CoRER: Coherent Relay Evidence Restoration for Initial Hidden-Human Contour Observation in Corner-Occluded Scenes

Anonymous authors
Paper under double-blind review

Abstract

Can faint reflections on a visible relay wall support an initial observation of a hidden person's basic region and main contour? An ordinary RGB camera captures these reflections under flashlight illumination, with the person completely hidden around a corner. Diffuse relay transport leaves weak target-associated variation beneath dominant local wall appearance, while contour support is distributed across decoder scales. We present the Coherent Relay Evidence Restoration Model (CoRER), whose Relay Diffuse-Coherent Representation (RDCR) organizes and preserves weak evidence for deeper encoding by applying scale-separated smoothing over short and long supports to projected local features and using their disagreement to regulate a bounded residual update. Its Scale-Coherent Relay-State Restoration (SCRR) measures smooth and detail agreement at decoder skip fusions, organizes multiscale support and reliability, and recalibrates preliminary foreground-background logits to restore the main contour. An explicit positional calibration supplies a bounded shape prior before the final argmax readout. Extensive comparisons, ablation studies, robustness evaluations, and qualitative results validate CoRER across four real corner scenes in the ReflectHuman benchmark involving different people, corner structures, illumination conditions, and scene settings. CoRER outperforms representative hidden-scene perception and image segmentation methods overall, recovering the hidden person's basic region and main contour from faint reflected-light spots on the relay wall captured by an ordinary camera. We show that faint wall reflections retain usable human-contour evidence and establish the feasibility of this initial contour observation.

1 Introduction

A corner can hide a person completely from an ordinary camera just when spatial awareness matters. In this setting, weak reflected light from the hidden region reaches a visible relay wall and is recorded by the camera. The observation is a red, green, and blue (RGB) image of the illuminated relay-wall spot, while the person remains outside the direct field of view. The task maps this weak wall observation to a binary hidden-human contour mask, with two-class contour logits as the primary learned output. For corner-hidden target awareness, the mask provides a basic region and main contour for locating the hidden person. This setting asks whether an ordinary camera can preserve usable spatial evidence after diffuse relay transport and convert it into an initial contour observation.

Time-resolved systems recover hidden geometry by measuring path-dependent transients Velten et al. (2012); O'Toole et al. (2018). Passive intensity approaches exploit spatial changes induced by occluders or indirect transport Bouman et al. (2017); Klein et al. (2016); Saunders et al. (2019). The RGB relay-wall observation provides spatial intensity information and requires minimal geometric calibration. The resulting task has two coupled difficulties: wall appearance dominates target-related variation, and contour support is distributed inconsistently across decoder scales. The central challenge is to recover contour-level readouts of people hidden around corners from relay-wall observations captured by an ordinary camera.

To turn this weak wall signal into a hidden-human contour, we propose the Coherent Relay Evidence Restoration Model (CoRER), which couples a Relay Diffuse-Coherent Representation (RDCR) front

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



Figure 1: Relay-wall contour setting. (a) Conceptual corner geometry: a camera and active source observe an illuminated relay wall while the target remains outside the direct line of sight. (b) Representative indoor relay-wall observation, in which the visible wall response is dominated by illumination and wall appearance while carrying a weak hidden-target return. (c) Controlled acquisition geometry with the wall-facing camera, active source, relay wall, and occluded target; the synchronized direct-view camera provides the manually verified contour ground truth.

end with Scale-Coherent Relay-State Restoration (SCRR) in the decoder. RDCR separates short-support and long-support responses before deeper encoding and uses their residual and disagreement to organize weak evidence that remains stable after diffuse transport. Contextual condensation carries this organized weak evidence into the deeper features used for contour recovery. SCRR gates encoder skips, measures smooth and detail agreement at successive decoder fusions, and organizes multiscale support and reliability before recalibrating auxiliary and main logits. The resulting decoder path restores the main contour from distributed support through staged restoration. The experimental section evaluates this relay-to-contour process through comparisons, module removals, robustness analyses, and qualitative cases. As illustrated in Figure 1, the task couples corner-occluded geometry with a visible relay-wall observation and a synchronized reference for contour annotation; the network maps the relay-wall image to a contour mask, providing an initial spatial observation of the hidden target. Across four real corner scenes involving different people, corner structures, illumination conditions, and scene settings, CoRER overall outperforms representative hidden-scene perception and image segmentation methods and recovers the hidden person's basic region and main contour from faint reflected-light spots on the relay wall captured by an ordinary camera, demonstrating the feasibility of this initial contour observation.

The main contributions are summarized as follows.

(1) We propose CoRER and validate across four real corner scenes that faint wall reflections recorded by an ordinary camera support recovery of a hidden person's basic region and main contour, establishing initial contour observation around corners.

(2) We introduce RDCR to organize diffuse-coherent weak evidence, separate short- and long-support responses to regulate residual updates through their disagreement, and propagate cross-scale-consistent evidence through front-end encoding.

(3) We introduce SCRR to organize multiscale contour support, aggregate reliability and agreement cues, and guide staged restoration of the main contour.

2 Related Work

Two research directions support this study: extracting hidden-scene information from indirect light, and converting image evidence into spatial masks.

Active NLOS Reconstruction. Active NLOS reconstruction uses controlled illumination and time-resolved measurements to estimate hidden geometry. Velten et al. proposed an ultrafast time-of-flight method, which uses transient light transport to recover hidden shape Velten et al. (2012). O'Toole et al. combine pulsed-laser relay-wall scanning and time-resolved detection with a light-cone transform for confocal reconstruction O'Toole et al. (2018). Recent learned systems extend that transient setting, and Li et al. proposed Non-Line-of-Sight Imaging with Transformer (NLOST), which uses a transformer representation to realize transient NLOS reconstruction Li et al. (2023).

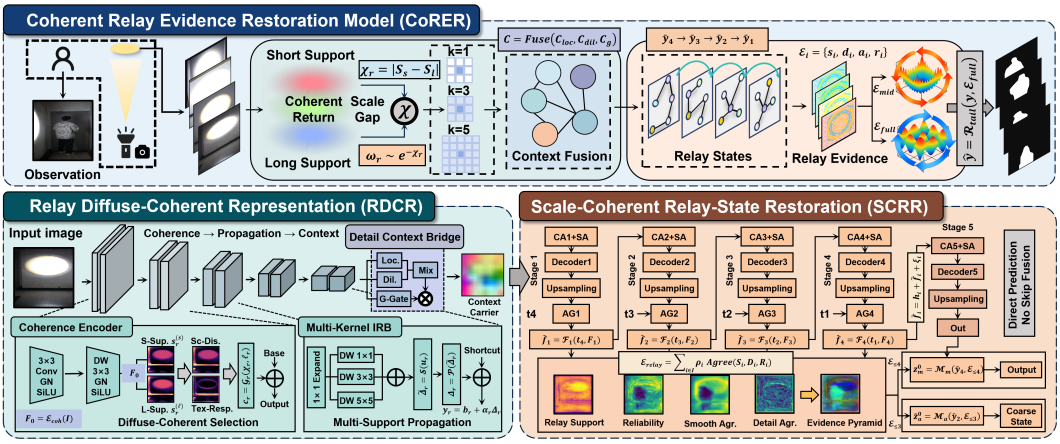

Figure 2: Overview of CoRER. **(a)** The relay-wall image passes through RDCR, contextual condensation, and SCRR before coordinate-calibrated contour readout. **(b)** RDCR compares short- and long-support relay responses to form a diffuse-coherent feature and guide contextual condensation. **(c)** SCRR computes support, detail, agreement, and reliability at multiscale fusions, aggregates them across stages, and restores auxiliary and main contour states.

Klinghoffer et al. proposed PlatoNeRF, which uses two-bounce lidar and neural rendering to handle hidden three-dimensional geometry Klinghoffer et al. (2024).

Ordinary-Camera Inference and Spatial Mask Prediction. Passive NLOS methods infer hidden information from ordinary intensity observations or naturally formed transport patterns. Bouman et al. proposed a corner-camera method, which uses indirect spatial transport to handle hidden-scene inference from ordinary intensity observations Bouman et al. (2017). Saunders et al. proposed computational periscopy, which uses ordinary camera measurements to realize passive scene recovery Saunders et al. (2019). Dense segmentation methods map an image representation directly to a mask, and Ronneberger et al. proposed U-Net, which uses an encoder-decoder with skip connections to realize dense segmentation Ronneberger et al. (2015). Xie et al. proposed SegFormer, which uses hierarchical transformer features and a lightweight decoder to handle semantic segmentation Xie et al. (2021). These observation and readout families provide useful precedents. CoRER connects intensity observations to hidden-human masks through scale-separated relay evidence and agreement-derived restoration of the basic region and main contour for initial observation around corners.

3 Method

3.1 Overall Architecture

When direct sight is blocked around a corner, the relay-wall spot contains a superposition of wall appearance, direct illumination, and a weak hidden-human return. CoRER first organizes diffuse-coherent weak evidence in this mixed wall observation through agreement across spatial supports, then restores the hidden person's basic region and main contour. RDCR organizes front-end weak evidence and condenses its context into a relay basis that stays stable under diffusion. SCRR organizes multiscale support and reliability to restore the final contour state. In compact form,

$$\hat{y} = \arg\max \mathcal{R}_{\mathrm{SCRR}}(\mathcal{R}_{\mathrm{RDCR}}(x)) . \tag{1}$$

Here x denotes the relay-wall spot image. The first stage extracts a relay basis that remains stable under diffusion, and the second stage organizes multiscale support and reliability into a scale-consistent contour state. Calibration refines the contour logits, and $\arg\max$ yields the binary mask. An intermediate decoder head produces an auxiliary prediction, while the main branch completes the remaining restoration stages and final readout. Figure 2 traces this relay-to-contour path and expands two mechanism stages that organize and restore the relay state.

3.2 Relay Diffuse-Coherent Representation (RDCR)

Diffuse transport leaves the hidden-human return as weak cross-scale-consistent evidence beneath dominant local wall appearance. RDCR therefore organizes the wall observation into a relay basis, evaluates that basis under two diffusion supports, and uses their disagreement to form a bounded residual correction.

We write the relay basis and its local re-encoding as

$$b_r(\mathbf{p}) = \int_{\Omega_x} \Psi_r(\mathbf{p}, \mathbf{q})\, \phi(x)(\mathbf{q})\, d\mathbf{q}, \qquad \ell_r = \mathcal{Q}_r(b_r), \tag{2}$$

where b_r is the relay-space lift of the observation, Ψ_r is the basis kernel, and ℓ_r is the locally re-encoded relay field at the corresponding representation scale. The local dimensions H_r and W_r describe the spatial resolution at which this relay state is organized.

Two diffusion supports are extracted from the same relay field by short and long receptive kernels:

$$s_r^{(\kappa)}(\mathbf{p}) = \int_{\Omega_r} K_{\sigma_\kappa}(\mathbf{p} - \mathbf{q})\, \ell_r(\mathbf{q})\, d\mathbf{q}, \qquad \kappa \in \{s, \ell\}. \tag{3}$$

The short support captures locally coherent relay structure, while the long support captures the broader relay envelope; comparing them provides the diffusion-consistency cue for soft weighting. RDCR uses the shared response across both diffusion spans to estimate how much of the local response remains stable. The local disagreement is summarized by

$$\chi_r(\mathbf{p}) = \left|s_r^{(s)}(\mathbf{p}) - s_r^{(\ell)}(\mathbf{p})\right| + \mu_r \left|\ell_r(\mathbf{p}) - s_r^{(s)}(\mathbf{p})\right|, \qquad \omega_r(\mathbf{p}) = \frac{\exp[-\lambda_r \chi_r(\mathbf{p})]}{\int_{\Omega_r} \exp[-\lambda_r \chi_r(\mathbf{u})]\, d\mathbf{u}}.$$

The RDCR update is then formed by weighting the local residual field with the coherence between those two supports:

$$y_r(\mathbf{p}) = b_r(\mathbf{p}) + \alpha_r \int_{\Omega_r} \omega_r(\mathbf{q}) \left[\frac{s_r^{(s)}(\mathbf{q}) + s_r^{(\ell)}(\mathbf{q})}{2} + \ell_r(\mathbf{q}) - \chi_r(\mathbf{q}) \right] d\mathbf{q}. \tag{4}$$

Here ω_r is the normalized coherence weight based on the local residual magnitude and cross-scale disagreement; it retains shared diffuse structure while reducing wall-dominant fluctuation. The resulting representation organizes the intermediate relay stream by separating diffusion-consistent structure from nuisance texture before contextual condensation. RDCR therefore preserves and propagates weak front-end evidence in a form that supports later contour reasoning.

3.3 Scale-Coherent Relay-State Restoration (SCRR)

Recovering the main contour from RDCR's organized weak evidence requires support across decoder scales to be aligned, separated, and pooled before the final decision. SCRR therefore treats each decoder stage as a relay-state restoration step: the skip return is aligned with the current decoder state, the aligned return is split into smooth support and residual detail, and the resulting support, reliability, and agreement are organized across scales.

At stage i, the decoder state and the skip return are fused into a single relay carrier

$$\bar{f}_i(\mathbf{p}) = \mathcal{F}_i\big(\Pi_i^{\uparrow}(\bar{y}_{i-1})(\mathbf{p}), f_i(\mathbf{p})\big) = h_i(\mathbf{p}) + \tilde{f}_i(\mathbf{p}) + \xi_i(\mathbf{p}), \tag{5}$$

where $\mathcal{F}_i$ denotes the aligned fusion law at that stage, h_i is the lifted decoder carrier, $\tilde{f}_i$ is the support-bounded skip return, and ξ_i is the conservative residual correction. At each fusion, the gated skip return provides smooth support s_i and residual detail d_i; their alignment with the decoder state generates stagewise agreement and reliability measures.

The stagewise evidence is then pooled over the active decoder set I:

$$\mathcal{E}_I(\mathbf{p}) = \frac{1}{|I|} \sum_{i \in I} \Pi_i^{\uparrow} \left[\rho_s\big(s_i(\mathbf{p})\big) + \rho_d\big(d_i(\mathbf{p})\big) + \rho_a\big(a_i(\mathbf{p})\big) + \rho_r\big(r_i(\mathbf{p})\big) \right]. \tag{6}$$

Here s_i, d_i, a_i, and r_i denote smooth support, residual detail, agreement, and reliability, respectively, and $\rho.$ maps them into a common evidence scale before pooling. The auxiliary head uses the early pooled span for mid-scale prediction and training supervision; the main branch uses the full pooled span for final relay correction.

The coarse relay state and the final contour projection are then driven by the pooled evidence together with a compact spatial prior:

$$z_a^0 = \mathcal{M}_a(\bar{y}_2, \mathcal{E}_{\leq 3}), \qquad z_m^0 = \mathcal{M}_m(\bar{y}_4, \mathcal{E}_{\leq 4}),$$

and the two branches are finally calibrated by

$$z_j = z_j^0 + \mathcal{B}_j\big(\mathcal{E}_j, \pi_j, 4\pi_j(1-\pi_j), \|\nabla \mathcal{S}_5(\pi_j)\|_2\big), \qquad j \in \{a, m\}. \tag{7}$$

Here $\mathcal{E}_a = \mathcal{E}_{\leq 3}$ and $\mathcal{E}_m = \mathcal{E}_{\leq 4}$: the auxiliary and main branches use evidence pooled over three and four decoder stages, respectively. π_j is the current foreground belief carried by the pre-corrected logits. The auxiliary prediction provides intermediate supervision during training; the main branch projects the final contour and produces the readout. SCRR thus preserves a staged restoration path from coarse region support to final contour logits through evidence pooled across decoder stages.

4 Experiment

4.1 Experimental Setup

We evaluate CoRER on ReflectHuman's four real indoor corner scenes: Reflect-Corridor (4,958 images), Reflect-Room (5,232 images), Reflect-Basement (10,862 images), and Reflect-Lab (3,729 images), covering different people, corner structures, illumination conditions, and scene settings. Each subset pairs 256×256 RGB relay-wall spot images with equally sized binary hidden-human contour masks. Synchronized direct-view images support manual mask annotation and review; the ordinary wall-facing camera image is the model input. Each scene has a fixed training–validation split (counts in Table A1). We report scene-level results and four-scene macro averages.

All models use the same optimization recipe in PyTorch 2.5 on one NVIDIA RTX 3080 GPU. Inputs are normalized 3-channel sRGB images resized to 256×256. Training uses AdamW with batch size 16, learning rate 10^{-4}, weight decay 10^{-4}, 50 epochs, StepLR with step size 20 and decay factor 0.5, AMP, and seed 0. CoRER's backbone uses channels 16, 32, 64, 96, and 160; depths 1, 1, 1, 1, and 1; and expansion factor 2.0. We select each run's best checkpoint by validation Dice.

Dice and IoU measure region overlap Dice (1945); Jaccard (1912), and Boundary-F measures boundary precision and recall Perazzi et al. (2016). Higher is better for all three. The appendix reports mean absolute error (MAE), Matthews correlation coefficient (MCC), and the 95th-percentile Hausdorff distance (HD95). Lower MAE and HD95 (in pixels) and higher MCC indicate better performance. Boundary-F uses the radius-2 boundary evaluator; all metrics are computed per image and then averaged within each scene.

4.2 Contour Recovery across Four Corner Scenes

Table 1 compares CoRER with 17 contour baselines from physics- and transport-guided, learning-based hidden-scene, and general segmentation families; Figure 3 shows corresponding sample-level predictions. All methods are evaluated under a common preprocessing and metric protocol on the relay-wall inputs.

CoRER attains the highest Dice, IoU, and Boundary-F values in all four scenes. Reflect-Room presents the closest overlap values, whereas Reflect-Lab shows a larger separation together with more fragmented relay evidence and more selective boundary recovery. The table-level gains and the visual continuity jointly show that faint reflected-light spots captured by an ordinary camera carry usable information about the hidden person's basic region and main contour.

The comparison advantage follows the relay-to-contour path established by CoRER. RDCR organizes diffuse-consistent weak evidence before deeper encoding, keeping the downstream representation aligned with the hidden-human return, while SCRR organizes cross-scale support and reliability during fusion and restores the main contour through logit calibration. The agreement between the quantitative ranking and the sample-level contour structure shows how the two stages convert a weak wall response into an initial hidden-human contour observation.

Table 1: Quantitative comparison across the four ReflectHuman scene subsets, grouped by baseline family. Each scene triplet reports Dice, IoU, and Boundary-F, for which higher values are better; the Source column gives the venue-year tag. Row backgrounds highlight CoRER (light red) and NLOS-LTM (light blue). Red and blue numbers indicate the best and second-best distinct values in each metric column; ties at the displayed precision share the same text color.

Model	Source	Reflect-Corridor			Reflect-Room			Reflect-Basement			Reflect-Lab		
		Dice ↑	IoU ↑	BF ↑	Dice ↑	IoU ↑	BF ↑	Dice ↑	IoU ↑	BF ↑	Dice ↑	IoU ↑	BF ↑
Physics- and Transport-Guided Methods													
NLOS-LTM Zhang et al. (2025)	TIP'25	0.684	0.534	0.173	0.699	0.549	0.188	0.722	0.572	0.201	0.755	0.620	0.171
NLOS-OT Geng et al. (2022)	TIP'22	0.695	0.550	0.183	0.698	0.548	0.171	0.722	0.577	0.233	0.720	0.577	0.137
NLOS-LPP Sun et al. (2025)	ICCV'25	0.680	0.531	0.167	0.709	0.561	0.172	0.726	0.579	0.225	0.693	0.547	0.125
RISE Torcolacci et al. (2024)	ICC'24	0.684	0.535	0.177	0.707	0.561	0.187	0.729	0.583	0.241	0.686	0.539	0.118
CurvNLOS Ding et al. (2024)	TPAMI'24	0.645	0.499	0.174	0.700	0.552	0.180	0.722	0.574	0.220	0.692	0.545	0.126
PlatoNeRF Klinghoffer et al. (2024)	CVPR'24	0.641	0.496	0.148	0.709	0.562	0.186	0.729	0.582	0.217	0.641	0.493	0.104
Learning-Based Hidden-Scene Methods													
ST-Mamba Li et al. (2024)	NeurIPS'24	0.681	0.533	0.153	0.703	0.552	0.167	0.719	0.572	0.213	0.698	0.554	0.127
TRT-NLOS Li et al. (2025)	arXiv'25	0.671	0.523	0.163	0.712	0.563	0.181	0.708	0.559	0.217	0.693	0.548	0.128
NLOSFormer Ye et al. (2026)	ACM TOG'26	0.656	0.504	0.174	0.710	0.562	0.200	0.728	0.582	0.229	0.685	0.540	0.124
DG-NLOS Su et al. (2025)	AAAI'25	0.645	0.500	0.161	0.722	0.576	0.199	0.725	0.579	0.222	0.684	0.536	0.116
TransDiff Cui et al. (2025)	TIP'25	0.663	0.516	0.167	0.708	0.561	0.201	0.721	0.574	0.212	0.676	0.529	0.122
NLOST Li et al. (2023)	CVPR'23	0.645	0.495	0.175	0.705	0.555	0.196	0.712	0.560	0.204	0.704	0.558	0.145
General Segmentation Methods													
MK-UNet Rahman & Marculescu (2025)	ICCV'25	0.626	0.471	0.143	0.697	0.545	0.159	0.730	0.583	0.229	0.737	0.597	0.165
Swin-Unet Cao et al. (2023)	ECCV'23	0.647	0.499	0.161	0.707	0.559	0.177	0.733	0.588	0.232	0.689	0.543	0.125
SegFormer Xie et al. (2021)	NeurIPS'21	0.639	0.492	0.147	0.707	0.561	0.191	0.723	0.575	0.222	0.706	0.564	0.130
MISSFormer Huang et al. (2023)	TMI'23	0.653	0.503	0.163	0.702	0.554	0.175	0.721	0.572	0.222	0.686	0.541	0.128
DUCK-Net Dumitru et al. (2023)	Sci Rep'23	0.652	0.507	0.162	0.691	0.541	0.181	0.731	0.585	0.227	0.750	0.610	0.154
CoRER	Ours	0.712	0.570	0.215	0.731	0.587	0.212	0.760	0.619	0.248	0.786	0.657	0.201



Figure 3: Representative clean-scene predictions. Each row shows a relay-wall input (Input), its direct-view reference image (Real GT) and reference mask (Mask GT), and predictions from CoRER and selected comparison methods. The direct-view image provides the annotation reference, and the relay-wall image is the network input. Table 1 gives corresponding quantitative results.

4.3 Ablation Study

Ablations isolate how early weak-evidence organization and decoder-scale restoration contribute to region coverage and contour fidelity under the same training and evaluation recipe. Table 2 groups component removal, design variant, and replacement studies. Figure 4 shows component removals and design variants; Figure A3 shows replacements.

Removing RDCR lowers the four-scene macro Dice, IoU, and Boundary-F by 0.044, 0.051, and 0.039, while removing SCRR lowers them by 0.031, 0.038, and 0.055. On this four-scene macro average, RDCR contributes more to region overlap and SCRR to main contour fidelity. The variant rows recover part of the full-model scores, and the replacement rows reveal how the two mechanisms respond to different scene conditions. The more boundary-sensitive scenes show a clearer SCRR drop, while the cleaner overlap-oriented scenes show a larger RDCR drop.

The ablation behavior matches the visual contour changes. Removing RDCR mainly weakens foreground support for the basic region, whereas removing SCRR more often produces main-contour

Table 2: Ablation of RDCR and SCRR across the four ReflectHuman scene subsets, grouped as component removal, design variant, and replacement studies. Each scene triplet reports Dice, IoU, and Boundary-F, for which higher values are better. Row backgrounds highlight CoRER (light red) and the SCRR variant (light blue). Red and blue numbers indicate the best and second-best distinct values in each metric column; ties at the displayed precision share the same text color.

Model	Reflect-Corridor			Reflect-Room			Reflect-Basement			Reflect-Lab		
	Dice ↑	IoU ↑	BF ↑	Dice ↑	IoU ↑	BF ↑	Dice ↑	IoU ↑	BF ↑	Dice ↑	IoU ↑	BF ↑
Component Removal												
w/o RDCR	0.674	0.529	0.183	0.692	0.541	0.192	0.715	0.566	0.195	0.733	0.592	0.150
w/o SCRR	0.696	0.548	0.159	0.706	0.556	0.148	0.725	0.578	0.203	0.740	0.600	0.145
Design Variant												
RDCR (variant)	0.691	0.550	0.206	0.702	0.551	0.173	0.722	0.574	0.237	0.739	0.604	0.165
SCRR (variant)	0.699	0.557	0.208	0.709	0.561	0.201	0.729	0.584	0.239	0.748	0.611	0.165
Replacement												
RDCR→VMamba	0.664	0.511	0.177	0.686	0.533	0.169	0.717	0.567	0.206	0.733	0.595	0.163
RDCR→MambaVision	0.666	0.516	0.150	0.693	0.542	0.200	0.724	0.578	0.231	0.736	0.600	0.170
RDCR→SeNaTra	0.663	0.511	0.175	0.694	0.540	0.157	0.720	0.571	0.222	0.736	0.601	0.169
SCRR→SAMRefiner	0.698	0.555	0.199	0.701	0.552	0.203	0.727	0.581	0.248	0.743	0.604	0.162
SCRR→Mask-Adapter	0.699	0.556	0.206	0.700	0.551	0.185	0.726	0.578	0.233	0.745	0.607	0.169
SCRR→QueryMeldNet	0.694	0.549	0.203	0.707	0.557	0.159	0.729	0.582	0.236	0.749	0.610	0.159
CoRER	0.712	0.570	0.215	0.731	0.587	0.212	0.760	0.619	0.248	0.786	0.657	0.201

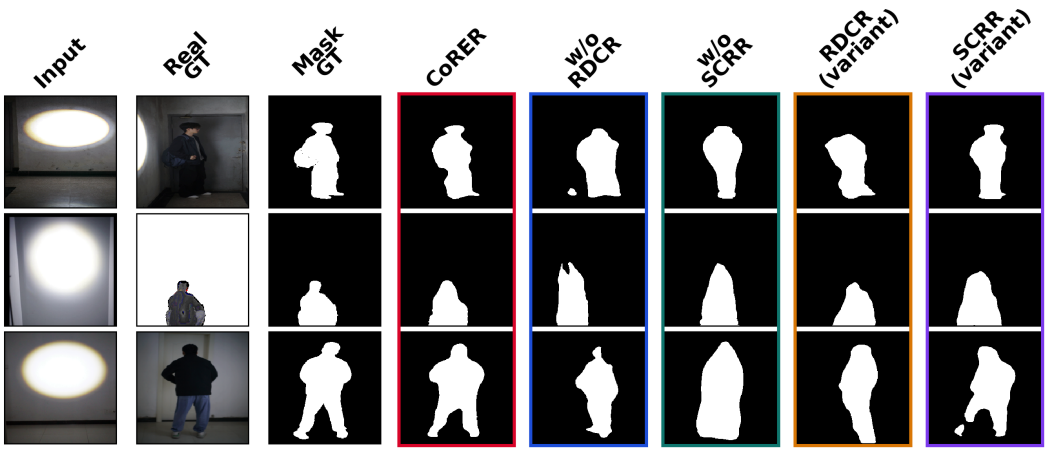

Figure 4: Representative qualitative ablation comparison. Each row shows a relay-wall input, its reference image and reference mask, and contour predictions from CoRER, the two component removals, and the two design variants. Table 2 gives the corresponding quantitative results.

breaks or spillover; the design variants retain part of the complete profile, and the replacement rows show the cost of substituting the coordinated evidence path with an alternative module. Together, the table and qualitative predictions establish complementary contributions from front-end evidence organization and back-end scale-coherent restoration.

4.4 Robustness Analysis

Figure 5 reports Dice across the real-scene brightness sweep in four scenes; Figure 6 shows predictions at corresponding brightness levels. The appendix tests distance, angle, wall material, and noise perturbations.

From the brightest to the darkest endpoint, CoRER changes by 0.002, less than 0.001, 0.006, and 0.001 Dice in Corridor, Room, Basement, and Lab, respectively; on Corridor, CoRER changes from 0.712 to 0.710, whereas NLOS-LTM Zhang et al. (2025) changes from 0.684 to 0.618. The curves and representative predictions show stable basic-region support and main-contour readout across the tested illumination attenuation.

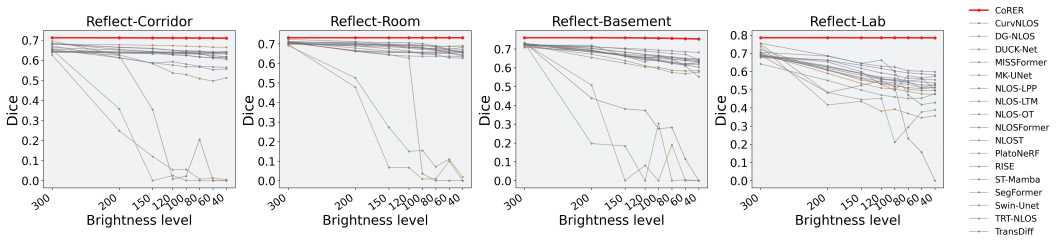

Figure 5: Dice under the real-scene brightness sweep; horizontal axes show brightness in nit. Each subplot corresponds to one ReflectHuman scene subset and compares CoRER with the selected baselines; higher values indicate more accurate contour segmentation.

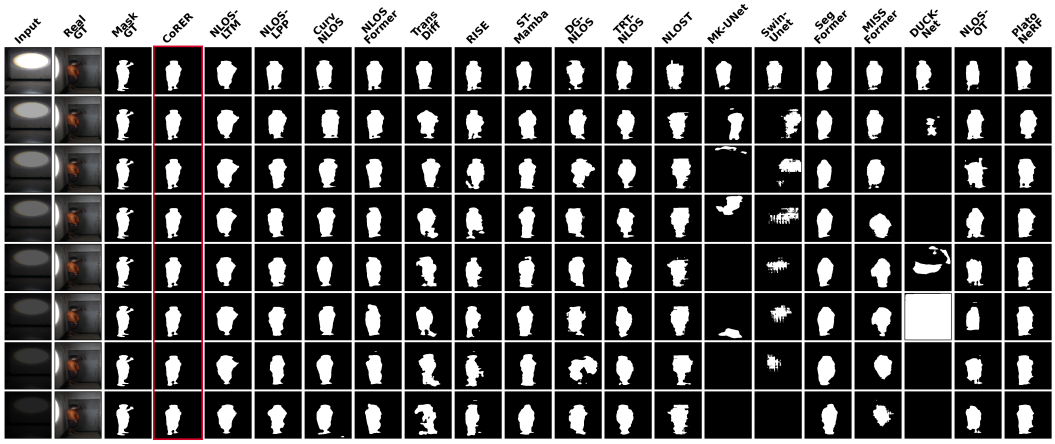

Figure 6: Representative predictions across the brightness levels in Figure 5. Successive rows show the same scene instance at different brightness levels. Columns show the relay-wall input, reference image, reference mask, and method predictions.

This stability follows the two-stage design: RDCR organizes the early response around diffusion-consistent structure, while SCRR organizes multiscale support and reliability to restore the main contour as the signal changes. The persistence of these visual and quantitative cues across the tested conditions supports the feasibility of initial hidden-human contour observation.

4.5 RDCR Mechanism Study

RDCR's contribution begins before contour prediction: it organizes weak wall responses into target-aligned evidence for deeper encoding. Figure 7 evaluates evidence localization over retained top-q percentiles, including top-q precision, recall, IoU, in-mask evidence mass, and cross-scale agreement; the experiment compares CoRER with its RDCR removal under the same evaluation protocol.

RDCR raises retained top-q IoU, in-mask evidence mass, and cross-scale agreement, and the gains become more useful after contextual condensation. Figures A33 and A34 further show that the organized response remains spatially coherent under representative signal changes and supports a more complete basic region at the output.

These results explain the front-end role of RDCR in the complete model. By separating short- and long-support responses and using their agreement to regulate the residual update, RDCR concentrates diffuse-consistent weak evidence before deeper encoding, giving SCRR a more stable carrier for subsequent contour restoration.

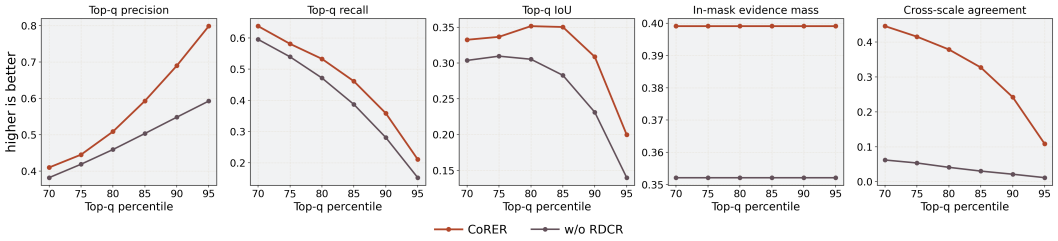

Figure 7: RDCR evidence localization over retained top-q percentiles. From left to right, the panels report top-q precision, recall, IoU, in-mask evidence mass, and cross-scale agreement for CoRER and its RDCR removal; higher values are better throughout.

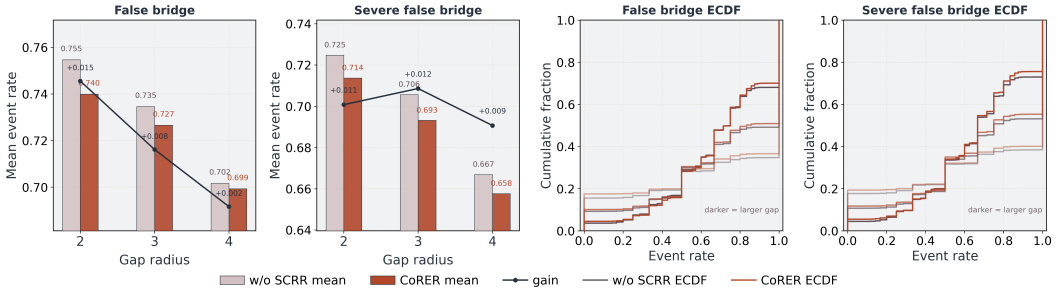

Figure 8: Pooled paired false-bridge analysis of SCRR over gap radii 2–4. The two left panels compare mean false-bridge and severe-false-bridge rates, with positive gain denoting a lower rate for CoRER; the two right panels show the corresponding empirical cumulative distributions.

4.6 SCRR Mechanism Study

Recovering a coherent main contour from distributed decoder evidence requires both boundary continuity and separation between nearby regions. Figure 8 evaluates pooled paired false-bridge behavior over gap radii 2–4, comparing CoRER with the SCRR removal through mean false-bridge rates and their empirical cumulative distributions.

The pooled gains remain positive across the displayed radii, and the empirical distributions shift toward lower false-bridge rates for CoRER. Together with the ablation table, qualitative predictions, and Figures A38–A40, these results show that the complete model preserves foreground support while reducing contour breaks, spillover, and unintended connections around local gaps.

SCRR provides the back-end mechanism for this behavior. It organizes smooth support, detail, agreement, and reliability across decoder scales, then uses the resulting relay state to recalibrate the contour logits, allowing the restoration path to preserve supported regions and repair ambiguous local boundaries.

5 Conclusion

We demonstrate initial hidden-human contour recovery with CoRER from faint wall reflections recorded by an ordinary RGB camera under flashlight illumination. RDCR organizes diffuse-coherent weak evidence before deeper encoding, while SCRR organizes multiscale support and reliability during decoder fusion and restores the main contour through final logit calibration. Across four real scenes in the ReflectHuman benchmark with different people, corner structures, illumination conditions, and scene settings, CoRER attains the highest reported scores on the main metrics and outperforms representative hidden-scene perception and image segmentation methods overall. The ablation and robustness results are consistent with RDCR organizing weak evidence and SCRR organizing multiscale support and reliability to restore the main contour. These results establish the initial feasibility of recovering the basic region and main contour of a completely hidden person from faint relay-wall reflections in the tested corner scenes.

REFERENCES

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

Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *Computer Vision – ECCV 2022 Workshops*, volume 13803, pp. 205–218, 2023. doi: 10.1007/978-3-031-25066-8_9.

Xingyu Cui, Huanjing Yue, Shida Sun, Yue Li, Yusen Hou, Zhiwei Xiong, and Jingyu Yang. Transdiff: Unsupervised non-line-of-sight imaging with aperture-limited relay surfaces. *IEEE Transactions on Image Processing*, 34:8018–8031, 2025. doi: 10.1109/TIP.2025.3637694.

Lee R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3): 297–302, 1945. doi: 10.2307/1932409.

Rui Ding, Juntian Ye, Qifeng Gao, Feihu Xu, and Yuping Duan. Curvature regularization for non-line-of-sight imaging from under-sampled data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):8474–8485, 2024. doi: 10.1109/TPAMI.2024.3409414.

Razvan-Gabriel Dumitru, Darius Peteleaza, and Catalin Craciun. Using duck-net for polyp image segmentation. *Scientific Reports*, 13:9803, 2023. doi: 10.1038/s41598-023-36940-5.

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.

Ali Hatamizadeh and Jan Kautz. MambaVision: A hybrid mamba-transformer vision backbone. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 25261–25270, 2025. doi: 10.1109/CVPR52734.2025.02352.

Xiaohong Huang, Zhifang Deng, Dandan Li, Xueguang Yuan, and Ying Fu. Missformer: An effective transformer for 2d medical image segmentation. *IEEE Transactions on Medical Imaging*, 42 (5):1484–1494, 2023. doi: 10.1109/TMI.2022.3230943.

Paul Jaccard. The distribution of the flora in the alpine zone. *New Phytologist*, 11(2):37–50, 1912. doi: 10.1111/j.1469-8137.1912.tb05611.x.

Jonathan Klein, Christoph Peters, Jaime Martin, Martin Laurenzis, and Matthias B. Hullin. Tracking objects outside the line of sight using 2d intensity images. *Scientific Reports*, 6:32491, 2016. doi: 10.1038/srep32491.

Tzofi Klinghoffer, Xiaoyu Xiang, Siddharth Somasundaram, Yuchen Fan, Christian Richardt, Ramesh Raskar, and Rakesh Ranjan. Platonerf: 3d reconstruction in plato's cave via single-view two-bounce lidar. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14565–14574, 2024. doi: 10.1109/CVPR52733.2024.01380.

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

Yue Li, Yi Sun, Shida Sun, Juntian Ye, Yueyi Zhang, Feihu Xu, and Zhiwei Xiong. Toward dynamic non-line-of-sight imaging with mamba enforced temporal consistency. In *Advances in Neural Information Processing Systems*, volume 37, pp. 126452–126473, 2024. doi: 10.52202/079017-4016.

Yue Li, Shida Sun, Yu Hong, Feihu Xu, and Zhiwei Xiong. 3d reconstruction from transient measurements with time-resolved transformer. *arXiv preprint arXiv:2510.09205*, 2025. doi: 10.48550/arXiv.2510.09205.

Yuqi Lin, Hengjia Li, Wenqi Shao, Zheng Yang, Jun Zhao, Xiaofei He, Ping Luo, and Kaipeng Zhang. SAMRefiner: Taming segment anything model for universal mask refinement. In *International Conference on Learning Representations*, 2025. doi: 10.48550/arXiv.2502.06756.

Yue Liu, Yunjie Tian, Yuzhong Zhao, Hongtian Yu, Lingxi Xie, Yaowei Wang, Qixiang Ye, Jianbin Jiao, and Yunfan Liu. VMamba: Visual state space model. In *Advances in Neural Information Processing Systems*, volume 37, pp. 103031–103063, 2024. doi: 10.52202/079017-3273.

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

Federico Perazzi, Jordi Pont-Tuset, Brian McWilliams, Luc Van Gool, Markus Gross, and Alexander Sorkine-Hornung. A benchmark dataset and evaluation methodology for video object segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 724–732, 2016.

Md Mostafijur Rahman and Radu Marculescu. MK-UNet: Multi-kernel lightweight cnn for medical image segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pp. 1053–1062, 2025. doi: 10.1109/ICCVW69036.2025.00114.

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

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

Xiongfei Su, Tianyi Zhu, Lina Liu, Zheng Chen, Yulun Zhang, Siyuan Li, Juntian Ye, Feihu Xu, and Xin Yuan. Dual-branch graph feature learning for nlos imaging. In *AAAI Conference on Artificial Intelligence*, volume 39, pp. 7051–7059, 2025. doi: 10.1609/aaai.v39i7.32757.

Shida Sun, Yue Li, Yueyi Zhang, and Zhiwei Xiong. Generalizable non-line-of-sight imaging with learnable physical priors. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1–10, 2025. doi: 10.1109/ICCV51701.2025.02322.

G. Torcolacci, A. Guerra, H. Zhang, F. Guidi, Q. Yang, Y. C. Eldar, and D. Dardari. Ris-empowered near-field imaging in nlos scenarios. In *2024 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 378–383, 2024. doi: 10.1109/ICCWorkshops59551.2024.10615296.

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

Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. In *Advances in Neural Information Processing Systems*, volume 34, pp. 12077–12090, 2021.

Ruilin Ye, Yijun Zhou, Jianwei Zeng, Chen Dai, Wenqing Hong, Wenwen Li, Jun Zhao, and Feihu Xu. Thermal non-line-of-sight imaging through rough surfaces. *ACM Transactions on Graphics*, 45(5):1–21, 2026. doi: 10.1145/3811030.

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.

APPENDIX

SUPPLEMENTARY MATERIAL FOR CORER: COHERENT RELAY EVIDENCE RESTORATION FOR INITIAL HIDDEN-HUMAN CONTOUR OBSERVATION IN CORNER-OCCLUDED SCENES

A EXPERIMENTAL SETUP AND DATASET CONSTRUCTION

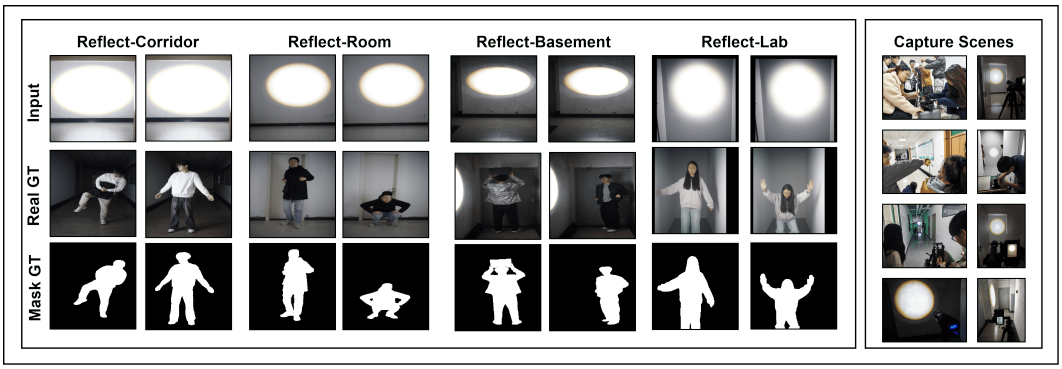

Figure A1: ReflectHuman samples and capture scenes. Left: two examples from each of Reflect-Corridor, Reflect-Room, Reflect-Basement, and Reflect-Lab. Within each scene group, rows show the relay-wall input (Input), synchronized direct-view annotation reference (Real GT), and binary contour mask (Mask GT). Right: photographs of the acquisition process and relay-wall setup. Only relay-wall images enter the model.

This appendix develops the evidence behind the four-scene ReflectHuman evaluation in the main paper. Figure A1 shows representative paired samples and capture scenes. Each sample pairs a 256×256 RGB image of faint reflected-light spots on the relay wall with a binary mask of the hidden person's basic region and main contour. A synchronized direct-view image supports manual annotation and review of the ground truth, and the ordinary wall-facing camera supplies the model input. Reflect-Corridor, Reflect-Room, Reflect-Basement, and Reflect-Lab contain 4,958, 5,232, 10,862, and 3,729 images, respectively, covering different people, corner structures, illumination conditions, and scene settings.

The real captures use a common relay-wall acquisition setup across the four scenes. A Sony A7SII wall-facing camera records at 30 fps with a fixed-focus 50 mm lens, an exposure time of $1/60$ s, ISO 100, and an $f/1.4$ aperture. A flashlight with an approximately 49 mm emitting aperture and 0.4 rad beam divergence illuminates a white-painted concrete relay wall, and the room illumination is reduced during acquisition. The synchronized wall-facing and direct-view cameras record the reflected-light observation and its annotation reference, respectively.

The scene geometries provide distinct relay conditions. Reflect-Corridor uses a T-shaped university-building corridor with a 600 lm flashlight, a 2 m illuminated-wall extent, a 7 m main corridor, a 2 m sensing branch, and a 3.2 m side-wall depth. Reflect-Room uses an apartment living room with a 500 lm flashlight, a 2 m illuminated-wall extent, a 4 m main direction, a 2 m sensing branch, and a 2 m side-wall depth; here, 500 lm denotes the total luminous flux of the source. Reflect-Basement uses a basement corridor with a 1000 lm flashlight, a 1 m illuminated-wall extent, a 5 m corridor direction, a 2 m sensing branch, and a 1.8 m side-wall depth. Reflect-Lab follows the common camera, illumination, synchronized-reference, and manual annotation procedure. This variation changes the size, intensity, and spatial distribution of the observed wall response while retaining the same hidden-human contour objective.

Each scene subset uses the fixed training–validation partition summarized in Table A1. The appendix retains Dice, IoU, and Boundary-F from the main paper and adds MAE, MCC, and HD95, allowing basic-region overlap, contour fidelity, foreground-background agreement, and boundary displacement to be examined together. The Source column records the venue-year tag of each com-

Table A1: Scene-wise training and validation counts for ReflectHuman. These fixed partitions are used throughout the paper.

Dataset subset	Scene role	Train pairs	Validation pairs	Total pairs
Reflect-Corridor	corridor scene	3,965	993	4,958
Reflect-Room	room scene	4,222	1,010	5,232
Reflect-Basement	basement scene	8,689	2,173	10,862
Reflect-Lab	lab-style scene	2,983	746	3,729
Total	ReflectHuman suite	19,859	4,922	24,781

parator. Shared preprocessing, resizing, and evaluation rules connect every appendix result directly to the main-paper evaluation.

All models are trained in PyTorch 2.5 on a single NVIDIA RTX 3080 GPU. Unless otherwise stated, optimization uses AdamW with batch size 16, learning rate 10^{-4}, weight decay 10^{-4}, 50 epochs, StepLR with step size 20 and decay factor 0.5, AMP, and seed 0. The best checkpoint on validation Dice is selected for each run.

Table A1 shows that the 24,781 image pairs span four scene roles and preserve substantial validation sets in every scene. Reflect-Basement supplies the largest subset and Reflect-Lab the smallest, while all four retain hundreds or thousands of validation pairs for scene-wise evaluation. This distribution supports the central test across four real corner settings: an ordinary camera observes faint reflected-light spots on the relay wall, and CoRER converts the surviving evidence into an initial contour observation under scene-dependent geometry and wall appearance.

B Mechanistic Derivation

We trace the computation from wall reflection to hidden-human contour: the relay-wall observation is lifted into a relay basis, RDCR organizes and condenses diffusion-consistent weak evidence, and SCRR restores the basic region and main contour from multiscale support. Equations (A1)–(A4) expand this sequence while preserving the notation and RDCR-to-SCRR dependency established in the main paper.

The first stage examines which weak responses persist across short and long spatial supports after diffuse relay transport. The observed spot combines direct illumination, wall appearance, and a weak target-dependent return, while target-related structure remains more consistent across nearby spatial supports than local wall fluctuation. RDCR therefore projects the observation into a relay basis, probes the same carrier at short and long diffusion spans, and uses their disagreement to regulate the residual update. The coherence envelope assigns greater support to responses that persist across both spans:

$$
\begin{aligned}
b_r(\mathbf{p}) &= \int_{\Omega_x} \Psi_r(\mathbf{p},\mathbf{q})\,\phi(x)(\mathbf{q})\,d\mathbf{q}, \qquad \ell_r = \mathcal{Q}_r(b_r),\\
s_r^{(\kappa)}(\mathbf{p}) &= \int_{\Omega_r} K_{\sigma_\kappa}(\mathbf{p}-\mathbf{q})\,\ell_r(\mathbf{q})\,d\mathbf{q}, \qquad \kappa \in \{s,\ell\},\\
\chi_r(\mathbf{p}) &= \left|s_r^{(s)}(\mathbf{p}) - s_r^{(\ell)}(\mathbf{p})\right| + \mu_r \left|\ell_r(\mathbf{p}) - s_r^{(s)}(\mathbf{p})\right|,\\
\omega_r(\mathbf{p}) &= \frac{\exp[-\lambda_r \chi_r(\mathbf{p})]}{\int_{\Omega_r} \exp[-\lambda_r \chi_r(\mathbf{u})]\,d\mathbf{u}},\\
y_r(\mathbf{p}) &= b_r(\mathbf{p}) + \alpha_r \int_{\Omega_r} \omega_r(\mathbf{q}) \left[\frac{s_r^{(s)}(\mathbf{q}) + s_r^{(\ell)}(\mathbf{q})}{2} + \ell_r(\mathbf{q}) - \chi_r(\mathbf{q})\right] d\mathbf{q}.
\end{aligned}
\tag{A1}
$$

In Equation (A1), b_r and ℓ_r denote the lifted and locally re-encoded relay fields. The short and long responses $s_r^{(s)}$ and $s_r^{(\ell)}$ expose the spatial persistence of the observed return, and χ_r measures both cross-span disagreement and local residual magnitude. The normalized coherence weight ω_r pools

the mean short- and long-support response with the disagreement-adjusted local field $\ell_r - \chi_r$ into y_r, as in Equation (4). This sequence organizes faint target-aligned structure before wall appearance dominates deeper representations.

Contextual condensation carries locally organized weak evidence into the scene-level relay representation used for contour recovery. It combines the RDCR output with broader context and recalibrates the resulting carrier across channels and spatial positions:

$$\begin{aligned} h_b^0 &= \mathcal{E}(y_r), \qquad v_b = \mathcal{B}_r(h_b^0), \qquad h_b = h_b^0 + \eta_b \odot v_b, \\ \vartheta_b^c &= \sigma\Big(\Phi_b^c\big(\operatorname{pool}_c(h_b)\big)\Big), \qquad \vartheta_b^s = \sigma\Big(\Phi_b^s\big(\operatorname{pool}_s(h_b)\big)\Big), \\ \bar{h}_b &= h_b \odot \vartheta_b^c \odot \vartheta_b^s. \end{aligned} \tag{A2}$$

Here, h_b^0 is the encoded RDCR state, v_b supplies the broader contextual response, and η_b regulates their residual integration. The channel and spatial terms ϑ_b^c and ϑ_b^s concentrate the condensed carrier $\bar{h}_b$ around evidence that remains coherent after hierarchical compression. Equation (A2) therefore completes RDCR by transporting locally organized weak evidence into the context available to the restoration path.

SCRR next combines dispersed evidence for the person's overall region with the finer support needed to place its main contour. At each restoration stage, the current state is aligned with the corresponding multiscale return, decomposed into smooth support and residual detail, and evaluated by support, reliability, and agreement. Pooling these quantities across stages lets the decoder use evidence that recurs at several scales when restoring the hidden person's basic region and main contour:

$$\begin{aligned} u_i &= \Pi_i^{\uparrow}(\bar{y}_{i-1}), \qquad \tilde{f}_i = \mathcal{A}_i(u_i, f_i), \qquad \bar{f}_i = \mathcal{F}_i(u_i, f_i), \\ s_i &= \mathcal{S}(\tilde{f}_i), \qquad d_i = \tilde{f}_i - s_i, \qquad h_i^s = \mathcal{S}(u_i), \\ q_i^s &= \cos(h_i^s, s_i), \qquad q_i^d = \cos(u_i - h_i^s, d_i), \\ U_i &= \frac{\langle |s_i| \rangle}{\langle |s_i| \rangle + \langle |d_i| \rangle + \epsilon}, \qquad R_i = \sigma\big(\xi_1 q_i^s + \xi_2 q_i^d - \xi_3 \langle |d_i| \rangle + \xi_4 \langle |s_i| \rangle\big)\,, \\ \mathcal{E}_I &= \frac{1}{|I|} \sum_{i \in I} \Pi_i^{\uparrow}\big[U_i, R_i, \tfrac{1}{2}(1+q_i^s), \tfrac{1}{2}(1+q_i^d)\big]\,. \end{aligned} \tag{A3}$$

In Equation (A3), u_i denotes the aligned decoder state, f_i the encoder skip feature, and $\tilde{f}_i$ the gated skip return formed by $\mathcal{A}_i$. The fusion operator $\mathcal{F}_i$ includes this gating, as in Equation (5). Smooth support s_i and residual detail d_i are extracted from the gated skip return; their agreements q_i^s and q_i^d compare them with the aligned decoder components. The support ratio U_i and reliability R_i summarize the stage evidence, and $\mathcal{E}_I$ places all active stages in a common spatial frame. The resulting evidence stack connects broad region support with finer contour evidence while retaining a reliability estimate for both.

The final step converts the pooled relay evidence into staged contour restoration. The auxiliary state uses the first three restoration stages to establish a coarse region hypothesis, and the main state uses the complete four-stage span to integrate the contour support. Both contour corrections use pooled evidence, current foreground belief, uncertainty, and boundary variation; the main branch then applies positional calibration:

$$\begin{aligned} z_a^0 &= \mathcal{M}_a(\bar{y}_2, \mathcal{E}_{\le 3})\,, \qquad z_m^0 = \mathcal{M}_m(\bar{y}_4, \mathcal{E}_{\le 4})\,, \\ z_j &= z_j^0 + \mathcal{B}_j\big(\mathcal{E}_j, \pi_j, 4\pi_j(1-\pi_j), \|\nabla \mathcal{S}_5(\pi_j)\|_2\big)\,, \qquad j \in \{a, m\}, \\ \hat{y} &= \arg\max z_m. \end{aligned} \tag{A4}$$

In Equation (A4), $\mathcal{E}_a = \mathcal{E}_{\le 3}$ and $\mathcal{E}_m = \mathcal{E}_{\le 4}$ pool evidence over the first three and all four restoration stages, respectively, as in Equation (7). The correction operators update the auxiliary and main logits with evidence appropriate to their decoder spans; $\mathcal{S}_5$ denotes smoothing over a 5×5 neighborhood before boundary variation is evaluated. The main-branch operator $\mathcal{B}_m$ includes positional calibration after contour correction. This calibration uses the coordinate basis $[\xi, \upsilon, e^{-\mu_1 \xi^2}, 1 - |\upsilon|, e^{-\mu_2 \xi^2 - \mu_3 \upsilon^2}]$, where $\xi, \upsilon \in [-1, 1]$ are normalized horizontal and vertical coordinates and $(\mu_1, \mu_2, \mu_3) = (2, 1.5, 0.8)$. Its uncertainty-weighted correction uses a bounded

parameterization. Finally, $\arg\max$ converts the restored main logits into the discrete mask $\hat{y}$. For consistency with Eq. (1), $\hat{y}$ denotes the final binary mask throughout this derivation, whereas y_r denotes the intermediate RDCR relay field. The complete derivation thus closes the relay-to-contour chain: RDCR locates and transports weak diffusion-consistent evidence, and SCRR organizes its multiscale support into the basic region and main contour used for initial contour observation.

The auxiliary logits provide supervision at an intermediate decoder stage, while subsequent stages continue from the decoder features. During evaluation, the main logits supply the contour prediction through $\arg\max$.

C Extended Experimental Results

C.1 Clean Comparison

We evaluate the basic regions and main contours recovered from faint wall reflections across four real corner scenes. Tables A2–A5 evaluate 17 contour baselines from the same three baseline families as the main paper and retain the complete six-metric profile. Dice and IoU measure the recovered basic region, Boundary-F and HD95 characterize the main contour, and MAE and MCC provide complementary foreground-background agreement. The shared row order exposes both the overall method ranking and scene-dependent changes.

Reflect-Corridor establishes a consistent advantage across region and boundary measures. In Table A2, NLOS-OT is the closest overlap competitor, while CoRER raises Dice, IoU, and Boundary-F by 0.017, 0.020, and 0.032 and simultaneously records lower MAE and HD95. The joint movement of overlap, boundary, and error metrics shows that the gain arises from a better supported region together with a more accurate contour, providing an initial contour observation in the corridor geometry.

Reflect-Room provides the closest comparison and confirms that the advantage remains measurable in a compact indoor layout. Table A3 places DG-NLOS nearest in Dice and IoU, with CoRER ahead by 0.009 and 0.011; CoRER also raises Boundary-F by 0.011 over the strongest boundary competitor and records the lowest HD95. Agreement across these metric families indicates that the wall response supports both the hidden person's basic region and the placement of the main contour under the room geometry.

Reflect-Basement shows the clearest basic-region separation among the first three scenes. In Table A4, CoRER improves Dice and IoU over the nearest rows by 0.027 and 0.031 and attains the strongest Boundary-F, MAE, MCC, and HD95 profile. The overlap gain is accompanied by a 3.872-pixel HD95 reduction from the closest HD95 competitor, linking stronger region support to a contour that remains spatially closer to the reference in the basement corridor.

Reflect-Lab extends the comparison to a fourth scene setting and preserves the same ordering. Table A5 shows gains of 0.031 Dice, 0.037 IoU, and 0.030 Boundary-F over the closest alternatives, together with the lowest MAE and HD95. The larger overlap separation and consistent boundary improvement demonstrate that CoRER extracts a usable basic region and main contour from the different wall-response distribution of this scene.

The four tables connect the comparison advantage to the two-stage mechanism. RDCR organizes target-aligned evidence before hierarchical compression, which is reflected in the consistent Dice, IoU, MAE, and MCC gains across wall layouts. SCRR then aggregates support, reliability, and agreement during restoration, matching the Boundary-F and HD95 gains in every scene. Their complementary effects explain why CoRER maintains the leading six-metric profile across different people, corner structures, illumination conditions, and scene settings.

Figure A2 provides the sample-level counterpart to the four quantitative tables. Across the displayed relay-wall observations, CoRER recovers a contiguous basic region while following the reference main contour more closely than the selected hidden-scene perception and image segmentation methods. The comparison rows also expose the two main failure modes summarized by the metrics: incomplete foreground support reduces Dice and IoU, while fragmented or displaced edges reduce Boundary-F and increase HD95.

Table A2: Clean comparison on Reflect-Corridor, grouped by baseline family. Dice, IoU, Boundary-F, and MCC are higher-is-better metrics, whereas MAE and HD95 are lower-is-better. Row backgrounds highlight CoRER (light red) and NLOS-OT (light blue). Red and blue numbers indicate the best and second-best distinct values in each metric column; ties at the displayed precision share the same text color. HD95 is measured in pixels.

Model	Source	Dice ↑	IoU ↑	BF ↑	MAE ↓	MCC ↑	HD95 ↓
Physics- and Transport-Guided Methods							
NLOS-LTM Zhang et al. (2025)	TIP'25	0.684	0.534	0.173	0.130	0.607	36.037
NLOS-OT Geng et al. (2022)	TIP'22	0.695	0.550	0.183	0.125	0.622	35.212
NLOS-LPP Sun et al. (2025)	ICCV'25	0.680	0.531	0.167	0.127	0.606	37.166
CurvNLOS Ding et al. (2024)	TPAMI'24	0.645	0.499	0.174	0.146	0.558	39.234
RISE Torcolacci et al. (2024)	ICC'24	0.684	0.535	0.177	0.127	0.609	36.707
PlatoNeRF Klinghoffer et al. (2024)	CVPR'24	0.641	0.496	0.148	0.145	0.554	41.966
Learning-Based Hidden-Scene Methods							
ST-Mamba Li et al. (2024)	NeurIPS'24	0.681	0.533	0.153	0.130	0.605	37.433
TRT-NLOS Li et al. (2025)	arXiv'25	0.671	0.523	0.163	0.134	0.593	37.742
NLOSFormer Ye et al. (2026)	ACM TOG'26	0.656	0.504	0.174	0.137	0.575	36.761
DG-NLOS Su et al. (2025)	AAAI'25	0.645	0.500	0.161	0.146	0.558	40.009
TransDiff Cui et al. (2025)	TIP'25	0.663	0.516	0.167	0.135	0.583	38.603
NLOST Li et al. (2023)	CVPR'23	0.645	0.495	0.175	0.158	0.555	43.785
General Segmentation Methods							
MK-UNet Rahman & Marculescu (2025)	ICCV'25	0.626	0.471	0.143	0.152	0.540	40.415
Swin-Unet Cao et al. (2023)	ECCV'23	0.647	0.499	0.161	0.147	0.561	39.402
SegFormer Xie et al. (2021)	NeurIPS'21	0.639	0.492	0.147	0.151	0.551	42.700
MISSFormer Huang et al. (2023)	TMI'23	0.653	0.503	0.163	0.141	0.571	37.789
DUCK-Net Dumitru et al. (2023)	Sci Rep'23	0.652	0.507	0.162	0.144	0.569	41.392
CoRER	Ours	0.712	0.570	0.215	0.111	0.647	33.378

Table A3: Clean comparison on Reflect-Room, grouped by baseline family. Dice, IoU, Boundary-F, and MCC are higher-is-better metrics, whereas MAE and HD95 are lower-is-better. Row backgrounds highlight CoRER (light red) and DG-NLOS (light blue). Red and blue numbers indicate the best and second-best distinct values in each metric column; ties at the displayed precision share the same text color. HD95 is measured in pixels.

Model	Source	Dice ↑	IoU ↑	BF ↑	MAE ↓	MCC ↑	HD95 ↓
Physics- and Transport-Guided Methods							
NLOS-LTM Zhang et al. (2025)	TIP'25	0.699	0.549	0.188	0.125	0.625	33.922
NLOS-OT Geng et al. (2022)	TIP'22	0.698	0.548	0.171	0.125	0.624	35.398
NLOS-LPP Sun et al. (2025)	ICCV'25	0.709	0.561	0.172	0.122	0.636	34.750
CurvNLOS Ding et al. (2024)	TPAMI'24	0.700	0.552	0.180	0.128	0.623	33.325
RISE Torcolacci et al. (2024)	ICC'24	0.707	0.561	0.187	0.125	0.631	32.922
PlatoNeRF Klinghoffer et al. (2024)	CVPR'24	0.709	0.562	0.186	0.123	0.635	33.916
Learning-Based Hidden-Scene Methods							
ST-Mamba Li et al. (2024)	NeurIPS'24	0.703	0.552	0.167	0.122	0.633	37.334
TRT-NLOS Li et al. (2025)	arXiv'25	0.712	0.563	0.181	0.120	0.641	32.483
NLOSFormer Ye et al. (2026)	ACM TOG'26	0.710	0.562	0.200	0.123	0.636	33.558
DG-NLOS Su et al. (2025)	AAAI'25	0.722	0.576	0.199	0.118	0.651	32.816
TransDiff Cui et al. (2025)	TIP'25	0.708	0.561	0.201	0.122	0.635	32.881
NLOST Li et al. (2023)	CVPR'23	0.705	0.555	0.196	0.125	0.629	31.513
General Segmentation Methods							
MK-UNet Rahman & Marculescu (2025)	ICCV'25	0.697	0.545	0.159	0.128	0.621	34.899
Swin-Unet Cao et al. (2023)	ECCV'23	0.707	0.559	0.177	0.123	0.634	35.531
SegFormer Xie et al. (2021)	NeurIPS'21	0.707	0.561	0.191	0.126	0.630	34.361
MISSFormer Huang et al. (2023)	TMI'23	0.702	0.554	0.175	0.129	0.624	33.356
DUCK-Net Dumitru et al. (2023)	Sci Rep'23	0.691	0.541	0.181	0.129	0.613	36.324
CoRER	Ours	0.731	0.587	0.212	0.118	0.659	31.331

The qualitative pattern follows the same mechanism as the scene-wise scores. RDCR supplies the restoration path with a more localized target-aligned carrier, helping the prediction retain the hidden person's basic region when the wall spot is dominated by illumination. SCRR then integrates support across decoder scales, closing gaps and limiting spill around the main contour. Figure A2

Table A4: Clean comparison on Reflect-Basement, grouped by baseline family. Dice, IoU, Boundary-F, and MCC are higher-is-better metrics, whereas MAE and HD95 are lower-is-better. Row backgrounds highlight CoRER (light red) and Swin-Unet (light blue). Red and blue numbers indicate the best and second-best distinct values in each metric column; ties at the displayed precision share the same text color. HD95 is measured in pixels.

Model	Source	Dice ↑	IoU ↑	BF ↑	MAE ↓	MCC ↑	HD95 ↓
Physics- and Transport-Guided Methods							
NLOS-LTM Zhang et al. (2025)	TIP'25	0.722	0.572	0.201	0.100	0.666	32.472
NLOS-OT Geng et al. (2022)	TIP'22	0.722	0.577	0.233	0.095	0.668	32.159
NLOS-LPP Sun et al. (2025)	ICCV'25	0.726	0.579	0.225	0.094	0.672	32.270
CurvNLOS Ding et al. (2024)	TPAMI'24	0.722	0.574	0.220	0.098	0.667	32.798
RISE Torcolacci et al. (2024)	ICC'24	0.729	0.583	0.241	0.092	0.677	31.479
PlatoNeRF Klinghoffer et al. (2024)	CVPR'24	0.729	0.582	0.217	0.093	0.676	32.495
Learning-Based Hidden-Scene Methods							
ST-Mamba Li et al. (2024)	NeurIPS'24	0.719	0.572	0.213	0.097	0.664	34.123
TRT-NLOS Li et al. (2025)	arXiv'25	0.708	0.559	0.217	0.101	0.651	32.956
NLOSFormer Ye et al. (2026)	ACM TOG'26	0.728	0.582	0.229	0.093	0.675	32.557
DG-NLOS Su et al. (2025)	AAAI'25	0.725	0.579	0.222	0.094	0.671	31.846
TransDiff Cui et al. (2025)	TIP'25	0.721	0.574	0.212	0.096	0.667	33.032
NLOST Li et al. (2023)	CVPR'23	0.712	0.560	0.204	0.100	0.656	32.610
General Segmentation Methods							
MK-UNet Rahman & Marculescu (2025)	ICCV'25	0.730	0.583	0.229	0.087	0.682	31.849
Swin-Unet Cao et al. (2023)	ECCV'23	0.733	0.588	0.232	0.093	0.681	31.607
SegFormer Xie et al. (2021)	NeurIPS'21	0.723	0.575	0.222	0.097	0.668	33.044
MISSFormer Huang et al. (2023)	TMI'23	0.721	0.572	0.222	0.097	0.666	32.376
DUCK-Net Dumitru et al. (2023)	Sci Rep'23	0.731	0.585	0.227	0.092	0.679	30.961
CoRER	Ours	0.760	0.619	0.248	0.084	0.712	27.089

Table A5: Clean comparison on Reflect-Lab, grouped by baseline family. Dice, IoU, Boundary-F, and MCC are higher-is-better metrics, whereas MAE and HD95 are lower-is-better. Row backgrounds highlight CoRER (light red) and NLOS-LTM (light blue). Red and blue numbers indicate the best and second-best distinct values in each metric column; ties at the displayed precision share the same text color. HD95 is measured in pixels.

Model	Source	Dice ↑	IoU ↑	BF ↑	MAE ↓	MCC ↑	HD95 ↓
Physics- and Transport-Guided Methods							
NLOS-LTM Zhang et al. (2025)	TIP'25	0.755	0.620	0.171	0.097	0.699	30.061
NLOS-OT Geng et al. (2022)	TIP'22	0.720	0.577	0.137	0.112	0.658	35.412
NLOS-LPP Sun et al. (2025)	ICCV'25	0.693	0.547	0.125	0.124	0.623	37.489
CurvNLOS Ding et al. (2024)	TPAMI'24	0.692	0.545	0.126	0.123	0.622	38.020
RISE Torcolacci et al. (2024)	ICC'24	0.686	0.539	0.118	0.126	0.615	38.118
PlatoNeRF Klinghoffer et al. (2024)	CVPR'24	0.641	0.493	0.104	0.144	0.559	41.462
Learning-Based Hidden-Scene Methods							
ST-Mamba Li et al. (2024)	NeurIPS'24	0.698	0.554	0.127	0.121	0.629	37.176
TRT-NLOS Li et al. (2025)	arXiv'25	0.693	0.548	0.128	0.122	0.623	37.385
NLOSFormer Ye et al. (2026)	ACM TOG'26	0.685	0.540	0.124	0.125	0.615	37.941
DG-NLOS Su et al. (2025)	AAAI'25	0.684	0.536	0.116	0.125	0.612	38.057
TransDiff Cui et al. (2025)	TIP'25	0.676	0.529	0.122	0.128	0.604	37.921
NLOST Li et al. (2023)	CVPR'23	0.704	0.558	0.145	0.117	0.637	35.448
General Segmentation Methods							
MK-UNet Rahman & Marculescu (2025)	ICCV'25	0.737	0.597	0.165	0.097	0.682	31.705
Swin-Unet Cao et al. (2023)	ECCV'23	0.689	0.543	0.125	0.126	0.616	36.189
SegFormer Xie et al. (2021)	NeurIPS'21	0.706	0.564	0.130	0.120	0.638	35.799
MISSFormer Huang et al. (2023)	TMI'23	0.686	0.541	0.128	0.125	0.614	37.400
DUCK-Net Dumitru et al. (2023)	Sci Rep'23	0.750	0.610	0.154	0.105	0.694	31.418
CoRER	Ours	0.786	0.657	0.201	0.081	0.739	28.096

therefore shows how the statistical advantage becomes an initial contour observation in representative samples.

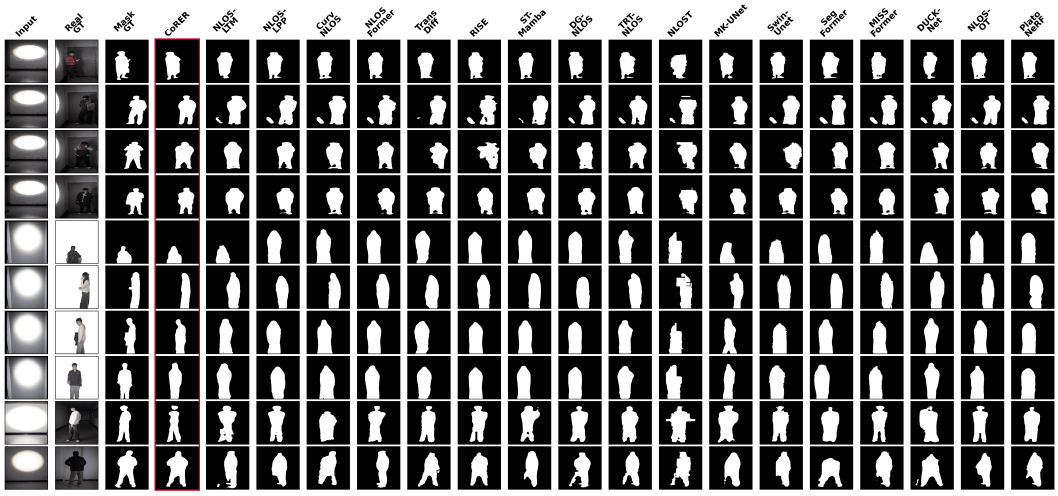

Figure A2: Extended clean qualitative comparison across additional relay-wall observations. Each row shows a relay-wall input, its reference image and reference mask, and predictions from CoRER and the selected baselines, complementing the scene-wise results in Tables A2–A5.

C.2 Ablation

The ablation study isolates how the complete relay-to-contour path produces its comparison advantage. Tables A6–A9 retain the common training and six-metric evaluator while organizing the rows into component removal, design variant, and replacement studies. Component removal measures the contribution of each named mechanism, design variants test simplified forms of the same stage, and replacements compare RDCR or SCRR with alternative entries.

The component-removal results establish complementary roles in the four-scene macro average, with scene-dependent boundary sensitivity. Removing RDCR lowers Dice by 0.038, 0.039, 0.045, and 0.053 from Reflect-Corridor through Reflect-Lab and lowers IoU by 0.041, 0.046, 0.053, and 0.065. Removing SCRR produces smaller overlap reductions and a larger macro-average Boundary-F reduction, with scene-wise Boundary-F losses of 0.056, 0.064, 0.045, and 0.056. The complete model therefore gains basic-region coverage from front-end evidence organization and main-contour fidelity from back-end scale-coherent restoration.

Reflect-Corridor makes this division visible in both removal and replacement rows. VMamba gathers image context through four-direction selective scans Liu et al. (2024), while MambaVision combines Mamba mixing with self-attention Hatamizadeh & Kautz (2025). Table A6 shows that the RDCR removal and all three RDCR replacements achieve lower Dice and IoU than the complete model, whereas the SCRR removal produces the largest Boundary-F drop among the component removals. The RDCR and SCRR design variants recover part of the full-model profile through their respective evidence-organization roles.

Reflect-Room strengthens the contour-restoration evidence. In Table A7, removing SCRR lowers Boundary-F from 0.212 to 0.148, while removing RDCR lowers Dice from 0.731 to 0.692. The replacement rows preserve the same distinction: SCRR alternatives retain portions of the region score yet leave a measurable contour gap, and RDCR alternatives remain below CoRER on the basic-region metrics. This scene links the two modules to separate, complementary parts of the prediction.

Reflect-Basement exposes the cost of the SAMRefiner replacement. SAMRefiner derives point, box, and mask prompts from coarse masks to guide SAM refinement Lin et al. (2025). Table A8 shows that $\mathrm{SCRR} \rightarrow \mathrm{SAMRefiner}$ matches CoRER at 0.248 Boundary-F, but has lower Dice (0.727 versus 0.760) and higher HD95 (44.328 versus 27.089 pixels). This metric combination pairs local boundary agreement with a larger contour-distance error. The complete SCRR path integrates multiscale support throughout restoration, preserving local boundary agreement while achieving lower overall contour-distance error.

Table A6: Ablation results on Reflect-Corridor, grouped as component removal, design variant, and replacement studies. Dice, IoU, Boundary-F, and MCC are higher-is-better metrics, whereas MAE and HD95 are lower-is-better. Row backgrounds highlight CoRER (light red) and the SCRR variant (light blue). Red and blue numbers indicate the best and second-best distinct values in each metric column; ties at the displayed precision share the same text color. HD95 is measured in pixels.

Variant	Dice ↑	IoU ↑	BF ↑	MAE ↓	MCC ↑	HD95 ↓
Component Removal						
w/o RDCR	0.674	0.529	0.183	0.128	0.599	36.553
w/o SCRR	0.696	0.548	0.159	0.122	0.627	36.554
Design Variant						
RDCR (variant)	0.691	0.550	0.206	0.117	0.623	34.529
SCRR (variant)	0.699	0.557	0.208	0.119	0.629	34.810
Replacement						
RDCR→VMamba	0.664	0.511	0.177	0.133	0.587	36.735
RDCR→MambaVision	0.666	0.516	0.150	0.147	0.583	39.465
RDCR→SeNaTra	0.663	0.511	0.175	0.135	0.584	37.276
SCRR→SAMRefiner	0.698	0.555	0.199	0.118	0.628	34.451
SCRR→Mask-Adapter	0.699	0.556	0.206	0.115	0.632	34.899
SCRR→QueryMeldNet	0.694	0.549	0.203	0.119	0.624	34.949
CoRER	0.712	0.570	0.215	0.111	0.647	33.378

Table A7: Ablation results on Reflect-Room, grouped as component removal, design variant, and replacement studies. Dice, IoU, Boundary-F, and MCC are higher-is-better metrics, whereas MAE and HD95 are lower-is-better. Row backgrounds highlight CoRER (light red) and the SCRR variant (light blue). Red and blue numbers indicate the best and second-best distinct values in each metric column; ties at the displayed precision share the same text color. HD95 is measured in pixels.

Variant	Dice ↑	IoU ↑	BF ↑	MAE ↓	MCC ↑	HD95 ↓
Component Removal						
w/o RDCR	0.692	0.541	0.192	0.126	0.617	35.335
w/o SCRR	0.706	0.556	0.148	0.129	0.628	35.089
Design Variant						
RDCR (variant)	0.702	0.551	0.173	0.124	0.629	37.961
SCRR (variant)	0.709	0.561	0.201	0.121	0.636	33.706
Replacement						
RDCR→VMamba	0.686	0.533	0.169	0.138	0.602	36.195
RDCR→MambaVision	0.693	0.542	0.200	0.131	0.612	33.674
RDCR→SeNaTra	0.694	0.540	0.157	0.138	0.612	36.595
SCRR→SAMRefiner	0.701	0.552	0.203	0.124	0.626	33.136
SCRR→Mask-Adapter	0.700	0.551	0.185	0.131	0.619	32.725
SCRR→QueryMeldNet	0.707	0.557	0.159	0.128	0.631	35.570
CoRER	0.731	0.587	0.212	0.118	0.659	31.331

Reflect-Lab produces the largest RDCR overlap reduction and a clear SCRR boundary reduction. Table A9 shows a 0.053 Dice and 0.065 IoU loss after RDCR removal, together with a 0.056 Boundary-F loss after SCRR removal. The complete model leads every variant and replacement across the six metrics, showing that the fourth scene benefits from both early weak-evidence organization and later contour-state restoration.

The four ablation tables close the mechanism argument established by the clean comparison. RDCR converts faint reflected-light spots on the relay wall into a target-aligned carrier that supports the basic region, and SCRR converts distributed multiscale support into the main contour. Their gains appear in different metric families and combine in CoRER, explaining the complete model's scene-wise advantage over the removal, variant, and replacement rows.

Figure A3 shows how the three ablation families alter representative contours. RDCR removals and replacements reduce or displace foreground support, producing incomplete basic regions that correspond to the Dice and IoU reductions in Tables A6–A9. SCRR removals and replacements

Table A8: Ablation results on Reflect-Basement, grouped as component removal, design variant, and replacement studies. Dice, IoU, Boundary-F, and MCC are higher-is-better metrics, whereas MAE and HD95 are lower-is-better. Row backgrounds highlight CoRER (light red) and SCRR→SAMRefiner (light blue). Red and blue numbers indicate the best and second-best distinct values in each metric column; ties at the displayed precision share the same text color. HD95 is measured in pixels.

Variant	Dice ↑	IoU ↑	BF ↑	MAE ↓	MCC ↑	HD95 ↓
Component Removal						
w/o RDCR	0.715	0.566	0.195	0.098	0.660	33.926
w/o SCRR	0.725	0.578	0.203	0.090	0.675	33.218
Design Variant						
RDCR (variant)	0.722	0.574	0.237	0.088	0.674	31.684
SCRR (variant)	0.729	0.584	0.239	0.089	0.679	29.983
Replacement						
RDCR→VMamba	0.717	0.567	0.206	0.097	0.662	32.907
RDCR→MambaVision	0.724	0.578	0.231	0.093	0.672	32.427
RDCR→SeNaTra	0.720	0.571	0.222	0.095	0.666	32.085
SCRR→SAMRefiner	0.727	0.581	0.248	0.096	0.675	44.328
SCRR→Mask-Adapter	0.726	0.578	0.233	0.090	0.676	30.865
SCRR→QueryMeldNet	0.729	0.582	0.236	0.090	0.677	30.452
CoRER	0.760	0.619	0.248	0.084	0.712	27.089

Table A9: Ablation results on Reflect-Lab, grouped as component removal, design variant, and replacement studies. Dice, IoU, Boundary-F, and MCC are higher-is-better metrics, whereas MAE and HD95 are lower-is-better. Row backgrounds highlight CoRER (light red) and the SCRR variant (light blue). Red and blue numbers indicate the best and second-best distinct values in each metric column; ties at the displayed precision share the same text color. HD95 is measured in pixels.

Variant	Dice ↑	IoU ↑	BF ↑	MAE ↓	MCC ↑	HD95 ↓
Component Removal						
w/o RDCR	0.733	0.592	0.150	0.104	0.673	32.554
w/o SCRR	0.740	0.600	0.145	0.100	0.684	33.302
Design Variant						
RDCR (variant)	0.739	0.604	0.165	0.098	0.682	31.277
SCRR (variant)	0.748	0.611	0.165	0.099	0.690	30.338
Replacement						
RDCR→VMamba	0.733	0.595	0.163	0.097	0.678	33.445
RDCR→MambaVision	0.736	0.600	0.170	0.097	0.681	31.319
RDCR→SeNaTra	0.736	0.601	0.169	0.102	0.677	30.795
SCRR→SAMRefiner	0.743	0.604	0.162	0.096	0.687	30.077
SCRR→Mask-Adapter	0.745	0.607	0.169	0.092	0.693	32.169
SCRR→QueryMeldNet	0.749	0.610	0.159	0.102	0.692	31.541
CoRER	0.786	0.657	0.201	0.081	0.739	28.096

preserve more of the coarse region while introducing contour breaks, local adhesion, or exterior spill, matching the larger Boundary-F and HD95 changes.

The design-variant rows recover more complete basic regions and main contours than the component removals, yet CoRER maintains the most complete region and the most coherent main contour in the displayed cases. This visual ordering confirms the quantitative module split: RDCR determines the quality of the evidence entering restoration, and SCRR determines how that evidence is organized into a coherent final boundary. Their composition turns the ordinary-camera wall response into a more reliable initial contour observation.

C.3 Robustness Overview

The robustness evaluation asks whether the relay evidence remains usable when acquisition geometry, wall appearance, and signal quality change. The same CoRER checkpoint and per-image metric



Figure A3: Extended qualitative ablation comparison. Each row shows a relay-wall input, its reference image and reference mask, and predictions from CoRER and the component removals, design variants, and replacements. Tables A6–A9 report the corresponding per-scene results.

computation used for the clean comparison are applied throughout. Tables A10–A12 use Reflect-Room to isolate camera angle, acquisition distance, and relay-wall material; the subsequent four-scene figures evaluate relative brightness attenuation, additive corruption, multiplicative corruption, phase irregularity, and spatial softening.

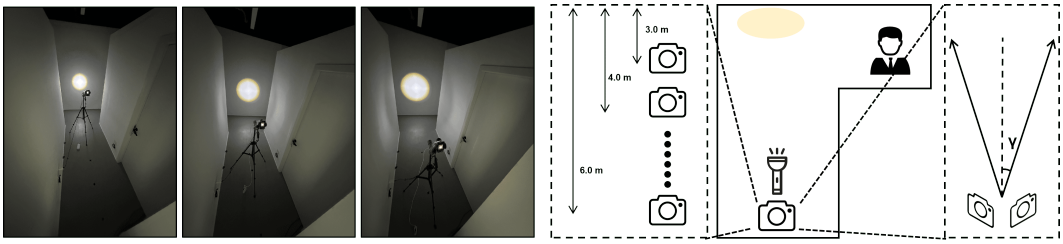

Figure A4: Acquisition geometry for the distance and camera-angle robustness experiments. From left to right: representative photographs of the distance acquisition setup, an acquisition-distance schematic, the corner layout, and a schematic angular offset (γ). Table A10 specifies the tested pitch and yaw offsets.

C.3.1 Camera-Angle Robustness

The real-scene camera-angle experiment evaluates CoRER on Reflect-Room under seven controlled offsets, as summarized in Table A10. Figure A4 illustrates the angular offset schematically. The nominal view uses zero pitch and yaw, followed by separate pitch or yaw offsets of 2°, 3°, and 5°. The trained checkpoint, input resolution, contour targets, and per-image evaluator remain fixed across these conditions, and the table reports Dice, IoU, Boundary-F, MAE, MCC, and HD95 for each offset.

Table A10: Real-scene camera-angle robustness of CoRER on Reflect-Room. Dice, IoU, Boundary-F, and MCC are higher-is-better metrics, whereas MAE and HD95 are lower-is-better. The rows with the highest and second-highest displayed Dice values are highlighted in red and blue; displayed ties share the same highlighting. HD95 is measured in pixels.

Camera offset	Dice ↑	IoU ↑	BF ↑	MAE ↓	MCC ↑	HD95 ↓
Pitch 0° / Yaw 0°	0.731	0.586	0.212	0.118	0.659	31.369
Pitch 2° / Yaw 0°	0.729	0.583	0.208	0.120	0.656	31.682
Pitch 0° / Yaw 2°	0.729	0.583	0.207	0.120	0.655	31.779
Pitch 3° / Yaw 0°	0.727	0.581	0.205	0.120	0.654	31.907
Pitch 0° / Yaw 3°	0.727	0.580	0.204	0.121	0.653	32.087
Pitch 5° / Yaw 0°	0.724	0.577	0.200	0.122	0.649	32.421
Pitch 0° / Yaw 5°	0.722	0.574	0.197	0.123	0.646	32.775

The overlap and boundary measures follow a gradual offset-dependent change. At the 2° offsets, Dice remains within 0.0024 of the nominal value and Boundary-F remains within 0.0052, indicating limited variation near the reference view. Increasing the pitch to 5° changes Dice from 0.7309 to 0.7236 and Boundary-F from 0.2120 to 0.1998, while the corresponding yaw offset gives 0.7217 Dice and 0.1974 Boundary-F. The error measures move in the complementary direction: MAE increases by 0.0044 for the largest pitch offset and by 0.0053 for the largest yaw offset, while HD95 increases by 1.05 and 1.41 pixels, respectively. The full trajectory therefore shows a smooth reduction in contour agreement as the observation departs from the reference angle, with yaw producing the larger endpoint change.

The angle response is consistent with a controlled displacement of the wall-carried spatial pattern. RDCR compares short- and long-support responses before deeper encoding, allowing a moderately shifted diffuse return to remain represented through local coherence across image positions. SCRR subsequently aggregates support and agreement across decoder scales, which preserves the shared contour structure while the offset changes the alignment of fine boundary evidence. The larger change at 5° reflects the increasing spatial mismatch of the relay response, whereas the retained overlap and boundary scores show that the two-stage path continues to provide an initial contour readout across the tested angular range.

C.3.2 Acquisition-Distance Robustness

The real-scene acquisition-distance experiment evaluates the same Reflect-Room checkpoint at relay-wall distances from 3.0 to 6.0 m, with the scene's 4.0 m main direction serving as the reference acquisition geometry. Figure A4 shows the distance geometry and representative capture photographs. Table A11 reports the six contour metrics at 3.0, 4.0, 4.5, 5.0, and 6.0 m. The Reflect-Room setup uses a 500 lm source, a 2 m illuminated-wall extent, a 2 m sensing branch, and a 2 m side-wall depth; 500 lm records source luminous flux. The common checkpoint and evaluator isolate how the contour readout changes as the wall-response footprint and strength vary with acquisition distance.

The distance sweep remains within a narrow metric envelope with small local fluctuations. Dice spans 0.7250–0.7341, IoU spans 0.5789–0.5898, and Boundary-F spans 0.2120–0.2211 over the complete range. Relative to the 4.0 m reference, the 6.0 m endpoint changes Dice by -0.0059 and IoU by -0.0074, while Boundary-F remains 0.0011 higher. MAE increases by 0.0016, MCC changes by -0.0057, and HD95 increases by 0.28 pixels at the farthest condition. The 3.0 and 4.5 m rows produce slightly higher overlap scores than the reference, whereas the two longer distances show a mild reduction, indicating that the usable relay evidence is maintained throughout the tested interval.

Distance changes jointly affect response amplitude, projected spatial scale, and the width of the diffuse support on the relay wall. RDCR organizes evidence across short and long spatial supports, so modest changes in the response footprint can be absorbed by the same coherence relation before contextual condensation. SCRR then pools smooth support, detail, agreement, and reliability across decoder stages, allowing contour evidence observed at a changed scale to contribute to the final state. This interaction explains the small non-monotone fluctuations at intermediate distances

Table A11: Real-scene acquisition-distance robustness of CoRER on Reflect-Room. Dice, IoU, Boundary-F, and MCC are higher-is-better metrics, whereas MAE and HD95 are lower-is-better. The rows with the highest and second-highest displayed Dice values are highlighted in red and blue; displayed ties share the same highlighting. HD95 is measured in pixels.

Distance	Dice ↑	IoU ↑	BF ↑	MAE ↓	MCC ↑	HD95 ↓
3.0 m	0.734	0.590	0.221	0.117	0.663	30.962
4.0 m	0.731	0.586	0.212	0.118	0.658	31.375
4.5 m	0.733	0.588	0.215	0.118	0.661	31.425
5.0 m	0.727	0.582	0.215	0.118	0.656	31.567
6.0 m	0.725	0.579	0.213	0.120	0.653	31.658

Table A12: Real-scene relay-wall material robustness of CoRER on Reflect-Room. Dice, IoU, Boundary-F, and MCC are higher-is-better metrics, whereas MAE and HD95 are lower-is-better. The rows with the highest and second-highest displayed Dice values are highlighted in red and blue; displayed ties share the same highlighting. HD95 is measured in pixels.

Relay-wall material	Dice ↑	IoU ↑	BF ↑	MAE ↓	MCC ↑	HD95 ↓
Matte white wall	0.731	0.586	0.212	0.118	0.659	31.369
Rough textured wallpaper	0.732	0.587	0.213	0.118	0.659	31.338
Diffuse-scattering wallpaper	0.729	0.585	0.211	0.118	0.657	31.417
Contaminated wallpaper	0.730	0.586	0.215	0.118	0.658	31.460
Large-area two-tone wall	0.731	0.587	0.212	0.118	0.659	31.495

and the gradual overlap reduction at the far endpoint while boundary fidelity remains close to the reference condition.

C.3.3 Relay-Wall Material Robustness

The real-scene wall-material experiment measures the Reflect-Room checkpoint under five relay surfaces: the matte white reference wall, rough textured wallpaper, diffuse-scattering wallpaper, contaminated wallpaper, and a large-area two-tone boundary condition. Table A12 reports the same six metrics for every surface. The scene layout, target masks, checkpoint, and evaluator are shared across rows, making wall appearance and scattering behavior the controlled acquisition factor.

The five material conditions produce closely grouped overlap and error values. Dice varies by 0.0024 across the full table, IoU by 0.0026, and MCC by 0.0021. Rough textured wallpaper gives the highest Dice, IoU, and MCC and the lowest HD95, while contaminated wallpaper gives the highest Boundary-F and lowest MAE. The diffuse-scattering condition produces the lowest overlap values, yet its Dice remains 0.7292 and its Boundary-F remains 0.2109. The two-tone boundary condition stays within 0.0002 Dice of the reference and within 0.00002 Boundary-F, while the complete HD95 range is 0.16 pixels. These narrow ranges show that the initial contour readout is maintained across the tested wall appearances, including surfaces with texture, scattering variation, contamination, and a large photometric boundary.

Wall materials primarily alter the local texture, contrast, and diffusion profile superimposed on the hidden-human return. RDCR addresses this variation at the representation entrance by comparing the relay response over two spatial supports and weighting the residual according to their coherence, which limits the influence of material-specific local structure. SCRR evaluates the resulting contour support across decoder scales and restores evidence that remains spatially consistent through the fusion path. The combination aligns with the small spread across all six metrics: material appearance changes the observed wall pattern, while the coherent component needed for the basic region remains available to the restoration path.

C.3.4 Brightness

The real-scene brightness sweep uses separately captured relay-wall images at each brightness level, with scene geometry and the reference contour held fixed. Brightness is reported in nit ($1\ \mathrm{nit} = 1\ \mathrm{cd/m^2}$), with 300 nit as the reference. Each acquisition is evaluated with the same trained model.

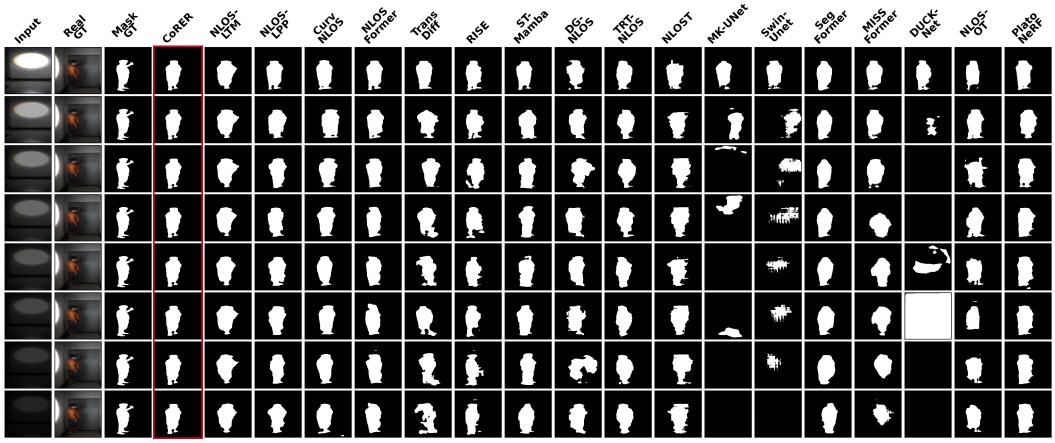

Figure A5: Representative predictions across the brightness sweep. Successive rows show the same scene instance at different brightness levels. Columns show the relay-wall input, reference image, reference mask, and method predictions.

Figure A5 shows representative predictions, and Figures A6–A9 report all six metrics for the four scenes across the common brightness levels.

The representative rows in Figure A5 retain the hidden person's basic region as the observed wall response becomes dimmer. Fine contour sections change earlier than the central foreground support, which matches the distinction between overlap and boundary metrics in the scene grids. CoRER preserves the basic region and main contour throughout the displayed conditions, while selected comparison methods lose basic-region support or produce fragmented boundaries earlier in the same rows.

Reflect-Corridor shows a small and ordered attenuation response. In Figure A6, CoRER moves from 0.7122 to 0.7098 Dice and from 0.5699 to 0.5663 IoU, while Boundary-F remains above 0.211 at the darkest endpoint. CoRER also preserves a 0.045 Dice and 0.041 Boundary-F advantage over the strongest displayed alternatives at that endpoint, showing that the initial contour observation remains supported as the wall spot weakens.

Reflect-Room provides the flattest cross-metric trajectory. Figure A7 shows Dice changing by less than 0.001, IoU remaining near 0.586, and HD95 changing by about 0.07 pixels from the reference endpoint. This consistency indicates that brightness attenuation changes the response amplitude while leaving the target-aligned spatial organization available to the model in the room setting.

Reflect-Basement exhibits the largest overlap change and preserves a clear comparison margin. In Figure A8, Dice changes from 0.7596 to 0.7532 and IoU from 0.6191 to 0.6115, while Boundary-F stays near its reference value. At the darkest endpoint, CoRER remains 0.071 Dice and 0.060 Boundary-F above the closest comparison traces and records a substantially lower HD95, demonstrating stable region and contour support under the strongest tested attenuation.

Reflect-Lab combines the highest overlap scores with minimal brightness sensitivity. Figure A9 keeps Dice near 0.785 and IoU near 0.657 across all levels, and Boundary-F slightly increases from the reference endpoint to the darkest condition. The representative mask remains spatially organized despite the attenuated input, indicating that the wall response retains a target-aligned pattern that CoRER can use in this scene.

The four-scene pattern matches the division of labor between RDCR and SCRR. Brightness attenuation weakens the observed response; RDCR compares local evidence across spatial supports and transports the coherent component into deeper context. SCRR then aggregates the remaining support and reliability across decoder scales, preserving the basic region and restoring the main contour. Their composition explains why the trajectories remain compact and why CoRER retains its comparison advantage at the lowest tested brightness.

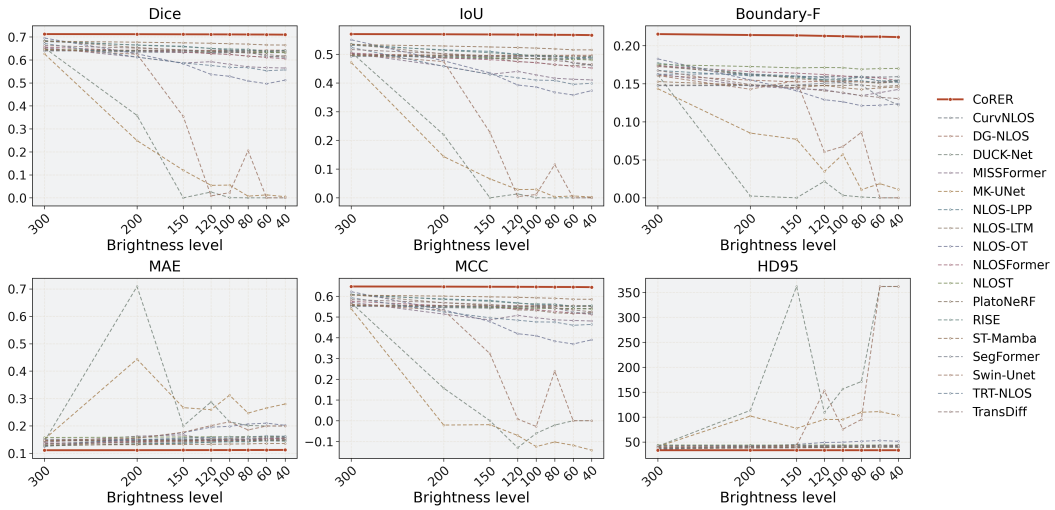

Figure A6: Metric trajectories under the brightness sweep on Reflect-Corridor; horizontal axes show brightness in nit. The six panels report Dice, IoU, Boundary-F, MAE, MCC, and HD95; higher is better except for MAE and HD95.

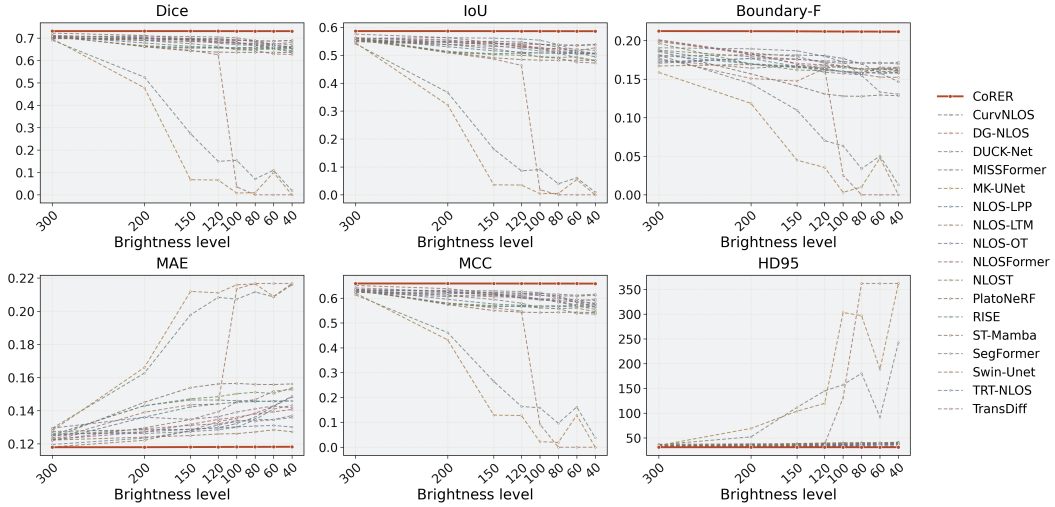

Figure A7: Metric trajectories under the brightness sweep on Reflect-Room; horizontal axes show brightness in nit. The six panels report Dice, IoU, Boundary-F, MAE, MCC, and HD95; higher is better except for MAE and HD95.

C.3.5 Gaussian Noise

The Gaussian-noise experiment evaluates additive local corruption across the four scene validation sets. Figure A10 follows fixed representative samples through the tested SNR levels, and Figures A11–A14 trace Dice, IoU, Boundary-F, MAE, MCC, and HD95 for CoRER and the comparison methods. The experiment asks whether random pointwise fluctuations obscure the spatially persistent evidence required for an initial contour observation.

Figure A10 shows that increasing additive noise first affects fine edge placement while the basic region remains supported. CoRER retains a connected basic region and follows the main contour through the displayed levels, whereas several comparison outputs accumulate small islands or lose parts of the basic region. This visual progression corresponds to the earlier movement of Boundary-F and HD95 relative to Dice and IoU in the metric grids.

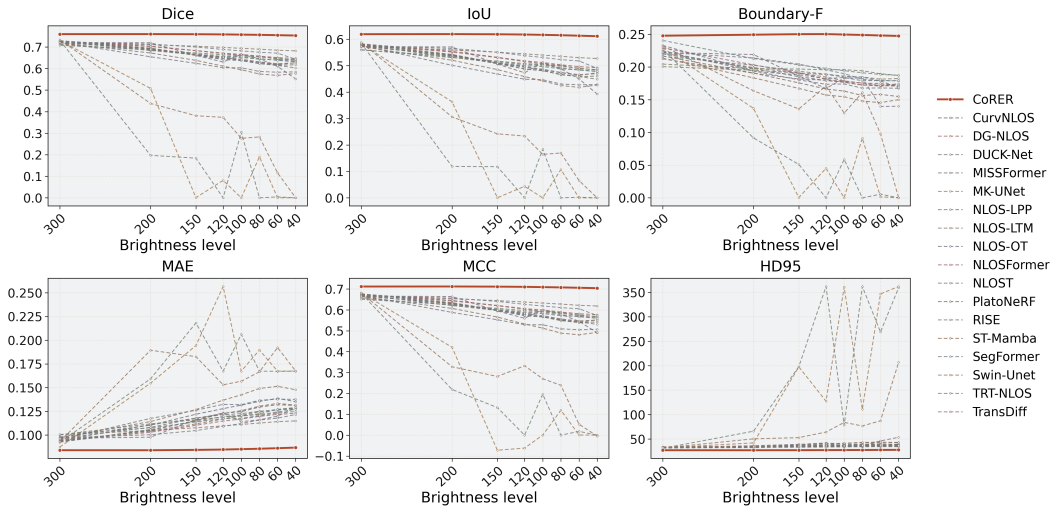

Figure A8: Metric trajectories under the brightness sweep on Reflect-Basement; horizontal axes show brightness in nit. The six panels report Dice, IoU, Boundary-F, MAE, MCC, and HD95; higher is better except for MAE and HD95.

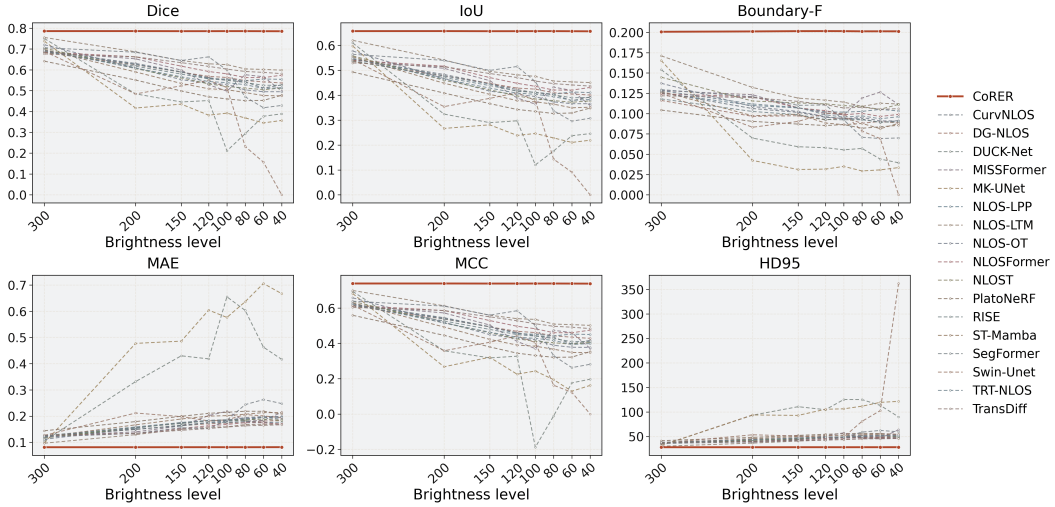

Figure A9: Metric trajectories under the brightness sweep on Reflect-Lab; horizontal axes show brightness in nit. The six panels report Dice, IoU, Boundary-F, MAE, MCC, and HD95; higher is better except for MAE and HD95.

Reflect-Corridor shows a gradual additive-noise response with a persistent comparison margin. Figure A11 records a Dice change from 0.7120 to 0.7097 and an IoU change from 0.5696 to 0.5666; at 75 dB, CoRER remains 0.086 Dice and 0.026 Boundary-F above the closest comparison traces. The stable HD95 indicates that the main contour position is retained together with the foreground region.

Reflect-Room remains nearly invariant across the Gaussian sweep. In Figure A12, Dice stays near 0.731, IoU near 0.586, and HD95 within a narrow range around 31.36 pixels. CoRER retains measurable advantages in both overlap and boundary fidelity at the strongest level, showing that the room-scene wall pattern continues to support the same main contour under additive fluctuation.

Reflect-Basement presents the strongest absolute contour profile and the smallest perturbation-induced change. Figure A13 shows less than 0.001 variation in Dice and IoU across the sweep, while Boundary-F remains near 0.248 and HD95 near 27.08 pixels. The gap over the closest com-

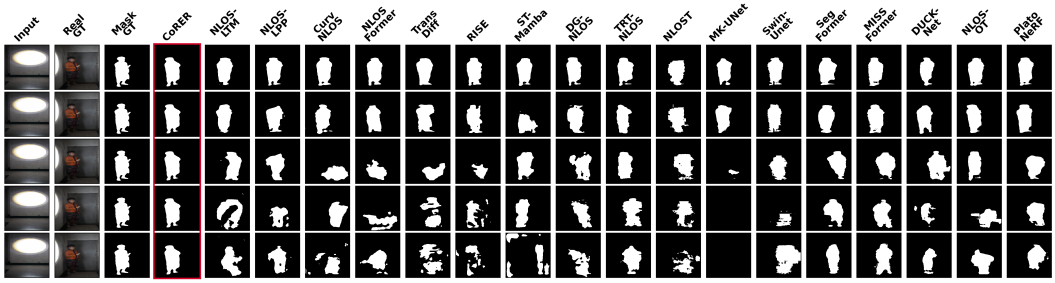

Figure A10: Representative predictions under Gaussian noise. Successive rows show the same scene instance at different Gaussian-noise levels. Columns show the relay-wall input, reference image, reference mask, and method predictions.

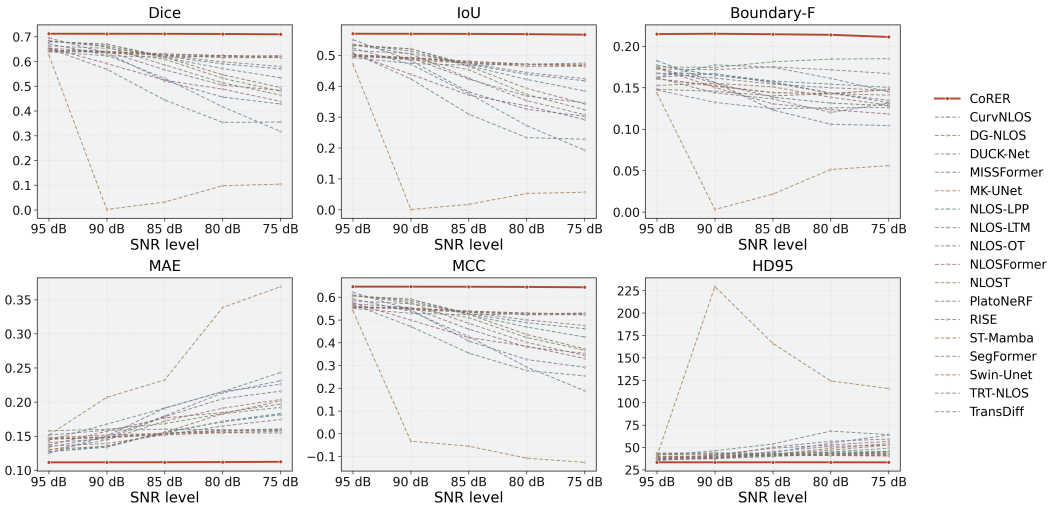

Figure A11: Metric trajectories under Gaussian noise on Reflect-Corridor. The six panels report Dice, IoU, Boundary-F, MAE, MCC, and HD95; higher is better except for MAE and HD95.

parison trace exceeds 0.170 Dice at the strongest level, linking stable target support to the model's cross-scene comparison advantage.

Reflect-Lab also retains a compact six-metric envelope. Figure A14 shows less than 0.0005 Dice variation from the first to the last Gaussian level and only a small Boundary-F change. The resulting mask preserves the basic region and main contour while additive noise changes local pixel values, confirming that the fourth scene follows the same evidence-stability pattern.

These results align with the two-stage mechanism. RDCR organizes responses that persist over short and long spatial supports, reducing the influence of unstructured local oscillation before deeper encoding. SCRR then restores the contour from agreement distributed across decoder scales, so isolated noise contributes less than recurring region and boundary support. The four scene grids therefore connect additive-noise stability to the same weak-evidence organization that drives the clean comparison.

C.3.6 Speckle Noise

The speckle experiment tests signal-dependent changes in the local wall-response amplitude. Figure A15 displays fixed representative samples across the tested SNR levels, and Figures A16–A19 report the four scene trajectories. This perturbation complements Gaussian noise by scaling the observed carrier locally, directly testing whether CoRER follows spatial organization across speckle variation.

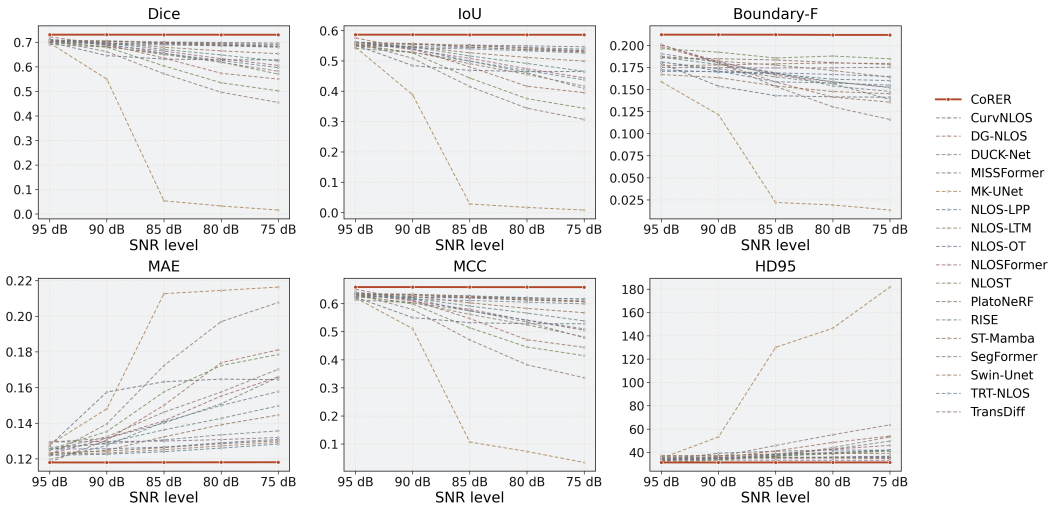

Figure A12: Metric trajectories under Gaussian noise on Reflect-Room. The six panels report Dice, IoU, Boundary-F, MAE, MCC, and HD95; higher is better except for MAE and HD95.

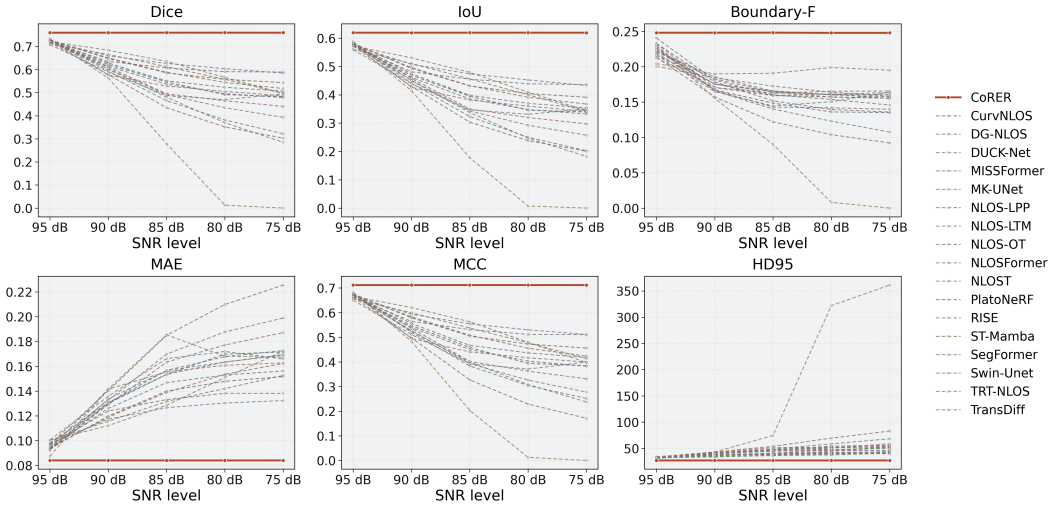

Figure A13: Metric trajectories under Gaussian noise on Reflect-Basement. The six panels report Dice, IoU, Boundary-F, MAE, MCC, and HD95; higher is better except for MAE and HD95.

The representative predictions in Figure A15 preserve the same basic region and main contour as speckle increases. Local contour sections change before the central support, yet CoRER retains a connected basic region and main contour across the rows. The visual behavior is mirrored by compact overlap curves and slightly more responsive boundary curves.

Reflect-Corridor shows the largest relative change within this family. Figure A16 records a Dice change from 0.7118 to 0.7096 and an IoU change from 0.5693 to 0.5664, while CoRER remains 0.080 Dice and 0.050 Boundary-F above the closest comparison traces at the strongest level. The retained HD95 advantage shows that speckle has limited influence on global contour placement.

Reflect-Room remains tightly clustered across all six metrics. In Figure A17, Dice changes by about 0.0003, Boundary-F remains near 0.212, and HD95 changes by less than 0.06 pixels. This flat trajectory indicates that the target-aligned spatial support survives local amplitude scaling in the room scene.

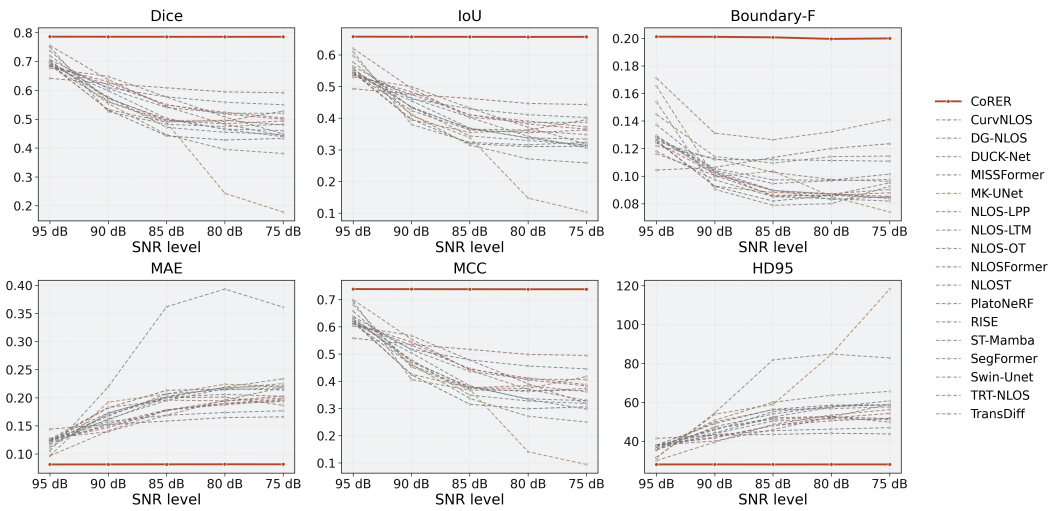

Figure A14: Metric trajectories under Gaussian noise on Reflect-Lab. The six panels report Dice, IoU, Boundary-F, MAE, MCC, and HD95; higher is better except for MAE and HD95.

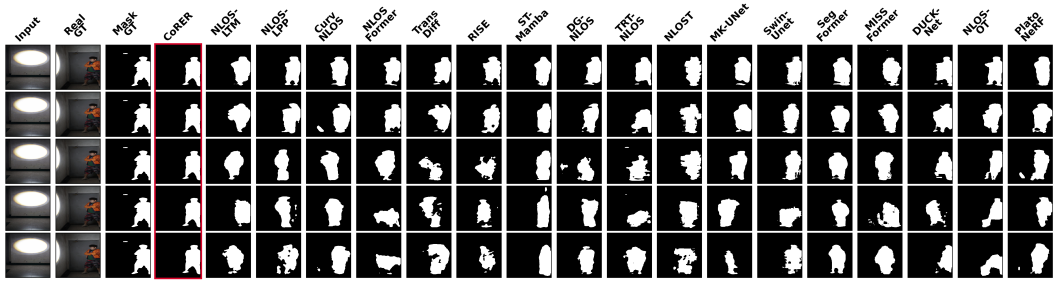

Figure A15: Representative predictions under speckle noise. Successive rows show the same scene instance at different speckle-noise levels. Columns show the relay-wall input, reference image, reference mask, and method predictions.

Reflect-Basement exhibits nearly identical overlap scores across the complete sweep. Figure A18 keeps Dice near 0.7597, IoU near 0.6192, and Boundary-F near 0.248. At the strongest level, CoRER leads the closest comparison trace by more than 0.136 Dice and 0.061 Boundary-F, demonstrating that its initial contour observation remains distinct under signal-dependent corruption.

Reflect-Lab maintains the same basic region while showing only small boundary fluctuations. Figure A19 ends with 0.7858 Dice and 0.2006 Boundary-F, close to the first speckle level, and records the lowest HD95 among the displayed methods. This scene therefore extends the same speckle stability to a different wall appearance and scene setting.

RDCR and SCRR explain the shared trend across these figures. RDCR evaluates coherence across spatial supports, so the response retains its influence under local gain changes when its spatial structure remains consistent. SCRR then pools support and reliability across restoration stages, tying the final contour to recurrent structure across scales. The combination preserves the basic region and main contour while the local magnitude of faint reflected-light spots changes.

C.3.7 Phase Jitter

The phase-jitter experiment perturbs the regularity of the relay pattern while retaining its overall intensity scale. Figure A20 shows the resulting prediction sequence, and Figures A21–A24 report the six metrics across the four scenes. The experiment isolates whether increasingly irregular local structure changes the target-aligned evidence used for the initial contour observation.

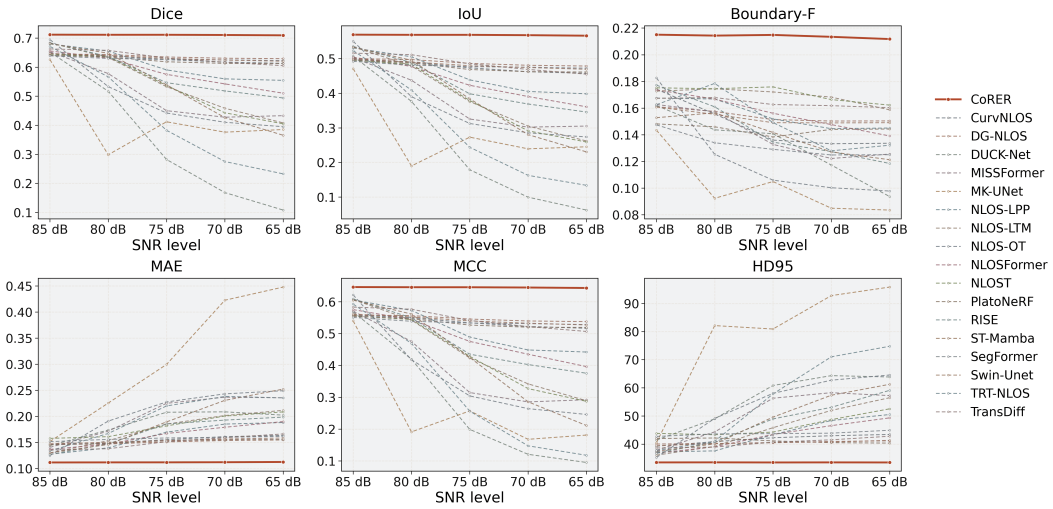

Figure A16: Metric trajectories under speckle noise on Reflect-Corridor. The six panels report Dice, IoU, Boundary-F, MAE, MCC, and HD95; higher is better except for MAE and HD95.

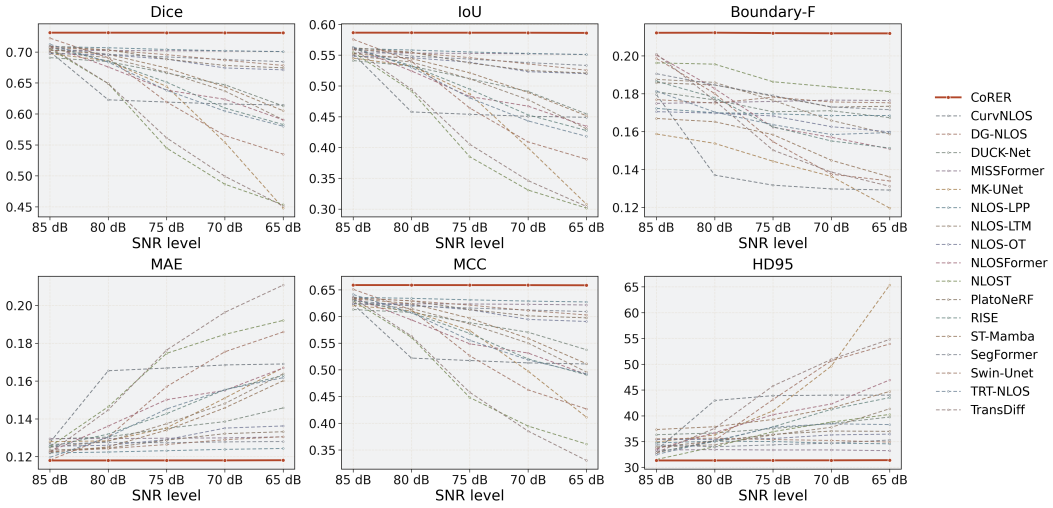

Figure A17: Metric trajectories under speckle noise on Reflect-Room. The six panels report Dice, IoU, Boundary-F, MAE, MCC, and HD95; higher is better except for MAE and HD95.

Figure A20 shows that phase disorder introduces local edge variation before it changes the basic region. CoRER retains the basic region and main contour across the displayed levels, and the recovered contour remains spatially aligned with the reference. The corresponding metric grids show modest boundary changes together with compact overlap trajectories.

Reflect-Corridor exhibits a gradual phase response. Figure A21 changes from 0.7122 to 0.7085 Dice and from 0.5699 to 0.5651 IoU, while Boundary-F remains above 0.209 at the largest jitter. CoRER preserves a 0.017 Dice and 0.024 Boundary-F advantage over the closest comparison traces at that endpoint, demonstrating that the irregular wall pattern still carries usable contour evidence.

Reflect-Room remains close to its reference trajectory. Figure A22 shows less than 0.0005 change in Dice and IoU and keeps HD95 near 31.37 pixels throughout. The smaller numerical margin in this scene is paired with consistent leadership across overlap and boundary measures, supporting the same initial contour observation under phase jitter.

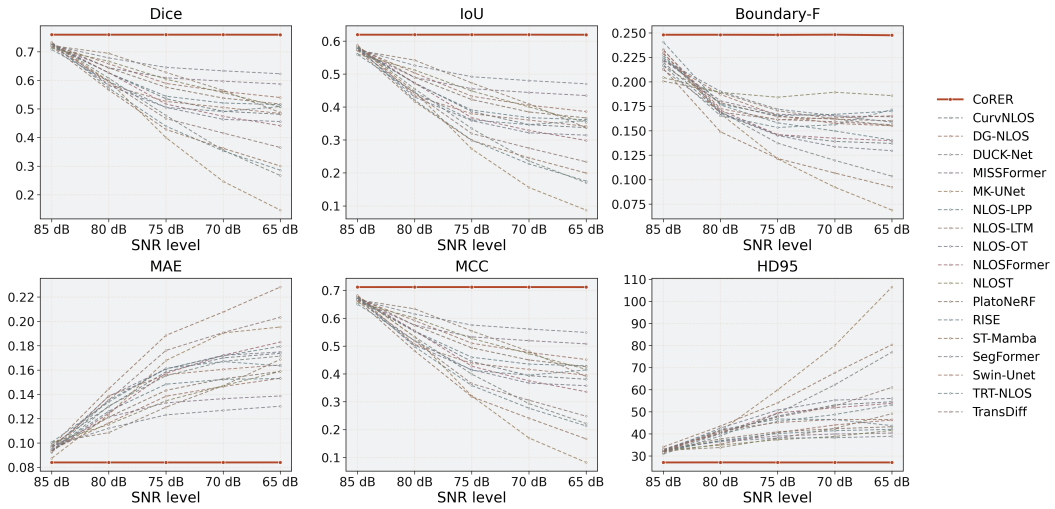

Figure A18: Metric trajectories under speckle noise on Reflect-Basement. The six panels report Dice, IoU, Boundary-F, MAE, MCC, and HD95; higher is better except for MAE and HD95.

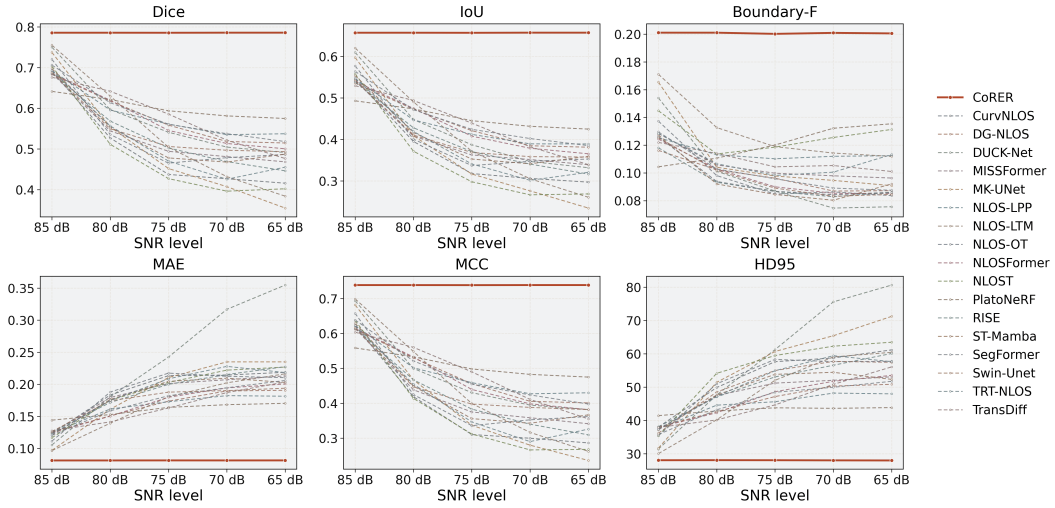

Figure A19: Metric trajectories under speckle noise on Reflect-Lab. The six panels report Dice, IoU, Boundary-F, MAE, MCC, and HD95; higher is better except for MAE and HD95.

Reflect-Basement shows a modest endpoint drift while retaining a clear six-metric advantage. In Figure A23, Dice reaches 0.7582, IoU 0.6173, and Boundary-F 0.2466 at the largest jitter. CoRER remains 0.051 Dice above the closest comparison trace and retains a lower HD95 by more than five pixels, indicating that the main contour stays globally aligned.

Reflect-Lab responds more visibly at the strongest phase setting and preserves the leading contour profile. Figure A24 changes from 0.7857 to 0.7832 Dice and from 0.2008 to 0.1978 Boundary-F, while the closest comparison trace remains more than 0.100 Dice lower. The fourth scene therefore maintains its basic region and main contour despite the increased local irregularity.

The common mechanism is cross-support persistence. RDCR organizes early responses according to their agreement over short and long supports, allowing coherent target structure to survive local phase changes. SCRR then aggregates support and agreement across decoder scales, reinforcing contour sections that recur in the staged relay state. This two-stage organization explains the compact phase-jitter trajectories and the retained comparison advantage across all four scenes.



Figure A20: Representative predictions under phase jitter. Successive rows show the same scene instance at different phase-jitter levels. Columns show the relay-wall input, reference image, reference mask, and method predictions.

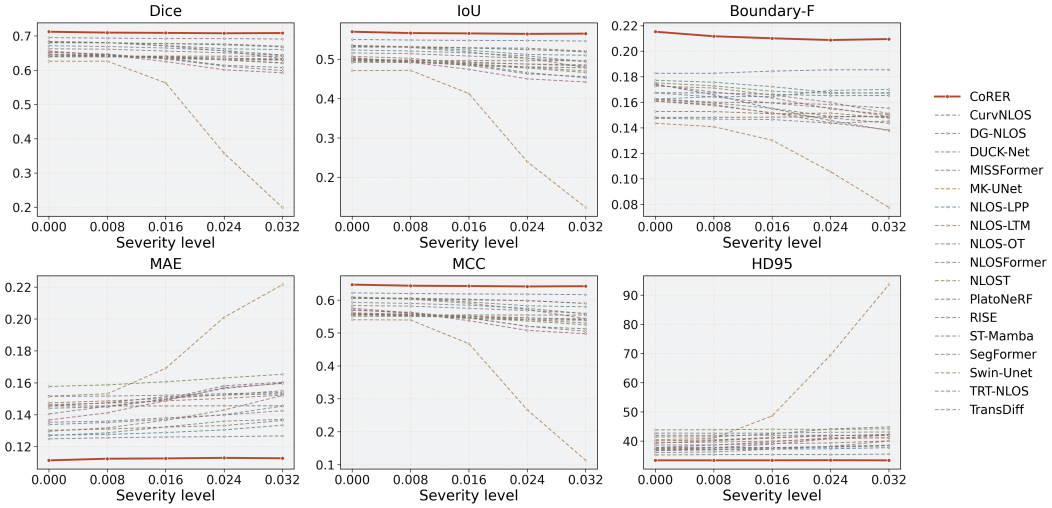

Figure A21: Metric trajectories under phase jitter on Reflect-Corridor. The six panels report Dice, IoU, Boundary-F, MAE, MCC, and HD95; higher is better except for MAE and HD95.

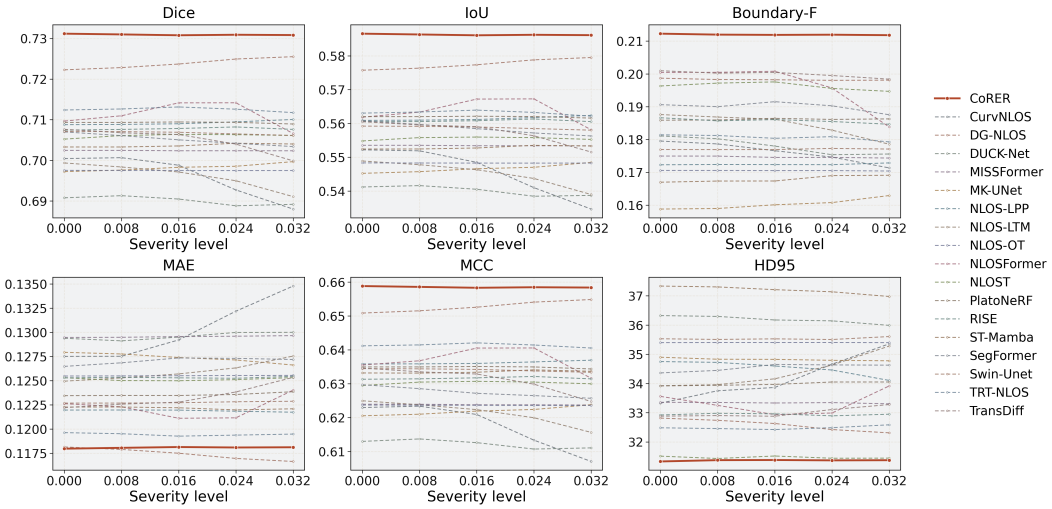

Figure A22: Metric trajectories under phase jitter on Reflect-Room. The six panels report Dice, IoU, Boundary-F, MAE, MCC, and HD95; higher is better except for MAE and HD95.

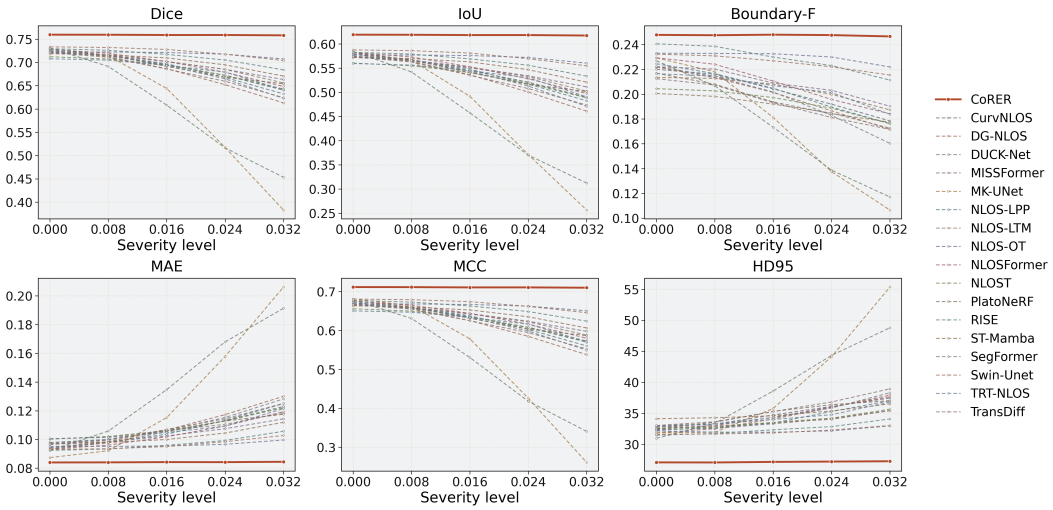

Figure A23: Metric trajectories under phase jitter on Reflect-Basement. The six panels report Dice, IoU, Boundary-F, MAE, MCC, and HD95; higher is better except for MAE and HD95.

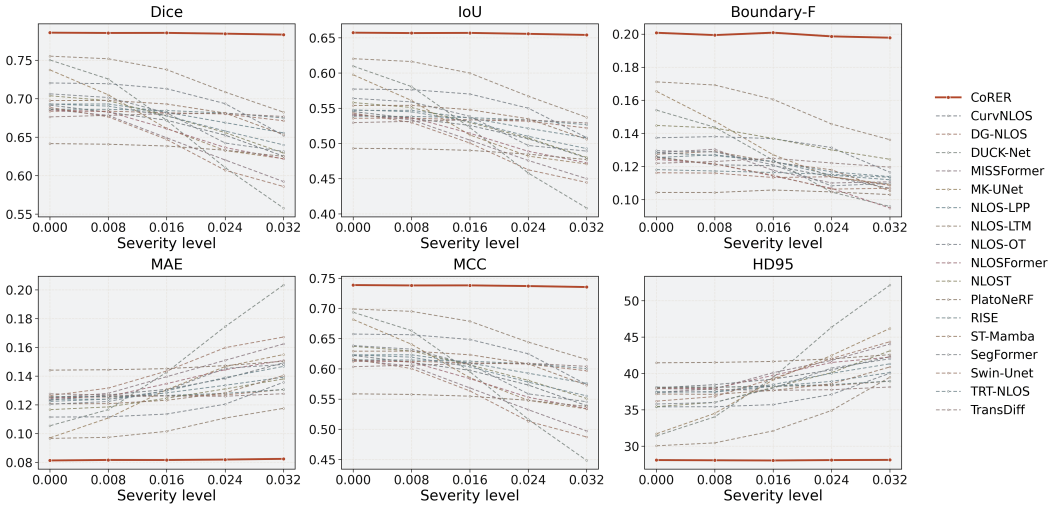

Figure A24: Metric trajectories under phase jitter on Reflect-Lab. The six panels report Dice, IoU, Boundary-F, MAE, MCC, and HD95; higher is better except for MAE and HD95.

C.3.8 Transport Blur

The transport-blur experiment progressively softens the spatial structure of the relay-wall response. Figure A25 visualizes representative predictions, and Figures A26–A29 report the scene-wise trajectories. Because blur suppresses high-frequency structure while retaining broader support, this experiment directly tests the transition from diffuse wall evidence to the hidden person's basic region and main contour.

The qualitative sequence in Figure A25 shows that contour sharpness changes before basic-region support. CoRER retains the basic region across the displayed blur levels, while narrow boundaries and local separations become progressively harder to resolve. The metric grids capture the same ordering: Boundary-F responds earlier than Dice and IoU, and HD95 records the accumulated displacement of softened contour sections.

Reflect-Corridor displays the largest blur response relative to its reference score. Figure A26 changes from 0.7122 to 0.7063 Dice, from 0.5699 to 0.5617 IoU, and from 0.2153 to 0.2073

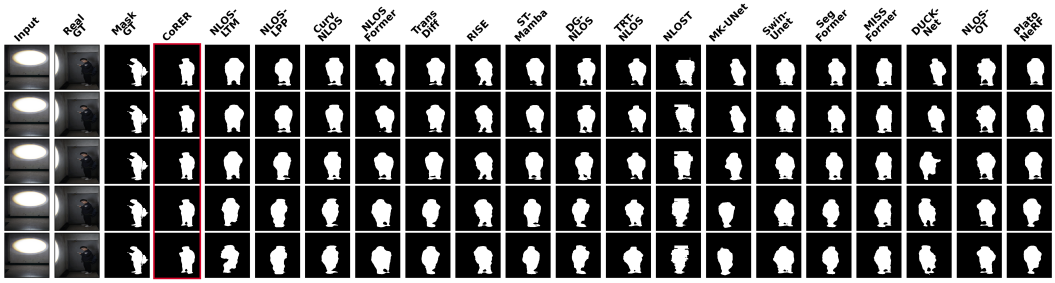

Figure A25: Representative predictions under transport blur. Successive rows show the same scene instance at different blur levels. Columns show the relay-wall input, reference image, reference mask, and method predictions.

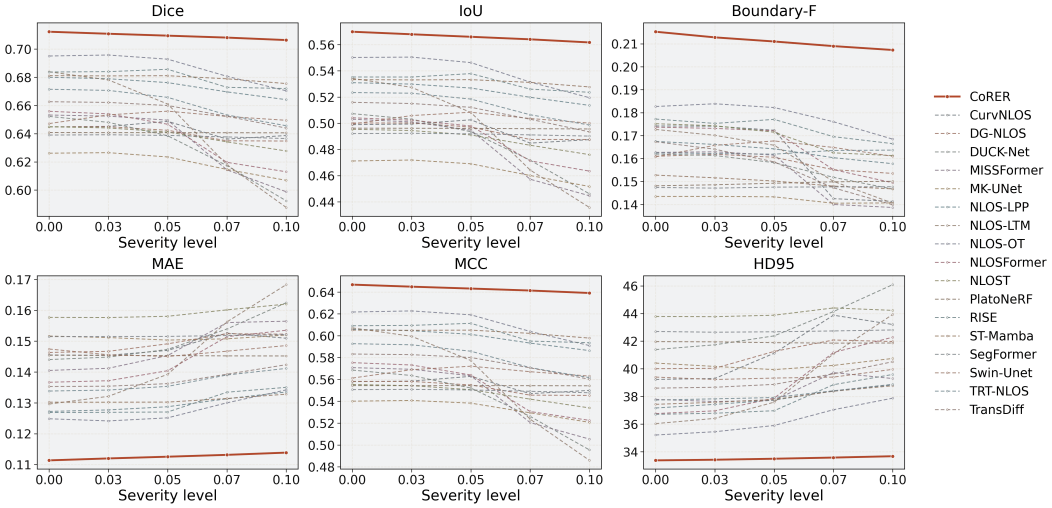

Figure A26: Metric trajectories under transport blur on Reflect-Corridor. The six panels report Dice, IoU, Boundary-F, MAE, MCC, and HD95; higher is better except for MAE and HD95.

Boundary-F. At the strongest blur, CoRER remains 0.031 Dice and 0.039 Boundary-F above the closest comparison traces, preserving both the basic region and its main contour.

Reflect-Room remains stable even as the relay pattern is spatially softened. Figure A27 shows less than 0.0005 Dice change, a Boundary-F reduction below 0.001, and a compact HD95 range. The room-scene response therefore retains broad target support that SCRR can organize into a consistent contour after fine structure is reduced.

Reflect-Basement shows a modest boundary-sensitive change and stable overlap. Figure A28 reaches 0.7573 Dice and 0.2427 Boundary-F at the strongest blur, with HD95 increasing by about 0.16 pixels. CoRER remains ahead of the closest comparison traces on all three quantities, demonstrating that broad relay support still provides a usable initial contour observation.

Reflect-Lab has the most visible endpoint change in this family. Figure A29 moves from 0.7857 to 0.7794 Dice and from 0.2008 to 0.1953 Boundary-F, while retaining a 0.076 Dice and 0.066 Boundary-F advantage over the closest comparison traces. The preserved margin shows that the scene's broader target response remains informative even as local contour cues are softened.

Transport blur makes the RDCR-to-SCRR dependency explicit. RDCR organizes the broad response shared across short and long supports, maintaining a target-aligned carrier as fine spatial detail decreases. SCRR then integrates the surviving support across decoder stages and restores the main contour from recurrent multiscale structure. The resulting trajectories show that ordinary-camera wall observations continue to support initial contour observation throughout the tested blur range.

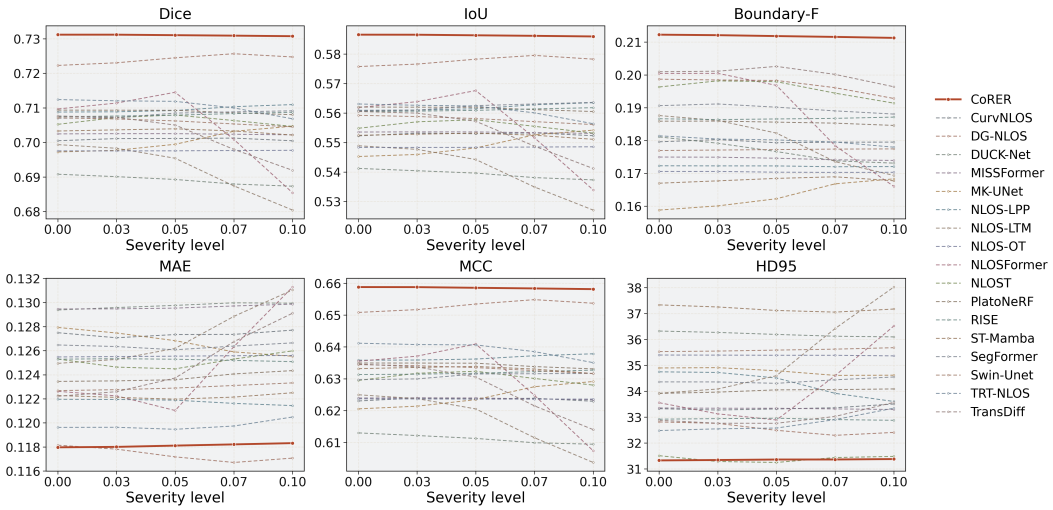

Figure A27: Metric trajectories under transport blur on Reflect-Room. The six panels report Dice, IoU, Boundary-F, MAE, MCC, and HD95; higher is better except for MAE and HD95.

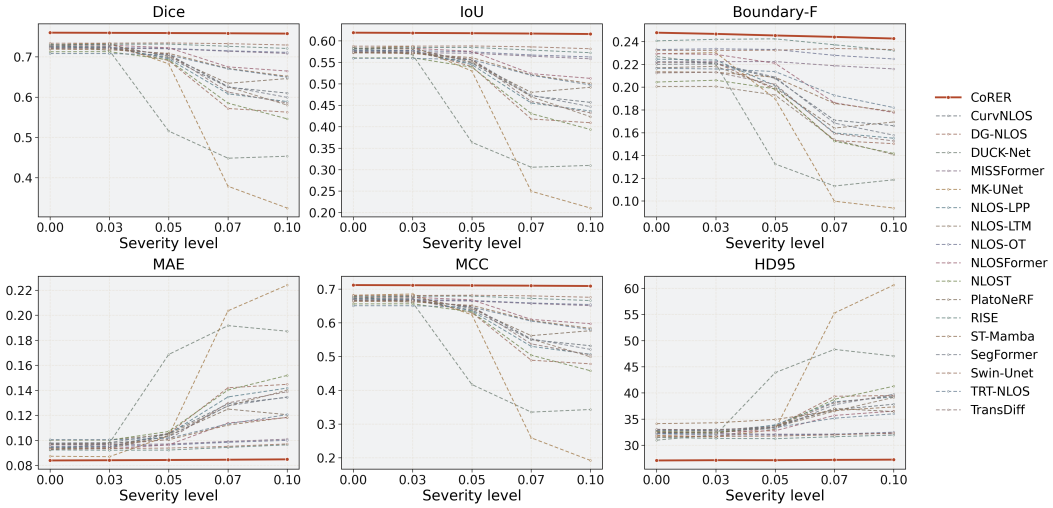

Figure A28: Metric trajectories under transport blur on Reflect-Basement. The six panels report Dice, IoU, Boundary-F, MAE, MCC, and HD95; higher is better except for MAE and HD95.

The following diagnostics trace where the useful support appears in the learned representation and how it relates to the final hidden-human region and contour.

C.4 Qualitative Evidence-to-Contour Diagnostic

This experiment examines how a localized relay-wall response is organized into spatial support for a hidden-human contour. Figure A30 presents one curated example from each of Reflect-Corridor, Reflect-Room, Reflect-Basement, and Reflect-Lab. For each scene, the diagnostic sequence contains the input overlaid with reference and predicted contours, a reference-guided mixed foreground–boundary response map, foreground confidence, a low-dimensional visualization of the tail feature, uncertainty, and three derived support views describing the bright-wall prior, boundary evidence, and contour support, followed by the final soft mask. These views serve as complementary probes of the same prediction, allowing the spatial relationship between the observed light response and the recovered contour to be inspected across distinct corner layouts and illumination patterns.

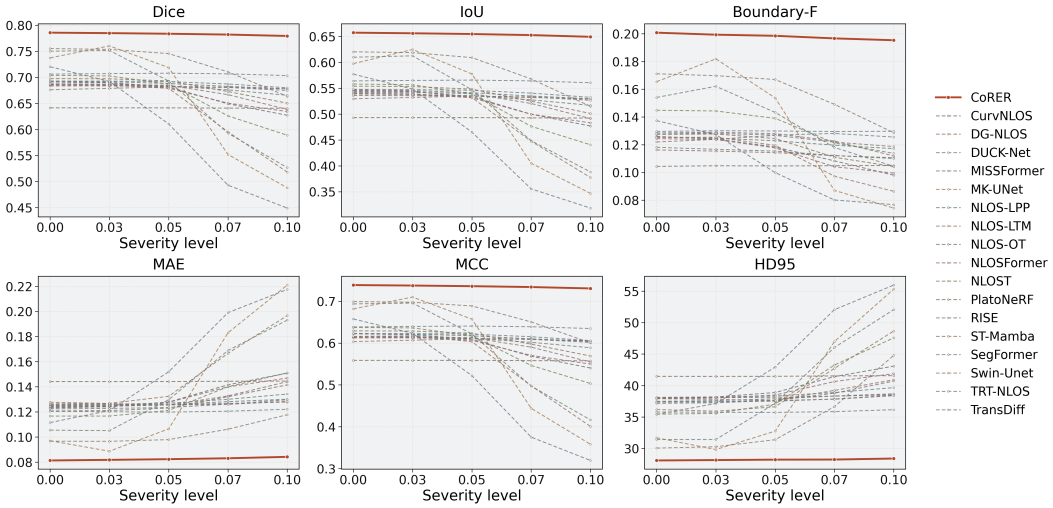

Figure A29: Metric trajectories under transport blur on Reflect-Lab. The six panels report Dice, IoU, Boundary-F, MAE, MCC, and HD95; higher is better except for MAE and HD95.

Reflect-Corridor demonstrates the transition from a broad horizontal wall spot to a compact basic-region hypothesis. The mixed response spans most of the illuminated ellipse, while foreground confidence and the projected tail features concentrate support within the basic region. Uncertainty and boundary evidence trace the main contour and interior gaps, while the final soft mask preserves the basic region and main contour.

Reflect-Room separates the foreground from a compact elliptical wall response. Its confidence and tail-feature views recover the basic region and an interior gap, while uncertainty follows both the exterior and interior main contour. The boundary-evidence and contour-support views delineate the separation retained in the final mask, illustrating the recovery of both the basic region and local contour structure in the room geometry.

Reflect-Basement shows the most asymmetric basic region and wall appearance among the four rows. The projected tail features localize the response toward the basic region, and the uncertainty field follows the sloped and irregular sections of the main contour. Boundary evidence highlights these local contour variations, while the contour-support view and final mask retain a connected basic region despite the offset illumination and textured background.

Reflect-Lab provides a narrow, vertically organized response. Foreground confidence and the projected tail features concentrate along the central axis of the basic region, while uncertainty and boundary evidence define the upper and lateral main contour. The contour-support view delineates the main contour retained in the final mask, extending the same evidence-to-contour pattern to the fourth scene.

The diagnostic is consistent with the two-stage mechanism of CoRER. RDCR organizes the early wall response around spatially coherent evidence, providing a localized carrier from which target-related structure can be retained across scene-dependent wall appearances. SCRR then consolidates foreground and boundary support across decoding scales, concentrating uncertainty and correction near the contour while preserving the supported interior. Together, the response, confidence, boundary, contour-support, and final-mask views document a qualitative correspondence between the observed relay response and the restored prediction across the four tested scenes.

C.5 RDCR Diagnostics

The RDCR diagnostics trace whether early wall responses become target-aligned evidence before contour restoration. Figure A31 follows the representation from the first two encoder stages to contextual condensation and compares CoRER with the RDCR removal through top-q IoU and in-mask evidence mass. Figure A32 then separates target and nuisance regions, Figure A33 evaluates

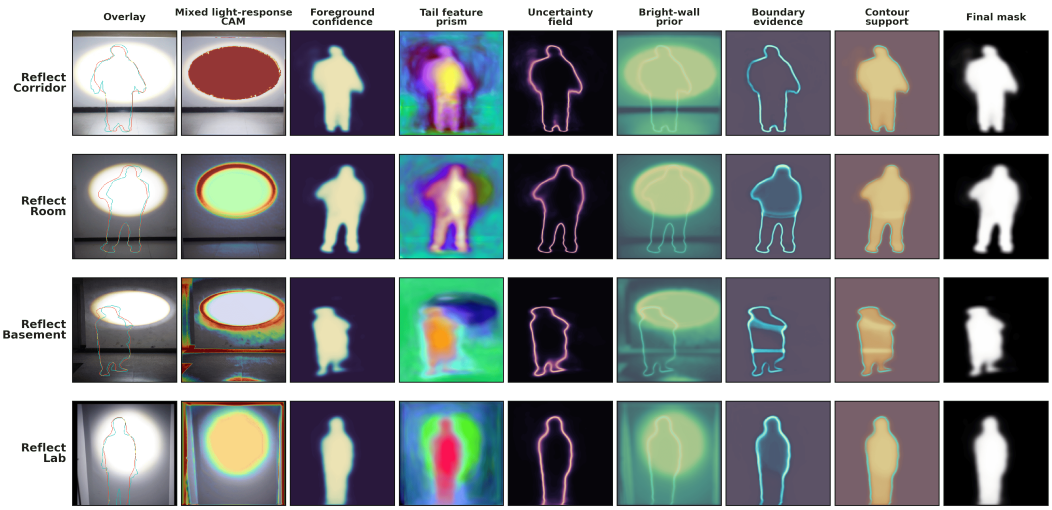

Figure A30: Qualitative evidence-to-contour diagnostics across Reflect-Corridor, Reflect-Room, Reflect-Basement, and Reflect-Lab (rows). From left to right, the columns show the input with reference and predicted contours, a reference-guided mixed foreground–boundary response map, foreground confidence, a low-dimensional tail-feature visualization, uncertainty, a derived bright-wall prior, derived boundary evidence, derived contour support, and the final soft mask. The columns provide complementary diagnostic views of the spatial response associated with each prediction.

evidence stability under signal changes, and Figure A34 measures the downstream geometric consequences. For a selected percentile q, the top-q region retains locations whose response is at or above the q-th percentile of the map; increasing q selects more concentrated evidence.

Figure A31 shows that RDCR improves both evidence localization and its propagation into context. In the four-scene summary, contextual-condensation in-mask evidence mass reaches 0.573 for CoRER and 0.374 for the RDCR removal, while top-q IoU reaches 0.415 and 0.287. The stage trajectories indicate that this advantage is present early and becomes more useful after contextual condensation, matching RDCR's role as a complete front-end representation path that organizes weak evidence and transports it toward the restoration stage.

Figure A32 shows that the localized response is preferentially associated with the hidden target. Across pooled paired samples, the mean CoRER-minus-removal change is 0.038 for in-mask evidence mass and 0.032 for the target-over-nuisance margin, with positive fractions of 0.726 and 0.604. The band profile and paired distributions therefore connect higher retained evidence to greater target separation, supporting the interpretation that RDCR organizes faint target-aligned responses beyond local wall appearance.

Figure A33 establishes that RDCR's target-aligned evidence remains spatially coherent under brightness attenuation, Gaussian noise, phase jitter, and transport blur. Evidence NCC and top-q overlap remain high over the displayed settings, while centroid drift stays small. The strongest changes appear when attenuation or spatial softening directly alters the response footprint, yet the evidence continues to occupy a consistent target-related location. This stability connects the front-end representation to the compact robustness trajectories in Figures A6–A29.

Figure A34 closes the RDCR chain at the final mask. In the four-scene macro average, CoRER records support loss of 0.233 versus 0.280, centroid offset of 0.239 versus 0.310, largest-component coverage of 0.765 versus 0.718, mid-band error of 0.368 versus 0.414, and deep-interior error of 0.149 versus 0.200 for the RDCR removal. Improved localization in Figure A31, stronger target separation in Figure A32, and stable evidence in Figure A33 therefore culminate in a more complete basic region with better interior support. Together, the four diagnostics show how RDCR turns faint reflected-light spots on the relay wall into the carrier required for downstream contour restoration.

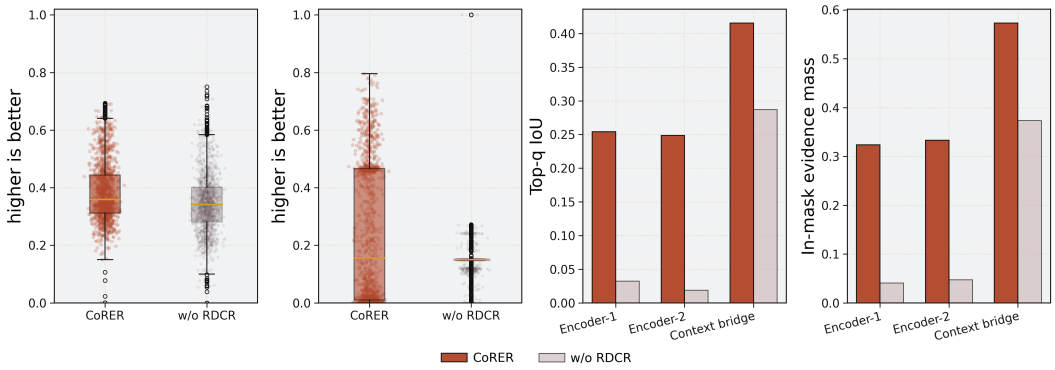

Figure A31: RDCR evidence progression from early evidence organization through contextual condensation. The two left panels show the distribution of retained evidence statistics, while the two right panels compare stage-wise top-q IoU and in-mask evidence mass for CoRER and the RDCR removal.

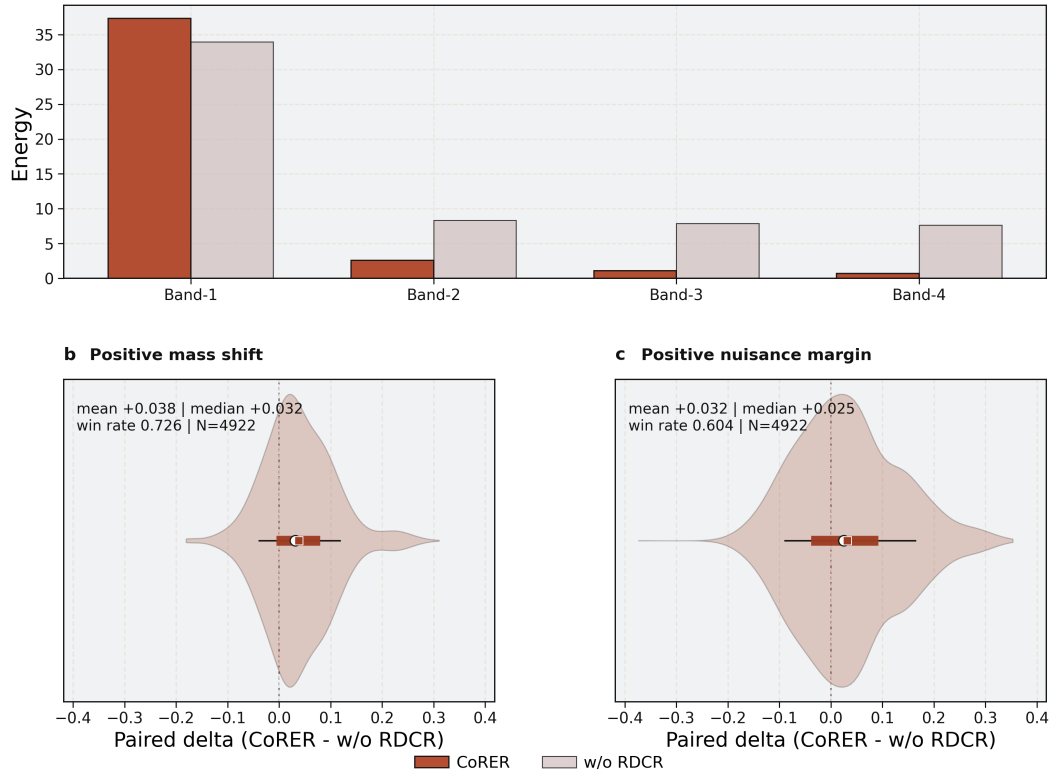

Figure A32: RDCR target-versus-nuisance separation. The upper panel reports energy across four bands, and the lower paired-delta distributions quantify shifts in in-mask evidence mass and target-over-nuisance margin relative to the RDCR removal.

C.6 SCRR Diagnostics

The SCRR diagnostics test how the back-end restoration path converts distributed relay support into a coherent main contour. Figure A35 examines state semantics, spatial organization, and cross-scale coherence; Figure A36 tests whether conflict evidence aligns with observed errors. Figures A37–A39 then connect high-risk restoration to boundary recovery and local separation, while Figures A40 and A41 extend the analysis across boundary radii and shape complexity.

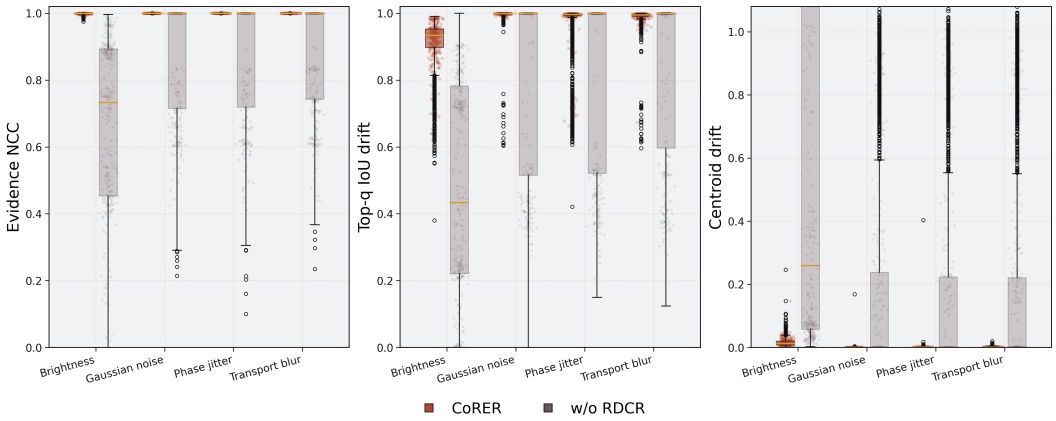

Figure A33: RDCR evidence stability under brightness, Gaussian noise, phase jitter, and transport blur. The panels report evidence NCC, top-q IoU drift, and centroid drift. The middle panel, labeled top-q IoU drift in the plot, reports overlap between the selected top-q evidence regions before and after the signal change; higher NCC and overlap, together with lower centroid drift, indicate more stable evidence localization.

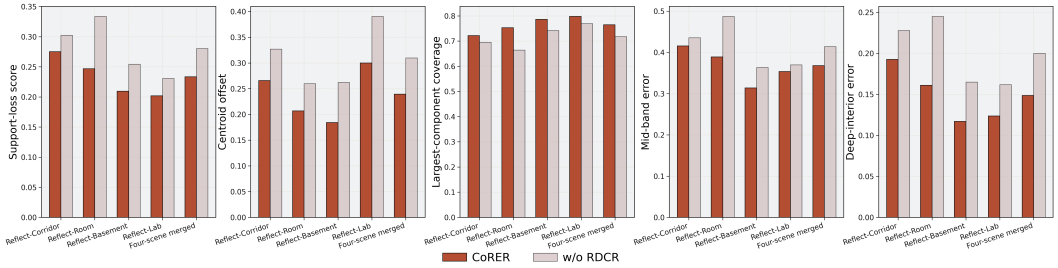

Figure A34: Downstream consequences of RDCR removal across the four scenes and their merge. The panels compare support loss, centroid offset, largest-component coverage, mid-band error, and deep-interior error; lower is better except for largest-component coverage.

Figure A35 shows that the SCRR states carry distinct restoration roles. Returned represents prediction support weighted by the support envelope; Conflict combines smooth/detail disagreement with edge evidence; Coarse warrant combines smooth agreement, support, and reliability; and Writable gates correction according to uncertainty and supporting evidence. Returned has its strongest response in true-positive regions, whereas Writable responds more strongly to false-positive and false-negative regions than to correctly classified regions. The scale-coherence panels retain high NCC and top-q overlap for Returned, Conflict, Coarse warrant, and Writable, indicating that these semantics recur across restoration levels. This organization supplies the decoder with separate signals for preserving supported foreground and correcting uncertain contour regions.

Figure A36 confirms that the conflict state is spatially associated with prediction error. As the retained conflict percentile becomes more selective, error precision rises, while error-region top-q IoU reaches its strongest range at intermediate selections before the support becomes sparse. The paired movement of overlap and precision shows that SCRR identifies regions where correction has the greatest opportunity to improve the main contour.

Figure A37 localizes SCRR's gains to regions carrying the greatest contour risk. The merged high-risk tier reaches a balanced gain of 0.322 and a score-win rate of 0.822. As cumulative coverage expands, balanced gain and score-win rate decrease because lower-risk regions enter the selected set; the strongest scene-wise values therefore occur near the selective end of the curves. This concentration matches the conflict-error alignment in Figure A36 and shows that SCRR directs restoration toward ambiguous local structure.

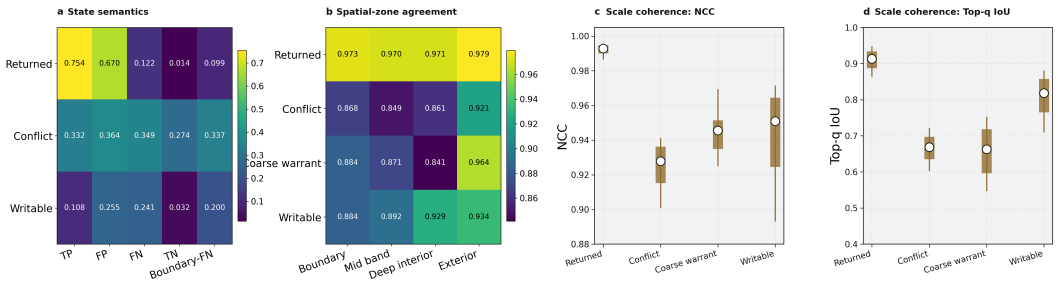

Figure A35: SCRR state organization. The first heatmap reports mean state responses in TP, FP, FN, TN, and Boundary-FN regions, and the second reports mid-to-tail NCC within boundary, mid-band, deep-interior, and exterior regions. The remaining panels summarize cross-scale NCC and top-q IoU for Returned, Conflict, Coarse warrant, and Writable.

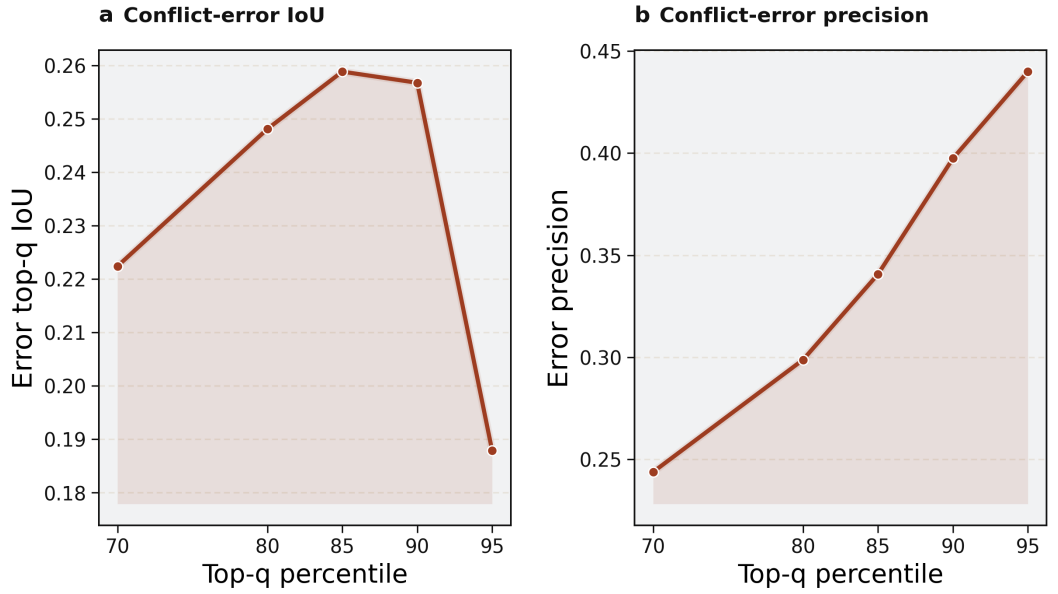

Figure A36: Alignment between SCRR conflict evidence and observed errors. The left panel reports error-region top-q IoU and the right panel reports error-region precision as the retained conflict percentile increases.

Figure A38 connects selective restoration to global boundary recovery. Across pooled validation images, largest-component coverage reaches 0.768 for CoRER versus 0.728 for the SCRR removal, while the boundary false-negative rate decreases from 0.876 to 0.833. The four-scene macro-average Boundary-F rises from 0.164 to 0.219, and the macro-average near-boundary error decreases from 0.468 to 0.459. These aligned changes show that SCRR restores missing contour support while retaining a larger share of the predicted foreground in its largest connected component.

Figure A39 resolves the same effect in closely spaced contour regions. Collision measures foreground intrusion into local gaps; center clearance describes the background interval preserved along the gap centerline, while positive span describes foreground extension along that line. Each width bin is defined by the distance from the gap center to the nearest reference foreground, and positive span is the length of the continuous predicted foreground along the centerline divided by the reference-background length on that line. Together with center-block, adhesion, and exterior-spill measures, these diagnostics assess local separation and unintended connections. The false-bridge rates in Figure 8 summarize unwanted foreground filling across reference gaps. Collision-event reductions grow from 0.017 to 0.046 across width bins 2–4, and center-clearance gains grow from 0.017 to 0.051. The accompanying reductions in positive-span and center-block events show that

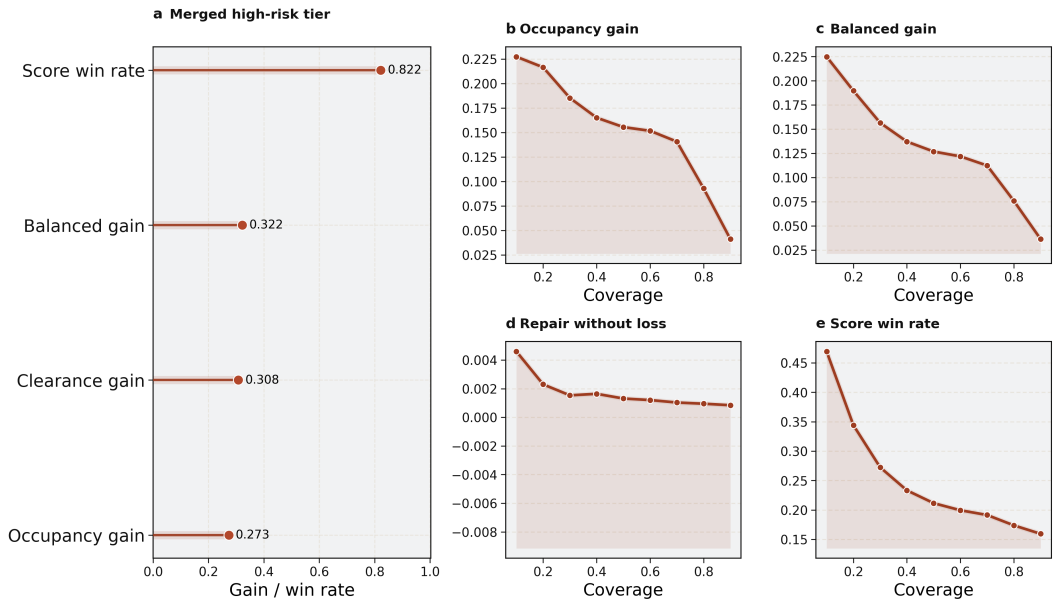

Figure A37: SCRR gains concentrated in high-risk local regions. The left panel summarizes the merged high-risk tier, and the remaining panels trace occupancy gain, balanced gain, repair-without-loss, and score-win rate as the coverage of risk-ranked windows expands.

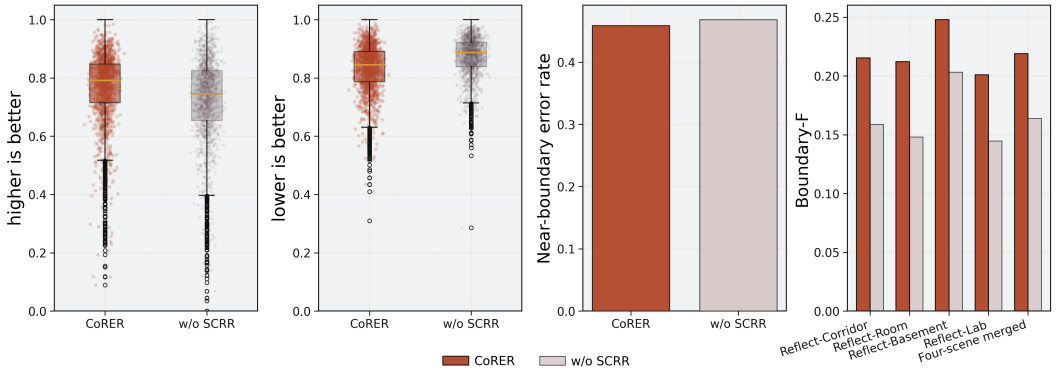

Figure A38: Paired boundary-recovery summary for SCRR. The panels compare largest-component coverage, boundary false-negative behavior, near-boundary error, and Boundary-F between CoRER and the SCRR removal; higher is better for coverage and Boundary-F, whereas lower is better for the error measures.

the separation is maintained across the local gap, while adhesion decreases from 0.287 to 0.262 and exterior spill from 0.073 to 0.065. SCRR therefore improves both sides of local contour restoration: it preserves foreground support and reduces unintended connections to nearby regions.

Figure A40 shows that the restoration advantage persists across boundary-matching tolerances and all four scenes. Against the SCRR removal, the merged Boundary-F gain grows from 0.035 at radius 1 to about 0.094 at radius 5, while the boundary false-negative rate is reduced by 0.026 at radius 1 and 0.081 at radius 5 (signed changes of approximately -0.026 and -0.081). HD95 remains lower for CoRER across the scene columns. The Boundary-F advantage persists from strict to more permissive matching tolerances, while lower HD95 confirms that the recovered contours remain closer to the reference. Together, these results show that SCRR improves local contour agreement and the spatial accuracy of the recovered main contour.

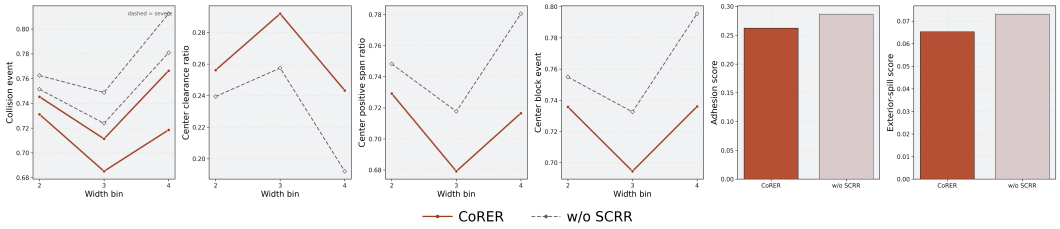

Figure A39: SCRR local separation in closely spaced contour regions. The first four panels compare collision, clearance, positive-span, and center-block events over width bins, and the final two panels report adhesion and exterior-spill errors.

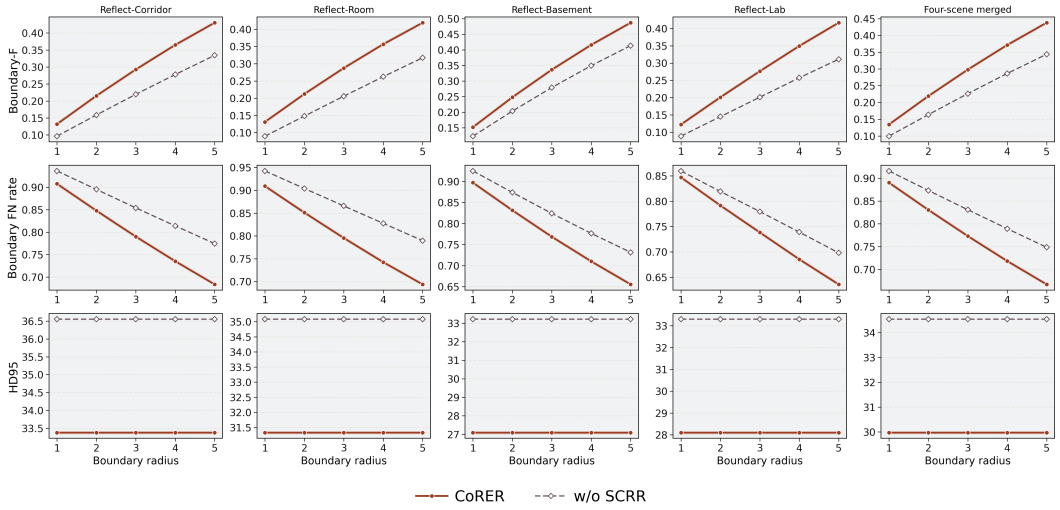

Figure A40: SCRR boundary-band scan. The columns correspond to Reflect-Corridor, Reflect-Room, Reflect-Basement, Reflect-Lab, and their four-scene merge; rows report Boundary-F, boundary false-negative rate, and HD95. The radius sets the boundary-matching tolerance; HD95 is unchanged by this tolerance.

Figure A41 extends the analysis from boundary scale to mask geometry. Shape complexity is computed from the reference mask by averaging the standardized values of $P/\sqrt{A}$, $\sqrt{\lambda_{\max}/\lambda_{\min}}$, and $1 - 4\pi A/P^2$, where A is foreground area, P is the boundary-pixel count, and the eigenvalues describe the covariance of foreground coordinates. Standardization and tertile grouping use the pooled four-scene validation set, yielding 1,641 low-, 1,640 mid-, and 1,641 high-complexity samples. Relative to the SCRR removal, Dice gains are 0.027, 0.032, and 0.034 for low-, mid-, and high-complexity masks, while Boundary-F gains are 0.048, 0.051, and 0.059. The increasing Boundary-F gain over the SCRR removal highlights the value of multiscale restoration support for more complex target contours. Together, Figures A35–A41 trace SCRR from organized decoder states to connected, locally separated, and geometrically consistent hidden-human contours.

C.7 Shared Paired-Difference Summary

Figure A42 aggregates sample-paired differences between CoRER and the RDCR or SCRR removal across all four scenes and their 4,922-image merge. The three panels report Dice and Boundary-F gains and HD95 reductions with a common positive direction: higher values indicate stronger region overlap, boundary fidelity, or a larger reduction in contour distance. This paired view complements the scene means in Tables A6–A9 by showing how consistently each mechanism contributes across individual samples.

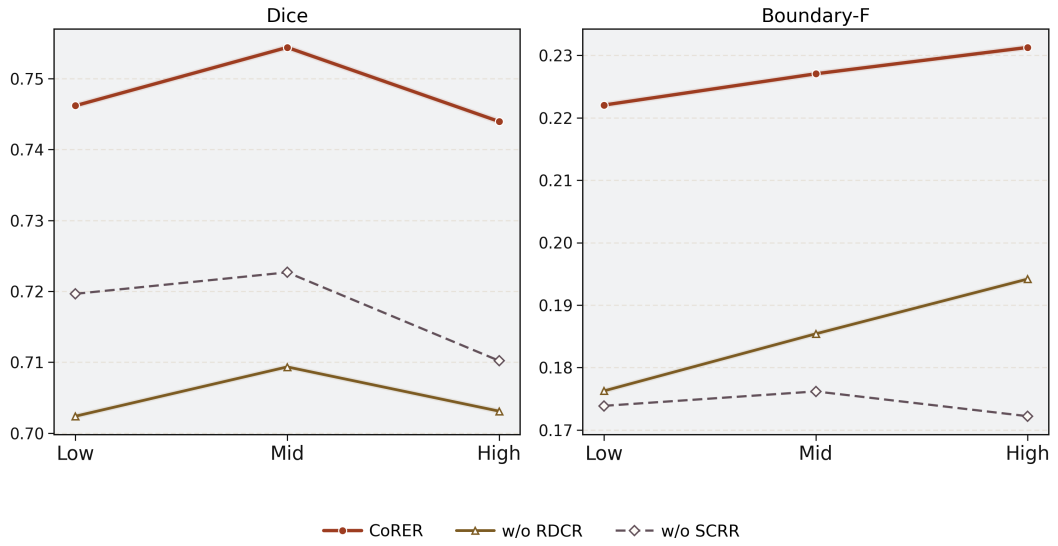

Figure A41: Contour quality stratified by shape complexity. Dice and Boundary-F are reported for low-, mid-, and high-complexity masks, comparing CoRER with the RDCR and SCRR removals.

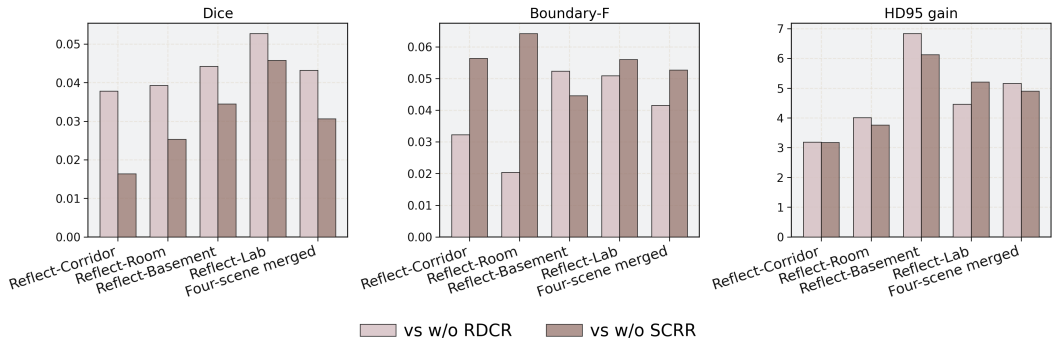

Figure A42: Mean paired performance improvements of CoRER over the RDCR and SCRR removals, shown for each scene and the four-scene merge. Dice and Boundary-F gains are computed as CoRER minus the removal model; HD95 reduction is computed as the removal model minus CoRER. Positive values favor the complete model.

The merged distribution separates the two module roles. Against the RDCR removal, the mean paired gain is 0.043 in Dice and 0.042 in Boundary-F, with win rates of 0.684 and 0.655; the corresponding mean per-image paired HD95 reduction is 5.159 pixels with a win rate of 0.709. Against the SCRR removal, the mean Dice gain is 0.031 and the mean Boundary-F gain rises to 0.053, with win rates of 0.639 and 0.689; the mean per-image paired HD95 reduction is 4.905 pixels with a win rate of 0.713. Scene-wise means remain positive in every displayed metric for both comparisons, showing that the aggregate direction is shared across Reflect-Corridor, Reflect-Room, Reflect-Basement, and Reflect-Lab.

The paired differences complete the appendix mechanism chain. RDCR contributes more strongly to basic-region overlap by organizing weak target-aligned evidence before compression, while SCRR contributes more strongly to boundary fidelity by restoring support and separation across decoder scales. Their positive paired gains across scenes explain the complete model's comparison advantage and reinforce the central conclusion: under ordinary flashlight illumination, faint wall reflections recorded by an ordinary RGB camera contain usable evidence for initial recovery of a completely hidden person's basic region and main contour.