PCAR-FLOW: PHYSICS-CALIBRATED ALLOCATION OF RESIDUALS FOR METALENS IMAGING

Anonymous authors
Paper under double-blind review

ABSTRACT

Metalens measurements exhibit highly nonuniform recoverability across field position, wavelength, spatial frequency, and signal level. Generative priors can recover detail in weakly constrained regions, but may also introduce content that is not uniquely supported by the measurement. We propose Physics-Calibrated Adaptive Residual Flow(PCAR-Flow), which allocates generative freedom according to *effective local observability*, a calibrated measure combining optical transfer, sensor noise, and local image statistics. PCAR-Flow recovers measurement-supported content through a deterministic path and applies conditional flow matching only to unresolved high-frequency residuals. Under the local Gaussian model, the complement of observability equals normalized posterior band uncertainty and, for channel-separable transfer, approaches the classical null-space limit as measurement noise vanishes. Experiments show that PCAR-Flow improves reconstruction quality while keeping forward inconsistency close to the deterministic solution.

1 INTRODUCTION

Metalens cameras compress complex optical functionality into a thin form factor and enable compact computational imaging systems (Tseng et al., 2021; Chakravarthula et al., 2023). Their measurements, however, do not constrain all image components equally. Wide-field operation introduces spatially varying aberrations (Pinilla et al., 2023; Liu et al., 2024), while broadband imaging couples wavelength-dependent transfer with sensor response and chromatic defocus (Fröch et al., 2025b; Kim et al., 2024). High-frequency transmission is spatially variant, and weak optical signals are easily obscured by the noise floor. A single capture therefore mixes deterministic structures with details that remain uncertain despite calibration. The primary difficulty stems from this uneven recoverability across space and frequency, rather than mere degradation severity.

Generative priors are valuable where direct inversion becomes unreliable. Diffusion-based restoration has shown strong perceptual recovery under severe degradation (Wang et al., 2024; Lin et al., 2024; Wu et al., 2024b;a), while generative inverse methods incorporate known measurements into the reconstruction process (Chung et al., 2023; Wang et al., 2023b). Recent metalens reconstruction has begun to use similar priors directly (Zhang et al., 2025). The benefit comes with ambiguity. A fine texture or boundary may look plausible even when the measurement provides weak support for that particular realization. Measurement guidance can reject content that strongly violates the forward model, but it cannot select a unique solution along weakly constrained directions. The perception-distortion trade-off already separates perceptual realism from signal fidelity (Blau & Michaeli, 2018). For calibrated imaging, the practical question is therefore how much prior-driven correction should be admitted when measurement support is incomplete.

Existing methods assign different roles to the measurement and the learned prior. Blind perceptual restoration relies heavily on image statistics when the degradation model is only weakly specified (Lin et al., 2024; Yu et al., 2024; Wu et al., 2024b). Posterior-guided diffusion keeps the full image stochastic and steers generation toward measurement consistency (Chung et al., 2023). Range and null-space methods preserve measured components and use the prior for directions that are exactly unobservable (Wang et al., 2023b). Calibrated metalens imaging does not fit cleanly into any of these regimes. Finite noise creates a broad region of partial recoverability in which weak nonzero transfer is neither reliably observed nor truly absent. Spatial variation further makes recoverability depend on position, frequency, signal level, and image statistics. Figure 1 illustrates the resulting mismatch

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.

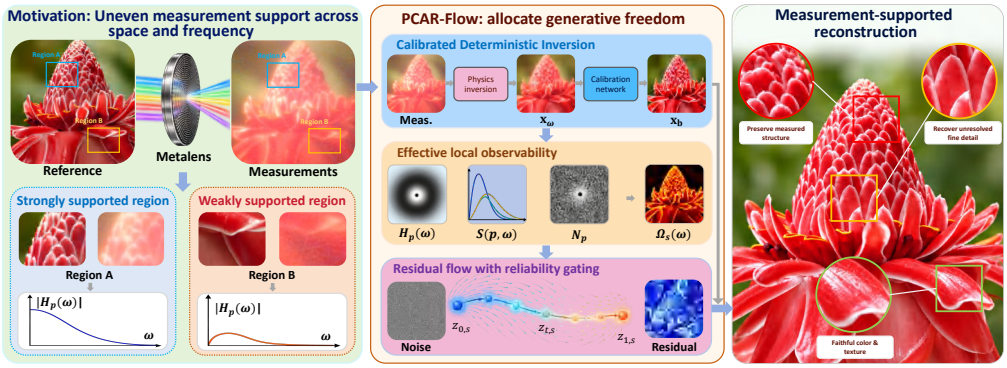

Figure 1: Motivation. Metalens recoverability varies across field position and spatial frequency. PCAR-Flow leaves well-observed content to calibrated inversion and assigns greater prior freedom to weakly observed residual components.

between measurement support and generative freedom. Well-measured components require minimal intervention, while weakly supported ones require greater prior involvement. The core challenge is therefore dynamically scaling generative freedom based on the measurement.

We propose *effective local observability* as a continuous measure of calibrated recoverability. Unlike optical transfer alone, it jointly accounts for the spatially varying cross-channel response, signal-dependent sensor noise, and local image statistics. Under the local Gaussian model, its complement equals posterior band uncertainty normalized by prior band energy, directly linking measurement support to unresolved uncertainty. Our **Physics-Calibrated Adaptive Residual Flow(PCAR-Flow)** uses the quantity to separate calibrated estimation from generative completion. A deterministic base reconstructor recovers measurement-supported content, conditional flow matching (Lipman et al., 2023; Liu et al., 2023; Albergo & Vanden-Eijnden, 2023) proposes high-frequency residual detail, and a reliability-controlled admission model determines how much of each proposal enters the final image from observability, predictive uncertainty, local reconstruction confidence, and forward disagreement.

Our main contributions are summarized as follows:

- We formulate generative reconstruction for calibrated metalens imaging as a problem of allocating prior freedom according to local measurement support, and introduce *effective local observability* to quantify recoverability from optical transfer, signal-dependent noise, and image statistics.
- We propose PCAR-Flow, which separates calibrated estimation from generative completion by restricting conditional flow matching to unresolved high-frequency residuals and controlling their admission with observability and reconstruction reliability.
- We establish a posterior interpretation of effective observability under the local Gaussian model. Its complement equals normalized posterior band uncertainty and, in the channel-separable case, approaches the classical null-space limit as measurement noise vanishes.
- Experiments show that PCAR-Flow improves reconstruction quality while maintaining low forward inconsistency across the evaluated conditions.

2 Related Work

Computational reconstruction for meta-optical imaging. Spatially varying optical degradation is commonly addressed by local deconvolution or calibrated inversion, which models field-dependent transfer instead of assuming a single shift-invariant kernel (Heide et al., 2013; Eboli et al., 2022; Kohli et al., 2025). General learned restorers such as Restormer, NAFNet, and MambaIR provide strong deterministic priors (Zamir et al., 2022; Chen et al., 2022; Guo et al., 2024), but do not explicitly quantify how strongly individual image components are constrained by a calibrated measurement. Metalens-specific restoration has progressed from end-to-end learned correction with DRMI (Seo et al., 2024) to degradation-aware generative reconstruction with DMDiff (Zhang et al., 2025). Recent systems further incorporate optical or acquisition cues for endoscopy, illumination variation, and

in-the-wild imaging (Li et al., 2025; Fang et al., 2026; Mandal et al., 2026). These developments complement joint optical and computational designs for compact broadband and wide-field imaging (Tseng et al., 2021; Chakravarthula et al., 2023; Pinilla et al., 2023; Kim et al., 2024; Liu et al., 2024; Fröch et al., 2025a). In PCAR-Flow, calibration is used not only as reconstruction context but also to regulate how much prior-driven correction is admitted locally.

Generative priors for measurement-constrained reconstruction. Generative restoration methods use learned priors in different ways. StableSR adapts a pretrained diffusion prior to real-world super-resolution (Wang et al., 2024). DiffBIR and SeeSR combine generative priors with restoration or semantic conditioning (Lin et al., 2024; Wu et al., 2024b), while OSEDiff reduces diffusion restoration to a single step (Wu et al., 2024a) and DiT-SR explores a diffusion-transformer architecture trained for super-resolution (Cheng et al., 2025). For measurement-constrained inverse problems, DPS applies likelihood guidance during posterior sampling (Chung et al., 2023), whereas DDNM separates measured range and null-space components when the linear operator is tractable (Wang et al., 2023b). ResShift models restoration through residual transport (Yue et al., 2023). Flow matching provides a continuous transport formulation with efficient ODE sampling (Lipman et al., 2023; Liu et al., 2023; Albergo & Vanden-Eijnden, 2023). PCAR-Flow applies flow matching only to residuals after calibrated deterministic estimation, restricting stochastic modeling to content not resolved by the measurement-supported base estimate.

3 METHOD

3.1 PROBLEM FORMULATION

Calibrated Image Formation. Let $\mathbf{x} \in \mathbb{R}^{H\times W\times C}$ denote the latent linear RGB image and let $\boldsymbol{\mu}$ denote the radiometrically corrected sensor measurement. Given calibration state κ, we write the image formation process as

$$\boldsymbol{\mu} = \mathcal{H}_\kappa \mathbf{x} + \boldsymbol{\epsilon}, \tag{1}$$

where $\mathcal{H}_\kappa$ denotes the spatially varying calibrated operator and $\boldsymbol{\epsilon}$ is sensor noise. Reconstruction can therefore be viewed as inference from $\boldsymbol{\mu}$ and $\mathcal{H}_\kappa$. The spectral PSF and sensor response yield a spatially varying cross-channel operator:

$$(\mathcal{H}_\kappa \mathbf{x})_c(p) = \sum_{c'=1}^{C} \sum_{q\in\mathcal{K}_p} h_{p,c,c'}(q,\kappa)\, x_{c'}(p-q), \tag{2}$$

where p indexes position, $\mathcal{K}_p$ the local PSF, and $h_{p,c,c'}$ the latent-to-sensor channel transfer ($c' \to c$).

Before radiometric correction, the sensor noise follows a calibrated Poisson–Gaussian model (Hasinoff, 2021; Foi et al., 2008). After correction, we use the signal-dependent Gaussian approximation

$$\sigma_c^2(p) = a_c\, \tilde{\mu}_c(p) + b_c, \qquad \tilde{\mu}_c(p) = \max(\mu_c(p), \epsilon_\mu)\,, \qquad \Sigma(\boldsymbol{\mu}) = \operatorname{diag}\{\sigma_c^2(p)\}_{p,c}\,, \tag{3}$$

where a_c and b_c share the dark-frame and flat-field corrections of the reconstruction (Section A.4).

Quantifying Local Observability. Because even weak responses can resolve strong signals, recoverability fundamentally depends on the prior energy relative to the noise floor. We therefore formulate observability via local Bayesian uncertainty reduction rather than raw magnitude. Given the local Fourier response $\mathbf{H}_p(\omega) \in \mathbb{C}^{C\times C}$, trained spectral covariance $\mathbf{S}(p,\omega) \succeq 0$, and calibrated noise $\mathbf{N}_p = \operatorname{diag}(\sigma_c^2(p))_{c=1}^C$, we model the latent state $\mathbf{z}$ and noise $\boldsymbol{\epsilon}$ as zero-mean Gaussians. Conditioning on the measurement resolves a portion of the prior covariance, denoted $\mathbf{R}_p(\omega)$:

$$\mathbf{z}(\omega) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{S}(p,\omega)), \qquad \boldsymbol{\epsilon}(\omega) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{N}_p), \tag{4}$$

$$\mathbf{R}_p(\omega) = \mathbf{S}(p,\omega)\mathbf{H}_p^H(\omega)[\mathbf{H}_p(\omega)\mathbf{S}(p,\omega)\mathbf{H}_p^H(\omega) + \mathbf{N}_p]^{-1}\mathbf{H}_p(\omega)\mathbf{S}(p,\omega). \tag{5}$$

Because residual reconstruction uses CDF 9/7 wavelet subbands instead of ideal Fourier partitions, we characterize band s by an effective second-order analysis response $G_s(\omega)$. The local observability $\Omega_s(p,c) \in [0,1]$ is defined as the fraction of prior band energy $P_{s,c}(p)$ resolved by the measurement:

$$P_{s,c}(p) = \int |G_s(\omega)|^2 [\mathbf{S}(p,\omega)]_{cc}\, d\omega, \qquad \Omega_s(p,c) = \frac{1}{P_{s,c}(p)} \int |G_s(\omega)|^2 [\mathbf{R}_p(\omega)]_{cc}\, d\omega. \tag{6}$$

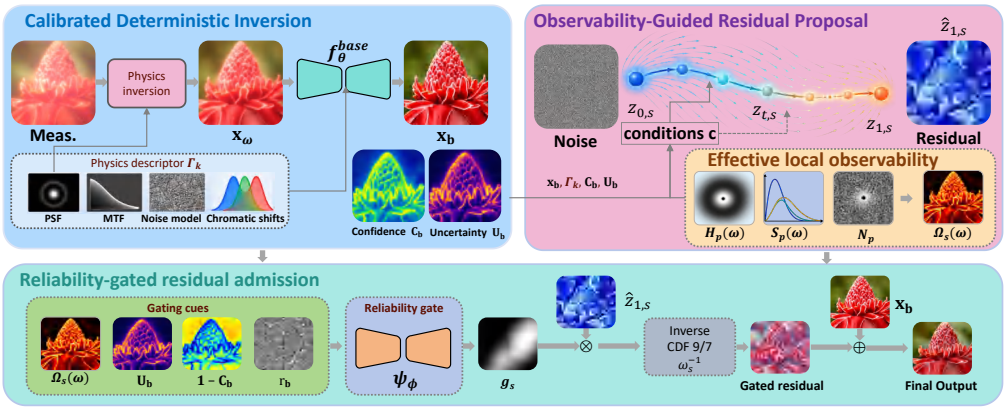

Figure 2: PCAR-Flow pipeline. Calibrated inversion produces the base estimate and reliability cues. Conditional flow proposes high-frequency residuals, and the admission gate combines observability, uncertainty, confidence, and forward disagreement before reconstruction.

To bypass residual generation, bands lacking sufficient prior energy ($P_{s,c}(p) < \epsilon_P$) default to $\Omega_s(p,c) = 1$. For a channel-separable (diagonal) $\mathbf{S}$, the observability Ω_s directly reduces to the $|G_s|^2 S_c$-weighted average of the classical Wiener support $\gamma_{p,c}(\omega)$:

$$\gamma_{p,c}(\omega) = \frac{|\mathcal{O}_{p,c}(\omega)|^2 S_c(p,\omega)}{|\mathcal{O}_{p,c}(\omega)|^2 S_c(p,\omega) + \sigma_c^2(p)}. \tag{7}$$

Optical transfer determines which frequencies reach the sensor, while signal level and noise determine how much uncertainty the measurement removes. In the channel-separable setting, the vanishing-noise limit recovers the classical observed versus null-space split.

3.2 PCAR-Flow

PCAR-Flow decouples deterministic inversion from stochastic generation. A calibrated base estimator recovers resolvable structure, conditional flow matching proposes high-frequency residuals, and an admission gate controls their contribution using observability and reconstruction reliability. Figure 2 shows how the calibrated operator enters both deterministic estimation and residual admission.

Calibrated Deterministic Inversion. The deterministic pathway explicitly inverts the calibrated physical operator to isolate resolvable structure. Starting from $\mathbf{x}^{(0)}$, we minimize the variance-weighted objective $\mathcal{E}_{\text{inv}}(\mathbf{x})$ via T learned gradient steps (with parameters $\alpha_t, \lambda_w \geq 0$) to yield the coarse estimate $\mathbf{x}_w = \mathbf{x}^{(T)}$:

$$\mathcal{E}_{\text{inv}}(\mathbf{x}) = \frac{1}{2}\left\|\Sigma^{-1/2}(\mathcal{H}_\kappa \mathbf{x} - \boldsymbol{\mu})\right\|_2^2 + \frac{\lambda_w}{2}\|\nabla \mathbf{x}\|_2^2, \tag{8}$$

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - \alpha_t\left[\mathcal{H}_\kappa^\top \Sigma^{-1}(\mathcal{H}_\kappa \mathbf{x}^{(t)} - \boldsymbol{\mu}) + \lambda_w \nabla^\top \nabla \mathbf{x}^{(t)}\right]. \tag{9}$$

To quantify data fidelity, we compute the whitened forward residual $\mathbf{r}_w$ and its local energy $e_w(p)$ over a spatial neighborhood $\mathcal{N}_p$. A NAFNet-style network f_θ^{base} then refines $\mathbf{x}_w$ conditioned on these cues (upsampled via $\mathcal{U}$) and the physical descriptors Γ_κ (PSF moments, MTF, chromatic shifts, noise). This final stage outputs the deterministic base $\mathbf{x}_b$, heteroscedastic uncertainty $\mathbf{U}_b > 0$, and tolerance probability $\mathbf{C}_b \in [0,1]$:

$$\mathbf{r}_w = \Sigma^{-1/2}(\mathcal{H}_\kappa \mathbf{x}_w - \boldsymbol{\mu}), \qquad e_w(p) = \text{mean}_{q \in \mathcal{N}_p}\|\mathbf{r}_w(q)\|_2^2, \tag{10}$$

$$(\mathbf{x}_b, \mathbf{C}_b, \mathbf{U}_b) = f_\theta^{\text{base}}(\boldsymbol{\mu}, \mathbf{x}_w, \Gamma_\kappa, \mathcal{U}(\mathbf{r}_w), \mathcal{U}(e_w)). \tag{11}$$

Observability-Guided Residual Proposal. To avoid regenerating deterministic coarse geometry, the stochastic model operates exclusively on high-frequency residuals. Excluding the LL band to preserve base structures, the target for the two-level CDF 9/7 high-frequency bands ($s \in \mathcal{S}_{\text{hf}}$) is

$$\mathbf{z}_{1,s} = \mathcal{W}_s(\mathbf{x} - \mathbf{x}_b), \qquad s \in \mathcal{S}_{\text{hf}}. \tag{12}$$

Given noise $\mathbf{z}_{0,s} \sim \mathcal{N}(0, \sigma_s^2 \mathbf{I})$, we define the linear probability path $\mathbf{z}_{t,s}$ and target velocity $\mathbf{v}_s^\star$. The velocity field $v_{\theta,s}$ is optimized via flow matching, conditioned on the base estimates, calibration, observability, and deterministic reliability (grouped as $\mathbf{c}$):

$$\mathbf{z}_{t,s} = (1-t)\mathbf{z}_{0,s} + t\mathbf{z}_{1,s}, \qquad \mathbf{v}_s^\star = \mathbf{z}_{1,s} - \mathbf{z}_{0,s}, \tag{13}$$

$$\mathbf{c} = \{\mathbf{x}_b, \Gamma_\kappa, \{\Omega_s\}_{s\in\mathcal{S}_{\mathrm{hf}}}, \mathbf{C}_b, \mathbf{U}_b\}, \tag{14}$$

$$\mathcal{L}_{\mathrm{FM}} = \mathbb{E}_{t,\mathbf{z}_0,\mathbf{z}_1}\left[\frac{1}{|\mathcal{S}_{\mathrm{hf}}|}\sum_{s\in\mathcal{S}_{\mathrm{hf}}}\frac{1}{N_s}\left\|v_{\theta,s}(\mathbf{z}_t, t \mid \mathbf{c}) - \mathbf{v}_s^\star\right\|_2^2\right]. \tag{15}$$

A shared three-scale U-Net with band embeddings jointly transports these frequencies. At inference, an ODE solver maps $\mathbf{z}_0$ to a residual proposal $\hat{\mathbf{z}}_{1,s}$.

Reliability-Gated Admission. Because a learned prior may propose details that are weakly supported by the current measurement, we dynamically gate residual proposals. The band-wise admission map $\mathbf{g}_s$ is conditioned on observability Ω_s, deterministic uncertainty maps $(\mathbf{U}_b, 1-\mathbf{C}_b)$, and the whitened forward disagreement $\mathbf{r}_b = \Sigma^{-1/2}(\mathcal{H}_\kappa \mathbf{x}_b - \boldsymbol{\mu})$ upsampled via $\mathcal{U}$:

$$\mathbf{g}_s = \psi_\phi\left(\Omega_s, \mathbf{U}_b, 1-\mathbf{C}_b, \mathcal{U}(\mathbf{r}_b)\right) \in [0,1], \qquad \hat{\mathbf{x}} = \mathbf{x}_b + \sum_{s\in\mathcal{S}_{\mathrm{hf}}} \mathcal{W}_s^{-1}\left(\mathbf{g}_s \odot \hat{\mathbf{z}}_{1,s}\right). \tag{16}$$

To discourage unnecessary residual admission under strong observability, we penalize counterfactual admission increases. Letting $\mathbf{g}_s^+$ denote the gate output under artificially increased observability $\Omega_s^+ = \mathrm{clip}(\Omega_s + \delta_\Omega, 0, 1)$, the monotonicity loss is:

$$\mathcal{L}_{\mathrm{mono}} = \mathbb{E}\left[\frac{1}{|\mathcal{S}_{\mathrm{hf}}|N_g}\sum_{s,p}\mathrm{ReLU}\left(g_s^+(p) - g_s(p)\right)\right]. \tag{17}$$

Finally, we explicitly bound the physical mismatch introduced by this gated admission. We define the normalized whitened forward error as

$$E_{\mathrm{phys}}(\mathbf{u}) = \frac{1}{N_{\mathrm{phys}}}\left\|\mathbf{D}_{\mathrm{phys}}\Sigma^{-1/2}(\mathcal{H}_\kappa \mathbf{u} - \boldsymbol{\mu})\right\|_1. \tag{18}$$

Here $\mathbf{D}_{\mathrm{phys}}$ broadcasts the valid PSF support mask. Defining the band-wise forward gain $K_s = \|\frac{1}{N_{\mathrm{phys}}}\mathbf{D}_{\mathrm{phys}}\Sigma^{-1/2}\mathcal{H}_\kappa\mathcal{W}_s^{-1}\|_{1\to1}$, linearity and the triangle inequality give:

$$E_{\mathrm{phys}}(\hat{\mathbf{x}}) \le E_{\mathrm{phys}}(\mathbf{x}_b) + \sum_{s\in\mathcal{S}_{\mathrm{hf}}} K_s\left\|\mathbf{g}_s \odot \hat{\mathbf{z}}_{1,s}\right\|_1 \le E_{\mathrm{phys}}(\mathbf{x}_b) + \sum_{s\in\mathcal{S}_{\mathrm{hf}}} K_s\|\mathbf{g}_s\|_\infty\|\hat{\mathbf{z}}_{1,s}\|_1. \tag{19}$$

3.3 Decoupled Training Strategy

To prevent downstream modules from corrupting upstream representations, the base estimator, residual flow, and admission gate are optimized sequentially.

Stage A: Deterministic Base. The estimator is trained to reconstruct the image while simultaneously exposing its own unreliability:

$$\mathcal{L}_{\mathrm{base}} = \|\mathbf{x}_b - \mathbf{x}\|_1 + \lambda_\nabla\|\nabla\mathbf{x}_b - \nabla\mathbf{x}\|_1 + \lambda_p E_{\mathrm{phys}}(\mathbf{x}_b) + \lambda_u\mathcal{L}_{\mathrm{NLL}}(\mathbf{U}_b) + \lambda_c\mathcal{L}_{\mathrm{conf}}(\mathbf{C}_b). \tag{20}$$

Here, $\mathcal{L}_{\mathrm{NLL}}$ is a heteroscedastic Gaussian negative log-likelihood, and $\mathcal{L}_{\mathrm{conf}}$ is a binary cross-entropy loss against the local tolerance target $c_b^\star(p) = \mathbb{1}[\bar{e}_b(p) \le \tau_C]$, where $\bar{e}_b(p)$ is the mean absolute error within a 9×9 spatial neighborhood $\mathcal{N}_p$.

Stage B: Residual Flow. With the base estimator frozen, the residual flow is optimized via $\mathcal{L}_{\mathrm{FM}}$ (Eq. 15). Because the upstream estimator is fixed, the residual targets remain stable while the flow is trained.

Stage C: Reliability-Gated Admission. Freezing both upstream modules, we train the admission gate using ODE-sampled proposals. To supervise the gate, we define a local oracle admission map as the clipped least-squares scale between the generated proposal $\hat{\mathbf{z}}_{1,s}$ and the target residual $\mathbf{z}_s^\star = \mathcal{W}_s(\mathbf{x} - \mathbf{x}_b)$:

$$\mathbf{g}_s^\star(p) = \mathrm{clip}\left(\frac{\langle\hat{\mathbf{z}}_{1,s}, \mathbf{z}_s^\star\rangle_{\mathcal{N}_p}}{\|\hat{\mathbf{z}}_{1,s}\|_{\mathcal{N}_p}^2 + \epsilon}, 0, 1\right). \tag{21}$$

Table 1: Datasets and evaluation protocols.

Protocol	Source	System	Size	Primary purpose
MetaCapture train	DIV2K/Flickr2K/WED	M1	7,020	training
MetaCapture val	DIV2K/Flickr2K/WED	M1	780	model selection
MetaCapture test	DIV2K/Flickr2K/WED	M1	3,000	main benchmark
Secondary metalens test	DIV2K subset	M2	500	cross-metalens generalization

Table 2: M1 reconstruction benchmark.

Methods	**PSNR↑**	**SSIM↑**	**LPIPS↓**	**DISTS↓**	**NIQE↓**	**MANIQA↑**	**MUSIQ↑**	**CLIP-IQA↑**
Ring deconv. (Kohli et al., 2025)	24.54	0.8017	0.3170	0.2870	8.5932	0.1923	34.63	0.3254
MambaIR (Guo et al., 2024)	26.83	0.8126	0.2960	0.2730	8.3476	0.2013	32.71	0.3308
Restormer (Zamir et al., 2022)	27.75	0.8412	0.2553	0.2204	7.9321	0.2146	37.64	0.3198
NAFNet (Chen et al., 2022)	26.18	0.8268	0.2489	0.2148	7.6845	0.2219	38.22	0.3274
StableSR (Wang et al., 2024)	28.58	0.8625	0.2558	0.1588	7.3405	0.2412	45.91	0.4097
DiffBIR (Lin et al., 2024)	28.91	0.8617	0.2386	0.1541	7.2243	0.2426	47.10	0.4215
SeeSR (Wu et al., 2024b)	27.95	0.8541	0.2315	0.1673	7.6028	0.2633	44.87	0.3913
OSEDiff (Wu et al., 2024a)	29.72	0.8477	0.2319	0.1696	**6.5159**	0.2544	49.35	0.4281
DRMI (Seo et al., 2024)	25.19	0.7623	0.3417	0.2707	8.1094	0.2102	36.17	0.3441
DMDiff ($\alpha = 0.75$) (Zhang et al., 2025)	30.31	0.8731	0.1705	0.1499	7.2751	0.2782	44.48	0.3869
DiT-SR (Cheng et al., 2025)	24.89	0.7466	0.3121	0.2103	7.9877	0.1803	38.51	0.2626
UFNet (Wu et al., 2026)	24.11	0.7385	0.3850	0.2416	8.3233	0.1737	36.04	0.2841
PCAR-Flow	**30.62**	**0.8841**	**0.1289**	**0.1218**	6.6429	**0.3304**	**52.42**	**0.4469**

Initial noise states are resampled during gate training, while the gate receives only the fixed reliability cues. The gate is optimized via:

$$\mathcal{L}_{\text{gate}} = \frac{\lambda_g}{|\mathcal{S}_{\text{hf}}|} \sum_{s \in \mathcal{S}_{\text{hf}}} \|\mathbf{g}_s - \mathbf{g}_s^\star\|_1 + \lambda_p E_{\text{phys}}(\hat{\mathbf{x}}) + \lambda_m \mathcal{L}_{\text{mono}}. \tag{22}$$

4 Experiments

4.1 Experimental Setup

Datasets and capture protocols. The primary benchmark follows the paired M1 MetaCamera protocol of Zhang et al. (2025), with 7,020 training, 780 validation, and 3,000 held out test pairs. All learnable components of PCAR-Flow are trained exclusively on the M1 training split, with the M1 validation split used for model selection. M2 follows the large-area PB-phase metalens configuration of Seo et al. (2024) and is calibrated independently. For cross-metalens evaluation, the M1-trained model is applied directly to M2 with all learned parameters frozen and without retraining or fine-tuning. Only the M2-specific calibration package is supplied at inference. This protocol evaluates generalization to a substantially different calibrated forward model. Table 1 summarizes the evaluated protocols, and Table 9 gives the optical specifications.

Baselines. The comparison covers spatially varying deconvolution with Ring (Kohli et al., 2025), general learned restoration with Restormer (Zamir et al., 2022), NAFNet (Chen et al., 2022), MambaIR (Guo et al., 2024), and UFNet (Wu et al., 2026), diffusion-based restoration with StableSR (Wang et al., 2024), DiffBIR (Lin et al., 2024), SeeSR (Wu et al., 2024b), OSEDiff (Wu et al., 2024a), and DiT-SR (Cheng et al., 2025), and metalens-specific reconstruction with DRMI (Seo et al., 2024) and DMDiff (Zhang et al., 2025). All methods use the same paired test set and the same preprocessing and metric pipeline for the reported comparison.

Metrics. PSNR and SSIM measure reference-based fidelity. LPIPS (Zhang et al., 2018) and DISTS (Ding et al., 2022) measure perceptual similarity. NIQE (Mittal et al., 2013), MANIQA (Yang et al., 2022), MUSIQ (Ke et al., 2021), and CLIP-IQA (Wang et al., 2023a) provide complementary no-reference quality assessment. Physical consistency is evaluated separately using the normalized whitened forward error E_{phys} in Equation (18). Additional uncertainty, regional, and efficiency analyses are reported in the supplement.

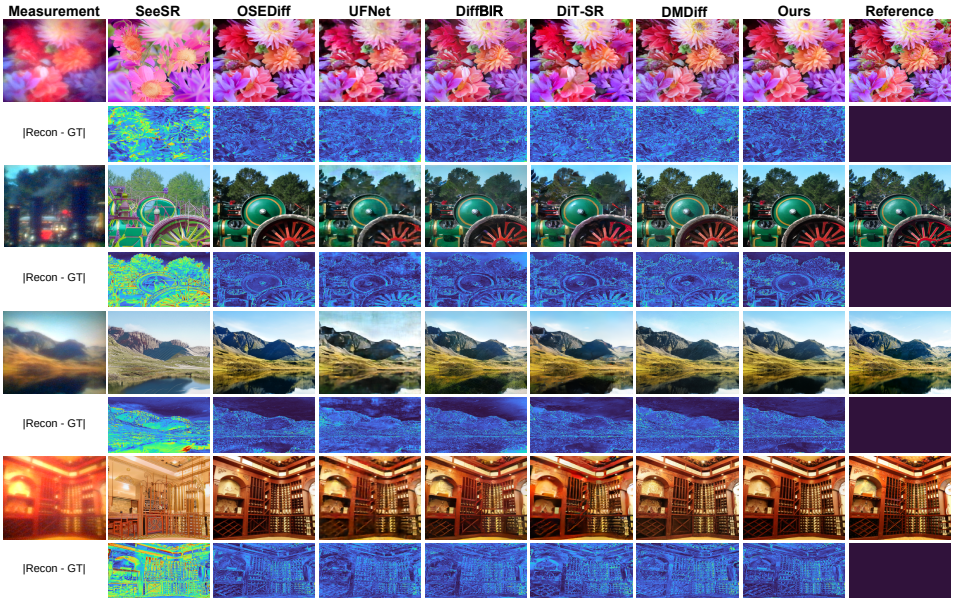

Figure 3: Representative M1 reconstructions with absolute error maps.



Figure 4: Spatial correspondence between effective observability and residual admission.

Implementation details. Models are implemented in PyTorch on one RTX A5000 GPU. The base reconstructor uses eight unrolled gradient steps and a four-scale NAFNet-style network. The residual flow uses a three-scale conditional U-Net over two-level CDF 9/7 detail bands. All learnable components are optimized exclusively on the M1 training split. The base, flow, and admission stages are trained for 200k, 240k, and 60k iterations with AdamW and cosine decay, with the M1 validation split used for model selection. No M2 reconstruction pairs are used for optimization, fine-tuning, or model selection. Inference uses 12 midpoint ODE evaluations unless stated otherwise. We use $\delta_\Omega = 0.05$ for the counterfactual monotonicity perturbation and a 9×9 local confidence window.

4.2 Reconstruction Performance

Table 2 compares reference-based, perceptual, and no-reference image quality on M1. Relative to DMDiff, PCAR-Flow improves PSNR from 30.31 to 30.62 dB, reduces LPIPS from 0.1705 to 0.1289, and reduces DISTS from 0.1499 to 0.1218. The no-reference metrics provide a complementary assessment of appearance quality, while physical consistency is reported separately through E_{phys} in Tables 4 to 6. The examples in Figure 3 show sharper local structure without changing the scene layout supported by the measurement.

4.3 Mechanism Analysis

Effective observability. Table 3 shows that optical transfer alone is insufficient to predict residual uncertainty. Adding signal-dependent noise and image statistics reduces calibration MAE from 0.118 to 0.047 and raises Spearman correlation from 0.69 to 0.92. The spatial maps in Figure 4 provide the corresponding local view. Lower-observability regions receive stronger residual admission, while well-observed content remains close to the deterministic base.

Continuous prior allocation. Table 4 isolates where the generative prior is allowed to act. The deterministic base gives the lowest E_{phys} at 0.0369 but retains more perceptual error. Full-image

Table 3: Observability components. MAE compares $1 - \Omega$ with normalized empirical posterior error.

Observability definition	Cal. MAE↓	r_s ↑
Optical transfer only	0.118	0.69
Optics + homoscedastic noise	0.094	0.76
Optics + signal-dependent noise	0.069	0.85
Full Ω	**0.047**	**0.92**

Table 4: Prior allocation on M1.

Strategy	PSNR↑	LPIPS↓	E_{phys} ↓
Base only	29.76	0.1568	**0.0369**
DDNM-style hard split	29.32	0.1572	0.0584
DPS-style full-image posterior	29.06	0.1510	0.0648
Full image conditional flow	29.88	**0.1248**	0.0748
HF residual flow without Ω	30.13	0.1435	0.0564
Analytic gate using $1 - \Omega$	30.40	0.1347	0.0453
PCAR-Flow	**30.62**	0.1289	0.0398

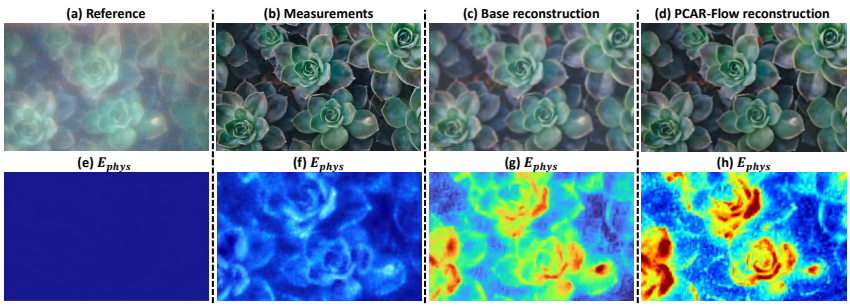

Figure 5: Representative reconstructions under different prior allocations.

Table 5: Reliability cue ablation. R_{mono} is the counterfactual monotonicity violation rate.

Gate configuration	LPIPS↓	E_{phys} ↓	R_{mono} ↓
w/o $\mathbf{U}_b$	0.1341	0.0460	0.027
w/o $\mathbf{C}_b$	0.1336	0.0452	0.025
w/o $\mathbf{r}_b$	0.1360	0.0493	0.030
w/o monotonic loss	0.1327	0.0439	0.083
Full gate	**0.1289**	**0.0398**	**0.018**

Table 6: Calibration mismatch on M1.

Condition	PSNR↑	LPIPS↓	E_{phys} ↓	Mean gate act.↓
Nominal calibration	30.62	0.1289	0.0398	0.382
Joint moderate drift	28.94	0.1544	0.0602	0.255
Physical remount	29.31	0.1512	0.0537	0.301
Thermal drift $+15^{\circ}$C	29.66	0.1438	0.0489	0.324

Table 7: Cross metalens generalization.

System	ρ	HF flow LPIPS↓	PCAR-Flow LPIPS↓	Rel. gain↑	HF flow E_{phys} ↓	PCAR-Flow E_{phys} ↓
M2 metalens	0.39	0.1460	0.1321	9.6%	0.0492	0.0410
M1 metalens	0.47	0.1435	0.1289	10.2%	0.0564	0.0398

conditional flow reaches the lowest LPIPS, 0.1248, but increases E_{phys} to 0.0748. PCAR-Flow

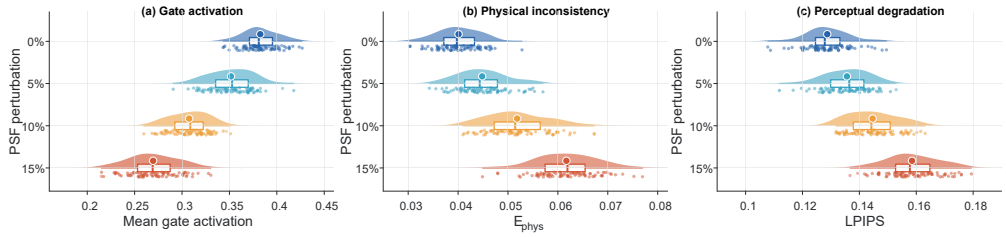

Figure 6: Controlled PSF-perturbation response.

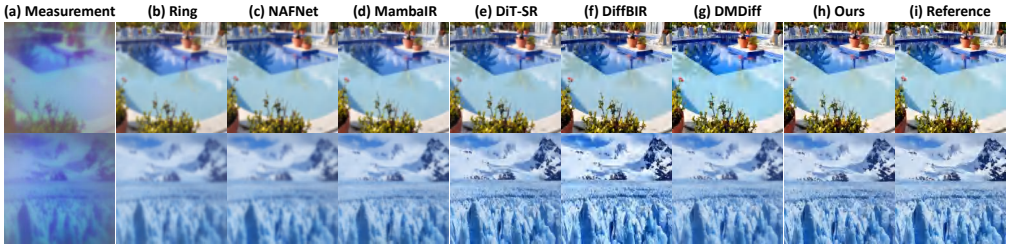

Figure 7: Qualitative cross-metalens generalization on M2 using the frozen M1-trained model.

reaches 0.1289 LPIPS with $E_{\text{phys}} = 0.0398$, reducing perceptual error relative to the base while keeping forward inconsistency close to the deterministic solution. Figure 5 shows representative reconstructions and the corresponding spatial forward-error maps.

Gate configuration. Table 5 separates the roles of the reliability cues. Removing the forward residual cue gives the largest increase in E_{phys}, from 0.0398 to 0.0493. Removing the monotonic loss raises R_{mono} from 0.018 to 0.083 while changing LPIPS more modestly. These results support the use of runtime forward disagreement together with monotonic control. Wavelet band selection, uncertainty calibration, runtime, memory, and solver depth are reported in the supplement.

4.4 Robustness and Cross-Metalens Generalization

Calibration mismatch. Table 6 evaluates physically motivated calibration mismatch conditions. Under joint moderate drift, E_{phys} rises from 0.0398 to 0.0602 while mean gate activation decreases from 0.382 to 0.255. Physical remounting and thermal drift show the same response direction. Figure 6 complements these cases with a controlled PSF-perturbation sweep, showing how gate activation, forward inconsistency, and perceptual error change as the supplied calibration departs from the nominal state.

Cross-metalens generalization. Table 7 and Figure 7 evaluate PCAR-Flow on the independently calibrated M2 system using the model trained exclusively on M1. All learned parameters remain frozen, and no M2 fine-tuning or parameter adaptation is performed. Only the M2-specific calibration package is substituted at inference to instantiate the corresponding forward operator and physical descriptors. Relative to the unguided high-frequency residual flow, PCAR-Flow reduces M2 LPIPS from 0.1460 to 0.1321 and E_{phys} from 0.0492 to 0.0410. Without retraining for new optics, this mechanism enables PCAR-Flow to qualitatively reconstruct sharper edges and textures.

5 Conclusion

This paper proposed PCAR-Flow for generative metalens reconstruction, formulating prior involvement as a continuous allocation problem governed by calibrated measurement support. Effective local observability measures resolved uncertainty and recovers the range and null space split as noise vanishes. Leveraging this theoretical foundation, PCAR-Flow seamlessly decouples deterministic estimation from high-frequency stochastic generation. To preserve physical fidelity, a reliability-guided admission model dynamically filters residuals to bound forward inconsistency. Extensive experiments demonstrate our method's superiority over existing baselines.

AI USE STATEMENT

Large language models were used for language editing, structural organization, and LaTeX assistance. The authors remain responsible for the technical claims, derivations, references, experimental design, and all reported measurements.

ETHICS STATEMENT

Our work focuses on computational imaging and does not involve human subjects or sensitive personal data. Since generative reconstruction may add details that are not fully supported by the measurement, reconstructed images should be interpreted together with the original measurements and calibration information, especially in applications where image fidelity is important.

REPRODUCIBILITY STATEMENT

Table 9 specifies the optical and capture parameters for the M1 and M2 metalens systems. Section A.4 documents PSF, spectral, flat-field, and noise calibration together with valid-region construction and radiometric alignment. Table 10 and Section A.5 give the M1-only training configuration, losses, and schedules. The M2 evaluation keeps all learned parameters frozen and uses only the M2-specific calibration package at inference. Code, calibration packages, evaluation scripts, and non-identifiable real captures will be released.

REFERENCES

Michael S. Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *International Conference on Learning Representations (ICLR)*, 2023.

Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6228–6237, 2018.

Praneeth Chakravarthula, Jipeng Sun, Xiao Li, Chenyang Lei, Gene Chou, Mario Bijelic, Johannes Fröch, Arka Majumdar, and Felix Heide. Thin on-sensor nanophotonic array cameras. *ACM Transactions on Graphics*, 42(6):1–18, 2023.

Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European Conference on Computer Vision (ECCV)*, pp. 17–33, 2022.

Kun Cheng, Lei Yu, Zhijun Tu, Xiao He, Liyu Chen, Yong Guo, Mingrui Zhu, Nannan Wang, Xinbo Gao, and Jie Hu. Effective diffusion transformer architecture for image super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 2455–2463, 2025. doi: 10.1609/aaai.v39i3.32247.

Hyungjin Chung, Jeongsol Kim, Michael T. McCann, Marc L. Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations (ICLR)*, 2023.

Keyan Ding, Kede Ma, Shiqi Wang, and Eero P. Simoncelli. Image quality assessment: Unifying structure and texture similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2567–2581, 2022.

Thomas Eboli, Jean-Michel Morel, and Gabriele Facciolo. Fast two-step blind optical aberration correction. In *European Conference on Computer Vision (ECCV)*, pp. 693–708, 2022.

Fen Fang, Xinan Liang, Muli Yang, Jinghong Zheng, Tobias Mass, Ying Sun, Xulei Yang, Xuewu Xu, and Zhengguo Li. Towards illumination-aware restoration of metalens-captured images: A new dataset and a strong baseline. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(25):21038–21046, 2026. doi: 10.1609/aaai.v40i25.39246.

Alessandro Foi, Mejdi Trimeche, Vladimir Katkovnik, and Karen Egiazarian. Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. *IEEE Transactions on Image Processing*, 17(10):1737–1754, 2008.

Johannes E. Fröch, Praneeth Chakravarthula, Jipeng Sun, Ethan Tseng, Shane Colburn, Alan Zhan, Forrest Miller, Anna Wirth-Singh, Quentin A. A. Tanguy, Zheyi Han, Karl F. Böhringer, Felix Heide, and Arka Majumdar. Beating spectral bandwidth limits for large aperture broadband nano-optics. *Nature Communications*, 16(1):3025, 2025a. doi: 10.1038/s41467-025-58208-4.

Johannes E. Fröch, Shane Colburn, David J. Brady, Felix Heide, Ashok Veeraraghavan, and Arka Majumdar. Computational imaging with meta-optics. *Optica*, 12(6):774, 2025b. doi: 10.1364/OPTICA.546382.

Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. MambaIR: A simple baseline for image restoration with state-space model. In *European Conference on Computer Vision (ECCV)*, pp. 222–241, 2024.

Samuel W. Hasinoff. Photon, Poisson noise. In *Computer Vision: A Reference Guide*, pp. 980–982. Springer, 2021.

Felix Heide, Mushfiqur Rouf, Matthias B. Hullin, Björn Labitzke, Wolfgang Heidrich, and Andreas Kolb. High-quality computational imaging through simple lenses. *ACM Transactions on Graphics*, 32(5):1–14, 2013.

Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. MUSIQ: Multi-scale image quality transformer. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5148–5157, 2021.

Youngjin Kim, Taeyeon Choi, Gun-Yeal Lee, Changhyun Kim, Junsuk Bang, Jaebum Jang, Yoonchan Jeong, and Byoungho Lee. Metasurface folded lens system for ultrathin cameras. *Science Advances*, 10(44):eadr2319, 2024. doi: 10.1126/sciadv.adr2319.

Amit Kohli, Anastasios N. Angelopoulos, David McAllister, Esther Whang, Sixian You, Kyrollos Yanny, Federico M. Gasparoli, Bo-Jui Chang, Reto Fiolka, and Laura Waller. Ring deconvolution microscopy: exploiting symmetry for efficient spatially varying aberration correction. *Nature Methods*, 22(6):1311–1320, 2025. doi: 10.1038/s41592-025-02684-5.

Wuyang Li, Wentao Pan, Xiaoyuan Liu, Zhendong Luo, Chenxin Li, Hengyu Liu, Din Ping Tsai, Mu Ku Chen, and Yixuan Yuan. Metascope: Optics-driven neural network for ultra-micro metalens endoscopy. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 25938–25950, 2025.

Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Wanli Ouyang, Yu Qiao, and Chao Dong. DiffBIR: Toward blind image restoration with generative diffusion prior. In *European Conference on Computer Vision (ECCV)*, pp. 414–429, 2024.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations (ICLR)*, 2023.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations (ICLR)*, 2023.

Yan Liu, Wenhui Li, Kaiwei Xin, Zeyu Chen, Zhengren Chen, Runzhe Chen, Xiaoyan Chen, Feng Zhao, Wei Zheng, and Jianji Dong. Ultra-wide FOV meta-camera with transformer-neural-network color imaging methodology. *Advanced Photonics*, 6(5):056001, 2024. doi: 10.1117/1.AP.6.5.056001.

Debabrata Mandal, Zhihan Peng, Yujie Wang, and Praneeth Chakravarthula. Enabling high-quality in-the-wild imaging from severely aberrated metalens bursts. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 849–859, 2026. doi: 10.1109/WACV61042.2026.00089.

Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2013.

Samuel Pinilla, Johannes E. Fröch, Seyyed Reza Miri Rostami, Vladimir Katkovnik, Igor Shevkunov, Arka Majumdar, and Karen Egiazarian. Miniature color camera via flat hybrid meta-optics. *Science Advances*, 9(21):eadg7297, 2023. doi: 10.1126/sciadv.adg7297.

Joonhyuk Seo, Jaegang Jo, Joohoon Kim, Joonho Kang, Chanik Kang, Seong-Won Moon, Eunji Lee, Jehyeong Hong, Junsuk Rho, and Haejun Chung. Deep-learning-driven end-to-end metalens imaging. *Advanced Photonics*, 6(6):066002, 2024. doi: 10.1117/1.AP.6.6.066002.

Ethan Tseng, Shane Colburn, James Whitehead, Luocheng Huang, Seung-Hwan Baek, Arka Majumdar, and Felix Heide. Neural nano-optics for high-quality thin lens imaging. *Nature Communications*, 12(1):6493, 2021.

Jianyi Wang, Kelvin C. K. Chan, and Chen Change Loy. Exploring CLIP for assessing the look and feel of images. In *AAAI Conference on Artificial Intelligence*, 2023a.

Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin C. K. Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. *International Journal of Computer Vision*, 132:5929–5949, 2024. doi: 10.1007/s11263-024-02168-7.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *International Conference on Learning Representations (ICLR)*, 2023b.

Rongyuan Wu, Lingchen Sun, Zhiyuan Ma, and Lei Zhang. One-step effective diffusion network for real-world image super-resolution. In *Advances in Neural Information Processing Systems*, volume 37, pp. 92529–92553, 2024a. doi: 10.52202/079017-2938.

Rongyuan Wu, Tao Yang, Lingchen Sun, Zhengqiang Zhang, Shuai Li, and Lei Zhang. SeeSR: Towards semantics-aware real-world image super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25456–25467, 2024b.

Ziang Wu, Qi Zhang, Xiaoli Sun, Yonglin Tian, and Chunwei Tian. A unified framework for noisy image super-resolution. *ACM Transactions on Autonomous and Adaptive Systems*, 2026.

Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. MANIQA: Multi-dimension attention network for no-reference image quality assessment. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1191–1200, 2022.

Fanghua Yu, Jinjin Gu, Zheyuan Li, Jinfan Hu, Xiangtao Kong, Xintao Wang, Jingwen He, Yu Qiao, and Chao Dong. Scaling up to excellence: Practicing model scaling for photo-realistic image restoration in the wild. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25669–25680, 2024.

Zongsheng Yue, Jianyi Wang, and Chen Change Loy. Resshift: Efficient diffusion model for image super-resolution by residual shifting. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

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

Jianing Zhang, Jiayi Zhu, Feiyu Ji, Xiaokang Yang, and Xiaoyun Yuan. Degradation-modeled multipath diffusion for tunable metalens photography. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 25914–25924, 2025.

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

A APPENDIX

A.1 NOTATION

Table 8 summarizes the notation used in the supplementary derivations and experiments.

Symbol	Meaning
$\mathbf{x}, \mathbf{y}, \boldsymbol{\mu}$	latent image, raw sensor frame, radiometrically corrected observation
$\mathcal{H}_\kappa$	calibrated forward operator with parameters κ (Equation (23))
κ	calibration package: PSF bank, MTF, chromatic shift, flat-field, noise model
$h_{p,c,c'}, \mathbf{H}_p(\omega)$	local cross-channel PSF and Fourier transfer matrix
a_c, b_c, Σ	Poisson–Gaussian coefficients, $\Sigma = \mathrm{diag}(a_c\tilde{\mu}_c + b_c)$
$\mathcal{W}_s, \mathcal{W}_s^{-1}$	forward and inverse wavelet transform restricted to band s
$\mathcal{S}_{\mathrm{hf}}$	generated bands: LH, HL, HH at levels 1 and 2 (LL excluded)
$\mathbf{S}(p,\omega), P_{s,c}(p)$	local RGB spectral covariance and wavelet band prior energy
$\mathbf{R}_p(\omega)$	covariance reduction supplied by the measurement (Equation (25))
$\Omega_s(p,c)$	prior-energy-weighted effective observability of band s (Equation (26))
$\rho(\kappa, \mathcal{D})$	dataset- and exposure-conditioned fraction of valid cells with $\Omega_s < 0.2$
$\mathbf{x}_w, \mathbf{r}_w, e_w$	coarse inverse (Equation (27)), whitened forward residual, and local energy
$\mathbf{x}_b, \mathbf{C}_b, \mathbf{U}_b$	base reconstruction, local tolerance probability, predicted standard deviation
Γ_κ	14-channel per-pixel calibration descriptor stack
$\mathcal{U}$	bilinear upsampling from tile resolution to pixel resolution
$\mathbf{z}_{t,s}, \mathbf{v}_s^\star$	interpolant state and target velocity of band s at time t
$\mathbf{g}_s$	residual admission map for band s
E_{phys}	mean whitened forward error over $\mathcal{M}_{\mathrm{phys}}$ (Equation (28))
$\mathcal{M}_{\mathrm{valid}}, \mathcal{M}_{\mathrm{align}}$	PSF truncation mask and registration quality mask

Table 8: Notation used in the supplementary derivations and experiments.

A.2 CORE DEFINITIONS

For completeness, we restate the quantities used throughout the supplement. Given calibration state κ, the radiometrically corrected measurement is

$$\boldsymbol{\mu} = \mathcal{H}_\kappa \mathbf{x} + \boldsymbol{\epsilon}, \qquad (\mathcal{H}_\kappa \mathbf{x})_c(p) = \sum_{c'=1}^{C} \sum_{q \in \mathcal{K}_p} h_{p,c,c'}(q;\kappa) x_{c'}(p-q). \tag{23}$$

The signal-dependent noise model is

$$\sigma_c^2(p) = a_c \tilde{\mu}_c(p) + b_c, \qquad \tilde{\mu}_c(p) = \max(\mu_c(p), \epsilon_\mu), \qquad \Sigma(\boldsymbol{\mu}) = \mathrm{diag}\{\sigma_c^2(p)\}_{p,c}. \tag{24}$$

For local spectral covariance $\mathbf{S}(p,\omega)$, transfer matrix $\mathbf{H}_p(\omega)$, and noise covariance $\mathbf{N}_p$, the covariance resolved by the measurement is

$$\mathbf{R}_p(\omega) = \mathbf{S}\mathbf{H}_p^H \left(\mathbf{H}_p \mathbf{S} \mathbf{H}_p^H + \mathbf{N}_p\right)^{-1} \mathbf{H}_p \mathbf{S}, \tag{25}$$

where the dependence of $\mathbf{S}$ on (p,ω) is omitted on the right hand side for compactness. For CDF 9/7 band s, effective observability is

$$\begin{aligned} P_{s,c}(p) &= \int |G_s(\omega)|^2 [\mathbf{S}(p,\omega)]_{cc}\, d\omega, \\ \Omega_s(p,c) &= \frac{1}{P_{s,c}(p)} \int |G_s(\omega)|^2 [\mathbf{R}_p(\omega)]_{cc}\, d\omega. \end{aligned} \tag{26}$$

The deterministic inverse follows

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - \alpha_t \left[\mathcal{H}_\kappa^\top \Sigma^{-1} (\mathcal{H}_\kappa \mathbf{x}^{(t)} - \boldsymbol{\mu}) + \lambda_w \nabla^\top \nabla \mathbf{x}^{(t)}\right]. \tag{27}$$

Physical consistency is evaluated in the sensor domain as

$$E_{\mathrm{phys}}(\mathbf{u}) = \frac{1}{N_{\mathrm{phys}}} \left\| \mathbf{D}_{\mathrm{phys}} \Sigma^{-1/2} \left(\mathcal{H}_\kappa \mathbf{u} - \boldsymbol{\mu}\right) \right\|_1, \tag{28}$$

where $\mathbf{D}_{\mathrm{phys}}$ denotes the valid PSF support mask and N_{phys} is the number of valid sensor samples.

A.3 PROOFS

A.3.1 POSTERIOR INTERPRETATION OF EFFECTIVE OBSERVABILITY

Fix position p and frequency ω. Conditioned on the plug-in covariance $\mathbf{N}_p(\boldsymbol{\mu})$, the local complex Gaussian model is

$$\mathbf{m}(\omega) = \mathbf{H}_p(\omega)\mathbf{z}(\omega) + \boldsymbol{\epsilon}(\omega), \qquad \mathbf{z} \sim \mathcal{N}_{\mathbb{C}}(0, \mathbf{S}), \qquad \boldsymbol{\epsilon} \sim \mathcal{N}_{\mathbb{C}}(0, \mathbf{N}_p). \tag{29}$$

Standard Gaussian conditioning gives

$$\begin{aligned} \mathbf{C}_{\text{post}} &= \mathbf{S} - \mathbf{S}\mathbf{H}_p^H(\mathbf{H}_p\mathbf{S}\mathbf{H}_p^H + \mathbf{N}_p)^{-1}\mathbf{H}_p\mathbf{S} \\ &= \mathbf{S} - \mathbf{R}_p. \end{aligned} \tag{30}$$

For wavelet band s and channel c,

$$\begin{aligned} V_{s|\mu,c}(p) &= \int |G_s(\omega)|^2 [\mathbf{C}_{\text{post}}(p,\omega)]_{cc}\, d\omega \\ &= P_{s,c}(p) - \int |G_s(\omega)|^2 [\mathbf{R}_p(\omega)]_{cc}\, d\omega \\ &= (1 - \Omega_s(p,c)) P_{s,c}(p). \end{aligned} \tag{31}$$

This identity holds under the local plug-in Gaussian model and describes measurement-induced uncertainty under that model. □

A.3.2 FORWARD CONSISTENCY BOUND

Let $\mathbf{d} = \sum_s \mathcal{W}_s^{-1}(\mathbf{g}_s \odot \hat{\mathbf{z}}_{1,s})$ so that $\hat{\mathbf{x}} = \mathbf{x}_b + \mathbf{d}$, and define $\mathbf{A}_{\text{phys}} = N_{\text{phys}}^{-1}\mathbf{D}_{\text{phys}}\Sigma^{-1/2}\mathcal{H}_\kappa$. By linearity and the triangle inequality,

$$\begin{aligned} E_{\text{phys}}(\hat{\mathbf{x}}) &= \|N_{\text{phys}}^{-1}\mathbf{D}_{\text{phys}}\Sigma^{-1/2}(\mathcal{H}_\kappa\mathbf{x}_b - \boldsymbol{\mu}) + \mathbf{A}_{\text{phys}}\mathbf{d}\|_1 \\ &\le E_{\text{phys}}(\mathbf{x}_b) + \sum_{s \in \mathcal{S}_{\text{hf}}} \|\mathbf{A}_{\text{phys}}\mathcal{W}_s^{-1}\|_{1\to 1} \|\mathbf{g}_s \odot \hat{\mathbf{z}}_{1,s}\|_1 \\ &\le E_{\text{phys}}(\mathbf{x}_b) + \sum_{s \in \mathcal{S}_{\text{hf}}} K_s \|\mathbf{g}_s\|_\infty \|\hat{\mathbf{z}}_{1,s}\|_1. \end{aligned} \tag{32}$$

The result follows from the induced $1 \to 1$ norm and $\|\mathbf{g} \odot \mathbf{z}\|_1 \le \|\mathbf{g}\|_\infty \|\mathbf{z}\|_1$. It bounds measurement inconsistency but does not identify a unique latent solution in weakly observed directions. □

A.4 CALIBRATION, CAPTURE PROTOCOL, AND VALID REGION

Table 9 summarizes the two metalens configurations. M1 serves as the sole training domain and the primary reconstruction benchmark. M2 follows the large-area PB-phase metalens configuration reported by Seo et al. (2024) and is calibrated independently. The two systems differ strongly in aperture, focal length, spectral characterization, and field-dependent transfer. M1 uses a 0.5 mm aperture and a 1.0 mm nominal focal length, whereas M2 uses a 10 mm aperture and a 24.5 mm focal length at 532 nm. Their low-observability fractions are also different, with $\rho = 0.47$ for M1 and $\rho = 0.39$ for M2.

The learned parameters of PCAR-Flow are trained only on M1 and remain frozen during M2 evaluation. No M2 reconstruction pairs are used for training, fine-tuning, or model selection. At inference, only the M2-specific calibration package replaces the M1 calibration package to instantiate the corresponding forward operator and physical descriptors. The M2 experiment therefore tests whether the learned allocation mechanism can operate under a new calibrated optical system without target domain parameter adaptation. Because the network weights are unchanged, the M2 results cannot be attributed to retraining on the target metalens.

The M2 nominal f-number is inferred from the reported 24.5 mm focal length at 532 nm and the 10 mm aperture. The remaining M2 optical values in Table 9 follow Seo et al. (2024), while radiometric and noise calibration are performed independently for our evaluation. The purpose of this separation is practical. The network supplies a learned image prior, while the calibration package specifies how a particular metalens and sensor transform that prior into a measurement. Cross metalens evaluation changes the latter while keeping the former fixed.

Table 9: Optical and evaluation specifications of M1 and M2. M2 follows the large-area PB-phase metalens reported by Seo et al. (2024). Sensor-side radiometry and evaluation are calibrated independently in our pipeline.

	M1	**M2**
Optical element	Si_3N_4 on SiO_2	PB phase TiO_2 metalens
Fabrication	e-beam lithography and ICP etching	nanoimprint lithography and TiO_2 deposition
Clear aperture	0.5 mm	10 mm
Nominal focal length	1.0 mm	24.5 mm at 532 nm
Chromatic focal length	measured spectral PSF field	29.0 / 24.5 / 20.5 mm at 450 / 532 / 635 nm
Focusing efficiency	measured during calibration	55.6% at 532 nm
Nominal f-number	f/2.0	≈f/2.45 at 532 nm
Spectral characterization	450 to 650 nm	450, 532, and 635 nm anchors
Evaluation resolution	400×400	400×400
Radiometric calibration	flat-field and noise fit	recalibrated independently
Low obs. fraction ρ	0.47	0.39

A.4.1 Measurement extraction

The calibration products in Figure 8 define the spatial, spectral, and noise terms used by the forward model. M1 is sampled on a 9×9 field grid with a 5 μm pinhole and 10 nm bandpass illumination from 450 to 650 nm. This grid is used to retain field dependence instead of collapsing the optics into one shift-invariant PSF. M2 uses the same calibration pipeline with spectral checks at 450, 532, and 635 nm to capture the chromatic focal shift of the reference configuration. For both systems, Poisson–Gaussian parameters (a_c, b_c) are fitted from flat-field exposure sweeps after dark frame correction. The wavelength-resolved PSFs are then combined with the measured sensor response to obtain the effective RGB kernels used by both inversion and forward-consistency evaluation.

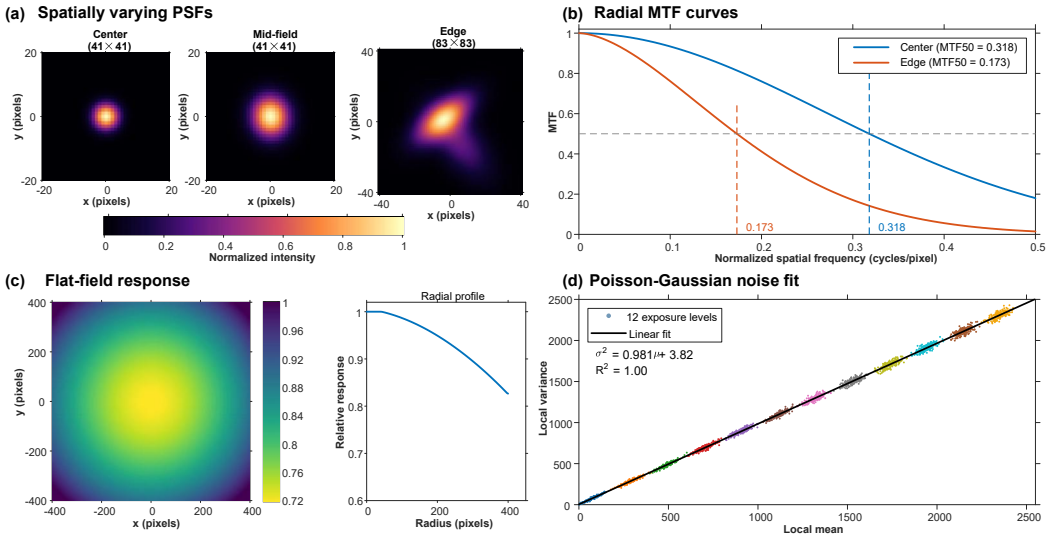

Figure 8: Calibration audit. The PSF examples expose field variation, the radial MTF summarizes frequency transfer, the flat-field records spatial sensitivity variation, and the variance mean fit estimates the signal-dependent noise model used by effective observability.

The two figures document complementary parts of the calibration. Figure 8 shows field-dependent PSFs, radial MTF, flat-field response, and the variance-mean relation used by the noise model. Figure 9 summarizes how these measurements are converted into the calibration package used at inference. Both Ω_s and E_{phys} are computed from this same calibration state.

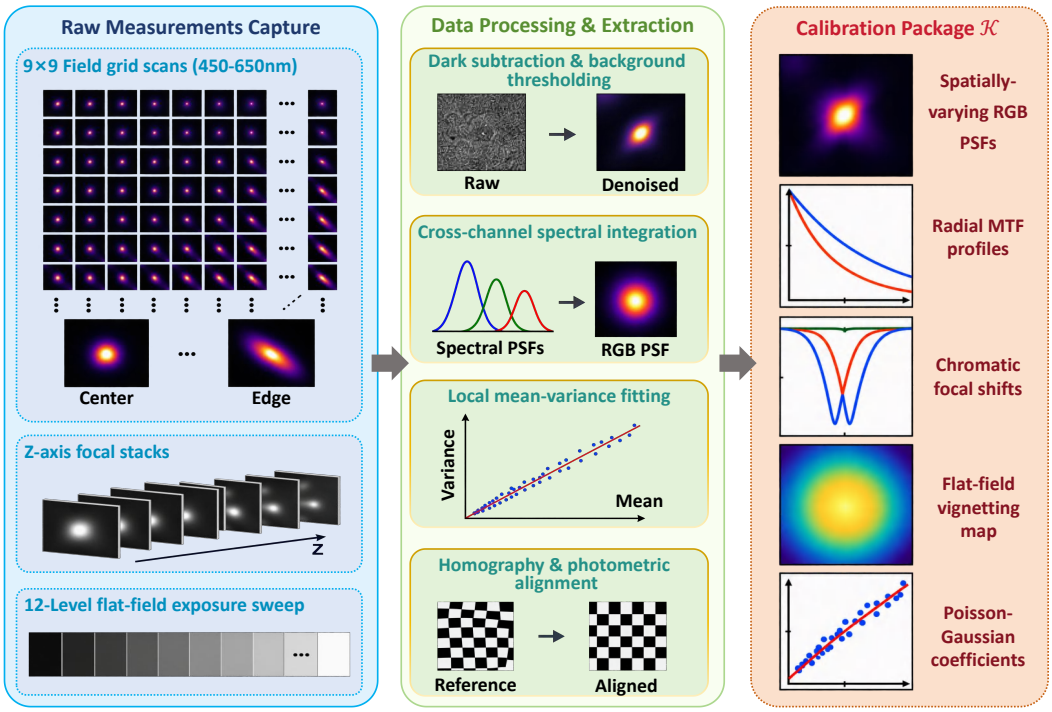

Figure 9: Calibration protocol. Measured PSFs, spectral response, flat-field correction, alignment, and sensor noise are converted into the calibration package that instantiates the forward model and the observability descriptors.

A.4.2 Capture protocols and masking

Images are displayed on an sRGB calibrated LCD at 0.4 m in a darkened environment. A 3×3 color correction is applied only after operator construction so that the physical forward model remains tied to the linear sensor response. Auto exposure is disabled because changing exposure would alter the relation between signal level and the calibrated noise variance. Pixel-aligned reference metrics are computed only after geometric registration. Patches with post alignment normalized cross correlation below 0.9 are excluded by $\mathcal{M}_{\text{align}}$ rather than being retained as difficult examples. This prevents registration error from being counted as reconstruction error.

A.5 Architecture and detailed training parameters

Table 10 summarizes the network capacity. All optimization reported in this subsection uses the M1 training split only. The M1 validation split is used for model selection, and the M2 test set is never used for network training, fine-tuning, or model selection. The base reconstructor contains 54.8 million parameters, while the velocity field contains 31.5 million. The admission gate adds only 0.4 million parameters.

Most learned capacity is assigned to deterministic reconstruction and residual transport. The admission gate accounts for less than 0.5% of the 86.7 M learned parameters.

Optimization and loss weighting. The confidence head is supervised by the local tolerance target

$$c_b^\star(p) = \mathbb{1}[\bar{e}_b(p) \le \tau_C], \tag{33}$$

where $\bar{e}_b(p)$ is the mean absolute reconstruction error in the 9×9 neighborhood $\mathcal{N}_p$.

We use $\lambda_\nabla = 0.2$, $\lambda_p = 0.5$, $\lambda_u = 0.1$, $\lambda_c = 0.2$, $\lambda_g = 0.3$, and $\lambda_m = 0.05$. The target in Equation (33) uses $\tau_C = 0.03$ in normalized linear RGB. The monotonicity offset is fixed to $\delta_\Omega = 0.05$, matching the setting used in the main experiments. The base, flow, and admission stages are optimized for 200k, 240k, and 60k iterations, respectively, with AdamW and cosine decay.

Table 10: Network configurations and parameter distribution for the model trained on M1 at 400×400 resolution.

Module	Configuration	Params
Coarse inverse	$T = 8$ unrolled preconditioned gradient steps, tile size 128, overlap 32	0.01 M
Base reconstructor	NAFNet-style encoder-decoder. 4 scales, widths $(48, 96, 192, 384)$. Γ_κ injected by FiLM. Generates $\mathbf{x}_b$, $\mathbf{C}_b$, and $\log \mathbf{U}_b^2$	54.8 M
Velocity field	three-scale U-Net at wavelet resolution. Widths $(96, 192, 384)$. Conditioning $\mathbf{c}$ injected via FiLM	31.5 M
Gate	five convolutional layers, width 64. Generates one sigmoid map per band	0.4 M

Table 11: Function evaluations versus reconstruction quality on the M1 test set.

NFE	PSNR↑	LPIPS↓	Time (s)
1	29.94	0.1660	0.42
4	30.34	0.1370	0.68
8	30.56	0.1300	0.88
12	**30.62**	**0.1289**	0.98
16	30.63	0.1285	1.21

Table 12: Runtime and peak memory at 400×400 on one RTX A5000 GPU.

Method	Params (M)	Mem (GB)	Time (s)
Wiener, per tile	N/A	0.3	0.09
Ring deconv.	2.1	1.4	0.19
DMDiff (1 step)	102.4	10.8	0.61
DMDiff (20 steps)	102.4	10.8	3.21
PCAR-Flow base only	54.8	5.2	0.37
PCAR-Flow, 12 NFE	**86.7**	**8.1**	**0.98**

Upstream modules are frozen before the next stage is optimized. This keeps the residual target fixed while training the flow and prevents the gate from changing the representation that generated its supervision.

A.6 Sampling cost, runtime, and auxiliary metrics

Unless stated otherwise, the auxiliary experiments in this subsection use the M1 test split with the model selected on M1 validation. Table 11 and Table 12 separate the cost of ODE sampling from the fixed cost of deterministic reconstruction.

Increasing NFE from 1 to 4 reduces LPIPS from 0.1660 to 0.1370. The change from 8 to 12 evaluations is smaller, reducing LPIPS by 0.0011 while adding 0.10 s. Increasing NFE from 12 to 16 changes LPIPS by 0.0004 and PSNR by 0.01 dB, while runtime rises from 0.98 to 1.21 s. We therefore use 12 evaluations in the reported model.

At 12 evaluations, PCAR-Flow uses 86.7 M parameters, 8.1 GB peak memory, and 0.98 s per 400×400 image. Relative to the 20-step DMDiff setting in Table 12, this corresponds to $3.28\times$ lower latency, 25% lower peak memory, and 15.3% fewer parameters. The deterministic base alone runs in 0.37 s.

Generating LH, HL, and HH gives 30.62 dB PSNR with $E_{\mathrm{phys}} = 0.0398$. Adding LL lowers LPIPS from 0.1289 to 0.1261, but PSNR falls by 0.48 dB and E_{phys} rises to 0.0619. We therefore keep LL in the deterministic path and restrict the residual flow to high-frequency bands.

The next two experiments examine the reliability signals used by the admission gate. In Table 14, ECE evaluates calibration of the tolerance probability $\mathbf{C}_b$, while NLL and 90% coverage error evaluate the heteroscedastic uncertainty $\mathbf{U}_b$ under the Gaussian model. Table 15 then measures how well the inference-time reliability signals rank difficult regions after reconstruction.

Table 13: Effect of generated wavelet bands on M1. LL is kept in the deterministic path in the default model.

Generated bands	PSNR↑	LPIPS↓	E_{phys} ↓
HH only	30.24	0.1427	0.0405
LH + HL	30.37	0.1376	0.0411
LH + HL + HH	**30.62**	0.1289	**0.0398**
LL + LH + HL + HH	30.14	**0.1261**	0.0619

Table 14: Calibration of base reconstruction reliability on M1.

Configuration	ECE↓	NLL↓	90% cov. err.↓
w/o NLL	0.069	2.81	0.118
Homoscedastic $\mathbf{U}_b$	0.052	2.19	0.091
w/o $\mathbf{C}_b$ head	N/A	1.94	0.066
Full base	**0.041**	**1.88**	**0.054**

Table 15: Coverage risk ranked by inference-time reliability signals on M1. Entries report LPIPS after retaining the indicated fraction of regions. Lower coverage removes regions with higher predicted risk first.

Ranking score	100%	90%	80%	70%
Random removal	0.1289	0.1287	0.1285	0.1282
$1 - \mathbf{C}_b$	0.1289	0.1218	0.1160	0.1110
$\|\mathbf{r}_b\|$	0.1289	0.1205	0.1132	0.1064
$\mathbf{U}_b$	**0.1289**	**0.1162**	**0.1054**	**0.0978**

The full heteroscedastic model reaches ECE 0.041, NLL 1.88, and 90% coverage error 0.054. Relative to a homoscedastic uncertainty model, ECE decreases by 21.2%, NLL by 14.2%, and coverage error by 40.7%. Removing the NLL objective degrades all three statistics. Removing the $\mathbf{C}_b$ head still permits NLL evaluation for $\mathbf{U}_b$, but the full model reduces coverage error from 0.066 to 0.054. These results support using uncertainty magnitude and local tolerance probability as distinct reliability cues rather than treating them as interchangeable confidence scores.

Random removal changes LPIPS only marginally in the reported evaluation. Ranking by $\mathbf{U}_b$ is more selective than ranking by the other tested cues. At 70% retained coverage, LPIPS is 0.0978 with $\mathbf{U}_b$, compared with 0.1064 for the forward residual and 0.1110 for $1 - \mathbf{C}_b$. This analysis does not alter the reconstruction. It measures whether the reliability cues rank larger reconstruction errors ahead of lower-risk regions.

A.7 Region-wise results and auxiliary benchmarks

Table 16 tests whether the method remains effective as optical degradation increases from the centre toward the field edge. A field-averaged score can obscure performance differences between central and peripheral regions with different transfer strength.

Compared with DMDiff, PCAR-Flow increases PSNR by 0.62 dB at the centre, 0.81 dB in the middle region, and 0.87 dB at the edge. LPIPS decreases by 10.5%, 13.6%, and 16.1% over the same regions, while the reduction in E_{phys} grows from 17.2% at the centre to 25.7% at the edge. The larger edge improvement is consistent with allocating more residual freedom where calibrated support is weaker.

A.8 Limitations

The main limitation of PCAR-Flow is its dependence on calibration. Effective observability inherits error from the forward operator, the signal-dependent noise estimate, and the image statistics model. The local stationarity approximation can also weaken near abrupt changes in the PSF field. The reliability gate can reduce residual admission when uncertainty or forward mismatch increases, but it cannot recover information that is absent from the measurement. Forward consistency alone does not

Table 16: Region-wise evaluation on M1. The edge region has the strongest optical attenuation.

Method	Region	PSNR↑	SSIM↑	LPIPS↓	E_{phys} ↓
DMDiff	Centre	30.46	0.8812	0.1217	0.0419
DMDiff	Middle	29.82	0.8625	0.1441	0.0513
DMDiff	Edge	28.31	0.8218	0.1936	0.0668
PCAR-Flow	Centre	**31.08**	**0.8974**	**0.1089**	**0.0347**
PCAR-Flow	Middle	**30.63**	**0.8846**	**0.1245**	**0.0392**
PCAR-Flow	Edge	**29.18**	**0.8527**	**0.1624**	**0.0496**

identify a unique latent solution in weakly observed directions, so the recovered detail should not be interpreted as directly measured content. Because the effective RGB operator integrates wavelength-resolved calibration through a finite spectral basis, chromatic transfer remains approximate under illumination spectra outside this basis. As a calibration-dependent framework, PCAR-Flow requires a system-specific calibration package for a new metalens design. The M2 experiment evaluates transfer to a second calibrated metalens without retraining. More extreme optical shifts, sensor changes, and calibration regimes remain outside the present evaluation.