Physics-Guided Conditional Flow Matching for Lensless Imaging

Qiwen Xiao* Edison Liu†
Department of Electrical Engineering and Computer Science (EECS)
University of California, Irvine
*qiwenx6@uci.edu †edisol1@uci.edu

Abstract—Lensless cameras encode scenes through a calibrated optical transfer function, producing highly multiplexed measurements that are typically ill-posed to invert. We study conditional flow matching (CFM) as a continuous-time conditional generative approach for reconstructing lensed images from diffuser-based lensless measurements. Our framework trains a conditional vector field via a simulation-free regression objective and generates reconstructions by integrating an ODE with a second-order Heun solver. We compare two practical parameterizations: direct velocity prediction (v-pred) and image prediction (x-pred) with a closed-form induced velocity. Experiments on the DiffuserCam MirFlickr benchmark show that CFM substantially outperforms an optimization-based ADMM solver and consistently improves over a supervised U-Net baseline, yielding reconstructions with sharper edges, more faithful textures, and fewer structured artifacts. Among the two parameterizations, x-prediction is consistently strongest, while an optional physics-guidance mechanism based on data-consistency updates provides modest additional gains, indicating that the conditional flow already captures much of the measurement structure while retaining a tunable lever to enforce fidelity at inference.

Index Terms—computational imaging, lensless imaging, diffuser camera, conditional flow matching, continuous normalizing flows, inverse problems

1 Introduction

Lensless cameras replace refractive optics with a coded element such as a diffuser, so the sensor records a highly multiplexed intensity pattern rather than a focused image. In diffuser-based systems like DiffuserCam, measurement formation is governed by a calibrated point spread function (PSF) together with practical non-idealities such as noise, sensor effects, and mild calibration mismatch [1]. Recovering a visually plausible lensed image from a single lensless measurement is therefore a challenging inverse problem: spatial information is entangled by the PSF, and small modeling errors can lead to structured artifacts.

Traditional lensless reconstruction approaches combine an explicit forward model with hand-designed priors and iterative optimization [2]. Such methods provide an interpretable way to enforce measurement consistency, but they often require careful tuning and can degrade under mismatch. Supervised neural networks offer fast amortized inference by learning a direct mapping from measurements to images; however, a single deterministic predictor may oversmooth fine details or introduce texture bias when the inverse problem is highly ill-posed.

This paper explores a conditional generative alternative based on Conditional Flow Matching (CFM). Continuous-time generative models are attractive in computational imaging because they define a transport from noise to data via an ODE, exposing a direct quality–compute trade-off through the choice of numerical solver and step budget. While classical continuous normalizing flows are commonly trained via likelihood objectives that require costly numerical simulation, Flow Matching (FM) replaces likelihood training with a simulation-free regression objective that matches a time-dependent vector field along a prescribed probability path [3]. Conditional Flow Matching (CFM) extends this framework to conditional generation by learning a vector field conditioned on an observation, enabling deterministic conditional sampling by ODE integration [4]. For lensless reconstruction, this provides a practical middle ground: we retain the flexibility of a generative formulation while keeping training stable and inference controllable.

Beyond the learned conditional flow, the lensless forward model can still be leveraged at inference time. Prior work on generative inverse problems shows that sampling trajectories can be guided by the measurement model to improve data fidelity [5]. In our setting, we study a lightweight physics-guidance mechanism based on data-consistency updates using the calibrated PSF. Importantly, our experiments indicate that the learned conditional flow already captures much of the measurement structure, so guidance acts primarily as an optional refinement that yields modest gains under conservative settings.

In summary, this work investigates physics-guided CFM for lensless-to-lensed reconstruction. Our contributions can be summarized as follows:

- To the best of our knowledge, we are the first to formulate lensless reconstruction as conditional generation with CFM and study its empirical advantages over optimization-based and supervised baselines.
- We compare two practical parameterizations, velocity prediction and image prediction with induced velocity, under a unified training and sampling pipeline.
- We analyze the inference-time trade-offs, including ODE step budget and the incremental effect of op-

tional physics guidance.[1]

2 Related Work

2.1 Lensless imaging and computational reconstruction

Diffuser- and mask-based lensless cameras trade optics for computation, where a scene is mapped to a highly multiplexed measurement through a PSF forward model, making inversion both ill-posed and sensitive to calibration and sensor non-idealities [1], [6], [7]. Classical reconstructions therefore rely on iterative optimization to combine measurement fidelity with hand-crafted priors, yielding strong and interpretable baselines but requiring careful tuning and typically producing a single point estimate [8], [9]. Learned approaches improve speed and quality by training a direct or unrolled mapping from measurements to images, but still commonly behave as deterministic predictors that can be brittle under mismatch or multi-modality [2], [10], [11], [12], [13]. Recent diffusion-prior lensless methods underscore a key lesson: pairing a physical forward model with a powerful generative prior can substantially improve reconstructions and mitigate artifacts, particularly when the inverse is ambiguous [5], [9], [14], [15], [16], [17], [18]. Motivated by this trajectory, we target a conditional *distribution* rather than a single regressor, while keeping the forward model central for evaluation and optional data-consistency refinement.

2.2 Flow matching and conditional flow matching

Continuous-time generative models represent sampling as integrating an ODE or SDE, providing a natural handle on step budgets and numerical stability [19], [20], [21]. Flow Matching (FM) makes such ODE-defined generators practical by replacing likelihood-based CNF training with a simulation-free regression objective on a prescribed probability path, often improving robustness and efficiency [3]. Conditional Flow Matching (CFM) extends FM to conditional generation by learning a time-dependent vector field conditioned on observations, enabling deterministic conditional sampling via ODE integration and offering a direct route to model target distribution for inverse problems [4]. Recent refinements further strengthen training stability and the quality–efficiency trade-off [22], [23]. Finally, diffusion-based posterior sampling for inverse problems illustrates how to incorporate measurement likelihoods as guidance during generative sampling [5], [18]. Our method adopts CFM as the core conditional generator and uses the known lensless forward operator as physics guidance, bridging continuous-time conditional transport with computational imaging constraints.

3 Methodology

3.1 Problem setup

We consider a paired dataset $\mathcal{D} = \{(y_i, x_i)\}_{i=1}^{N}$, where $y \in \mathbb{R}^{C\times H\times W}$ denotes a lensless measurement and $x \in \mathbb{R}^{C\times H\times W}$ denotes the corresponding lensed ground-truth image. A calibrated forward operator H_ϕ maps an image to a predicted measurement,

$$y \approx H_\phi(x), \tag{1}$$

where ϕ denotes calibration parameters such as the diffuser PSF. Our goal is to learn a conditional generative model $p_\theta(x \mid y)$ capable of producing high-fidelity reconstructions conditioned on a given measurement y.

3.2 Conditional flow matching objective

We adopt CFM to learn a continuous-time conditional vector field whose induced ODE defines a conditional generative process [3], [4]. CFM constructs a probability path that interpolates between a tractable reference distribution and the data distribution while conditioning on y. Concretely, we use a straight path between Gaussian noise and the data sample. For each training pair (x, y), we draw

$$x_1 \sim \mathcal{N}(0, \sigma_0^2 I), \tag{2}$$

and sample $t \sim \mathcal{U}[t_{\min}, t_{\max}]$. The intermediate state is defined by

$$x_t = (1-t)\, x_1 + t\, x. \tag{3}$$

The linear path (3) induces a closed-form target velocity field given by the path derivative,

$$v^\star(x_t, t; x, x_1) = \frac{d}{dt} x_t = x - x_1, \tag{4}$$

which is constant in t. We train a neural conditional vector field $v_\theta(x_t, y, t)$ by minimizing the regression loss

$$\mathcal{L}_{\mathrm{CFM}}(\theta) = \mathbb{E}_{(x,y)\sim\mathcal{D}}\mathbb{E}\left[\|v_\theta(x_t, y, t) - (x - x_1)\|_2^2\right]. \tag{5}$$

A key practical benefit of this formulation is that learning reduces to supervised regression and does not require differentiating through numerical ODE solvers during training [3], [4].

3.3 Parameterizations: v-prediction and x-prediction

Directly predicting clean data can be advantageous because natural images concentrate on a low-dimensional manifold, whereas noised targets are off-manifold and can be harder to model in high-dimensional pixel spaces [24]. Therefore, We consider two parameterizations that share the same conditioning mechanism and backbone architecture but differ in the predicted quantity and the induced velocity computation.

3.3.1 Velocity-field prediction (v-prediction)

In the v-prediction parameterization, the network directly outputs the conditional velocity field:

$$v_\theta(x_t, y, t) = f_\theta(x_t, y, t), \tag{6}$$

where f_θ is a conditional U-Net-like backbone that takes as input the noisy state x_t, the measurement condition y, and continuous time t.

1. Our code is publicly available at github.com/Charley-xiao/lensless-flow.

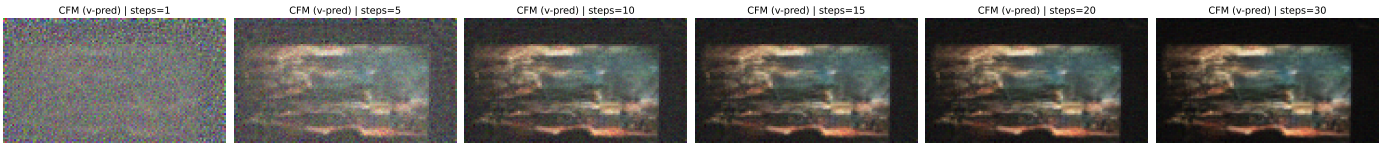

Fig. 1. Evolution of a v-prediction CFM reconstruction as the ODE step budget increases. Starting from near-Gaussian noise at very small step counts, the solver quickly recovers the global structure (5–10 steps) and then refines contrast and texture with diminishing returns beyond ∼15–20 steps.

3.3.2 *Image prediction (x-prediction)*

In the x-prediction parameterization, the network predicts a clean image estimate $\hat{x}_\theta$:

$$\hat{x}_\theta = f_\theta(x_t, y, t). \tag{7}$$

To remain consistent with the linear path (3), we convert $\hat{x}_\theta$ into an induced velocity using the corresponding closed-form relationship. From (3), one obtains

$$x_1 = \frac{x_t - tx}{1-t}, \qquad v^\star = x - x_1 = \frac{x - x_t}{1-t}. \tag{8}$$

Accordingly, we define the model velocity as

$$v_\theta(x_t, y, t) = \frac{\hat{x}_\theta - x_t}{\max(1-t,\, \delta)}, \tag{9}$$

where δ is a small positive constant used to clamp the denominator and avoid numerical instability as $t \to 1$.

3.4 Physics-guided sampling

At test time, given a measurement y, we sample an initial state

$$z_0 \sim \mathcal{N}(0, \sigma_0^2 I), \tag{10}$$

and generate a reconstruction by integrating the conditional ODE forward in time:

$$\frac{dz}{dt} = v_\theta(z, y, t), \quad t \in [0, 1]. \tag{11}$$

We employ Heun's method (second-order Runge–Kutta) with K uniform steps. Let $\Delta t = 1/K$ and $t_k = k\Delta t$. The solver update is

$$k_1 = v_\theta(z_k, y, t_k), \tag{12}$$
$$\tilde{z} = z_k + \Delta t\, k_1, \tag{13}$$
$$k_2 = v_\theta(\tilde{z}, y, t_{k+1}), \tag{14}$$
$$z_{k+1} = z_k + \frac{\Delta t}{2}(k_1 + k_2). \tag{15}$$

For v-prediction, v_θ is given directly by the network output; for x-prediction, v_θ is computed via (9). This deterministic ODE-based sampling provides a stable and efficient mechanism for conditional generation under a fixed step budget.

To encourage agreement with the observed measurement, we optionally incorporate a data-consistency (DC) refinement after each numerical integration step. We define the measurement-domain objective

$$\ell_{\mathrm{dc}}(z; y) = \|H_\phi(z) - y\|_2^2. \tag{16}$$

A single DC refinement applies a gradient step on ℓ_{dc}:

$$z \leftarrow z - \eta \nabla_z \ell_{\mathrm{dc}}(z; y) = z - 2\eta\, H_\phi^\top \big(H_\phi(z) - y\big), \tag{17}$$

where η controls the guidance strength; multiple refinement iterations may be applied per ODE step.

When an explicit η is not provided, we set a conservative default based on the spectral norm of $H^\top H$ for FFT-diagonalizable convolution:

$$\eta = \frac{\gamma}{2(L + \varepsilon)}, \qquad L = \max_\omega |\mathrm{OTF}(\omega)|^2, \tag{18}$$

where $\gamma \in (0, 1)$ is a safety factor and ε is a small constant for numerical stability. This choice controls the DC update magnitude while preserving compatibility with the ODE-based sampling procedure.

4 Experimental Setup

We examine how well our approaches perform in lensless image reconstruction compared to existing methods.

4.1 Dataset and preprocessing

The dataset consists of paired lensless and lensed images derived from the DiffuserCam Mirflickr benchmark using publicly available processing tools [1]. All pixel intensities were normalized to the range $[0, 1]$. To manage computational complexity, the images were downsampled by a factor of 4, resulting in a final dimensionality of $3 \times 67 \times 120$ $(C \times H \times W)$. The dataset was partitioned into a training set of 24,000 images and a test set of 999 images.

4.2 Methods compared

We evaluate two CFM variants: a v-prediction model and an x-prediction model. Each variant is tested in two inference configurations: (i) standard ODE sampling without physics guidance and (ii) ODE sampling augmented with data-consistency refinement. As baselines, we include an ADMM-based reconstruction and a supervised U-Net trained to regress the lensed image directly from the lensless measurement [25]. The U-Net uses GroupNorm+SiLU residual blocks with 3×3 convolutions, strided-convolution downsampling and transposed-convolution upsampling, and skip connections between matching resolutions. The channel schedule follows `base_channels`$=$ 32 and `channel_mults`$=$ $[1, 2, 4, 8]$ with `num_res_blocks`$=$ 2 per resolution. For RGB inputs ($C = 3$), this architecture contains approximately 14.9M parameters.

4.3 Metrics

To evaluate reconstruction quality, image-domain fidelity is assessed using Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM) [26]. Both

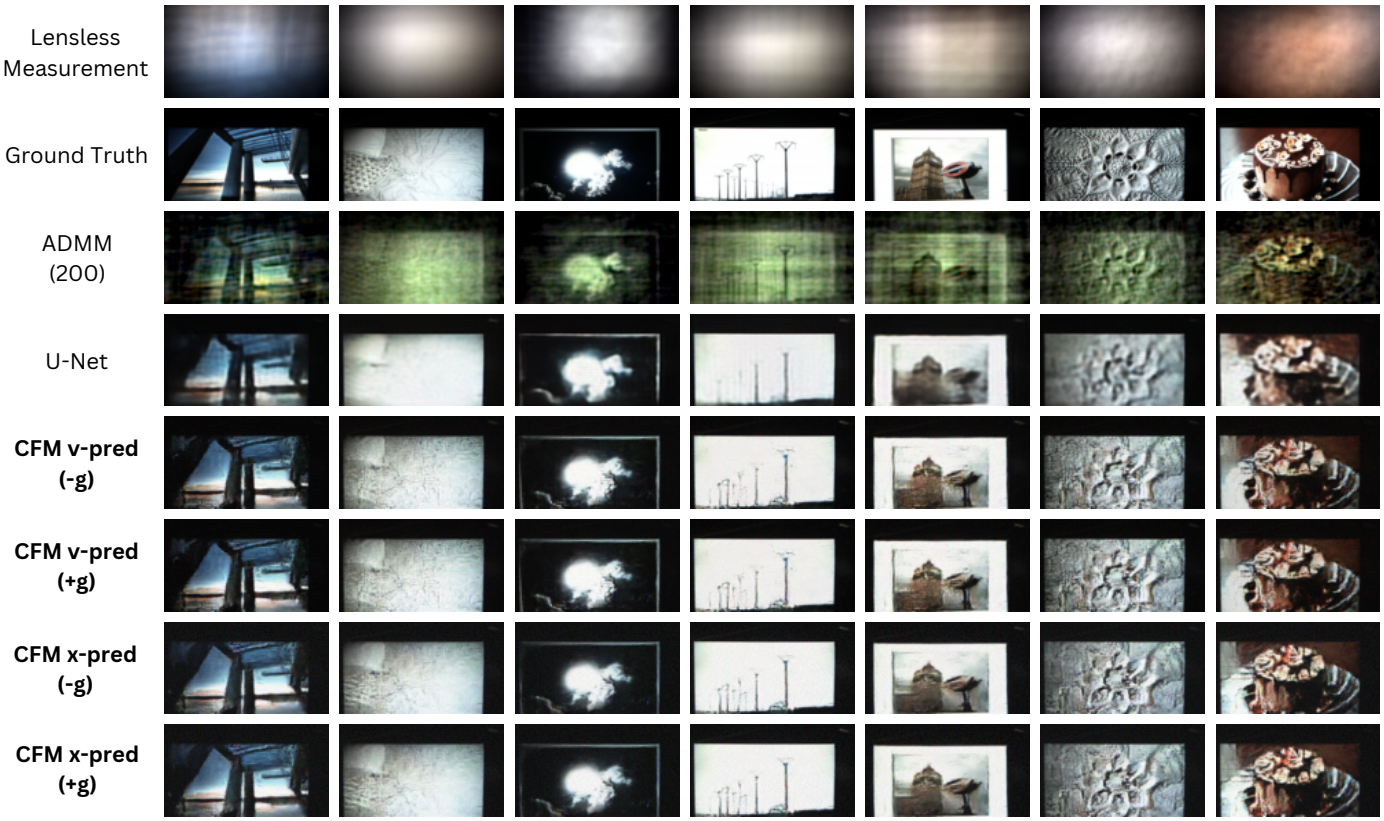

Fig. 2. Qualitative comparison on representative test examples. From top to bottom: lensless measurements, ground truth, ADMM (200 iterations), supervised U-Net baseline, and CFM variants (v-pred/x-pred) without guidance (-g) and with physics guidance (+g).

TABLE 1
Quantitative reconstruction performance on the test set. SSIM and PSNR are higher-is-better, while MSE is lower-is-better. Best values per metric are shown in **bold underline**.

Method	SSIM ↑	PSNR (dB) ↑	MSE ↓
ADMM	0.15	7.26	0.1879
U-Net	0.76	20.67	0.0086
CFM v-pred (-g)	0.79	22.26	0.0062
CFM v-pred (+g)	0.80	22.27	0.0062
CFM x-pred (-g)	**0.81**	23.00	0.0053
CFM x-pred (+g)	**0.81**	**23.01**	**0.0051**

metrics are computed across all color channels and averaged to produce a single scalar value per image.

The SSIM is implemented using a Gaussian sliding window of size 11×11 with a standard deviation $\sigma = 1.5$. We set the stability constants to $K_1 = 0.01$ and $K_2 = 0.03$ with a dynamic range $L = 1$. For PSNR, we assume a peak signal value of 1.0, consistent with our data normalization.

For completeness, we also report the (image-domain) mean squared error

$$\mathrm{MSE}(x, \hat{x}) = \frac{1}{CHW} \sum_{c=1}^{C} \sum_{h=1}^{H} \sum_{w=1}^{W} \left(x_{c,h,w} - \hat{x}_{c,h,w}\right)^2, \quad (19)$$

averaged over the dataset.

5 Results and Analysis

Table 1 summarizes the quantitative results, and Figs. 1–2 provide qualitative evidence. Overall, conditional flow matching (CFM) outperforms both ADMM and the supervised U-Net baseline. Among CFM variants, x-prediction achieves the best fidelity, while physics guidance yields small but consistent improvements in measurement agreement.

5.1 Quantitative comparison

ADMM performs poorly (SSIM 0.15, PSNR 7.26), reflecting the difficulty of single-shot diffuser inversion and the sensitivity of iterative solvers to model mismatch. The supervised U-Net provides a strong learned baseline (SSIM 0.76, PSNR 20.67, MSE 8.6×10^{-3}), but both CFM variants improve further. V-prediction reaches SSIM 0.79 and PSNR 22.26 (MSE 6.2×10^{-3}), while x-prediction attains the best overall scores (SSIM 0.81, PSNR 23.00, MSE 5.3×10^{-3}). With guidance, x-prediction achieves the lowest MSE (5.1×10^{-3}) and highest PSNR (23.01), indicating a modest but measurable gain from data consistency.

5.2 Qualitative analysis and artifact characterization

Figure 2 shows clear qualitative differences. ADMM exhibits strong structured artifacts and noticeable color bias as a low-frequency tint, consistent with reconstructions being dominated by the PSF imprint under weak priors. The U-Net recovers coarse scene layout but tends to oversmooth fine

textures and soften edges, a typical regression-to-the-mean behavior for deterministic predictors in ambiguous inverse problems. In contrast, both CFM variants produce sharper edges, improved contrast, and more faithful textures, with substantially fewer PSF-shaped artifacts. Across the shown examples, x-prediction is visually slightly cleaner than v-prediction, particularly in thin structures and textured regions, matching the metric improvements in Table 1.

5.3 Effect of ODE steps and guidance

Figure 1 illustrates the quality–compute trade-off. With very few steps, the result remains noise-like. By 5 steps, the reconstruction already captures global structure and plausible colors, and 10–15 steps refine geometry and contrast substantially. Beyond $\sim$20 steps, improvements are incremental, suggesting that most quality is achieved within a moderate step budget.

Physics guidance has a marginal effect on SSIM and PSNR, with slightly clearer benefits in MSE for x-prediction. This is consistent with the fact that the model is trained and sampled *conditionally* on y, so much of the measurement constraint is already embedded in the learned flow; conservative guidance then acts as a small corrective step rather than a dominant force. The small gap is also desirable from a robustness perspective: strong guidance could amplify calibration mismatch and reintroduce structured artifacts.

5.4 Takeaways

CFM provides the best reconstructions among all compared approaches, improving both perceptual quality and standard fidelity metrics over U-Net and far outperforming ADMM. X-prediction is consistently strongest, and the observed denoising trajectory indicates rapid convergence within 10–15 Heun steps. Physics guidance yields small but consistent improvements under conservative settings, suggesting that the learned conditional flow already captures most of the forward-model structure while retaining the option to enforce data consistency at inference.

6 Conclusion

We presented a physics-guided conditional flow matching framework for lensless-to-lensed image reconstruction. The method learns a continuous-time conditional generative model via regression-based CFM training and reconstructs images by solving the resulting ODE with a second-order Heun integrator. In experiments on the DiffuserCam MirFlickr benchmark, both CFM parameterizations outperform classical optimization and supervised regression baselines: compared to ADMM, CFM removes strong PSF-shaped artifacts and yields substantially higher fidelity, and compared to a supervised U-Net, CFM improves PSNR and SSIM while producing sharper edges and more faithful textures. Among variants, x-prediction is consistently best, up to PSNR 23.01, SSIM 0.81, indicating that predicting clean images and inducing velocities can be more stable than directly regressing velocities in this setting. We also observed that reconstruction quality saturates after a moderate number of ODE steps, offering a practical quality–compute trade-off, and that physics guidance yields small but consistent improvements under conservative settings.

Limitations include reliance on accurate calibration of the forward model and evaluation on a single dataset and downsampling configuration. Future work will study stronger and adaptively scheduled guidance and extend the approach to higher-resolution reconstructions and more diverse lensless hardware.

Roles of each teammate in final project

Qiwen Xiao proposed the use of CFM for diffuser-based lensless reconstruction, designed the overall pipeline, implemented the CFM training and sampling code, and ran the CFM experiments. He also produced key figures related to the flow sampling trajectory and contributed to the analysis and write-up of the method and results.

Edison Liu was responsible for establishing and validating strong baselines. He implemented and trained the ADMM reconstruction baseline and the supervised U-Net baseline, tuned their hyperparameters for the DiffuserCam MirFlickr setting, and generated the corresponding quantitative and qualitative results used for comparison. He also contributed to experiment execution and helped interpret baseline behaviors in the results and analysis section.

References

[1] N. Antipa, G. Kuo, R. Heckel, B. Mildenhall, E. Bostan, R. Ng, and L. Waller, "Diffusercam: lensless single-exposure 3d imaging," *Optica*, vol. 5, no. 1, pp. 1–9, 2018.

[2] K. Monakhova, K. Yanny, N. Aggarwal, and L. Waller, "Learned reconstructions for practical mask-based lensless imaging," *Optics Express*, vol. 27, no. 20, pp. 28 075–28 090, 2019.

[3] Y. Lipman, R. T. Q. Chen, H. Ben-Hamu, M. Nickel, and M. Levine, "Flow matching for generative modeling," *arXiv preprint arXiv:2210.02747*, 2022.

[4] A. Tong, K. Fatras, N. Malkin, G. Huguet, Y. Zhang, J. Rector-Brooks, G. Wolf, and Y. Bengio, "Improving and generalizing flow-based generative models with minibatch optimal transport," *arXiv preprint arXiv:2302.00482*, 2023.

[5] H. Chung, J. Kim, M. T. McCann, M. L. Klasky, and J. C. Ye, "Diffusion posterior sampling for general noisy inverse problems," *arXiv preprint arXiv:2209.14687*, 2022.

[6] M. S. Asif, A. Ayremlou, A. Sankaranarayanan, A. Veeraraghavan, and R. Baraniuk, "Flatcam: Thin, lensless cameras using coded aperture and computation," *IEEE Transactions on Computational Imaging*, vol. 3, no. 3, pp. 384–397, 2017.

[7] V. Boominathan, J. K. Adams, J. T. Robinson, and A. Veeraraghavan, "Phlatcam: Designed phase-mask based thin lensless camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 7, pp. 1618–1629, 2020.

[8] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[9] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *IEEE GlobalSIP*, 2013, pp. 945–948.

[10] P. Kingshott *et al.*, "Unrolled primal-dual networks for lensless cameras," *Optics Express*, vol. 30, no. 26, pp. 46 324–46 340, 2022.

[11] T. Zeng and E. Y. Lam, "Robust reconstruction with deep learning to handle model mismatch in lensless imaging," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 1080–1092, 2021.

[12] H. Qian *et al.*, "Robust unrolled network for lensless imaging with enhanced resistance to model mismatch and noise," *Optics Express*, vol. 32, no. 17, pp. 30 267–30 283, 2024.

[13] Y. Zheng *et al.*, "A simple framework for 3d lensless imaging with programmable masks," in *ICCV*, 2021.

[14] W. Wan *et al.*, "Multi-phase fza lensless imaging via diffusion model," *Optics Express*, vol. 31, no. 12, pp. 20 595–20 612, 2023.

[15] X. Cai, Z. You, H. Zhang, W. Liu, J. Gu, and T. Xue, "Phocolens: Photorealistic and consistent reconstruction in lensless imaging," in *NeurIPS*, 2024.

[16] D. Xiao *et al.*, "Stablecam: Lensless imaging with stable diffusion model-based reconstruction," *Optik*, 2025.
[17] E. Yosef *et al.*, "Difuzcam: replacing camera lens with a mask and diffusion priors for reconstruction," *Scientific Reports*, 2025.
[18] B. Kawar, M. Elad, S. Ermon, and J. Song, "Denoising diffusion restoration models," *arXiv preprint arXiv:2201.11793*, 2022.
[19] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud, "Neural ordinary differential equations," *arXiv preprint arXiv:1806.07366*, 2018.
[20] W. Grathwohl, R. T. Q. Chen, J. Bettencourt, I. Sutskever, and D. Duvenaud, "Ffjord: Free-form continuous dynamics for scalable reversible generative models," in *ICLR*, 2019.
[21] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," *arXiv preprint arXiv:2011.13456*, 2020.
[22] A.-A. Pooladian, H. Ben-Hamu, C. Domingo-Enrich, B. Amos, Y. Lipman, and R. T. Q. Chen, "Multisample flow matching: Straightening flows with minibatch couplings," *arXiv preprint arXiv:2304.14772*, 2023.
[23] A. Tong, N. Malkin, K. Fatras, L. Atanackovic, Y. Zhang, G. Huguet, G. Wolf, and Y. Bengio, "Simulation-free schrödinger bridges via score and flow matching," *arXiv preprint arXiv:2307.03672*, 2023.
[24] T. Li and K. He, "Back to basics: Let denoising generative models denoise," 2026. [Online]. Available: https://arxiv.org/abs/2511.13720
[25] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015. [Online]. Available: https://arxiv.org/abs/1505.04597
[26] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.