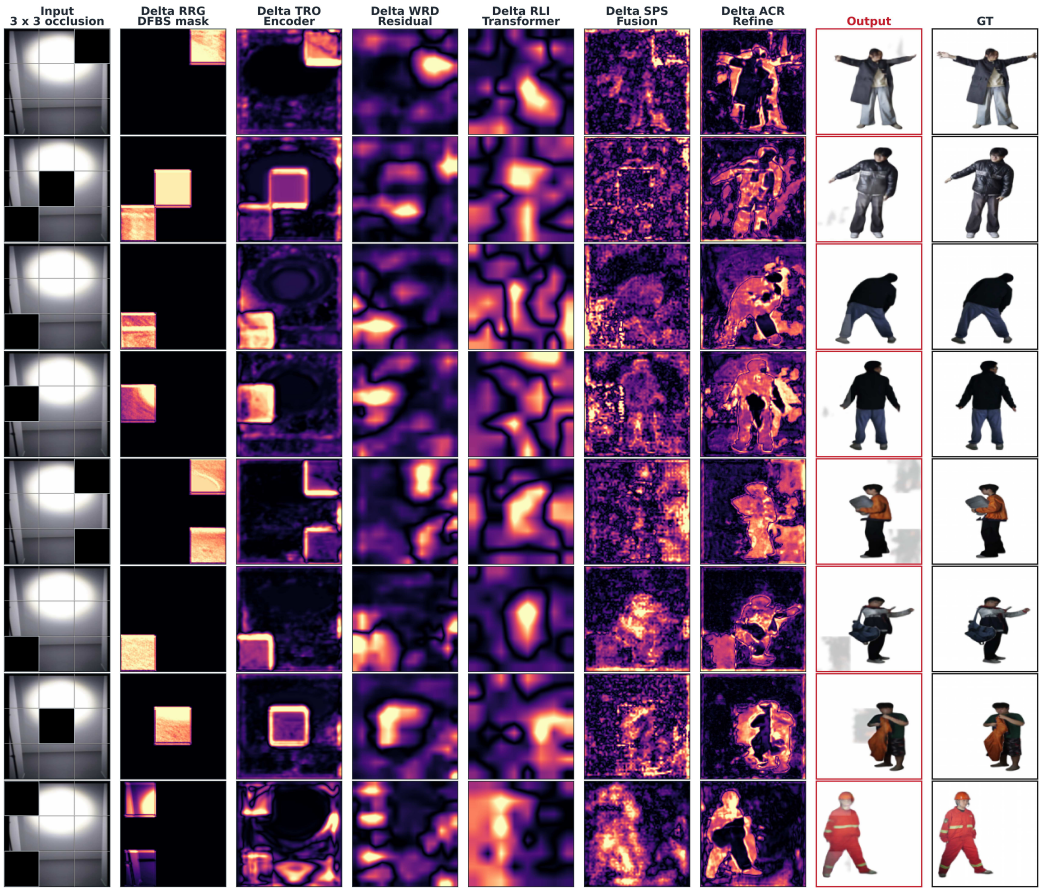

Figure S16: Local evidence removal response atlas. Inputs are partitioned into a 3×3 grid and one or two cells are set to RGB zero using a target-blind deterministic rule before fixed-weight inference. The first column shows the perturbation layout. The six module panels show absolute clean-to-masked response changes for RRG, TRO, WRD, RLI, SPS, and ACR, followed by the localized reconstruction and GT. Response maps use independent display normalization; the final column shows the corresponding direct-view reference. The panels are descriptive fixed-checkpoint response visualizations.

The atlas is read column-wise as an evidence path from the wall observation to the reconstructed outline. Early responses retain broad illumination structure and transport mixtures, whereas later maps are closer to the spatial support used by the output head. Because each response map is independently normalized for display, color intensity describes the spatial pattern within each stage; Table S17 provides the numerical cross-check. The GT column shows the corresponding direct-view reference.

The examples show how local evidence changes propagate through LUMEN across different body configurations. Table S17 retains the upper display percentile and alignment statistics for the response maps.

Each hook has its own response definition and display range, and the output column uses a different construction from channel-averaged intermediate activations. The numerical tables preserve these distinctions.

The clean-response measurements provide the reference state for the target-blind occlusion visualization. The same fixed checkpoint evaluates clean and masked inputs; the seeded mask rule defines the masked pass, so the displayed changes remain comparable.

Each numerical response column uses its corresponding component's definition. RRG is read from its reliability head, while later components use channel-aggregated response magnitude and the output uses image-space change. Upper percentiles summarize the displayed response, and alignment quantities describe spatial correspondence between the response and selected support.

Dataset	Stage	Hook	P98	IoU	MAE
R-C	ACR	Refinement	0.45	**0.93**	0.39
	RLI	Global attn.	0.36	0.43	**0.33**
	RRG	Reliability	0.45	0.18	0.63
	SPS	Screening	0.08	0.18	0.54
	TRO	Transport	**0.50**	0.34	0.61
	WRD	Residuals	0.29	0.20	0.57
	LUMEN	Image	**0.87**	**0.93**	**0.01**
R-R	ACR	Refinement	0.44	**0.98**	**0.35**
	RLI	Global attn.	0.37	0.21	0.40
	RRG	Reliability	0.45	0.22	0.45
	SPS	Screening	0.13	0.03	0.56
	TRO	Transport	**0.53**	0.13	0.61
	WRD	Residuals	0.34	0.17	0.64
	LUMEN	Image	**0.96**	**1.00**	**0.00**
R-N I	ACR	Refinement	0.48	**0.97**	0.41
	RLI	Global attn.	0.31	0.30	**0.30**
	RRG	Reliability	0.45	0.10	0.32
	SPS	Screening	0.24	0.31	0.48
	TRO	Transport	**0.56**	0.03	0.72
	WRD	Residuals	0.43	0.13	0.72
	LUMEN	Image	**0.96**	**0.99**	**0.00**
R-N II	ACR	Refinement	**0.55**	**0.97**	0.41
	RLI	Global attn.	0.39	0.44	**0.25**
	RRG	Reliability	0.45	0.10	0.28
	SPS	Screening	0.26	0.27	0.43
	TRO	Transport	0.53	0.04	0.73
	WRD	Residuals	0.34	0.07	0.25
	LUMEN	Image	**0.97**	**0.99**	**0.00**

Table S17: Fixed-checkpoint response statistics for the displayed held-out cases. R-N I and R-N II denote two distinct cases, each with one terminal LUMEN row. P98 reports the upper response-amplitude percentile used for display; red and blue identify the largest and second-largest displayed amplitudes, respectively. IoU and MAE are descriptive alignment measures, ordered high-to-low and low-to-high, respectively. Each column marks one red and one blue numeral; ties favor LUMEN, then table order. Hook names identify component functions.

The masking rule is target blind: it uses only the predefined grid and seed. Clean and masked observations are evaluated independently by the same weights, so their difference locates sensitivity to the selected input region.

Table S17 complements the component ablation section: ablation measures endpoint changes, whereas the response table records where fixed components react to local input changes.

H.2 Local Structured Noise Responses

Figure S17 examines local motion blur and scatter noise. The input is again partitioned into a 3×3 grid, and one or two randomly selected cells receive a low-amplitude structured perturbation. The six module columns show post-perturbation activations after RRG, TRO, WRD, RLI, SPS, and ACR, followed by a localized output visualization and GT. The output visualization retains the clean result outside the selected support and a controlled corrupted-minus-clean residual inside it.

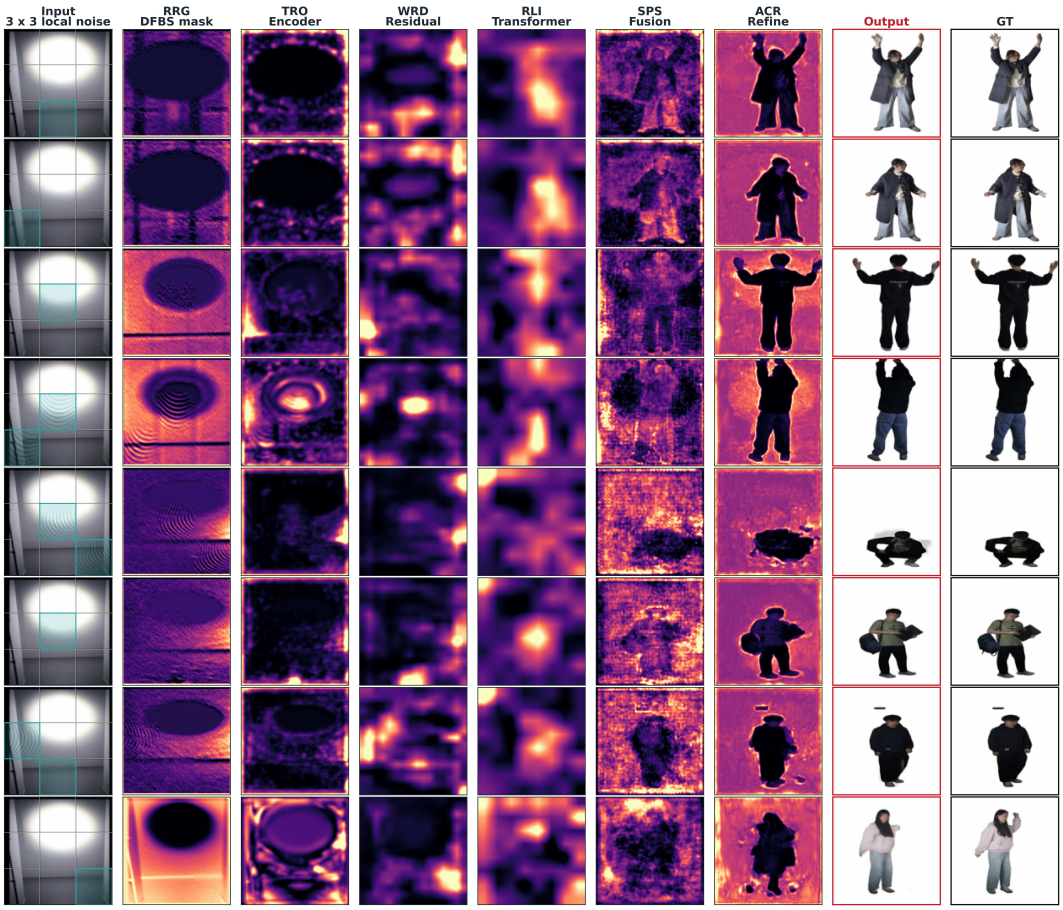

Figure S17: Local structured-noise response atlas. One or two cells in a 3×3 input partition receive localized motion blur or scatter noise, and the perturbed input is passed through the same fixed checkpoint. The first column shows the perturbation layout. The six module panels show post-perturbation activations for RRG, TRO, WRD, RLI, SPS, and ACR, followed by the constructed local-response output and the corresponding direct-view reference. RRG uses the reliability response and the other component panels use channel-mean activation magnitudes. Independent percentile normalization is used for component display. The output retains the clean result outside the selected support and a controlled corrupted-minus-clean residual inside it.

The structured-noise atlas complements the evidence-removal atlas by showing how localized corruption alters the same internal stages; the accompanying table supplies the fixed-checkpoint response statistics.

Motion blur redistributes contrast along a direction, whereas localized scatter noise introduces structured local intensity variation. Their component maps can differ even when both affect the same grid cell. Each map is independently normalized for legibility, so spatial extent and location are the relevant visual comparisons.

The unperturbed regions provide an internal spatial reference. In the component maps, concentrated changes near selected cells indicate limited response spread, while diffuse changes indicate propagation through later global operations. The constructed output displays the selected local residual against the clean body layout; GT supplies a spatial reference.

The output visualization uses the fixed clean result outside the selected support and a controlled corrupted-minus-clean residual inside it. This construction isolates the local response; full-image robustness is reported in the noise tables.

Across the two atlases, the evidence chain is read from left to right. Broad early maps are consistent with wall-level transport mixtures, intermediate maps organize the routed evidence, and later maps reflect the spatial interaction used for structural readout. This progression describes the spatial responses of the named stages. The fixed-checkpoint responses trace how a local change in the wall return passes through evidence routing and structural readout, complementing the reconstruction, ablation, and perturbation results.

I Evaluation Units and Evidence Synthesis

The benchmark rows describe four acquisition regimes; perturbation and intervention tables report their own measured conditions. All displayed values use direction-aware formatting and two-decimal presentation where stated; the source values retain their original precision. Surface connections and percentile projections visualize measured values. Ranks, endpoint changes, and pooled distributions summarize the underlying node-level records.

The evidence is layered. The comparison tables establish multi-dataset performance and rankings. The robustness and physical tables establish behavior over measured single and paired perturbations. The ablation and feature-response records distinguish endpoint sensitivity from descriptive internal response. Endpoint degradation supports component contributions under the tested configurations; the feature atlases locate response changes along the implemented module chain.

The appendix supports the claims at the level of the tested relay-wall setup, measured datasets, frozen comparison methods, and stated perturbation grids. It preserves crossings, ties, negative margins, non-monotone responses, and valid comparison records so that the reader can inspect the full evidence.

Reconstruction, intervention, and perturbation records retain different statistical units. A benchmark row is one protocol-level endpoint, an ablation row is a configuration-specific endpoint, a single-variable robustness row is tied to one severity coordinate, and a dual-variable record is tied to one Cartesian coordinate. Each experiment family retains its own statistical unit. Means, quartiles, percentiles, and ranks are formed only within the experiment family defined by their captions and tables.

All comparison directions are fixed before highlighting. Higher PSNR, SSIM, MS-SSIM, FSIM, IoU, and direction-aligned quality indicate improvement; lower LPIPS, GMSD, ER, DCT, MAE, and native error measures indicate improvement. Best and second-best typography follows these directions within the relevant dataset or condition block. Displayed values determine highlighting, with ties favoring LUMEN or Full and then table order.

Figures preserve the measured anchors even when lines, surfaces, ellipses, or wireframes are added for readability. These graphical connections describe the measured anchors. Likewise, independent activation normalization makes response location visible but prevents quantitative comparison of color magnitude across modules. The numerical summaries retain each display's measured comparison unit.

The experiment families describe complementary aspects of reconstruction. Raw measurements provide reference values, compact tables expose rankings, traces reveal node-wise changes, distributions reveal spread, surfaces reveal interactions, and response atlases reveal localized internal sensitivity. Matched inputs and frozen checkpoints connect these views of reconstruction behavior. Where they differ, the more specific measurement governs: a local crossing is retained even if an aggregate is favorable, and a descriptive activation pattern does not override an endpoint metric.

This reporting design supports component contributions, pathway sensitivity, measured robustness, and cross-protocol reconstruction quality. The measured effects relate the evidence-routing and structural-readout stages to reconstruction under the tested conditions. Within the recorded L-shaped-corner setup, the appendix provides a traceable path from paired observations to benchmark results, controlled interventions, perturbation responses, and localized feature evidence.

I.1 Comparison Units and Interpretation

Each comparison has a defined unit. A colored value belongs to one dataset, perturbation family, physical pair, or feature hook as declared by its table block. Comparisons are local to that unit. Method order is fixed for readability and LUMEN is placed last, but best and second-best status is recomputed from the retained measurements in the applicable block. This rule is shared by the benchmark, ablation, robustness, physical, and feature-response tables.

Metric directions distinguish similarity from error. PSNR, SSIM, MS-SSIM, FSIM, IoU, percentile quality, and the direction-aligned joint score increase with improved agreement. LPIPS, GMSD, ER, DCT, MAE, and native error measures decrease with improved agreement. Reverse-LPIPS visualizations are labeled explicitly, and tables retain raw LPIPS. When an aggregate combines metrics, the appendix also provides metric-specific values or surfaces so that compensation between directions remains visible.

Measured endpoints and derived summaries serve distinct purposes. A benchmark endpoint and a measured perturbation node are raw evaluation records. A mean, quartile, rank, percentile, PCA

coordinate, damage value, envelope margin, or joint score is a deterministic summary of those records. Derived quantities make repeated structure readable but do not add samples. Their captions and surrounding text state the source population so that each summary can be traced back to the measurements from which it was formed.

Visual connections describe the spacing and progression of measured nodes. Curves connect retained one-dimensional nodes, surfaces connect retained Cartesian anchors, and wireframes or ellipses describe displayed structure. The measured anchors define the sample count. Likewise, the response atlases normalize component maps independently for display, so hue and saturation are not a common unit across modules. Numerical tables accompany the visual panels wherever a cross-check is required.

Checkpoints and preprocessing define the evaluated reconstruction mapping. Benchmark, transfer, noise, and physical robustness results use the named frozen checkpoint and fixed preprocessing for their registered evaluation. Zero-shot transfer and perturbation studies keep weights, thresholds, and target-domain settings fixed. Direct-view references are withheld during inference and used only for supervision where specified or for post-inference scoring. The wall observation supplies the test-time input.

Each experiment supports a specific interpretation. Endpoint degradation supports component contributions under the tested configuration, and pathway interventions measure sensitivity to the selected path. Robustness curves and surfaces support behavior over their measured nodes. Feature atlases support localized response propagation under their stated construction. These measurements describe the tested configurations and acquisition regimes. Conclusions in the main paper are therefore tied to the four datasets and the stated experimental settings.

Finally, the evidence families should be read together. The benchmark establishes reconstruction quality, the ablation tables test dependence on named computations, the noise and physical studies test a frozen mapping under controlled shifts, and the atlases expose where localized changes appear along the implemented module chain. Agreement across these families supports a coherent account of LUMEN without asking any single plot or table to carry the entire claim. The appendix is organized in this order to connect each interpretation to its measurements.

The supplementary analyses follow the physical observation chain. Every figure and table is introduced by its scientific question, followed by the relevant comparison rule, and interpreted with the same dataset and metric boundaries used to generate it. Numerical summaries accompany the response plots and feature maps. The resulting appendix remains a single evidence chain: acquisition defines the observation, benchmarks establish the endpoint, interventions test dependence, perturbations test stability, and response maps expose localized behavior under fixed weights.