RECONSTRUCTION-AWARE DISTRIBUTION MATCHING FOR INVERSE PROBLEMS WITH UNPAIRED DATA

Anonymous authors
Paper under double-blind review

ABSTRACT

Recovering the forward operator of an imaging system from unpaired data avoids both calibration hardware and the clean/degraded pairs that supervision requires—pairs that, for a real lens, are often impossible to acquire. Existing unpaired approaches that explicitly estimate the forward operator by distribution matching evaluate a candidate only through the measurements it generates, whose distribution should match that of the real ones. Components suppressed by the operator are barely present in the degraded images, and therefore barely constrain the operator under such a criterion. Yet inversion precisely tries to recover them. As a result, two operators that are nearly indistinguishable as forward models can invert very differently. We address this by adding a comparison on the clean side: real degraded images, once restored, should be distributed like clean ones. Restorations are computed by a differentiable plug-and-play algorithm parameterized by the operator being learned. On spatially varying PSF calibration and blind super-resolution, the method recovers more accurate operators than degraded-side matching alone, and closes much of the gap to restoration with the true operator.

1 INTRODUCTION

Typical image inverse problems consist of recovering a clean image x from a measurement $y = Ax + \varepsilon$, where A describes the forward operator and ε is an additive noise term. Many methods assume knowledge of A or that paired examples (x, y) are available for supervised training. In practice, neither assumption is easy to satisfy, even in apparently simple scenarios. For instance, when A is a blur due to optics, accurate estimation of A typically requires a controlled acquisition setup often combined with specialized displays (Kee et al., 2011; Bauer et al., 2018) for photography or beads for microscopy (Speiser et al., 2021). Moreover, the procedure must be repeated for every optical system. Paired real data is also difficult to obtain and remains a bottleneck for real-world restoration: even for a simple denoising task, reference images are typically constructed from multiple exposures of the same scene, which can introduce brightness changes and subpixel misalignment (Zamir et al., 2020). Many methods consequently train on synthetic degradation pipelines containing a broad range of corruptions (Wang et al., 2021). These pipelines are, however, hand-designed: there is no guarantee that they cover the acquisition system at hand, and performance drops when they do not.

When A is unknown and paired data is unavailable, two common approaches are blind and unpaired methods. *Blind* methods estimate A and x jointly from a single degraded image, using priors on both the clean image and the degradation operator (Schuler et al., 2012; Chung et al., 2023; Laroche et al., 2024). They require only the degraded image y as input, but the operator prior is fixed before observing the acquisition system and may be a poor match. *Unpaired* methods (Yuan et al., 2018; Sim et al., 2020; Wolf et al., 2021; Tang et al., 2024) make a different trade-off: they require a collection $\mathcal{Y}$ of degraded images from the same system, together with a separate collection $\mathcal{X}$ of clean images, and learn a mapping between the two distributions without requiring per-image correspondence. In most of these approaches, the degradation is represented implicitly by a network and cannot be plugged into another non-blind solver. Meanti et al. (2025) instead recover an *explicit* forward operator through distribution matching, using conditional flow matching (Lipman et al., 2023; Liu et al., 2023). The recovered operator can then be combined with any non-blind solver (Venkatakrishnan et al., 2013; Zhang et al., 2021; Zhu et al., 2023), and is itself the quantity of interest in calibration problems. This is the setting we adopt.

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.

Yet, Meanti et al. (2025) only evaluate an operator through the measurements it produces, matching the distribution of images from $\mathcal{X}$ degraded by a candidate operator A to the real degraded images $\mathcal{Y}$. This can be problematic when the forward operator removes information. Components that are strongly suppressed by A are barely present in the degraded images and therefore provide little information for estimating A. Yet reconstruction attempts to recover these same components, so an operator error that is difficult to detect on the degraded side can still have a substantial effect after inversion. Efrat et al. (2013) study this sensitivity for super-resolution and show that an accurate blur model can be more important than a sophisticated image prior. Thus, two operators that produce similar degraded-image distributions can behave differently when used for reconstruction. In this paper, we make operator estimation reconstruction-aware, using only unpaired data.

Our approach In this work, we improve on the criterion of Meanti et al. (2025), which operates only on the degraded-image space, with a second comparison after reconstruction. This additional term encourages real measurements restored with the candidate operator to produce images distributed like clean ones. The two comparisons therefore evaluate the same operator from opposite sides of the inverse problem: the first through its forward action, and the second through the procedure which will be used for downstream reconstruction.

To allow end-to-end gradient-based learning, the restoration algorithm must depend explicitly on the candidate operator and be differentiable with respect to it. Most plug-and-play algorithms (Venkatakrishnan et al., 2013) satisfy both requirements: the operator enters explicitly through the data-fidelity step, while a fixed denoiser provides the image prior. We use DPIR (Zhang et al., 2021), which provides high-quality restoration with a small number of iterations.

Differentiating the reconstruction objective is nevertheless more expensive than differentiating the measurement objective, as it requires backpropagating through an iterative restoration algorithm which involves calling a deep denoiser at each step. Backpropagation therefore requires storing the activations of all denoiser evaluations, which is memory intensive. We use truncated backpropagation instead, retaining gradients only through the final iterations of the solver, following a strategy that has proved effective in other iterative optimization settings (Shaban et al., 2019).

A second difficulty lies in balancing the two objectives because their gradients can have very different scales: In the measurement branch, the operator is applied once to generate a synthetic degradation. In the reconstruction branch, it affects every iteration of the restoration algorithm, so small operator changes can be amplified or attenuated as they propagate through the solver. Consequently, the relative magnitude of the two gradients can vary substantially during training, making a fixed weight difficult to tune. We therefore downweight the reconstruction gradient whenever its norm exceeds that of the measurement gradient, so that it cannot dominate because of its numerical scale.

Contributions

- We show theoretically that degraded-side matching barely constrains how the operator acts on strongly attenuated image components, even though errors on these components still affect the reconstruction. The clean-side comparison remains sensitive to these components and acts as a prior on the operator, favoring operators whose restorations are distributed most like clean images.
- We introduce reconstruction-aware distribution matching for unpaired operator learning. We complement measurement matching with a clean-side comparison obtained by restoring real measurements using a differentiable plug-and-play algorithm parameterized by the operator.
- Across blind deblurring, space-varying PSF calibration, and blind super-resolution tasks, reconstruction-aware matching produces more accurate estimates than degraded-side matching alone and substantially reduces the gap to restoration with the true operator.

2 Related Work

Learning degradations from unpaired data. Unpaired methods learn from a collection of degraded images and a separate collection of clean ones, without per-image correspondence. RCOT (Tang et al., 2024) and UOTIP (Lee et al., 2026) learn a transport map from measurements to clean signals, respectively in the blind and non-blind setting. DeFlow (Wolf et al., 2021) uses a conditional normalizing flow, and cycle-consistency approaches (Zhu et al., 2017; Yuan et al., 2018; Sim et al., 2020) train forward and backward networks jointly; Chen et al. (2020); Maeda (2020) further develop these approaches for blind super-resolution. Such implicit models can synthesize

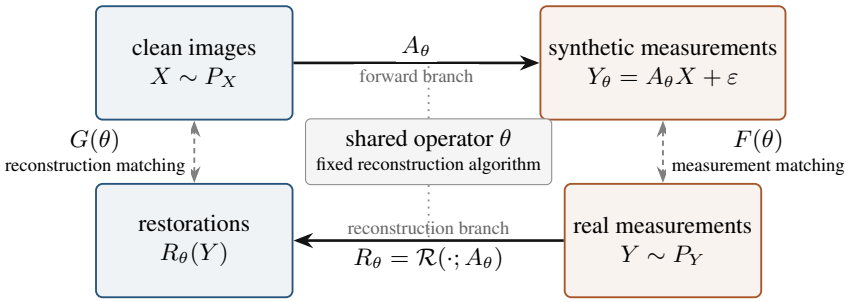

Figure 1: **Reconstruction-aware operator learning.** Measurement matching F encourages clean images degraded with A_θ to reproduce the distribution of real measurements. With the reconstruction-side objective G, real measurements are restored using the same candidate operator, and the resulting images are matched to the clean-image distribution.

training pairs, but they do not provide an explicit operator that can be plugged into a non-blind solver. In contrast, our goal is to learn an *explicit* forward operator. A few papers share the same goal: Lim et al. (2020) replace one of the two cycleGAN generators by a linear blur kernel for microscopy applications; Luo et al. (2022a) and Levac et al. (2025) learn distributions over blur kernels (and noise, for the former) with adversarial objectives, the latter building on AmbientGAN (Bora et al., 2018); and Meanti et al. (2025) estimate a single operator by diffusion distribution matching. In all of them, the operator is evaluated only through the measurements it generates, never through a reconstruction computed with it.

Blind restoration from a single image. Recent diffusion-based methods perform joint inference over the image and degradation (Chung et al., 2023; Murata et al., 2023; Laroche et al., 2024; Sanghvi et al., 2024; Li et al., 2025); blind super-resolution can similarly estimate the blur kernel from a single image (Bell-Kligler et al., 2019; Yang et al., 2024; Luo et al., 2022b). Estimating a high-dimensional operator, such as a space-varying PSF, is challenging from a single observation with strong priors. We instead exploit a collection of images acquired with the same system, which carries more information about the degradation.

Reconstruction-aware operator estimation. Several works take reconstruction into account for kernel estimation. Chen et al. (2025) learn spatially-varying PSFs from calibration patterns, and assess the estimation quality through the quality of a deblurring step. Lunz et al. (2021) correct an inaccurate operator using information from reconstruction, and Guan et al. (2024) correct model mismatch jointly with reconstruction. Yet, these approaches do not address the unpaired learning problem, which is the focus of our work.

Distribution matching and differentiable reconstruction. Score distillation (Poole et al., 2023) and its variational extensions (Wang et al., 2023; Luo et al., 2023) provide gradients for matching a parameterized distribution to a pretrained generative prior. Meanti et al. (2025) adapt these tools to explicit operator learning with conditional flow matching (Lipman et al., 2023; Liu et al., 2023); we use the same distribution-matching estimator for both our measurement- and reconstruction-side losses. Plug-and-play methods (Venkatakrishnan et al., 2013; Romano et al., 2017) combine an explicit data-fidelity term with a denoising prior. DPIR (Zhang et al., 2021) implements this principle with half-quadratic splitting. We use DPIR because the forward operator remains explicit throughout reconstruction, allowing the reconstruction loss to be differentiated with respect to it.

3 Method

Problem setting and formulation. We are given independent clean and degraded datasets $\mathcal{X}$ and $\mathcal{Y}$, with distributions P_X and P_Y. We assume that the clean set represents the latent image distribution underlying the measurements,

$$Y = A_{\theta_\star} X' + \varepsilon, \qquad X' \sim P_X, \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2 I), \tag{1}$$

where $A_{\theta_\star}$ is the unknown forward operator parametrized by $\theta_\star$. The clean image X' that generated a given measurement is never observed, so no pairing between $\mathcal{X}$ and $\mathcal{Y}$ is available.

For a candidate operator A_θ, we can nevertheless generate synthetic measurements

$$Y_\theta = A_\theta X + \varepsilon, \qquad X \sim P_X, \tag{2}$$

with distribution $P_{Y,\theta} := \mathrm{Law}(Y_\theta)$. Given a distributional discrepancy $\mathcal{D}$, measurement-side matching therefore minimizes

$$F(\theta) := \mathcal{D}(P_{Y,\theta} \| P_Y)\,. \tag{3}$$

Under the assumed model, $\theta_\star$ is a minimizer of F, although it does not need to be the unique one.

Our main idea is to evaluate the same candidate operator through a reconstruction procedure. Let $R_\theta(y) := \mathcal{R}(y; A_\theta)$ denote a non-blind reconstruction algorithm run with A_θ, differentiable with respect to θ. Applying it to real measurements induces $Q_\theta := \mathrm{Law}(R_\theta(Y))$, for $Y \sim P_Y$. We then compare these reconstructions with the clean-image distribution, and optimize

$$L(\theta) = F(\theta) + \lambda G(\theta) \quad \text{with} \quad G(\theta) := \mathcal{D}(Q_\theta \| P_X)\,. \tag{4}$$

Thus, F asks whether A_θ reproduces the observed degradation statistics, while G asks how the same candidate behaves after inversion. Figure 1 illustrates the measurement and reconstruction branches.

Complementary roles of the two objectives. Measurement matching is tied to the forward model: with the correct operator, synthetic and real measurements follow the same distribution. The reconstruction objective plays a different role. It measures how the candidate operator affects the distribution of the inverse solutions by comparing restored measurements with clean images. Even with the correct operator, these two distributions do not need to coincide exactly, because reconstruction cannot fully recover information lost to degradation and noise. For most restorer choices, G is even biased, but it nevertheless encourages a desirable property (the distribution of reconstructions should be similar to that of clean images) and remains sensitive to operator errors that may be almost invisible in measurement space. We therefore use F and G as complementary signals: F constrains the forward model, while G acts as a prior on the components that F barely constrains.

In Section 4, we study the sensitivity of the two objectives to an operator error on a single image component (say a frequency), with ρ denoting the SNR of that component after degradation. When $\rho \ll 1$, the component is strongly attenuated relative to the measurement noise and measurement-side information decays as $\Theta(\rho^2)$, while the effect on reconstruction decays only as $\Theta(\rho)$.

3.1 Practical distribution matching with IKL

The formulation above is independent of the discrepancy $\mathcal{D}$. Following Meanti et al. (2025), we instantiate both branches with the integrated KL (IKL) loss function. We present below its gradient, whereas the loss itself and further details are given in Appendix B. Specifically, consider matching a parameterized distribution P_θ to a fixed target distribution Q. Let $Z_1(\theta) \sim P_\theta$ be a sample that is differentiable with respect to θ, $Z_0 \sim \mathcal{N}(0, I)$, and $Z_t = (1-t)Z_0 + tZ_1(\theta)$. Conditional flow matching represents a distribution P by a velocity field

$$v_P(z,t) = \mathbb{E}[Z_1 - Z_0 \mid Z_t = z]. \tag{5}$$

Then, following Luo et al. (2023), the gradient of the chosen IKL loss can be written

$$\nabla_\theta \mathrm{IKL}(P_\theta \| Q) = \int_0^1 \mathbb{E}\big[(J_\theta Z_1)^\top \big(v_{P_\theta}(Z_t,t) - v_Q(Z_t,t)\big)\big]\,\mathrm{d}t, \tag{6}$$

where $J_\theta Z_1$ is the Jacobian of the generated sample with respect to θ. Following Luo et al. (2023); Meanti et al. (2025), we estimate v_Q with a *teacher* flow trained once on the target data and then frozen, and v_{P_θ} with an *auxiliary* flow trained online on samples from the current model distribution. The gradient (6) is then estimated by sampling t and Z_0, see Appendix B for more details.

Measurement and reconstruction branches. We apply the same estimator twice. For measurement matching, $Z_1(\theta) = A_\theta X + \varepsilon$: the teacher is trained on real degraded images and the auxiliary flow tracks the current synthetic measurements. For reconstruction matching, $Z_1(\theta) = R_\theta(Y)$: the teacher is trained on clean images and the auxiliary flow tracks the current restorations.

3.2 Algorithmic Recipe

We next describe the main ingredients of our approach, presented in Algorithm 1. Appendix A details the operator constraints and parametrization, notably fixing the classical translation ambiguity in convolution operator estimation.

Algorithm 1 Reconstruction-aware operator learning.

Require: clean set $\mathcal{X}$, degraded set $\mathcal{Y}$, frozen teachers $v_X^{\mathrm{tc}}, v_Y^{\mathrm{tc}}$, reconstruction steps K.
initialize θ and auxiliary fields $v_X^{\mathrm{aux}}, v_Y^{\mathrm{aux}}$
for $k = 1, \dots, N$ **do**
  sample $X \sim \mathcal{X}$, $Y \sim \mathcal{Y}$ and ε
  update v_Y^{aux} on $A_\theta X + \varepsilon$ with θ fixed
  get new samples $X \sim \mathcal{X}$, $Y \sim \mathcal{Y}$ and ε
  estimate $g_F \approx \nabla_\theta F(\theta_k)$ using Eq. (6) with $Z_1(\theta) = A_\theta X + \varepsilon$
  compute $R_{\theta_k}(Y)$ with K DPIR iterations and estimate $g_G \approx \nabla_\theta G(\theta_k)$ using Eq. (6).
  update v_X^{aux} on $R_\theta(Y)$ detached with θ fixed
  $c_k \leftarrow \min(1, \|g_F\|_2 / (\|g_G\|_2 + \epsilon))$
  update θ with a stochastic gradient descent solver (*e.g.*, Adam) using $g_F + \lambda\, c_k\, g_G$
  enforce operator constraints (Appendix A)
end for
return A_θ

Differentiable plug-and-play reconstruction We instantiate R_θ with DPIR (Zhang et al., 2021), a plug-and-play half-quadratic splitting algorithm whose iterations alternate a data-fidelity update involving A_θ with a fixed pretrained denoiser. Because the restorer and its schedule are fixed, the reconstruction branch depends on θ only through A_θ. We run K DPIR iterations and differentiate only through the final $K_{\mathrm{bp}} \le K$, reducing activation memory from $O(K)$ to $O(K_{\mathrm{bp}})$ (see Monga et al., 2021; Shaban et al., 2019). Additional solver details are given in Appendix C.

Balancing the two stochastic gradients The theory motivates combining the two branches but does not determine their relative scale. At operator-update step k, let

$$g_F \approx \nabla_\theta F(\theta_k), \qquad g_G \approx \nabla_\theta G(\theta_k), \tag{7}$$

denote the stochastic IKL gradient estimates obtained from Eq. (6) for the measurement and reconstruction branches, respectively. The former differentiates through one application of A_θ, whereas the latter differentiates through the iterative inverse solver, so their norms can differ substantially.

Because we did not manage to obtain satisfactory results by using a fixed weighting strategy, we use a gradient-norm balancing heuristic and update the operator by using the gradient estimate

$$g_k = g_F + \lambda\, c_k\, g_G, \quad \text{with} \quad c_k = \min\left(1, \frac{\|g_F\|_2}{\|g_G\|_2 + \epsilon}\right). \tag{8}$$

where c_k normalizes the scale of the reconstruction gradient. The rule is asymmetric: it downweights the reconstruction branch when its gradient is larger than the measurement gradient, but never amplifies it above its raw scale. Thus G cannot dominate merely because differentiation through the solver produces larger gradients, while F retains its role as the forward-model anchor. This is related in spirit to gradient normalization for multi-loss optimization (Chen et al., 2018).

In practice, we use $\lambda = 1$, which ensures $c_k\|g_G\| \le \|g_F\|$, hence $\langle g_F + c_k g_G, g_F\rangle \ge 0$: the combined update never opposes the gradient of F, and $\lambda = 1$ is the largest weight with this property. The clean-side term can thus steer the descent of F, but never trade data fit for its own preference. The update therefore does not minimize Eq. (4) with a fixed predefined weight: when $c_k < 1$, it is the gradient of $F + c_k G$, with a weight $c_k \le 1$ that adapts during training. With exact gradients, its stationary points are the points where $\nabla F + \mu \nabla G = 0$ for some $\mu \in [0, 1]$.

Practical heuristics. Algorithm 1 presents the method in its idealized form. A few heuristics are used in all our experiments and are described in Appendix C. Notably, the weight λ is ramped linearly from 0 to 1 over an initial fraction of training: early on the operator is close to its initialization, so its reconstructions are poor and the clean-side gradient says little about θ. Then, we report an exponential moving average of the operator parameters rather than the last iterate, which removes the step-to-step fluctuation of a low-dimensional parameter driven by stochastic gradients.

4 THEORY: WHY RECONSTRUCTION-AWARE MATCHING HELPS

Setup: Sensitivity analysis for one component. A forward operator can transmit different image components (*e.g.*, a spatial frequency) with different strengths. Here, we consider one scalar component and study the effects of the loss functions F and G when estimating the operator's action on this component, as well as the effect of an estimation error on the reconstruction.

After rescaling the clean component and the measurement noise to unit variance, we write its observation as

$$Y_\rho = \sqrt{\rho}\, X + N, \qquad N \sim \mathcal{N}(0,1), \qquad \mathbb{E}[X] = 0, \quad \mathbb{E}[X^2] = 1, \tag{9}$$

where $\rho > 0$ is the SNR of component X after degradation. When $\rho \ll 1$, the operator has attenuated the component such that its contribution to the measurement is small relative to noise.

Sensitivity analysis for estimating ρ. We define a perturbation of the component gain ρ by

$$\rho_\delta = \rho(1+\delta), \qquad \delta > -1. \tag{10}$$

This is a relative perturbation: δ represents the fractional error regardless of how visible the component is. For the analysis below, we consider the KL loss function for the measurement side:

$$F_\rho(\delta) = \mathrm{KL}\big(P_\rho \| P_{\rho(1+\delta)}\big) \quad \text{with} \quad P_\gamma := \mathrm{Law}(\sqrt{\gamma}X + N). \tag{11}$$

To define a loss function on the reconstruction side, we must define a reconstruction method. We use here the posterior mean which represents an "ideal" MMSE restorer:

$$\eta_\gamma(y) = \mathbb{E}[X \mid \sqrt{\gamma}X + N = y]. \tag{12}$$

Its reconstruction risk is then

$$\mathcal{R}_\rho(\delta) = \mathbb{E}\Big[\big(X - \eta_{\rho(1+\delta)}(Y_\rho)\big)^2\Big]. \tag{13}$$

Finally, we consider the reconstruction-side objective using the Wasserstein-2 loss function:

$$G_\rho(\delta) = W_2^2(Q_{\rho,\delta}, P_X), \qquad P_X = \mathrm{Law}(X) \quad \text{and} \quad Q_{\rho,\delta} = \mathrm{Law}\big(\eta_{\rho(1+\delta)}(Y_\rho)\big). \tag{14}$$

Note that we use ordinary KL, W_2, and posterior-mean reconstruction because they make the sensitivity analysis tractable, whereas the practical method uses IKL and DPIR. Under this choice of loss functions F_ρ and G_ρ, we can now state our main sensitivity results for low-SNR components.

Theorem 1 (Sensitivity of poorly observed components)**.** *Assume that X is bounded. As $\rho \to 0$,*

$$F_\rho''(0) = \frac{\rho^2}{2} + o(\rho^2), \quad \textit{and} \quad \frac{d^2}{d\delta^2}\big[\mathcal{R}_\rho(\delta) - \mathcal{R}_\rho(0)\big]\Big|_{\delta=0} = \frac{\rho}{2} + o(\rho). \tag{15}$$

Moreover, there exists a constant $\kappa_X \in (0,1)$, depending only on P_X, such that

$$G_\rho'(0) = -\kappa_X\sqrt{\rho} + O(\rho), \qquad G_\rho''(0) = \frac{\kappa_X}{2}\sqrt{\rho} + O(\rho). \tag{16}$$

Appendix D provides the proof of Theorem 1.

Interpretation. Since F_ρ is minimized at the true component gain ($\delta = 0$), its curvature $F_\rho''(0)$ measures the local sensitivity of measurement matching to a relative operator error. More precisely, it is the Fisher information about δ contained in one degraded observation. Thus, as ρ decreases, the operator's action on that component rapidly becomes difficult to estimate from degraded data alone. Its importance for reconstruction decreases more slowly. The excess MSE caused by the same relative operator error has curvature $\Theta(\rho)$, whereas the measurement information is only $\Theta(\rho^2)$. Hence the degraded data can become weakly informative about an operator component while errors in that component still matter after inversion.

The clean-side objective G_ρ remains sensitive in this regime, with slope and curvature of order $\sqrt{\rho}$ but this requires a cautious interpretation. G is sensitive in the low-SNR regime, but it is biased since like any point-estimate restorer, including DPIR, the posterior mean does not produce images distributed like clean ones, even with the true operator. G should thus be seen as a prior on components that the degraded data barely constrain. It acts as a selection criterion among operators that F finds nearly equivalent, encouraging solutions with a desirable property: reconstructions should be distributed like clean images, leading in practice to better estimation (see Section 5).

In Appendix E.7, we study an unbiased but more expensive variant, comparing restorations of real measurements with restorations of clean images degraded by the candidate operator. It performs comparably, suggesting that the bias of G is not harmful.

Table 1: **Deblurring.** Uniform blur on FFHQ-256 (mean over the nine degradations) and measured chromatic PSF field (8×8 grid) on the 76 DPDD test images. Kernel error: aligned kernel MAE ($\times 10^{-4}$); PSNR and LPIPS; SSIM is reported in Appendix E. Restormer and OmniLens are supervised single-image restoration networks: they output a restored image rather than an operator (hence the dashes), are applied directly to the degraded images and are run on the PSF field only.

	Uniform blur (nine kernels)			Measured chromatic PSF field		
Method	Kernel err. ↓	PSNR ↑	LPIPS ↓	Kernel err. ↓	PSNR ↑	LPIPS ↓
Restormer (Zamir et al., 2022)	—	—	—	—	28.57	0.205
OmniLens (Jiang et al., 2026)	—	—	—	—	29.06	0.203
Measurement-only	2.16	31.57	0.110	8.68	23.85	0.207
Ours	**1.56**	**32.02**	**0.106**	**5.52**	**32.41**	**0.132**
True operator	—	32.18	0.105	—	35.52	0.111

5 EXPERIMENTS

We evaluate reconstruction-aware matching on four problems of increasing realism. We follow the unpaired setting of Meanti et al. (2025), but make it harder in every respect: nine uniform degradations instead of two, a measured lens PSF field with strong chromatic aberration, blind super-resolution without any high-resolution reference, and a real lens calibrated from only fifteen photographs of a single scene, a simple setup that does not require extensive data acquisition.

Setup. The main controlled comparison is with degraded-side matching alone, i.e., the approach of Meanti et al. (2025) implemented in our pipeline with the reconstruction branch disabled ($\lambda = 0$). The configuration of this *measurement-only* baseline is fixed first and reused unchanged for our method ($\lambda = 1$), so that the reconstruction branch is the only difference. All experiments share the same optimization recipe: the optimizer, learning rates, batch size, network architectures, reconstruction weight $\lambda = 1$, and solver settings are identical across tasks, and the task-specific settings (Table 3) mainly concern the operator and the data, such as the kernel size, boundary handling, and noise level. We report the kernel error, i.e., the MAE between the estimated and true kernels up to the translation ambiguity (Appendix A), and the quality of the restorations obtained with DPIR (Zhang et al., 2021), which measures what an estimate is worth after inversion; restoration with the true operator serves as an upper bound. All results use the last checkpoint. All experiments run on a single RTX 6000 Ada GPU or an RTX 6000 PRO. Details are given in Appendix E. Our code and the scripts to reproduce the results are provided as supplementary material.

Summary of results. In all three quantitative settings, reconstruction-aware matching yields more accurate operators and better restorations than measurement-only matching, with the largest gain for the most complex operator: it closes 74% of the PSNR gap to the true operator on uniform blur, and turns a failed estimate of the chromatic PSF field (23.9 dB) into one within 3.1 dB of the true field. In single-image blind super-resolution, it outperforms state-of-the-art methods specialized for this problem in PSNR, with a gap to the true operator about 40% smaller than that of the strongest one.

5.1 SPATIALLY UNIFORM BLUR

We consider nine degradations of FFHQ-256 images (Appendix E.2): three irregular motion blurs, four analytic PSFs, and two channel-dependent operators. The clean and degraded training sets contain 1000 images each, drawn from disjoint partitions, with noise level $\sigma = 0.02$. Each operator is estimated as a free 33×33 kernel, with one kernel per channel for the channel-dependent operators.

Reconstruction-aware matching reduces the mean kernel error by 28% ($2.16 \to 1.56$), reduces the PSNR gap to the true operator from 0.61 to 0.16 dB, and improves LPIPS accordingly (Table 1, left). The gain is systematic rather than driven by a few kernels: restoration improves on eight of the nine degradations (paired sign-flip test, $p = 0.008$), and comes within 0.1 dB of the true operator on five of them. The only exception is the broad $\sigma = 3$ Gaussian, for which both estimates lose about 0.3 dB. Per-degradation results are given in Appendix E.3.

5.2 MEASURED PSF FIELD WITH CHROMATIC ABERRATION

We next estimate the PSF field of a real wide-aperture lens, a Canon EF 24 mm f/1.4L USM at f/1.4 measured by Bauer et al. (2018), modeled as an 8×8 grid of 27×27 RGB kernels. Towards the corners, the PSFs become broader, elongated, with strong chromatic aberrations (Appendix E.2). The two unpaired sets come from different corpora: measurements are built from DIV2K (Agustsson & Timofte, 2017), and clean images come from DPDD (Abuolaim & Brown, 2020).

Here, measurement-only matching performs poorly (23.85 dB, Table 1, right), whereas our approach reaches 32.41 dB, within 3.1 dB of the true field. Figure 5 in Appendix E.4 shows why. The measurement-only PSFs are over-smoothed, in particular along the borders of the field, where the PSF is broadest. This is consistent with Section 4: a broad PSF strongly attenuates many components, which the degraded data barely constrain and on which the reconstruction-side prior acts.

Table 1 also reports two supervised networks that restore the image directly, without estimating an operator: Restormer (Zamir et al., 2022), applied with its defocus-deblurring model, which was trained on DPDD, and OmniLens (Jiang et al., 2026), trained on the PSF fields of a large library of automatically designed lenses. They reach 28.57 and 29.06 dB, respectively. Restormer targets a different blur, defocus rather than lens aberrations, but OmniLens is trained precisely for lens aberrations; neither matches a calibrated operator for this lens.

5.3 BLIND SUPER-RESOLUTION

We evaluate blind $\times 2$ super-resolution on the 100 images of DIV2KRK (Bell-Kligler et al., 2019), each degraded by a different anisotropic kernel that must be estimated from the image alone. Following Bell-Kligler et al. (2019), crops of the low-resolution image serve as degraded samples and larger crops of the same image as clean ones, under the assumption of cross-scale patch recurrence, so that neither the true kernel nor any high-resolution image is used for training. We compare with three state-of-the-art methods specialized for this problem: KernelGAN (Bell-Kligler et al., 2019), the supervised DCLS (Luo et al., 2022b), and DIP-DKP (Yang et al., 2024). Our method, by contrast, is not specialized: it uses the same objective and restorer as for deblurring.

Comparison setup with a common Gaussian prior. All three baselines favor compact, near-Gaussian kernels: DIP-DKP draws its candidate kernels from an anisotropic Gaussian family, KernelGAN penalizes non-compact and off-center kernels, and DCLS is trained on synthetic Gaussian kernels. To factor out the effect of such priors in the evaluation, we apply the same Gaussian prior to every method, ours included: each estimated kernel is centered, its border is zeroed, and it is replaced by the anisotropic Gaussian with the same first two moments, before restoration with the same DPIR solver. This post-processing never degrades a method on average. It also explains why our KernelGAN results are higher than those reported by Bell-Kligler et al. (2019). Results without post-processing, where our method also yields the best PSNR, are given in Appendix E.5.

Results. Although not designed for super-resolution, our method achieves the best mean PSNR (Table 2), 1.04 dB above measurement-only matching, with a gap to the true operator of 0.6 dB against 1.1 dB for the strongest baseline, DIP-DKP. It outperforms KernelGAN and DCLS on 82 and 97 of the 100 images, respectively. Against DIP-DKP, the per-image comparison is close to even ($55/100$), but the mean gain is statistically significant ($+0.45$ dB, $p = 2 \cdot 10^{-4}$ two-sided paired permutation test): our advantage comes from avoiding failures rather than from winning most images. Consistently, the errors of the baselines grow with the sharpness of the true kernel, whereas ours depends only weakly on its shape.

5.4 PROOF OF CONCEPT FOR REAL LENS CALIBRATION

Finally, we calibrate a real lens from fifteen photographs of a single textured outdoor scene, taken with a consumer camera at f/16 (seven frames that are blurry due to diffraction, forming $\mathcal{Y}$) and at f/5.6 (eight center crops, which are sharp with no chromatic aberrations, forming $\mathcal{X}$), without paired images. The camera is reframed and rotated between shots, including upside-down, so that the two sets share no image correspondence and no orientation is tied to either aperture (Appendix E.6).

Table 2: **Blind super-resolution on DIV2KRK** ($\times 2$). All kernels receive the same centering and Gaussian post-processing and are restored with DPIR. Kernel error: aligned MAE ($\times 10^{-4}$). Bold denotes best entry and those not significantly different from it according to a paired permutation test.

	Kernel err. ↓	PSNR-Y ↑	PSNR ↑	SSIM ↑	LPIPS ↓
KernelGAN	3.42	34.22	32.46	0.8481	0.2714
DCLS	6.97	33.59	31.84	0.8160	0.3129
DIP-DKP	2.57	34.54	32.77	**0.8575**	**0.2599**
Measurement-only	2.96	33.95	32.20	0.8459	0.2677
Ours	**2.15**	**34.99**	**33.21**	**0.8561**	**0.2601**
True operator	—	35.62	33.81	0.8724	0.2435

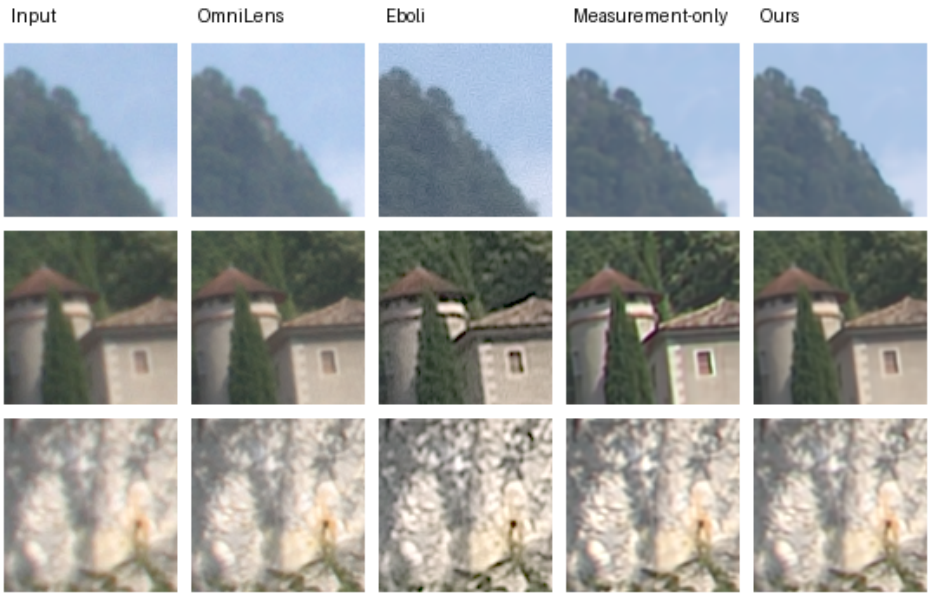

Figure 2: **Real lens calibration.** Chromatic aberrations are best removed by Eboli et al. (2022) but this method does not perform as well as others for deblurring, sometimes producing ringing (middle row) or residual noise artefacts (top row). Measurement-only sometimes fails (middle row). OmniLens is close to our approach, with slightly more residual chromatic aberrations (top row).

We estimate a 9×12 grid of 13×13 RGB kernels, all treated as free parameters. This generic model ignores that PSFs at nearby field positions are strongly dependent, and that accurate classical models exist, *e.g.*, for lateral chromatic aberration (Kang, 2007). Yet, without relying on any such model, our method recovers a plausible PSF field (Appendix E.6). Since no ground truth is available, we compare restorations visually with two single-image methods: the two-step method of Eboli et al. (2022), which estimates local Gaussian PSFs and corrects lateral chromatic aberration, and OmniLens (Jiang et al., 2026), a network already used in Section 5.2. Visual results are presented in Figure 2 and in Appendix E.6. The full images are provided as supplementary material.

6 Concluding Remarks

We have introduced a new approach for learning explicit imaging operators from unpaired data. By evaluating the same candidate operator both before and after inversion, the method complements degraded-side matching where, as our analysis shows, the measurements weakly constrain the forward model. Across various operator learning tasks, this improves estimation and typically reconstruction quality, with the largest gains in the high-dimensional space-varying setting. The main limitations are the additional calibration-time cost (about $\times 6$, see Appendix E.8) and the dependence on the chosen reconstruction algorithm used to define the clean-side prior.

AI USE STATEMENT

We used generative AI tools, including ChatGPT and Claude, during the development and preparation of this work. In accordance with the ICLR 2027 AI Policy for Authors, we disclose their use in several parts of the research workflow.

Generative AI was used to assist with the implementation of the method and the experimental workflow. An initial version of the experimental codebase was written without generative AI assistance; Claude Code was subsequently used to help modify, debug, and improve the implementation and to assist with running experiments. We also used LLMs to discuss and refine aspects of the methodology and theoretical analysis, including checking derivations and improving the presentation of mathematical arguments.

We additionally used ChatGPT and Claude for tasks for which disclosure is recommended by the policy: drafting and editing portions of the manuscript, improving clarity and organization, reviewing the paper for missing reproducibility details, and searching for potentially relevant prior work. LLM-assisted literature searches were used as a discovery tool only; relevant references and claims were subsequently checked by the authors against the original papers.

All AI-assisted code was reviewed and tested by the authors, and experimental results were obtained from the resulting implementation rather than generated by an LLM. Mathematical statements and derivations suggested or revised with LLM assistance were checked by the authors. The authors reviewed all AI-assisted text and remain responsible for the final content of the paper, including its code, experiments, mathematical claims, citations, and conclusions.

ETHICS STATEMENT

This work develops a method for estimating imaging operators and solving image inverse problems from unpaired data. It does not involve human-subject experiments or the collection or release of new personal data. Our experiments use existing image datasets and pretrained models.

As with other image restoration methods, the proposed approach may inherit biases or limitations present in the image distributions and pretrained denoisers on which it relies. This consideration is particularly important in applications in which reconstructed images may be used as evidence or support important decisions. When deployed in the real world, image restoration methods such as ours need to be thoroughly evaluated and tested for each practical use case.

REPRODUCIBILITY STATEMENT

The released code includes the configurations and scripts used to produce the reported experiments and figures.

REFERENCES

Abdullah Abuolaim and Michael S. Brown. Defocus deblurring using dual-pixel data. In *European Conference on Computer Vision (ECCV)*, 2020.

Eirikur Agustsson and Radu Timofte. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In *Conference on Computer Vision and Pattern Recognition (CVPR) Workshop*, 2017.

Matthias Bauer, Valentin Volchkov, Michael Hirsch, and Bernhard Schölkopf. Automatic estimation of modulation transfer functions. In *IEEE International Conference on Computational Photography*, 2018.

Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. Blind super-resolution kernel estimation using an internal-GAN. In *Adv. Neural Information Processing Systems (NeurIPS)*, 2019.

Ashish Bora, Eric Price, and Alexandros G. Dimakis. AmbientGAN: Generative models from lossy measurements. In *International Conference on Learning Representations (ICLR)*, 2018.

Liqun Chen, Yuxuan Li, Jun Dai, Jinwei Gu, and Tianfan Xue. A physics-informed blur learning framework for imaging systems. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

Shuaijun Chen, Zhen Han, Enyan Dai, Xu Jia, Ziluan Liu, Xing Liu, Xueyi Zou, Chunjing Xu, Jianzhuang Liu, and Qi Tian. Unsupervised image super-resolution with an indirect supervised path. In *Conference on Computer Vision and Pattern Recognition (CVPR) Workshop*, 2020.

Zhao Chen, Vijay Badrinarayanan, Chen-Yu Lee, and Andrew Rabinovich. GradNorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In *International Conference on Machine Learning (ICML)*, 2018.

Hyungjin Chung, Jeongsol Kim, Sehui Kim, and Jong Chul Ye. Parallel diffusion models of operator and image for blind inverse problems. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, Hoboken, NJ, 2 edition, 2006. ISBN 978-0-471-24195-9.

Alex Dytso, H. Vincent Poor, and Shlomo Shamai. Conditional mean estimation in gaussian noise: A meta derivative identity with applications. *IEEE Transactions on Information Theory*, 69(3): 1883–1898, 2023.

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

Netalee Efrat, Daniel Glasner, Alexander Apartsin, Boaz Nadler, and Anat Levin. Accurate blur models vs. image priors in single image super-resolution. In *International Conference on Computer Vision (ICCV)*, 2013.

Peimeng Guan, Naveed Iqbal, Mark A. Davenport, and Mudassir Masood. Solving inverse problems with model mismatch using untrained neural networks within model-based architectures. *Transactions on Machine Learning Research (TMLR)*, 2024.

Dongning Guo, Shlomo Shamai, and Sergio Verdú. Mutual information and minimum mean-square error in Gaussian channels. *IEEE Transactions on Information Theory*, 51(4):1261–1282, 2005.

Dongning Guo, Yihong Wu, Shlomo Shamai, and Sergio Verdú. Estimation in Gaussian noise: Properties of the minimum mean-square error. *IEEE Transactions on Information Theory*, 57(4): 2371–2385, 2011.

Qi Jiang, Yao Gao, Shaohua Gao, Zhonghua Yi, Xiaolong Qian, Hao Shi, Kailun Yang, Lei Sun, Kaiwei Wang, and Jian Bai. OmniLens: Towards universal lens aberration correction via LensLib-to-specific domain adaptation. *Optics & Laser Technology*, 198:114858, 2026.

Sing Bing Kang. Automatic removal of chromatic aberration from a single image. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

Eric Kee, Sylvain Paris, Simon Chen, and Jue Wang. Modeling and removing spatially-varying optical blur. In *IEEE International Conference on Computational Photography*, 2011.

Charles Laroche, Andrés Almansa, and Eva Coupete. Fast diffusion EM: A diffusion model for blind inverse problems with application to deconvolution. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2024.

Donggyu Lee, Taekyung Lee, and Jaewoong Choi. UOTIP: Unbalanced optimal transport map for unpaired inverse problems. In *International Conference on Machine Learning (ICML)*, 2026.

Brett Levac, Ajil Jalal, Kannan Ramchandran, and Jonathan I. Tamir. Double blind imaging with generative modeling. *arXiv preprint arXiv:2503.21501*, 2025.

Anqi Li, Weijie Gan, and Ulugbek S. Kamilov. Plug-and-play posterior sampling for blind inverse problems. In *2025 IEEE Statistical Signal Processing Workshop (SSP)*, 2025.

Sungjun Lim, Hyoungjun Park, Sang-Eun Lee, Sunghoe Chang, Byeongsu Sim, and Jong Chul Ye. CycleGAN with a blur kernel for deconvolution microscopy: Optimal transport geometry. *IEEE Transactions on Computational Imaging*, 6:1127–1138, 2020.

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.

Sebastian Lunz, Andreas Hauptmann, Tanja Tarvainen, Carola-Bibiane Schönlieb, and Simon Arridge. On learned operator correction in inverse problems. *SIAM Journal on Imaging Sciences*, 14(1):92–127, 2021.

Weijian Luo, Tianyang Hu, Shifeng Zhang, Jiacheng Sun, Zhenguo Li, and Zhihua Zhang. Diff-instruct: A universal approach for transferring knowledge from pre-trained diffusion models. In *Adv. Neural Information Processing Systems (NeurIPS)*, 2023.

Zhengxiong Luo, Yan Huang, Shang Li, Liang Wang, and Tieniu Tan. Learning the degradation distribution for blind image super-resolution. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022a.

Ziwei Luo, Haibin Huang, Lei Yu, Youwei Li, Haoqiang Fan, and Shuaicheng Liu. Deep constrained least squares for blind image super-resolution. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022b.

Shunta Maeda. Unpaired image super-resolution using pseudo-supervision. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

Giacomo Meanti, Thomas Ryckeboer, Michael Arbel, and Julien Mairal. Unsupervised imaging inverse problems with diffusion distribution matching. In *International Conference on Computer Vision (ICCV)*, 2025.

Vishal Monga, Yuelong Li, and Yonina C. Eldar. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine*, 38(2):18–44, 2021. doi: 10.1109/MSP.2020.3016905.

Naoki Murata, Koichi Saito, Chieh-Hsin Lai, Yuhta Takida, Toshimitsu Uesaka, Yuki Mitsufuji, and Stefano Ermon. GibbsDDRM: A partially collapsed gibbs sampler for solving blind inverse problems with denoising diffusion restoration. In *International Conference on Machine Learning (ICML)*, 2023.

Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5–6), 2019.

Ben Poole, Ajay Jain, Jonathan T. Barron, and Ben Mildenhall. DreamFusion: Text-to-3d using 2d diffusion. In *International Conference on Learning Representations (ICLR)*, 2023.

Vyacheslav V. Prelov and Sergio Verdú. Second-order asymptotics of mutual information. *IEEE Transactions on Information Theory*, 50(8):1567–1580, 2004. doi: 10.1109/TIT.2004.831784.

Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (RED). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844, 2017.

Yash Sanghvi, Yiheng Chi, and Stanley H. Chan. Kernel diffusion: An alternate approach to blind deconvolution. In *European Conference on Computer Vision (ECCV)*, 2024.

Christian J. Schuler, Michael Hirsch, Stefan Harmeling, and Bernhard Schölkopf. Blind correction of optical aberrations. In *European Conference on Computer Vision (ECCV)*, 2012.

Amirreza Shaban, Ching-An Cheng, Nathan Hatch, and Byron Boots. Truncated back-propagation for bilevel optimization. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2019.

Byeongsu Sim, Gyutaek Oh, Jeongsol Kim, Chanyong Jung, and Jong Chul Ye. Optimal transport driven CycleGAN for unsupervised learning in inverse problems. *SIAM Journal on Imaging Sciences*, 13(4):2281–2306, 2020.

Artur Speiser, Lucas-Raphael Müller, Philipp Hoess, Ulf Matti, Christopher J Obara, Wesley R Legant, Anna Kreshuk, Jakob H Macke, Jonas Ries, and Srinivas C Turaga. Deep learning enables fast and dense single-molecule localization with high accuracy. *Nature methods*, 18(9):1082–1090, 2021.

Xiaole Tang, Xin Hu, Xiang Gu, and Jian Sun. Residual-conditioned optimal transport: towards structure-preserving unpaired and paired image restoration. In *International Conference on Machine Learning (ICML)*, 2024.

Singanallur V. Venkatakrishnan, Charles A. Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *Global Conference on Signal and Information Processing (GlobalSIP)*, 2013.

Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. In *International Conference on Computer Vision (ICCV) Workshop*, 2021.

Zhengyi Wang, Cheng Lu, Yikai Wang, Fan Bao, Chongxuan Li, Hang Su, and Jun Zhu. ProlificDreamer: High-fidelity and diverse text-to-3D generation with variational score distillation. In *Adv. Neural Information Processing Systems (NeurIPS)*, 2023.

Valentin Wolf, Andreas Lugmayr, Martin Danelljan, Luc Van Gool, and Radu Timofte. DeFlow: Learning complex image degradations from unpaired data with conditional flows. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

Zhixiong Yang, Jingyuan Xia, Shengxi Li, Xinghua Huang, Shuanghui Zhang, Zhen Liu, Yaowen Fu, and Yongxiang Liu. A dynamic kernel prior model for unsupervised blind image super-resolution. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

Yuan Yuan, Siyuan Liu, Jiawei Zhang, Yongbing Zhang, Chao Dong, and Liang Lin. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In *Conference on Computer Vision and Pattern Recognition (CVPR) Workshop*, 2018.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. CycleISP: Real image restoration via improved data synthesis. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

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 *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. Plug-and-play image restoration with deep denoiser prior. *IEEE Transactions on Pattern Analysis and Machinte Intelligence (PAMI)*, 2021.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *International Conference on Computer Vision (ICCV)*, 2017.

Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (NTIRE)*, 2023.

A Operator constraints and kernel centering

Except for the three motion kernels (Appendix E.2), all kernel-based operators use the centering procedure described in this appendix. Kernels are constrained to be non-negative and unit-sum with a softplus parametrization, while their absolute position is fixed by a centering convention.

A.1 Parameterization and translation ambiguity

For a single kernel, we optimize an unconstrained tensor $\theta \in \mathbb{R}^{h\times w}$ and define

$$a_\theta = \mathrm{softplus}(\theta), \qquad \bar{k}_\theta = \frac{a_\theta}{\sum_u (a_\theta)_u}. \tag{17}$$

Thus non-negativity and unit sum are enforced by construction. The same parameterization is applied to each kernel for multi-channel and space-varying operators. The final forward kernel includes the sub-pixel correction described below and is renormalized afterward.

If the patch distribution and measurement noise are translation-invariant, shifting the kernel leaves the measurement distribution unchanged, which raises a classical translation ambiguity for kernel estimation in blind deblurring tasks. Following Meanti et al. (2025), we therefore fix the kernel position. Rather than adding their center-of-mass penalty, we explicitly recenter the kernel after each operator update, as explained below.

A.2 Centering procedure

Robust center estimate. A global center of mass can be sensitive to small amounts of mass far from the main kernel. We instead use a local mean-shift centroid. Starting from $\bar{k}_\theta$, we subtract its median background level, retain the non-negative residual K, initialize c at its global centroid, and iterate

$$c \leftarrow \frac{\sum_{u:\|u-c\|_\infty \le w} K_u\, u}{\sum_{u:\|u-c\|_\infty \le w} K_u}. \tag{18}$$

We use $w = 10$ pixels and five iterations. Restricting the centroid to a local window reduces the influence of distant residual mass while remaining suitable for flat or elongated kernels.

Integer recentering. Let δ move c to the geometric center of the support and let δ_{int} be its rounding to the nearest integer. After each operator update, we circularly shift θ by δ_{int}. This is a permutation of its entries and therefore introduces no interpolation. For Adam, the first- and second-moment buffers are shifted by the same amount so that the optimizer state remains aligned with the parameters. The exponential moving average of the operator parameters is shifted by the same amount before being updated, so that it never mixes differently translated kernels.

Sub-pixel correction. The residual $\delta_{\mathrm{sub}} = \delta - \delta_{\mathrm{int}}$ is at most half a pixel in each coordinate. At each forward pass, we bilinearly shift the positive pre-kernel and renormalize:

$$k_\theta = \frac{T_{\delta_{\mathrm{sub}}} a_\theta}{\sum_u (T_{\delta_{\mathrm{sub}}} a_\theta)_u}, \tag{19}$$

where $T_{\delta_{\mathrm{sub}}}$ denotes bilinear translation with zero padding. The center estimate is detached from the computation graph, so gradients flow through the shifted kernel values but not through the centering procedure. Because this residual shift is recomputed at each forward pass and never written back to θ, interpolation errors do not accumulate, although a nonzero residual shift slightly smooths the kernel.

A.3 Variants and evaluation

Chromatic operators. Centering each color channel independently would remove the relative displacements that encode chromatic aberration. We therefore estimate a single shift from the green channel and apply it to all channels, removing only their common translation.

Space-varying operators. For a grid of spatially varying PSFs, a common shift of all kernels corresponds to the global ambiguity above. In practice, we apply the same centering convention to each local kernel, using green as the anchor for color PSFs. This keeps each PSF inside its finite support and fixes a consistent convention across the grid. It also removes spatially varying shifts (geometric distortion), which are only weakly identifiable from unpaired data; the ground-truth PSFs are centered in the same way for evaluation.

Evaluation. We report kernel error after translation alignment: the estimated kernel is shifted at sub-pixel resolution so that its center of mass matches that of the ground truth before computing the mean absolute error. This factors out the common translation ambiguity and measures kernel shape more directly. For chromatic operators, alignment is anchored on a single channel so that errors in relative channel displacement remain penalized.

B IKL GRADIENT ESTIMATOR AND IMPLEMENTATION

For a distribution P, draw $Z_1 \sim P$, $Z_0 \sim \mathcal{N}(0, I)$ independently, sample $t \sim \mathrm{Unif}(0, 1)$, and define

$$Z_t = (1 - t)Z_0 + tZ_1. \tag{20}$$

For fixed t, let $P^{(t)}$ denote the law of Z_t. The integrated KL loss (Luo et al., 2023) is then

$$\mathcal{D}(P\|Q) = \mathrm{IKL}(P\|Q) = \int_0^1 w(t)\ \mathrm{KL}\left(P^{(t)}\|Q^{(t)}\right) \mathrm{d}t, \tag{21}$$

where w is a weighting function specified below.

Conditional flow matching represents P through a velocity field trained with

$$\mathcal{L}_{\mathrm{CFM}}(v; P) = \mathbb{E}\Big[\big\|v(Z_t, t) - (Z_1 - Z_0)\big\|_2^2\Big]. \tag{22}$$

Its minimizer is the conditional expectation

$$v_P(z, t) = \mathbb{E}[Z_1 - Z_0 \mid Z_t = z], \tag{23}$$

which appears in Eq. (6).

We choose $w(t) = (1 - t)/t^2$. Writing s_P for the score of $P^{(t)}$, the gradient of each KL term, for a target Q independent of θ, is

$$\nabla_\theta\, \mathrm{KL}\left(P_\theta^{(t)}\|Q^{(t)}\right) = \mathbb{E}\big[(s_{P_\theta} - s_Q)(Z_t)^\top\, \partial_\theta Z_t\big], \tag{24}$$

see Luo et al. (2023). Since $s_P - s_Q = \frac{t}{1-t}(v_P - v_Q)$ and $\partial_\theta Z_t = t\, J_\theta Z_1$, this gradient equals $\frac{t^2}{1-t}\,\mathbb{E}\big[(J_\theta Z_1)^\top (v_{P_\theta} - v_Q)(Z_t, t)\big]$; the weight w cancels the factor $t^2/(1 - t)$ and yields Eq. (6).

Teacher and auxiliary fields. For a fixed target distribution Q, we train a teacher field v^{tc} with Eq. (22) on samples from Q and then freeze it. The current model distribution P_θ changes as θ is updated, so a second field v^{aux} is trained online with the same CFM objective on samples $Z_1(\theta) \sim P_\theta$. Thus $v^{\mathrm{tc}} \approx v_Q$ while $v^{\mathrm{aux}} \approx v_{P_\theta}$, following the alternating distribution-matching construction of Luo et al. (2023), and used by Meanti et al. (2025).

For one draw of $Z_1(\theta)$, t and Z_0 (averaged over a mini-batch in practice), Eq. (6) is estimated by

$$\widehat{g}_\theta = (J_\theta Z_1)^\top \left[v^{\mathrm{aux}}(Z_t, t) - v^{\mathrm{tc}}(Z_t, t)\right]. \tag{25}$$

During the operator update, both velocity predictions are treated as constants with respect to θ. In practice, the same gradient is obtained by automatic differentiation of the stop-gradient surrogate

$$\widehat{\mathcal{J}}(\theta) = \left\langle \mathrm{sg}\left[v^{\mathrm{aux}}(Z_t, t) - v^{\mathrm{tc}}(Z_t, t)\right], Z_1(\theta)\right\rangle, \tag{26}$$

where sg is the stop-gradient operator. By the chain rule, $\nabla_\theta \widehat{\mathcal{J}}(\theta) = \widehat{g}_\theta$.

Although Z_t depends on $Z_1(\theta)$, this dependence is used only to query the two velocity fields; their outputs are detached when updating the operator. This is exact, since the velocity fields enter Eq. (6) only through their values at Z_t. Conversely, when updating v^{aux}, the sample $Z_1(\theta)$ is detached and θ is held fixed. Hence the alternating updates optimize either the auxiliary field or the operator, never both through the same backward pass.

The two branches. The estimator is identical in the measurement and reconstruction branches. For measurement matching,

$$Z_1(\theta) = A_\theta X + \varepsilon, \tag{27}$$

the teacher is trained on P_Y, and the auxiliary field tracks $P_{Y,\theta}$. For reconstruction matching,

$$Z_1(\theta) = R_\theta(Y), \tag{28}$$

the teacher is trained on P_X, and the auxiliary field tracks Q_θ. The only difference is therefore the differentiable map through which $Z_1(\theta)$ depends on the operator parameters; in the reconstruction branch, this backward pass is truncated as described in Section 3.2.

C Implementation and task-specific details

This section provides implementation details omitted from Algorithm 1, in addition to the IKL estimator and auxiliary-flow training described in Appendix B, and the kernel parameterization and centering procedure in Appendix A.

C.1 Optimization details

Reconstruction warm-up. Early in training, A_θ is still close to its initialization and $R_\theta(Y)$ is consequently unreliable. We therefore introduce the reconstruction branch gradually by replacing λ in Eq. (8) with

$$\lambda(k) = \min\left(1, \frac{k}{K_{\mathrm{warm}}}\right)\lambda. \tag{29}$$

After K_{warm} operator updates, the target weight λ is kept fixed. The clean-side auxiliary field continues to track the current reconstruction distribution throughout the warm-up.

Operator averaging. We maintain an exponential moving average $\bar{\theta}$ of the operator parameters and use $A_{\bar{\theta}}$ for evaluation. This reduces the step-to-step variation induced by the stochastic IKL gradients. The average is shifted together with the operator at each recentering, so that it does not mix differently translated kernel representatives; see Appendix A.

Auxiliary-field initialization and reuse. The auxiliary fields are initialized from their corresponding frozen teachers,

$$v_X^{\mathrm{aux}} \leftarrow v_X^{\mathrm{tc}}, \qquad v_Y^{\mathrm{aux}} \leftarrow v_Y^{\mathrm{tc}},$$

rather than from random weights. This provides a useful initialization for the online distributions they subsequently track.

Updating v_Y^{aux} only requires synthetic measurements $A_\theta X + \varepsilon$. For v_X^{aux}, we reuse the reconstruction already computed for the clean-side gradient: after estimating g_G, $R_\theta(Y)$ is detached and used as an auxiliary-flow training sample. This avoids a second call to the reconstruction solver and results in one reconstruction per operator-update cycle. All teacher and auxiliary fields use the conditional flow-matching objective of Eq. (22), with $t \sim \mathrm{Unif}(0,1)$.

C.2 Reconstruction and data handling

DPIR and truncated backpropagation. We implement R_θ with DPIR (Zhang et al., 2021), keeping its pretrained denoiser and iteration schedule fixed during operator learning. We run K iterations and differentiate only through the final K_{bp} ($K = 8$ and $K_{\mathrm{bp}} = 1$ in all experiments; see Appendix E.1). The state after iteration $K - K_{\mathrm{bp}}$ is detached and used to initialize the differentiated part of the solver. The data-fidelity update uses a closed-form proximal operator when available and conjugate gradients otherwise.

Noise levels. The measurement noise is not estimated jointly with the operator. The noise standard deviation used to synthesize measurements in the F branch and the noise level supplied to DPIR are fixed to the known value for synthetic data and treated as fixed hyperparameters for real data. Thus θ is the only quantity inferred from the unpaired datasets.

Patch sampling and boundary handling. Training uses patches sampled at random spatial locations. The measurement branch degrades clean patches, while the reconstruction branch restores degraded patches at the same training resolution. We use replicate padding unless the task uses a naturally periodic construction, in which case circular padding is used; blind super-resolution instead applies the operator as a normalized convolution, since its observed set consists of real crops for which neither convention is appropriate. The same boundary convention is used by the forward model and the reconstruction solver. Per-experiment choices are reported in Appendix E.1.

Spatial conditioning. Except in blind super-resolution, all teacher and auxiliary velocity fields are conditioned on the patch location. They therefore take the form

$$v(z_t, t, \ell), \tag{30}$$

where ℓ encodes the normalized spatial coordinates of the patch. This is important for spatially varying degradation, for which both image and measurement statistics can vary with position. We use the same conditioning in the spatially uniform experiments for consistency.

C.3 Task-specific forward models

Spatially varying blur. Following Meanti et al. (2025), we represent the spatially varying PSF by an $n_h \times n_w$ grid of local kernels. The kernel associated with a patch is obtained by bilinear interpolation of this grid at the patch center. Non-negativity, normalization, chromatic anchoring, and centering are described in Appendix A. Our method and the measurement-only baseline use exactly the same operator parameterization and constraints.

Blind super-resolution. Instead of generating synthetic measurements from a clean dataset as in Eq. (2), A_θ is applied to crops of the observed low-resolution image, and the resulting twice-degraded distribution is matched to that of smaller crops of the same image, following the cross-scale protocol of Bell-Kligler et al. (2019). Likewise, the reconstruction branch restores these smaller crops and matches the restorations to crops of twice their size, which play the role of clean images.

C.4 Initialization

The unconstrained operator parameters are initialized randomly. Combined with the softplus-and-normalize parameterization of Appendix A.1, this yields a diffuse initial kernel without imposing a preferred PSF shape.

D Technical details for Section 4

The results derived here are inspired by the analysis of minimum mean square error estimators for Gaussian denoising in the information theory literature (see Guo et al., 2011).

Theorem 1 concerns the scalar channel

$$Y_\rho = \sqrt{\rho}\, X + N, \qquad N \sim \mathcal{N}(0,1), \qquad \mathbb{E}[X] = 0, \quad \mathbb{E}[X^2] = 1. \tag{31}$$

The scalar variable X represents one component of the clean signal, while ρ measures how strongly that component remains observable after application of the forward operator and addition of noise.

We compare the true channel parameter ρ with the candidate

$$\rho_\delta = \rho(1+\delta). \tag{32}$$

For later use, define the minimum-mean square estimator

$$\eta_\gamma(y) = \mathbb{E}[X \mid \sqrt{\gamma} X + N = y], \tag{33}$$

the reconstruction distribution

$$Q_{\rho,\delta} = \mathrm{Law}\big(\eta_{\rho(1+\delta)}(Y_\rho)\big)\,, \tag{34}$$

and its reconstruction risk

$$\mathcal{R}_\rho(\delta) = \mathbb{E}\Big[\big(X - \eta_{\rho(1+\delta)}(Y_\rho)\big)^2\Big]\,. \tag{35}$$

D.1 Measurement-side sensitivity and Fisher information

Let p_ρ denote the density of $P_\rho = \mathrm{Law}(\sqrt{\rho}X + N)$. A local expansion of the KL divergence in the smooth family $\{P_\rho\}$ gives

$$\mathrm{KL}(P_\rho \| P_{\rho+\Delta}) = \frac{1}{2} J(\rho)\Delta^2 + o(\Delta^2), \tag{36}$$

where

$$J(\rho) = \mathbb{E}_{Y_\rho}\left[\left(\partial_\rho \log p_\rho(Y_\rho) \right)^2 \right] \tag{37}$$

is the Fisher information for the channel parameter ρ (see Cover & Thomas, 2006, Chapter 11). Since the relative perturbation corresponds to $\Delta = \rho\delta$,

$$F_\rho''(0) = \rho^2 J(\rho). \tag{38}$$

Low-SNR expansions for Gaussian channels are classical (Prelov & Verdú, 2004). Here the required leading term follows directly by expanding the Gaussian kernel:

$$p_\rho(y) = \mathbb{E}[\varphi(y - \sqrt{\rho}X)] = \varphi(y)\left[1 + \frac{\rho}{2}(y^2 - 1) + O(\rho^{3/2}) \right], \tag{39}$$

where φ is the standard-normal density. Therefore

$$\partial_\rho \log p_\rho(y)\big|_{\rho=0} = \frac{1}{2}(y^2 - 1), \tag{40}$$

and hence

$$J(0) = \frac{1}{4}\mathbb{E}[(N^2 - 1)^2] = \frac{1}{2}. \tag{41}$$

The same expansion gives $J(\rho) \to 1/2$, so

$$F_\rho''(0) = \frac{\rho^2}{2} + o(\rho^2). \tag{42}$$

D.2 Bounds on low-SNR MMSE estimators

Lemma 1. *Let X satisfy*

$$\mathbb{E}[X] = 0, \qquad \mathbb{E}[X^2] = 1, \qquad |X| \le M \quad \textit{a.s.},$$

and let $N \sim \mathcal{N}(0,1)$ be independent of X. For

$$Y_\beta = \beta X + N, \qquad m_\beta(y) = \mathbb{E}[X \mid Y_\beta = y],$$

there exists a constant $C = C(M)$ such that, for every $y \in \mathbb{R}$ and every $|\beta| \le 1$,

$$|m_\beta(y) - \beta y| \le C\beta^2(1 + y^2), \tag{43}$$

$$|\partial_\beta m_\beta(y) - y| \le C|\beta|(1 + y^2), \tag{44}$$

$$|\partial_\beta^2 m_\beta(y)| \le C(1 + y^2). \tag{45}$$

Proof. Let us introduce the standard Gaussian density

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.$$

Conditioned on $X = x$, Y_β has the density $p_{Y_\beta|X=x}(y) = \varphi(y - \beta x)$ and for every bounded measurable f, we introduce the quantity

$$E_{\beta,y}(f) := \mathbb{E}[f(X) \mid Y_\beta = y] = \frac{\int f(x)\varphi(y - \beta x)\, P_X(dx)}{\int \varphi(y - \beta x)\, P_X(dx)}. \tag{46}$$

Thus, if

$$Z_\beta(y) := \int \exp\left(\beta x y - \frac{\beta^2 x^2}{2} \right) P_X(dx),$$

then

$$E_{\beta,y}(f) = \frac{1}{Z_\beta(y)} \int f(x) \exp\left(\beta xy - \frac{\beta^2 x^2}{2}\right) P_X(dx). \tag{47}$$

We next differentiate (47), noting that if

$$h_{\beta,y}(x) := xy - \beta x^2,$$

then

$$\partial_\beta \exp\left(\beta xy - \frac{\beta^2 x^2}{2}\right) = h_{\beta,y}(x) \exp\left(\beta xy - \frac{\beta^2 x^2}{2}\right).$$

Then, a short calculation gives

$$\begin{aligned} \partial_\beta E_{\beta,y}(f) &= E_{\beta,y}(f h_{\beta,y}) - E_{\beta,y}(f) E_{\beta,y}(h_{\beta,y}) \\ &= \mathrm{Cov}(f(X),\, h_{\beta,y}(X) \mid Y_\beta = y)\,. \end{aligned} \tag{48}$$

The differentiation under the integral is justified by $|X| \le M$; for fixed y and $|\beta| \le 1$, both the integrand and its β-derivatives are bounded by integrable functions depending only on M and y.

We now apply (48) with $f(X) = X$. Define

$$V_\beta(y) := \mathrm{Var}(X \mid Y_\beta = y), \qquad K_\beta(y) := \mathrm{Cov}(X, X^2 \mid Y_\beta = y).$$

Since $h_{\beta,y}(X) = yX - \beta X^2$, we obtain the exact identity

$$\partial_\beta m_\beta(y) = y V_\beta(y) - \beta K_\beta(y). \tag{49}$$

Then, differentiating once more gives

$$\partial_\beta^2 m_\beta(y) = y\, \partial_\beta V_\beta(y) - K_\beta(y) - \beta\, \partial_\beta K_\beta(y).$$

We will now proceed by bounding each term, starting with an auxiliary result. For $j = 1, 2, 3$, let

$$a_{j,\beta}(y) := \mathbb{E}[X^j \mid Y_\beta = y].$$

Since $|X| \le M$, we have $|a_{j,\beta}(y)| \le M^j$. Moreover, Eq. (48) gives

$$\partial_\beta a_{j,\beta}(y) = \mathrm{Cov}\left(X^j,\, yX - \beta X^2 \mid Y_\beta = y\right).$$

For $|\beta| \le 1$,

$$|yX - \beta X^2| \le M|y| + M^2,$$

and therefore

$$|\partial_\beta a_{j,\beta}(y)| \le C(1 + |y|), \qquad j = 1, 2, 3, \tag{50}$$

where C depends only on M (note that we use the relation $\mathrm{Cov}(Z1, Z2) \le 2\|Z_1\|_\infty \|Z_2\|_\infty$ and the fact that necessarily $M \ge 1$ given the normalization assumption on X).

We now remark that

$$V_\beta = a_{2,\beta} - a_{1,\beta}^2, \qquad K_\beta = a_{3,\beta} - a_{1,\beta} a_{2,\beta},$$

and we can now obtain the following bounds (by adjusting the value of C)

$$|K_\beta(y)| \le C, \qquad |\partial_\beta V_\beta(y)| + |\partial_\beta K_\beta(y)| \le C(1 + |y|). \tag{51}$$

Using Eq. (51), and adjusting again the value of C, for $|\beta| \le 1$,

$$|\partial_\beta^2 m_\beta(y)| \le C(1 + y^2),$$

which proves (45). Note that we use here the fact that $|y| \le (1 + y^2)/2$.

It remains to integrate this bound. At $\beta = 0$, $Y_0 = N$ carries no information about X, so

$$m_0(y) = \mathbb{E}[X] = 0, \qquad V_0(y) = \mathrm{Var}(X) = 1.$$

Evaluating (49) at $\beta = 0$ therefore gives

$$\partial_\beta m_\beta(y)|_{\beta=0} = y.$$

By the fundamental theorem of calculus,

$$\partial_\beta m_\beta(y) - y = \int_0^\beta \partial_t^2 m_t(y)\, dt,$$

and hence

$$|\partial_\beta m_\beta(y) - y| \le C|\beta|(1+y^2),$$

which proves (44). Integrating once more,

$$m_\beta(y) - \beta y = \int_0^\beta \big(\partial_t m_t(y) - y\big)\, dt,$$

so that

$$|m_\beta(y) - \beta y| \le C\beta^2(1+y^2),$$

after adjusting once more the constant C. This proves (43).

□

D.3 Effect of operator error on reconstruction

The posterior mean in Eq. (33) is the minimum-MSE estimator for the Gaussian channel; we use standard properties of Gaussian-channel conditional-mean estimation throughout this section (Guo et al., 2005; 2011). In particular, $\eta_\rho(Y_\rho) = \mathbb{E}[X \mid Y_\rho]$ is the L^2 projection of X onto functions of Y_ρ. The resulting orthogonality identity gives

$$\mathcal{R}_\rho(\delta) - \mathcal{R}_\rho(0) = \mathbb{E}\left[\big(\eta_{\rho(1+\delta)}(Y_\rho) - \eta_\rho(Y_\rho)\big)^2\right]. \tag{52}$$

Thus the excess MSE is exactly the squared change in the reconstruction caused by using the wrong operator parameter.

To study the low-SNR regime, write $b = \sqrt{\rho}$ and, using the same notation as in the previous subsection,

$$m_b(y) = \mathbb{E}[X \mid bX + N = y],$$

we obtain from Lemma 1 that

$$|m_b(y) - by| \le Cb^2(1+y^2), \tag{53}$$

The amplitude after perturbation is then $b_\delta = b\sqrt{1+\delta}$, such that

$$\partial_\delta m_{b_\delta}(y)|_{\delta=0} = \frac{b}{2}\partial_b m_b(y). \tag{54}$$

Differentiating Eq. (52) twice under the expectation, which is justified by Lemma 1 and the finite fourth moment of Y_ρ, and using that the integrand vanishes at $\delta = 0$, gives

$$\frac{d^2}{d\delta^2}[\mathcal{R}_\rho(\delta) - \mathcal{R}_\rho(0)]\Big|_{\delta=0} = \frac{\rho}{2}\mathbb{E}\left[\big(\partial_b m_b(Y_\rho)\big)^2\right]. \tag{55}$$

By Lemma 1,

$$|\partial_b m_b(y) - y| \le Cb(1+y^2).$$

Evaluating this bound at $y = Y_\rho = N + bX$ and taking the L^2 norm gives

$$\|\partial_b m_b(Y_\rho) - Y_\rho\|_2 \le Cb\left(\mathbb{E}[(1+Y_\rho^2)^2]\right)^{1/2} = O(b),$$

since X is bounded and N is Gaussian, so $\mathbb{E}[Y_\rho^4]$ is uniformly bounded for $0 \le b \le 1$. Moreover,

$$\|Y_\rho - N\|_2 = b\|X\|_2 = b.$$

Hence, by the triangle inequality,

$$\|\partial_b m_b(Y_\rho) - N\|_2 = O(b).$$

It follows that

$$\mathbb{E}\left[\big(\partial_b m_b(Y_\rho)\big)^2\right] \longrightarrow \mathbb{E}[N^2] = 1$$

as $\rho \to 0$. Therefore,

$$\frac{d^2}{d\delta^2}[\mathcal{R}_\rho(\delta) - \mathcal{R}_\rho(0)]\Big|_{\delta=0} = \frac{\rho}{2} + o(\rho). \tag{56}$$

D.4 SENSITIVITY OF THE OBSERVABLE CLEAN-TARGET LOSS

We now study the reconstruction-side objective

$$G_\rho(\delta) = W_2^2(Q_{\rho,\delta}, P_X), \tag{57}$$

where $P_X = \mathrm{Law}(X)$ is the clean law. The goal is to understand how this loss changes when the candidate operator is perturbed by δ, particularly when ρ is small.

Writing the loss in terms of quantiles. Our first observation is that the MMSE estimator is non-decreasing in the observation. Indeed, from Prop. 1 of Dytso et al. (2023),

$$\eta'_\gamma(y) = \sqrt{\gamma}\ \mathrm{Var}(X \mid \sqrt{\gamma}X + N = y) \geq 0. \tag{58}$$

Let us now write $b = \sqrt{\rho}$, as in the previous subsection, and recall that $m_\beta(y) = \eta_{\beta^2}(y)$. Let F_X and F_{Y_ρ} denote the cdfs of the clean signal and observation, respectively, and define their generalized quantile functions by

$$x(u) = F_X^{-1}(u), \qquad y_b(u) = F_{Y_\rho}^{-1}(u). \tag{59}$$

Because $m_{b\sqrt{1+\delta}}$ is non-decreasing, as shown earlier, it preserves the ordering of observations: applying the reconstruction map to the u-quantile of Y_ρ gives the u-quantile of the reconstruction distribution. Hence,

$$F^{-1}_{Q_{\rho,\delta}}(u) = m_{b\sqrt{1+\delta}}(y_b(u)) \qquad \text{for almost every } u \in (0,1). \tag{60}$$

This observation allows us to exploit a classical representation of the Wasserstein metric in one dimension (Peyré & Cuturi, 2019, Sec. 2.6):

$$G_\rho(\delta) = \int_0^1 \left[x(u) - m_{b\sqrt{1+\delta}}(y_b(u))\right]^2 \mathrm{d}u. \tag{61}$$

This expression makes the comparison explicit: the loss matches each clean-signal quantile to the corresponding quantile of the reconstructed signal.

Low SNR analysis. As above, set $b = \sqrt{\rho}$ and write $z_\delta(u) := m_{b\sqrt{1+\delta}}(y_b(u))$. Equation (61) can be equivalently rewritten as $G_\rho(\delta) = \|x - z_\delta\|_2^2$, where $\|\cdot\|_2$ is the norm on $L^2(0,1)$. Note that the quantiles y_b depend on the true b but not on δ. Below, primes on z_δ denote derivatives with respect to δ.

By the chain rule, with derivatives of m_β taken with respect to β,

$$\begin{aligned} z'_0(u) &= \frac{b}{2}\ \partial_\beta m_\beta(y_b(u))|_{\beta=b}\,, \\ z''_0(u) &= \frac{b^2}{4}\ \partial^2_\beta m_\beta(y_b(u))\big|_{\beta=b} - \frac{b}{4}\ \partial_\beta m_\beta(y_b(u))|_{\beta=b}\,. \end{aligned}$$

Lemma 1 bounds the deviations from the linear terms and the second derivative by multiples of $1 + y^2$. Since y_b is the quantile function of $Y_\rho = N + bX$,

$$\int_0^1 (1 + y_b(u)^2)^2\,\mathrm{d}u = \mathbb{E}[(1 + Y_\rho^2)^2] = O(1)$$

as $b \to 0$, because X is bounded and N is Gaussian. Thus the three bounds in Lemma 1 give directly

$$\begin{aligned} \|z_0 - by_b\|_2 &= O(b^2), \\ \left\|z'_0 - \frac{b}{2}y_b\right\|_2 &= O(b^2), \\ \left\|z''_0 + \frac{b}{4}y_b\right\|_2 &= O(b^2). \end{aligned} \tag{62}$$

Differentiating the exact quantile identity in Eq. (61) yields

$$G'_\rho(0) = -2\int_0^1 (x - z_0)z'_0 \,\mathrm{d}u, \tag{63}$$

$$G''_\rho(0) = 2\int_0^1 (z'_0)^2 \,\mathrm{d}u - 2\int_0^1 (x - z_0)z''_0 \,\mathrm{d}u. \tag{64}$$

Differentiation under the integral is justified by the bounds in Lemma 1 and the finite fourth moment of Y_ρ. Since $\|x\|_2 = 1$ and $\|y_b\|_2^2 = \mathbb{E}[Y_\rho^2] = 1 + b^2$, Eq. (62) and Cauchy–Schwarz give

$$\begin{aligned} G'_\rho(0) &= -bC_b + O(b^2), \\ G''_\rho(0) &= \frac{b}{2}C_b + O(b^2), \end{aligned} \qquad C_b := \int_0^1 x(u)y_b(u) \,\mathrm{d}u. \tag{65}$$

The remaining issue is that C_b still depends on the observation SNR. We now identify a constant depending only on the clean law that approximates C_b with error $O(b)$. Let Φ denote the standard-normal cdf. Since $Y_\rho = N + bX$, the particular coupling (Y_ρ, N) and the one-dimensional Wasserstein quantile formula (Peyré & Cuturi, 2019, Sec. 2.6) imply

$$\begin{aligned} \|y_b - \Phi^{-1}\|_2 &= W_2(P_\rho, \mathcal{N}(0,1)) \\ &\le \left(\mathbb{E}|bX|^2\right)^{1/2} = b. \end{aligned} \tag{66}$$

Define the fixed cross-moment

$$\kappa_X := \int_0^1 x(u)\Phi^{-1}(u) \,\mathrm{d}u = \int_0^1 F_X^{-1}(u)\Phi^{-1}(u) \,\mathrm{d}u. \tag{67}$$

Because $\|x\|_2 = 1$, Cauchy–Schwarz now gives

$$|C_b - \kappa_X| \le \|x\|_2 \|y_b - \Phi^{-1}\|_2 \le b. \tag{68}$$

Thus $C_b = \kappa_X + O(b)$; its contribution to Eq. (65) has error $O(b^2)$. The two quantile functions in Eq. (67) are non-decreasing, centered and non-constant, so $\kappa_X > 0$; Cauchy–Schwarz gives $\kappa_X \le 1$, with equality if and only if X is standard Gaussian.

Proposition 1 (Clean-target sensitivity)**.** *Under the assumptions of Theorem 1, as $\rho \to 0$,*

$$G'_\rho(0) = -\kappa_X\sqrt{\rho} + O(\rho), \qquad G''_\rho(0) = \frac{\kappa_X}{2}\sqrt{\rho} + O(\rho). \tag{69}$$

Proof. Substitute Eq. (68) into Eq. (65) and use $b = \sqrt{\rho}$. □

E Experimental settings and additional results

This appendix details the experimental settings (Appendix E.1) and the operators (Appendix E.2), and gives additional results for each experiment, for the unbiased variant of the reconstruction loss (Appendix E.7), as well as the runtime (Appendix E.8).

E.1 Experimental settings

For each task, the measurement-only configuration is fixed first and reused for our method, with the reconstruction branch added at $\lambda = 1$. Teachers are conditional flow-matching velocity fields with 32 base channels, multipliers $(1, 2, 3, 4)$, one block per resolution, attention at 32×32; they are trained once per task on $8 \cdot 10^6$ images and then frozen. Auxiliary fields share this architecture and are initialized from the corresponding teachers.

Operator learning uses Adam with learning rate $3 \cdot 10^{-3}$ for the operator and 10^{-4} for the auxiliary fields, batch size 32, 64×64 patches, one auxiliary update per operator update, interpolation times drawn uniformly in $(0, 1)$, $K = 8$ DPIR iterations with the final $K_{\mathrm{bp}} = 1$ differentiated, $\epsilon = 10^{-12}$ in Eq. (8), and an operator EMA with $\beta = 0.999$. Table 3 lists the task-specific choices, and Appendix C gives further implementation details.

Table 3: **Per-task settings.** K_{warm} is the linear warm-up of λ in Eq. (29), expressed as a fraction of the training budget. Centering is used unless otherwise stated; see Appendix A.

	Uniform blur	Space-varying	Blind SR	Real capture
Clean set $\mathcal{X}$	FFHQ-256, 1k	DPDD, sharp	LR image	f/5.6 crops
Degraded set $\mathcal{Y}$	FFHQ-256, 1k	DIV2K + PSF	LR crops, 32^2	f/16 frames
Evaluation	FFHQ-256, 1k	DPDD test, 76	DIV2KRK, 100	—
Operator	33×33	8×8 grid, 27^2	11×11, $\times 2$	9×12 grid, 13^2
Channels	1 (3 for color)	3	1	3
Boundary	circ. / repl.	replicate	norm. conv.	replicate
Centering	off / on	on, green	on	on, green
Location conditioning	yes	yes	no	yes
Forward σ	$2 \cdot 10^{-2}$	10^{-2}	0	$2 \cdot 10^{-2}$
Restorer σ	$2 \cdot 10^{-2}$	10^{-2}	10^{-2}	$2 \cdot 10^{-2}$
λ / K_{warm}	1 / 1/4	1 / 1/4	1 / 1/2	1 / 1/4
Budget (images)	$4 \cdot 10^6$	$4 \cdot 10^6$	$5 \cdot 10^5$	$4 \cdot 10^6$
Operator lr schedule	cosine	cosine	cosine	cosine

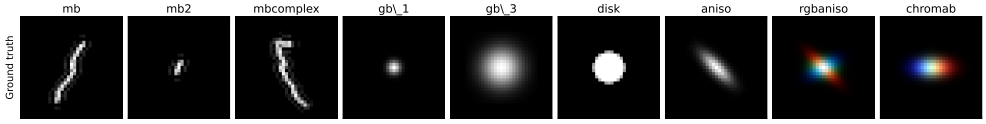

Figure 3: **Uniform degradations used in our experiments.** The set contains three motion kernels, four analytic PSFs, and two channel-dependent operators. Each kernel is shown on a square-root intensity scale and normalized to its own maximum. The final two are displayed as RGB composites to reveal differences between color channels.

Task-specific choices. The three motion-blur experiments use circular padding and disable recentering; all other experiments follow the centering convention of Appendix A. Blind super-resolution keeps centering, because the operator position selects a decimation phase, and applies the forward operator as a normalized convolution: the observation is zero-padded and then divided by the response of the same kernel to an all-ones mask, so that border pixels are renormalized by the support they actually received. Forward and restorer noise levels coincide for the deblurring experiments; blind super-resolution is noiseless, and DPIR uses 10^{-2} as a regularization level. Since super-resolution estimates one operator per image, it uses a shorter budget and a longer warm-up. In blind super-resolution, a Gaussian prior is applied to every method, ours included, through the post-processing of Section 5.3; no other kernel regularizer is used in any task.

E.2 Operators used in the experiments

This appendix visualizes the ground-truth operators used in Sections 5.1 and 5.2. Kernels are displayed on a square-root intensity scale and normalized to their own maximum to make low-amplitude structure visible.

Spatially uniform kernels Figure 3 shows the nine uniform degradations of Table 4, represented on the 33×33 support used for operator estimation.

The first three operators are irregular motion-blur kernels generated from random trajectories of increasing complexity and spatial extent. The analytic PSFs contain isotropic Gaussians with $\sigma \in \{1, 3\}$, a defocus disk of radius 5, and an anisotropic Gaussian with $\sigma = (3, 1)$ rotated by $45°$. Together they cover smooth isotropic blur, directional blur, defocus, and irregular motion. The motion kernels use circular padding and do not apply the centering correction; all other kernel experiments follow the convention of Appendix A.

The two color operators test channel-dependent degradation. The first uses a different anisotropic Gaussian in each channel while keeping their centers aligned. The second models lateral chromatic

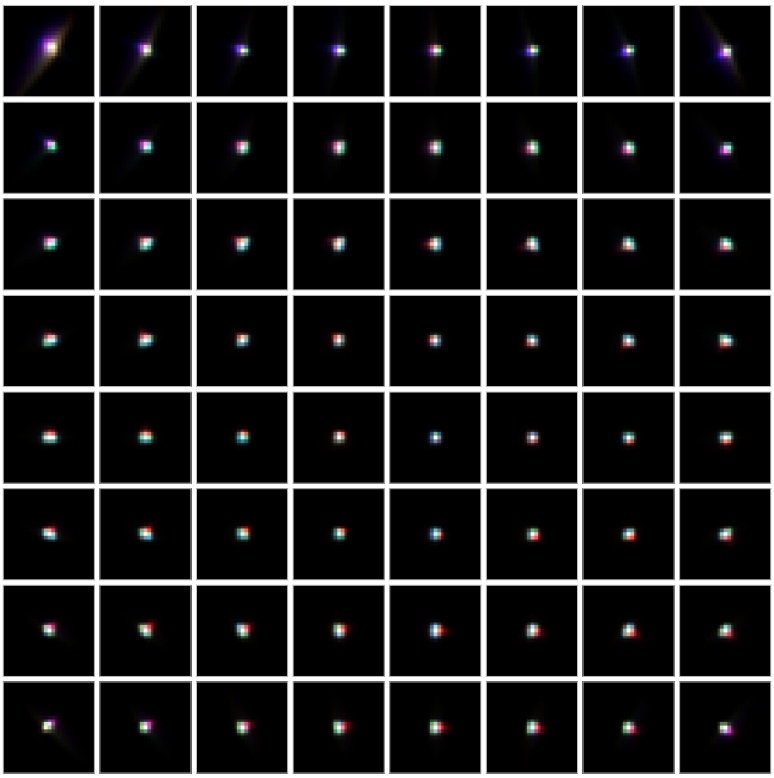

Figure 4: **Measured** 8×8 **PSF field** from a Canon EF 24 mm f/1.4L USM lens at f/1.4 (Bauer et al., 2018). Each tile shows the 27×27 RGB PSF at one field position and is normalized to its own maximum for visualization.

aberration by shifting the red and blue kernels by ± 2 pixels relative to green. We therefore use the green channel as the centering anchor so that the common translation is removed without eliminating the relative chromatic displacement; see Appendix A.

Measured space-varying PSF The operator of Section 5.2 is derived from the measured PSF field of a Canon EF 24 mm f/1.4L USM lens at full aperture released by Bauer et al. (2018). We sample the measured field on an 8×8 regular grid, crop each PSF to 55×55, resample it to 27×27, and normalize each channel to unit sum. The resulting grid is centered using the same green-anchored convention as the estimated operator.

E.3 Spatially uniform blur

The clean, degraded, and evaluation sets contain 1000 FFHQ-256 images each, drawn from disjoint partitions. Table 4 gives the per-degradation results summarized in Table 1, and Table 5 the corresponding SSIM and LPIPS.

E.4 Measured PSF field with chromatic aberration

The degraded set consists of DIV2K images blurred with the measured field of Appendix E.2 and corrupted by Gaussian noise with $\sigma = 10^{-2}$; the clean set consists of the sharp DPDD training images. Restoration is evaluated on the 76 DPDD test images at full resolution, degraded with the true field. Table 6 adds SSIM to the PSNR and LPIPS of Table 1: all metrics rank the methods identically.

Table 4: **Spatially uniform blur on FFHQ-256.** Kernel error: shift-aligned kernel MAE ($\times 10^{-4}$); restoration is RGB PSNR (dB) after DPIR with the estimated operator. *Input* is the degraded image and *True op.* uses the ground-truth operator.

	Kernel error ↓		Restoration PSNR ↑			
Degradation	Meas.-only	Ours	Input	Meas.-only	Ours	True op.
Motion `mb`	3.49	**1.74**	22.12	31.23	**31.63**	31.71
Motion `mb2`	**2.02**	2.94	28.76	34.12	**34.56**	34.57
Motion `mbcomplex`	3.67	**1.71**	21.68	31.64	**32.18**	32.19
Gaussian σ=1	1.74	**0.87**	30.01	33.19	**34.75**	34.85
Gaussian σ=3	**1.20**	1.25	24.98	**29.23**	29.21	29.52
Defocus disk	2.50	**2.16**	25.12	30.39	**30.76**	31.17
Anisotropic Gaussian	1.77	**1.47**	26.41	30.88	**31.05**	31.36
Per-channel anisotropic	1.71	**0.95**	26.91	31.87	**32.26**	32.35
Chromatic aberration	1.37	**0.99**	25.16	31.54	**31.77**	31.88
Mean	2.16	**1.56**	25.68	31.57	**32.02**	32.18

Table 5: **Spatially uniform blur: SSIM and LPIPS** at the last checkpoint.

	SSIM ↑			LPIPS ↓		
Degradation	Meas.-only	Ours	True op.	Meas.-only	Ours	True op.
Motion `mb`	0.8744	**0.8791**	0.8779	0.0867	**0.0771**	0.0789
Motion `mb2`	**0.9274**	0.9272	0.9267	0.0397	**0.0352**	0.0405
Motion `mbcomplex`	0.8845	**0.8915**	0.8915	0.0819	**0.0699**	0.0720
Gaussian σ=1	0.9249	**0.9292**	0.9281	0.0492	**0.0467**	0.0464
Gaussian σ=3	**0.8257**	0.8192	0.8272	**0.2061**	0.2197	0.2053
Defocus disk	**0.8551**	0.8537	0.8671	0.1435	**0.1287**	0.1281
Anisotropic Gaussian	**0.8737**	0.8686	0.8732	0.1339	**0.1325**	0.1332
Per-channel anisotropic	**0.8895**	0.8885	0.8903	**0.1145**	0.1172	0.1113
Chromatic aberration	**0.8815**	0.8797	0.8812	**0.1301**	0.1310	0.1284

E.5 BLIND SUPER-RESOLUTION

Protocol. For each image, the degraded set consists of 32×32 crops of the low-resolution image and the clean set of 64×64 crops of the same image, so that the objective matches the law of $A \cdot x_{64}$ to that of x_{32}, following the cross-scale recurrence assumption of Bell-Kligler et al. (2019). We estimate one 11×11 kernel per image, the native size of the DIV2KRK ground truth, apply it with the normalized-convolution boundary condition of Appendix E.1, and restore with DPIR, using the estimated kernel as an explicit $\times 2$ downsampling operator. We report PSNR on the luminance channel (PSNR-Y) and in RGB, SSIM, LPIPS, and the center-of-mass-aligned kernel MAE on a common 33×33 canvas, which is needed to compare methods that predict kernels of different supports. KernelGAN's kernels, stored on a 17×17 grid, are first cropped to 11×11, so that the zeroed border covers the same fraction of the support for every method. Paired comparisons use two-sided sign-flip permutation tests on the per-image differences (20,000 permutations).

Post-processing. The common Gaussian prior of Section 5.3 is applied identically to every estimated kernel before restoration: the kernel is centered, its outer one-pixel border is set to zero, it is renormalized, and it is replaced by the anisotropic Gaussian matching its first two moments. The border is cropped before the projection, since outer mass would otherwise bias the moments. The projection needs no tuning: blending the kernel with its projection ($\alpha = 0.75$) and replacing it entirely ($\alpha = 1$) are statistically indistinguishable ($+0.02$ dB, $p = 0.25$), and we use $\alpha = 1$. For the baselines, centering is a half-pixel correction of the kernel position. The position sets the decimation phase, so this correction matters for restoration while leaving the kernel shape untouched (Appendix A); applying it only to our own estimate would be an unfair advantage.

Full results Table 7 presents the results of every method with its raw kernel, and after post-processing. The centering correction alone is worth $+1.31$ dB to KernelGAN and $+0.43$ dB to

Table 6: **Measured PSF field: secondary restoration metrics** on the 76 DPDD test images.

Method	SSIM ↑	LPIPS ↓
Restormer (Zamir et al., 2022)	0.854	0.205
OmniLens (Jiang et al., 2026)	0.862	0.203
Measurement-only	0.838	0.207
Ours	**0.925**	**0.132**
True operator	0.939	0.111

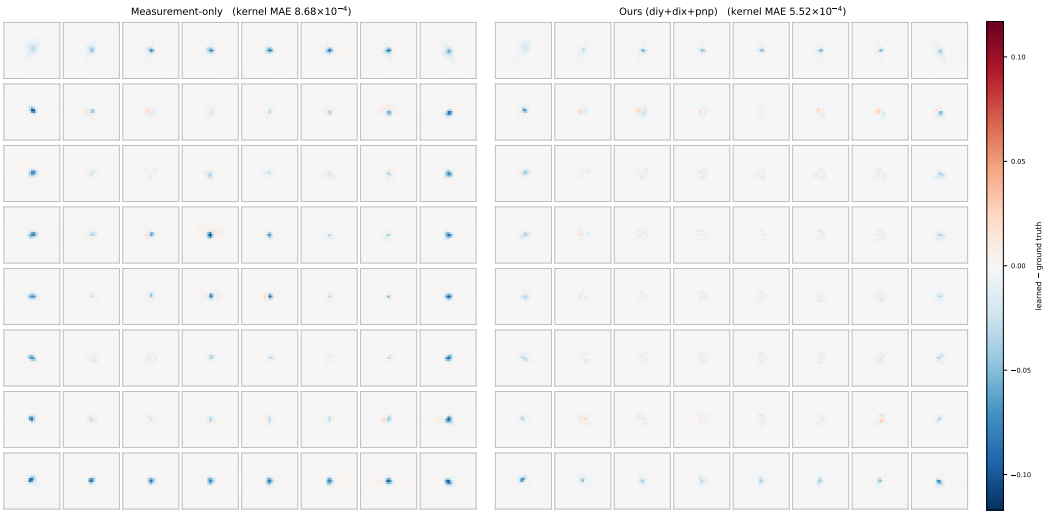

Figure 5: **Error of the recovered PSF field** (estimate − true PSF, each normalized to unit sum) at the 8×8 field positions, for measurement-only matching (left) and ours (right), on a shared color scale.

DIP-DKP (34.45 dB when centered only), while moving their aligned kernel error by at most 0.05. The Gaussian part of the post-processing is then strongly method-dependent and inversely related to how much each method already constrains its estimate: +0.61 dB for ours, +0.31 for KernelGAN, +0.14 for DCLS, and +0.09 for DIP-DKP (+1.73 for measurement-only matching). The post-processing never harms a method on average, which is why the main text reports post-processed numbers; without it, ours still has the best PSNR. For KernelGAN, centering and projection together add 1.62 dB in PSNR-Y (32.60 → 34.22); with restoration by DPIR instead of ZSSR, they explain why our KernelGAN results are higher than those reported by Bell-Kligler et al. (2019).

E.6 Proof of concept for real lens calibration

Acquisition. A single textured outdoor scene is photographed with a consumer camera (Panasonic Lumix GX9 with Leica DG Summilux 25mm f/1.4 II ASPH lens) at two apertures. The seven f/16 frames, which exhibit stronger diffraction blur, are used at full sensor resolution (5200×3904) and form $\mathcal{Y}$; the eight f/5.6 frames are center-cropped to 768×768 and form $\mathcal{X}$ (Figure 6). The camera is reframed between apertures, so that no image in one set corresponds to any image in the other, and the content of the two sets differs: the f/16 frames cover the whole field, while the f/5.6 crops are taken from its center. The camera is also rotated between shots, including upside-down, so that no orientation is systematically associated with either aperture.

Estimation and baselines. We estimate a 9×12 grid of 13×13 RGB kernels, matching the sensor aspect ratio. The kernels are treated as free parameters, without any lens or chromatic-aberration model, and are learned with the recipe of the other space-varying experiment and the values of Table 3; the noise level is treated as a fixed hyperparameter. Both baselines restore each f/16 frame on its own: the two-step method of Eboli et al. (2022), which estimates local Gaussian PSFs and

Table 7: **Blind SR on DIV2KRK ($\times 2$), full results.** Each method is shown with its raw kernel and after the post-processing of Section 5.3 (border zeroed, recentred, renormalized, then replaced by the anisotropic Gaussian matching its first two moments; for the baselines, this also includes the half-pixel centering correction). Kernel error is COM-aligned MAE ($\times 10^{-4}$) on a common 33×33 canvas. **Bold** marks the best entry in a column together with every entry not significantly different from it (two-sided paired permutation test, $p \geq 0.05$).

Method		Kernel err. ↓	PSNR-Y ↑	PSNR ↑	SSIM ↑	LPIPS ↓
KernelGAN	raw	3.50	32.60	30.89	0.8172	0.2769
	centered only	3.50	33.91	32.17	0.8438	0.2715
	post-processed	3.42	34.22	32.46	0.8481	0.2714
DCLS	raw	8.74	33.45	31.71	0.8111	0.3219
	post-processed	6.97	33.59	31.84	0.8160	0.3129
DIP-DKP	raw	3.02	34.02	32.28	0.8497	**0.2611**
	post-processed	2.57	34.54	32.77	**0.8575**	**0.2599**
Measurement-only	raw	5.43	32.22	30.52	0.7958	0.2931
	post-processed	2.96	33.95	32.20	0.8459	0.2677
Ours	raw	3.76	34.38	32.62	0.8486	0.2705
	post-processed	**2.15**	**34.99**	**33.21**	**0.8561**	**0.2601**
True operator		—	35.62	33.81	0.8724	0.2435

corrects lateral chromatic aberration, is run with its official code; OmniLens (Jiang et al., 2026), a network trained on the PSF fields of a large library of automatically designed lenses, is run with the released pretrained weights.

Evaluation. No paired sharp image or calibrated PSF is available for this capture, so we report neither PSNR nor kernel error. We also avoid reprojection consistency as a primary metric: in our tests, it can be small for operators that still visibly under-deblur the image. Figure 7 shows the recovered PSF fields. Figure 8 presents additional qualitative results on several crops.



Figure 6: **The fifteen images of the real capture.** Top: the seven f/16 frames forming $\mathcal{Y}$, at full sensor resolution (5200×3904). Bottom: the eight f/5.6 frames forming $\mathcal{X}$, center-cropped to 768^2. The camera is reframed and rotated between shots, so that the two sets share no image correspondence and no common orientation; images are shown as captured.

E.7 Unbiased reconstruction loss

Loss. As shown in Section 4, the clean-side loss G is biased: with a point-estimate restorer, restorations of real measurements are not distributed like clean images, even with the true operator. The bias can be removed by comparing restorations of real measurements with restorations of synthetic measurements generated by the candidate operator:

$$\widetilde{G}(\theta) := \mathcal{D}\big(Q_\theta \,\|\, Q_\theta^{\mathrm{syn}}\big), \qquad Q_\theta^{\mathrm{syn}} := \mathrm{Law}\left(R_\theta(A_\theta X + \varepsilon)\right), \quad X \sim P_X, \tag{70}$$

where $Q_\theta = \mathrm{Law}(R_\theta(Y))$ and $\mathcal{D}$ is the weighted IKL of Appendix B. Under the model of Eq. (1), $A_{\theta^\star} X + \varepsilon$ has the law of Y, so $Q_{\theta^\star} = Q_{\theta^\star}^{\mathrm{syn}}$ and $\widetilde{G}(\theta^\star) = 0$: since $\mathcal{D} \geq 0$, the true operator minimizes $\widetilde{G}$, whatever the restorer. The two losses differ only in their target: G compares restorations

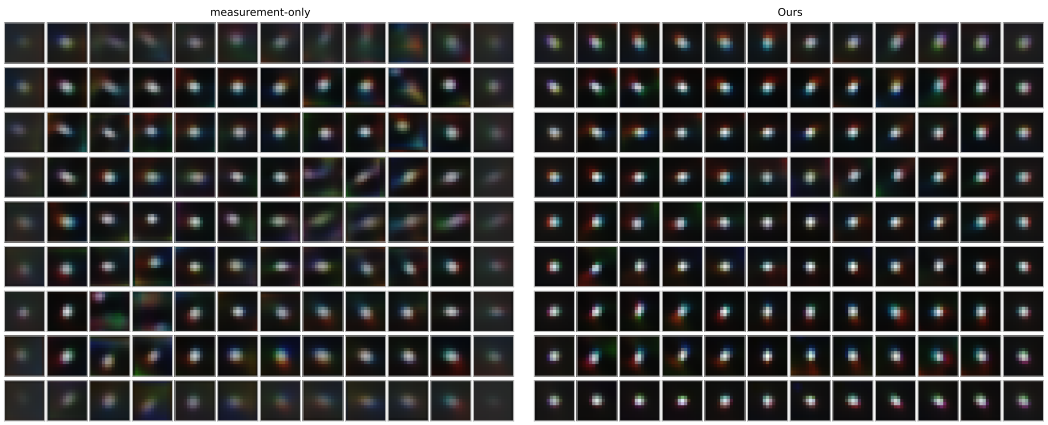

Figure 7: **Recovered PSF fields on the real capture.** Learned 9×12 grid of 13×13 RGB kernels after $4 \cdot 10^6$ training images, for measurement-only matching (left) and ours (right), laid out by field position. Each PSF is normalized to unit sum and displayed as an RGB composite, so a colored fringe indicates a channel-dependent shift.

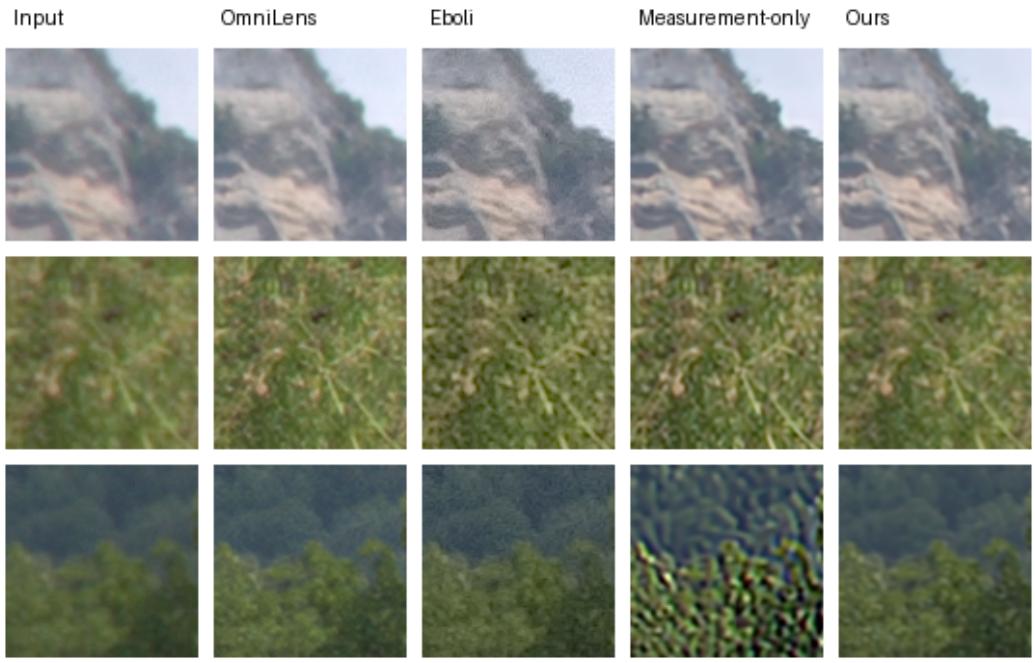

Figure 8: **Real lens calibration. Additional crops.** First row: note the green fringe between vegetation and sky. OmniLens does not remove it whereas other approaches do. On the last row, measurement-only catastrophically fails whereas other approaches perform well. Residual noise artefact are still present in the method of Eboli et al. (2022) despite manually trying many parameters.

of real measurements with clean images, whereas $\widetilde{G}$ compares them with clean images as rendered by the candidate degradation and the restorer. They coincide for a restorer that preserves the clean distribution, such as posterior sampling, for which $Q_\theta^{\mathrm{syn}} = P_X$ for every θ.

Implementation. We estimate the gradient of $\widetilde{G}$ with the estimator of Eq. (6), taking $Z_1(\theta) = R_\theta(Y)$ as for G. Since the target Q_θ^{syn} depends on θ, it cannot be represented by a frozen teacher: the clean-image teacher v_X^{tc} is replaced by a second auxiliary field, with the same architecture and

training as v_X^{aux}, trained online on detached restorations $R_\theta(A_\theta X+\varepsilon)$ and treated as fixed during the operator update. Each operator update therefore requires a second restoration and a second auxiliary-field update, which makes $\widetilde{G}$ more expensive than G. All other settings are those of Appendix E.1.

Table 8: **Unbiased loss $\widetilde{G}$ versus the default** G on the deblurring tasks, at the last checkpoint. The kernel-error change is relative to G; ΔPSNR is computed before rounding.

Degradation	Δ kernel err.	PSNR, G	PSNR, $\widetilde{G}$	ΔPSNR
Motion `mb`	$+19.5\%$	31.63	31.61	-0.02
Motion `mb2`	$+14.3\%$	34.56	34.58	$+0.02$
Motion `mbcomplex`	$+39.2\%$	32.18	32.25	$+0.08$
Gaussian σ=1	$+32.2\%$	34.75	34.33	-0.42
Gaussian σ=3	-35.9%	29.21	29.43	$+0.23$
Defocus disk	-2.0%	30.76	30.72	-0.04
Anisotropic Gaussian	-24.9%	31.05	31.24	$+0.19$
Per-channel anisotropic	$+9.9\%$	32.26	32.17	-0.09
Chromatic aberration	-10.7%	31.77	31.75	-0.02
Measured PSF field	-5.2%	32.41	32.35	-0.06
Mean	—	32.06	32.04	-0.01

Results. We rerun the nine uniform degradations and the measured PSF field with $\widetilde{G}$ in place of G, as well as blind super-resolution on the same 100 DIV2KRK images, and report the last checkpoint. On average, replacing G by $\widetilde{G}$ leaves restoration unchanged (-0.01 dB over the ten deblurring settings, better on four; Table 8), which supports the statement of Section 4 that the bias of G does not degrade performance in practice. Individual settings move in both directions, most visibly for the Gaussian-type kernels (-0.42 dB for $\sigma = 1$, $+0.23$ dB for $\sigma = 3$, $+0.19$ dB for the anisotropic Gaussian). Large changes in kernel error need not reach the restored image: on the three motion kernels, kernel MAE increases by 14 to 39% while PSNR changes by at most 0.08 dB. On blind super-resolution, paired over the 100 images, $\widetilde{G}$ improves restoration with raw kernels ($+0.10$ dB, median $+0.02$ dB, $56/100$ images, $p = 0.001$) but not after the post-processing of Section 5.3 (-0.04 dB, median -0.04 dB, $31/100$ images, $p = 0.06$), suggesting that both correct the same defect of the raw estimates. Since $\widetilde{G}$ brings no measurable benefit at a higher cost (roughly doubling the computational time), we keep G as the default.

E.8 Runtime Analysis

The reconstruction branch adds one solver call per operator update. On identical hardware, it increases the time spent in operator updates by $5.7\times$ to $6.6\times$ (Table 9). This cost is incurred only while estimating the operator; once learned, both methods produce the same type of explicit operator and therefore have identical restoration cost.

Table 9: **Cost of the reconstruction branch.** Mean time spent in the optimization steps per $131K$ training images (4096 operator updates at batch size 32), excluding data loading, validation and checkpointing, and discarding the first $131K$ images; all runs on identical GPUs.

Experiment	Meas.-only (s)	Ours (s)	Overhead
Uniform, `mb`	212	1401	$6.6\times$
Uniform, Gaussian σ=1	235	1492	$6.3\times$
Space-varying PSF	309	1762	$5.7\times$