PhoCoLens: Photorealistic and Consistent Reconstruction in Lensless Imaging

Xin Cai[1,2], **Zhiyuan You**[1], **Hailong Zhang**[3], **Wentao Liu**[2,4], **Jinwei Gu**[1], **Tianfan Xue**[1,2]

[1]The Chinese University of Hong Kong, [2]Shanghai Artificial Intelligence Laboratory,
[3]Tsinghua University, [4]SenseTime

`{cx023, yz023, tfxue}@ie.cuhk.edu.hk, jwgu@cuhk.edu.hk,`
`{zhanghl21}@mails.tsinghua.edu.cn, liuwentao@sensetime.com`

Abstract

Lensless cameras offer significant advantages in size, weight, and cost compared to traditional lens-based systems. Without a focusing lens, lensless cameras rely on computational algorithms to recover the scenes from multiplexed measurements. However, current algorithms struggle with inaccurate forward imaging models and insufficient priors to reconstruct high-quality images. To overcome these limitations, we introduce a novel two-stage approach for consistent and photorealistic lensless image reconstruction. The first stage of our approach ensures data consistency by focusing on accurately reconstructing the low-frequency content with a spatially varying deconvolution method that adjusts to changes in the Point Spread Function (PSF) across the camera's field of view. The second stage enhances photorealism by incorporating a generative prior from pre-trained diffusion models. By conditioning on the low-frequency content retrieved in the first stage, the diffusion model effectively reconstructs the high-frequency details that are typically lost in the lensless imaging process, while also maintaining image fidelity. Our method achieves a superior balance between data fidelity and visual quality compared to existing methods, as demonstrated with two popular lensless systems, PhlatCam and DiffuserCam. Project website: `phocolens.github.io`.

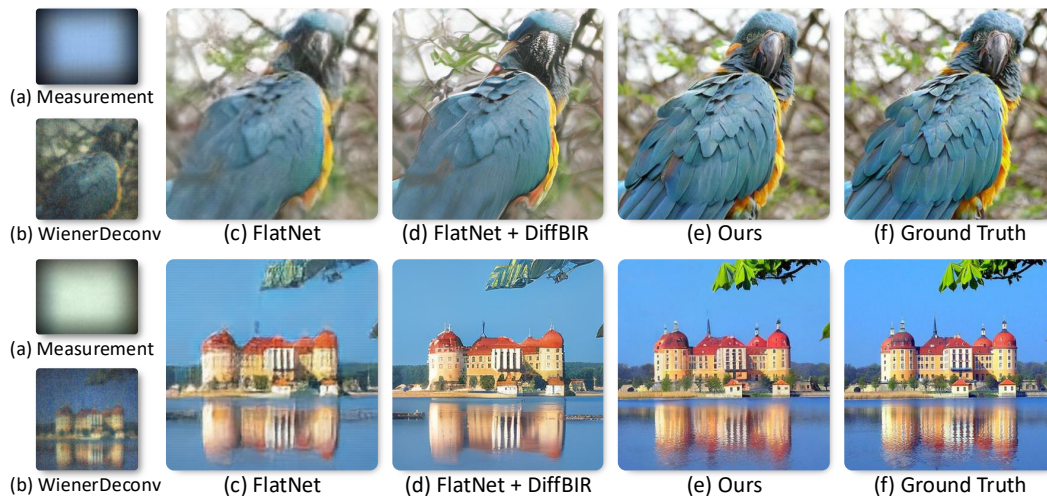



Figure 1: We introduce PhoCoLens, a lensless reconstruction algorithm that achieves both better visual quality and consistency to the ground truth than existing methods. Our method recovers more details compared to traditional reconstruction algorithms (b) and (c), and also maintains better fidelity to the ground truth compared to the generative approach (d).

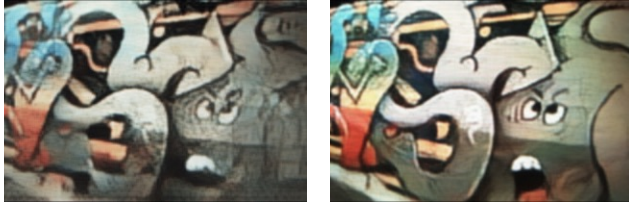
38th Conference on Neural Information Processing Systems (NeurIPS 2024).

w/o spatially varying deconv   with spatially varying deconv

Figure 2: Comparison between results without (left) and with (right) our spatially varying deconvolution.

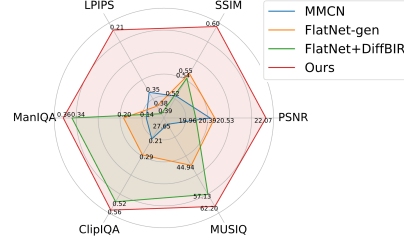



Figure 3: Evaluation of fidelity and authenticity on PhlatCam [17] dataset.

1 Introduction

Lensless imaging systems [6] have emerged as a groundbreaking solution for ultra-compact, lightweight, and cost-effective imaging. They replace lenses by amplitude [3, 4] or phase masks [1, 5, 20] placed close to the sensor to modulate incoming light. This design significantly reduces camera size and weight and enables innovative sensor shapes such as spherical [14] or cylindrical.

Considering the raw measurements from lensless cameras are typically blurry and unrecognizable, it is hard to recover a high-quality image of the original scene. An example of lensless measurement and recovered images by WienerDeconv [44] is shown in the left column of Fig. 1. Due to the lack of focusing elements, lensless cameras cannot directly record the scene but encode it into a complex diffraction pattern. This encoding process can be approximated as a convolution with a large Point Spread Function (PSF). The PSF acts like a low-pass filter applied to the scene thus introducing ambiguity, which means there could be multiple possible recoveries for a single measurement. Therefore, powerful algorithms are critical for high-quality reconstruction in lensless systems.

The primary challenge of a lensless reconstruction is to achieve both photorealism and consistency. Photorealism requires high-quality reconstruction with rich details, while consistency further requires reconstructed contents to be consistent with the original scene. Traditional techniques, like WienerDeconv, can reconstruct images that align with the ground truth but visual quality is significantly degraded (Fig. 1b). Learning-based approaches such as FlatNet-gen [17] attempt to enhance visual quality by training with paired images and lensless measurements, yet they often fail to recover high-frequency details (Fig. 1c). The visual quality can be further improved by the restoration algorithms using generative priors, like DiffBIR [23] (Fig. 1d). While these restoration methods can inject rich details, they may also alter image content or insert non-existent objects, breaking consistency. For example, in Fig. 1d, the beak in the top row has the wrong shape compared to the ground truth (Fig. 1f), and the leaves in the bottom row look fake.

Moreover, an inaccurate imaging process simulation may also hurt data consistency. Most existing reconstruction algorithms [17, 27] simplify the imaging process as a convolution with a shift-invariant PSF. However, in practice, PSFs are spatially varying, particularly when the angle of incidence increases. Consequently, these areas experience a noticeable drop in the reconstructed similarity to the original scene, especially in the peripheral field of view.

To achieve both photorealism and consistency, we propose a two-stage lensless reconstruction based on range-null space decomposition [36]. According to this decomposition, the reconstructed image consists of two components, one from "range space", which can be directly calculated from the lensless measurement (Fig. 1a), and the other from "null space", which are the detailed textures lost during the lensless imaging process. Therefore, the first stage prioritizes data consistency by recovering the "range space" component, which is the low-frequency content that matches the ground truth. The second stage focuses on photorealism by adding more high-frequency details from the "null space" while maintaining the consistency established in the first stage.

In the first stage, to improve consistency, we propose a novel spatially varying deconvolution to reconstruct the "range space" content. Unlike traditional methods that assume a shift-invariant PSF, our approach automatically adapts to spatial variations in the PSF across the camera's field of view in a data-driven manner. This innovation more accurately models the forward imaging process, leading to better reconstruction of structural integrity and detail in the low-frequency components. Fig. 2 shows improvement with our method (right) against those using a single kernel for deconvolution (left).

2

In the second stage, to improve photorealism, we integrate a generative prior using a pre-trained diffusion model, to insert realistic details into the first-stage output. This model specifically targets the recovery of high-frequency details lost in the lensless imaging process. We condition the diffusion model on the low-frequency content reconstructed in the first stage, guiding it to incorporate missing high-frequency elements from the "null space". Our supervised approach ensures the final images are consistent with actual measurements and also achieve photorealistic quality.

Combining these two stages, our lensless reconstruction achieves a good balance between data consistency and visual quality in the reconstructed images, surpassing existing methods, as shown in Fig. 1e. To validate this, we compare our method with others on two types of lensless cameras: PhlatCam [17] and DiffuserCam [27], using metrics that assess both fidelity and visual quality. The qualitative comparison on the PhlatCam dataset, demonstrating improvements in both aspects, is shown in Fig. 3, highlighting the superiority of our technique over current methods.

2 Related Work

Lensless imaging is traditionally addressed by solving regularized least squares problems for convolutional imaging models, often incorporating sparsity constraints like those in the gradient or frequency domain [1, 5, 21, 31]. Deep learning has revolutionized lensless imaging by enabling learnable deconvolution parameters and perception enhancement through image-to-image networks. Recent innovations include deep unrolled techniques [19, 27, 49], alongside feed-forward deconvolution methods in image space [17] or in feature space [22]. However, these methods generally assume a constant point spread function in the imaging process. Our proposal introduces a spatially varying deconvolution approach to overcome this limitation and achieve more precise image reconstruction.

Spatially-varying deconvolution has been a well-studied area [11, 24, 46] due to the prevalence of imaging systems with PSFs that vary across the FoV. However, these methods [2, 20, 25, 29, 47] are often slow, computationally intensive, and result in poor image quality, especially in complex systems. Recently, MultiWienerNet [48] introduced a deep learning approach for fast, spatially varying deconvolution, but it requires tedious multi-location PSF calibrations. In contrast, our spatially varying deconvolution method, designed for lensless imaging, leverages a single initial PSF and automatically learns variations across the image, eliminating the need for extensive calibration.

Inverse imaging with diffusion models can be categorized as supervised or zero-shot. Supervised methods train a conditional diffusion model [8, 13, 12, 32, 50] with paired images to bridge the gap between input and desired outputs, leveraging generative priors for inverse imaging [35, 39] and controlled generation [28, 34]. On the other hand, zero-shot methods [37] employ guidance to address a wide range of general inverse problems [9, 38]. These techniques typically rely on predefined conditions for guidance. Specific models like DDRM [15] and DDNM [41] focus on mathematical decompositions to improve the diffusion process. Our work integrates the strengths of both approaches, combining supervised fine-tuning with a theoretical framework based on range-null space decomposition, aiming for a robust solution to inverse imaging challenges.

3 Preliminary

3.1 Range-Null Space Decomposition

The lensless imaging process can be formulated as a linear transformation [10]. Specifically, given a scene plane $\mathbf{x} \in \mathbb{R}^{M^2}$, the sensor measurements $\hat{\mathbf{y}} \in \mathbb{R}^{N^2}$ is expressed as $\hat{\mathbf{y}} = \mathbf{A}\mathbf{x} + \mathbf{n}$. Here, $\mathbf{A} \in \mathbb{R}^{N^2 \times M^2}$ is the transfer matrix of the lensless imaging system, and $\mathbf{n}$ is the sensor noise. The transfer matrix $\mathbf{A}$ essentially encodes how light from each point on the scene plane contributes to each sensor pixel. Given a calibrated camera system, the transfer matrix $\mathbf{A}$ is known and the objective of lensless reconstruction is to recover the scene plane image $\mathbf{x}$ (like Fig. 1f) from the measurement (like Fig. 1a). For simplicity, below we consider a noise-free case as $\mathbf{y} = \mathbf{A}\mathbf{x}$.

We then introduce range-null space decomposition. Let $\mathbf{A}^{\dagger} \in \mathbb{R}^{M^2 \times N^2}$ be the pseudo-inverse of the linear matrix $\mathbf{A}$, which satisfies $\mathbf{A}\mathbf{A}^{\dagger}\mathbf{A} \equiv \mathbf{A}$. The operation $\mathbf{A}^{\dagger}\mathbf{A}$ can be interpreted as a projection onto the range space of $\mathbf{A}$ because for any sample $\mathbf{x}$, we have $\mathbf{A}\mathbf{A}^{\dagger}\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x}$. In contrast, the operation $(\mathbf{I} - \mathbf{A}^{\dagger}\mathbf{A})$ acts as a projection onto the null space of $\mathbf{A}$, given that $\mathbf{A}(\mathbf{I} - \mathbf{A}^{\dagger}\mathbf{A})\mathbf{x} = \mathbf{0}$. Therefore, any sample $\mathbf{x}$ can be decomposed into two orthogonal components: one that lies in the

3

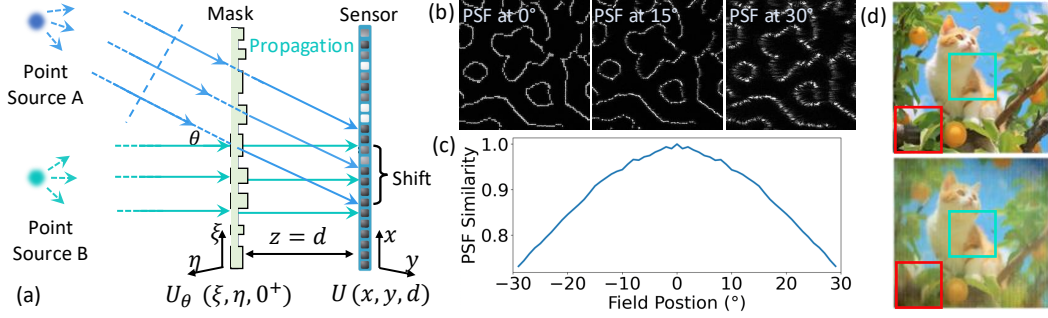



Figure 4: Characterization of PSFs in Lensless Camera. (a) Illustration of light propagation in the lensless camera: two point sources A and B at infinity emitting parallel light beams. Source A emits at angle θ relative to the optical axis, causing a PSF shift on the sensor plane This PSF shift depends on both the incident angle θ and the distance d between the sensor and the mask. (b) Simulated PSFs for light sources at angles of 0°, 15°, and 30°. (c) Inner product similarity between the on-axis PSF and off-axis PSFs at different field positions. (d) Reconstruction using PSF at 0°, degradation is more significant at the periphery (red box) than the center (green box).

range space of $\mathbf{A}$, and the other in the null space of $\mathbf{A}$. Mathematically, this decomposition is:

$$\mathbf{x} \equiv \mathbf{A}^\dagger \mathbf{A} \mathbf{x} + (\mathbf{I} - \mathbf{A}^\dagger \mathbf{A})\mathbf{x}, \tag{1}$$

where $\mathbf{A}^\dagger \mathbf{A} \mathbf{x}$ is the range space component, and $(\mathbf{I} - \mathbf{A}^\dagger \mathbf{A})\mathbf{x}$ is the null space component.

Through this decomposition, we derive two distinct elements indicative of consistency and photo-realism. The "range space" term $\mathbf{A}^\dagger \mathbf{A} \mathbf{x}$ ensures consistency, as its multiplication by $\mathbf{A}$ yields the measurement $\mathbf{y}$, fulfilling the consistency condition $\mathbf{A}\mathbf{x} = \mathbf{y}$. Conversely, the "null space" term $(\mathbf{I} - \mathbf{A}^\dagger \mathbf{A})\mathbf{x}$ ensures that the reconstructed image $\hat{\mathbf{x}}$ aligns with natural image statistics.

3.2 Mismatch in Convolutional Lensless Imaging Model

In lensless literature, most researchers [1, 5, 17, 27] simplify the imaging process as a convolution, because the full transfer matrix $\mathbf{A} \in \mathbb{R}^{N^2 \times M^2}$ is too large to compute. Specifically, the measurement $\mathbf{y}$ is obtained as $\mathbf{y} = \mathbf{h} * \mathbf{x}$, where $\mathbf{h}$ is the point spread function (PSF) of the lensless system and $*$ denotes the convolution. Most previous reconstruction algorithms are based on this assumption.

However, the real lensless imaging model is not simply a spatially invariant convolution as shown in Fig. 4a. Consider a lensless camera consisting of a mask placed at the longitudinal position $z = 0$ and a sensor placed at $z = d$. Let $U(x, y, z)$ be a complex scalar wave field, a complex function of transverse coordinates x, y, and longitudinal position z. Denote the wave field immediately after propagating through the mask from a point source at infinity with an incident angle θ as $U_\theta(\xi, \eta, 0^+)$. Using the Huygens-Fresnel principle [10], the intensity pattern $p_\theta(x, y)$ captured by the sensor is:

$$p_\theta(x, y) = |U_\theta(x, y, d)|^2 = \left| \frac{d}{j\lambda r^2} \iint U_\theta(\xi, \eta, 0^+) \exp(jkr) d\xi d\eta \right|^2, \tag{2}$$

where the distance r is given by $r = \sqrt{d^2 + (x-\xi)^2 + (y-\eta)^2}$ and λ is the wavelength of light.

Therefore, according to Eqn. 2, the spatially invariant convolutional imaging model only satisfies under the Fresnel approximation [5, 10]:

$$r = \sqrt{d^2 + (x-\xi)^2 + (y-\eta)^2} \approx d + (x-\xi)^2/(2d) + (y-\eta)^2/(2d). \tag{3}$$

This approximation is valid only when the distance d between the lensless mask and the sensor is large enough to satisfy $d \gg \sqrt{(x-\xi)^2 + (y-\eta)^2}$. However, most lensless masks are very close to the sensor ($d = 2\text{mm}$ in a typical lensless camera), breaking this assumption.

This mismatch can lead to inaccuracies [26, 43, 52] in the convolutional imaging model. To demonstrate that, we simulate the PSF of a typical lensless camera (Phlatcam [5]) across various incident angles θ using Eqn. (2). As shown in Fig. 4b, PSFs at different angles (0°, 15°, and 30°) are visually different. Quantitatively, Fig. 4c shows the similarity between PSFs at the center and -30° drops from 1.0 to 0.7.

4

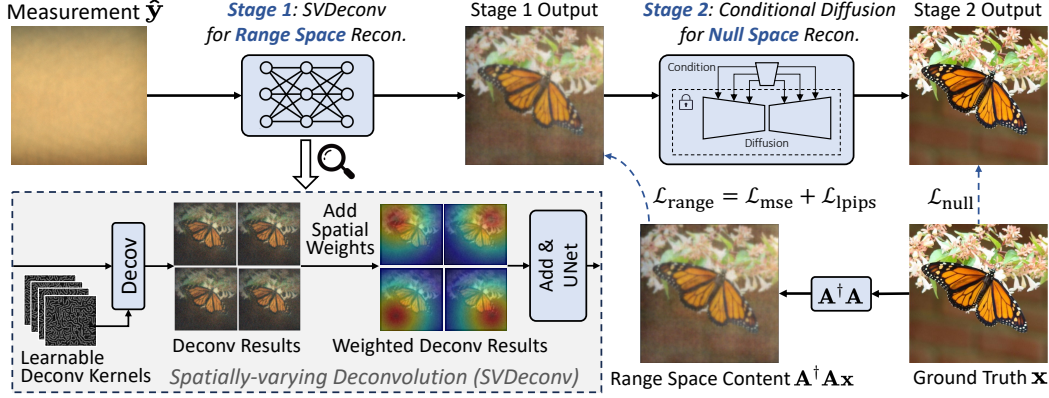



Figure 5: System Overview. The two-stage pipeline begins with a spatially varying deconvolution network mapping lensless measurements to range space. Then a conditional diffusion model for null space recovery refines details using the first stage output, achieving the final reconstruction.

To further show the consequences of this mismatch, we simulate lensless capture from a clean scene image (Fig. 4d top), with the accurate spatially-varying PSF based on Eqn. (2), but solve the inverse imaging problem using a simple convolutional model using the PSF at the center. Fig. 4d bottom shows the result, and there are more artifacts at the boundary (red box) than at the center (green box). This is because the mismatch between the actual PSF and the single PSF used for deconvolution is more significant at the periphery field of view.

4   Method

In this section, we introduce **PhoCoLens**, a method for achieving both **Pho**torealistic and **Co**nsistent reconstruction in **Lens**less imaging. As illustrated in Fig. 5, PhoCoLens consists of two main stages: range space reconstruction and null space recovery. We will first introduce the whole framework and then provide a detailed explanation of each stage.

Given an input lensless measurement $\hat{\mathbf{y}} = \mathbf{A}\mathbf{x} + \mathbf{n}$, where $\mathbf{y} = \mathbf{A}\mathbf{x}$ is the noise-free part of the lensless measurement, $\mathbf{A}$ is the transfer matrix of the lensless camera, $\mathbf{x}$ is the original scene we aim to reconstruct, and $\mathbf{n}$ is the sensor noise. The objective in lensless image reconstruction is to recover a photo-realistic image $\hat{\mathbf{x}}$ which satisfies $\mathbf{A}\hat{\mathbf{x}} = \mathbf{y}$.

Inspired by the range null space decomposition in section 3.1, we can decompose any potential solution $\hat{\mathbf{x}}$ into two orthogonal components: the range space component that maintains consistency and null space component that maximizes photorealism. Formally, this decomposition is given by $\hat{\mathbf{x}} = \mathbf{A}^\dagger\mathbf{A}\mathbf{x} + (\mathbf{I} - \mathbf{A}^\dagger\mathbf{A})\bar{\mathbf{x}}$, where $\mathbf{A}^\dagger\mathbf{A}\mathbf{x}$ is the range space component and $(\mathbf{I} - \mathbf{A}^\dagger\mathbf{A})\bar{\mathbf{x}}$ is the null space component. Note that any choice of $\bar{\mathbf{x}}$ satisfies the equation $\mathbf{A}\hat{\mathbf{x}} = \mathbf{y}$, as $\mathbf{A}(\mathbf{I} - \mathbf{A}^\dagger\mathbf{A})\bar{\mathbf{x}} = \mathbf{0}$.

With this decomposition, we reconstruct the range space and null space components in two stages, as illustrated in Fig. 5. In the first stage, we recover the range space content represented by $\mathbf{A}^\dagger\mathbf{A}\mathbf{x}$ from the noisy lensless measurement $\hat{\mathbf{y}}$. Specifically, we propose **SVDeconv**, a physics-inspired **S**patially-**V**arying **Deconv**olution Network that learns to reconstruct the range space content effectively. The second stage focuses on adding the null space content represented by $(\mathbf{I} - \mathbf{A}^\dagger\mathbf{A})\bar{\mathbf{x}}$. For this purpose, we introduce *null space diffusion*, a conditional diffusion model designed for null space recovery. By conditioning on the range space reconstruction obtained from the first stage, the null space diffusion enhances these images by incorporating the null space content, thereby rendering them with a more realistic appearance. Below we introduce each stage.

4.1   Spatially-varying Deconvolution Network

In the first stage, we use SVDeconv, a novel network designed to invert the forward lensless imaging model and reconstruct the range space content $\mathbf{A}^\dagger\mathbf{A}\mathbf{x}$ from a lensless measurement $\hat{\mathbf{y}} = \mathbf{A}\mathbf{x} + \mathbf{n}$ with noise $\mathbf{n}$. SVDeconv is composed of two main components: a differentiable multi-kernel deconvolution layer and a refinement U-Net, as depicted in the gray box at the bottom left corner of Fig. 5.

5

Traditional lensless imaging methods often employ a single PSF to model the imaging system, which is inaccurate for large field of view (FoV) scenarios as analyzed in section 3.2. To address this, SVDeconv utilizes a set of learnable $K \times K$ PSF kernels, accounting for spatial variations across the field of view. Specifically, we partition the target image region into a grid of $K \times K$ segments, and for each, we apply an individual PSF kernel. This process results in $K \times K$ deconvolution operations, producing $K \times K$ intermediate deconvolved images, as shown in Fig. 5. The multi-kernel deconvolution is mathematically described as a Hadamard product in the Fourier domain:

$$\mathbf{x}_{\mathrm{de}}^{(i)} = \mathcal{F}^{-1}(\mathcal{F}(\mathbf{p}^{(i)}) \odot \mathcal{F}(\hat{\mathbf{y}})), i = 1, 2, ..., K \times K, \tag{4}$$

where $\mathbf{x}_{\mathrm{de}}^{(i)}$ is the i-th deconvolution results of the i-th learnable PSF kernel $\mathbf{p}^{(i)}$, $\mathcal{F}(.)$ and $\mathcal{F}^{-1}(.)$ are the DFT and the Inverse DFT operations, and $\odot$ denotes the Hadamard product.

From this deconvolution operation, we obtain $K \times K$ intermediate deconvolved images. Assuming that each image accurately reconstructs a specific region of the target image corresponding to its PSF's field point, we propose to integrate these images into a single, unified intermediate image using an innovative interpolation method. The interpolation is expressed as:

$$\mathbf{x}_{\mathrm{int}}(u, v) = \sum_{i=1}^{K^2} w_i(u, v) \mathbf{x}_{\mathrm{de}}^{(i)}(u, v), \tag{5}$$

where u, v are coordinates. The weight $w_i(u, v)$ for each deconvolved image $\mathbf{x}_{\mathrm{de}}^{(i)}$ is inversely proportional to the distance between the point (u, v) and the focal center of each corresponding PSF, formally defined as:

$$w_i(u, v) = \frac{d_i^{-\frac{1}{2}}(u, v)}{\sum_{j=1}^{K^2} d_j^{-\frac{1}{2}}(u, v)}, \text{ with } d_i(u, v) = (u - u_i)^2 + (v - v_i)^2. \tag{6}$$

In this way, it ensures that $w_i(u, v)$ normalizes the contribution of each deconvolved image based on inverse Euclidean distance to each point (u_i, v_i), which represents the center of the FoV corresponding to the i-th learnable PSF kernel $\mathbf{p}^{(i)}$. This weighted sum effectively interpolates the intermediate images to form a unified representation with enhanced clarity and detail across the entire FoV.

One advantage of this weight design is that it significantly simplifies the calibration process. Previous multi-kernel methods, such as [48], often require multiple calibrated PSFs at different focal centers, leading to a time-consuming calibration process. On the contrary, our approach only requires one calibrated PSF to initialize all kernels. This is because we pre-define the center of the PSF and allow the model to learn its variations automatically.

After the interpolation, the intermediate image $\mathbf{x}_{\mathrm{int}}$ is fed into the refinement U-Net [33], removing noise and artifacts in the reconstruction to approximate the range space content $\mathbf{A}^{\dagger}\mathbf{A}\mathbf{x}$. We employ a combination of MSE loss and LPIPS loss [51] to train both the learnable PSFs and U-Net.

4.2 Null Space Content Recovery

With the reconstructed range space content, we aim to recover a final image that maintains consistency and improves photorealism. To ensure consistency, the difference between the final image and the range space content (residual content) should reside in the null space, which ensures the final image aligns with the original lensless measurement. To improve photorealism, the combination of the null space and range space content should appear as a realistic real-world image.

To achieve both objectives, we propose null-space diffusion, which takes the reconstructed range space content as the condition and generates an image that adheres to both constraints. It ensures the residual content lies in the null space while maintaining its ability to produce realistic images.

To ensure the outputs from null-space diffusion meet the consistency requirement, we train the model such that the residual content of these samples falls within the null space. Mathematically, if we denote the output as $\widetilde{\mathbf{x}}$, it should satisfy the following equation when conditioned by $\mathbf{A}^{\dagger}\mathbf{A}\mathbf{x}$:

$$\mathbf{A}(\widetilde{\mathbf{x}} - \mathbf{A}^{\dagger}\mathbf{A}\mathbf{x}) = \mathbf{0} \implies \mathbf{A}\widetilde{\mathbf{x}} - \mathbf{A}\mathbf{x} = \mathbf{0} \implies \mathbf{A}(\widetilde{\mathbf{x}} - \mathbf{x}) = \mathbf{0}. \tag{7}$$

Recall that $\mathbf{A}$ is the transfer matrix of the lensless camera and $\mathbf{A}^{\dagger}$ is its pseudo-inverse. This equation indicates that applying the operator $\mathbf{A}$ to the difference between the generated sample $\widetilde{\mathbf{x}}$ and the original sample $\mathbf{x}$ results in zero. This ensures that the residual content resides in the null space of $\mathbf{A}$.

6

Based on this, we design the null space diffusion as follows. Given an image $\mathbf{x}$ and its corresponding range space content $\mathbf{c} = \mathbf{A}^\dagger \mathbf{A} \mathbf{x}$, we add noise progressively to the image resulting in a noisy image $\mathbf{x}_t$, where t is the noise addition iteration. Similar to other conditional diffusion models [50], we train a network ϵ_θ to predict the noise added on the noisy image $\mathbf{x}_t$ with the optimization objective:

$$\min_\theta \mathcal{L}_{\text{null}} = \min_\theta \mathbb{E}_{\mathbf{x}, t, \mathbf{c}, \epsilon \sim \mathcal{N}(0,1)} [||\mathbf{A}(\epsilon_\theta(\mathbf{x}_t, t, \mathbf{c}) - \epsilon)||_2^2], \qquad (8)$$

which is derived to align with the goal in Eqn. (7). More details are in the Appendix.

Additionally, to ensure the null-space diffusion produces realistic images according to given conditions, we utilize a pre-trained diffusion model such as Stable Diffusion [32] with its weights frozen. We focus on training supplementary conditioning modules to guide the generative process using range space conditions while preserving its powerful ability to generate realistic images. Specifically, we follow the StableSR [40] structure, which involves using a conditional encoder to extract multi-scale features from the range space condition and then using them to modulate the intermediate feature maps of the residual blocks in the diffusion model. This modulation can align the generated images with the range space conditions, enhancing the fidelity and realism of output images.

5 Experiments

In this section, we assess the performance of our proposed method using two lensless imaging datasets, collected in real-world environments by two different kinds of lensless cameras: PhlatCam [5] and DiffuserCam [1]. PhlatCam employs a phase mask, while DiffuserCam utilizes a diffuser for its lensless mask. We compare our approach with other methods and conduct a comprehensive ablation study to evaluate the effectiveness of our design.

5.1 Dataset and Metrics

The *PhlatCam* dataset [17] contains 10,000 images across 1,000 classes resized to 384×384 pixels. Images are displayed and captured using a lensless PhlatCam [5], generating images at 1280 × 1480 pixels. We use 990 classes for training and 10 classes for testing, following the original protocol.

The *DiffuserCam* dataset [27] contains 25,000 paired images captured simultaneously using a standard lensed camera (ground truth) and a mask-based lensless camera DiffuserCam [1]. These pairs are split into 24,000 images for training and 1,000 for testing. Both cameras utilize sensors with a native resolution of 1080×1920 pixels, which are first downsampled to 270×480 pixels and then further cropped to a final resolution of 210×380 pixels for proper display.

To evaluate both consistency (fidelity) and photorealism (visual quality), we employ two metric sets:

- **Full-reference metrics** to evaluate the consistency. Three full-reference metrics, PSNR, SSIM [42], and LPIPS [51], are used to evaluate the distance between the network output and ground truth, which indicates the fidelity of the reconstruction.
- **Non-reference metrics** to evaluate the photorealism. Three non-reference metrics, ManIQA [45], ClipIQA [39], and MUSIQ [16], are used to evaluate the visual quality of reconstructed images.

5.2 Implementation Details

For SVDeconv, we utilize 3×3 PSF kernels for deconvolution. We initialize the 9 kernels using a single calibrated PSF from both the DiffuserCam and PhlatCam datasets. The U-Net component in SVDeconv is adapted from the U-Net used in Le-ADMM-U for training on the DiffuserCam, and from the U-Net in FlatNet-gen for the PhlatCam. We train the network for 100 epochs with a batch size of 5, using the Adam optimizer [18]. The learning rate is set to 3e-5 for U-Net training in both datasets. For deconvolution kernel training, the learning rate is set at 4e-9 for PhlatCam and 3e-5 for DiffuserCam. We set the MSE and LPIPS loss weights to be 1 and 0.05, respectively. In the DiffuserCam dataset, where the measurements are cropped by the sensor's limited size, we employ replicate padding [17] which expands the width and height of the measurements by a factor of two.

For null-space diffusion, we use the range space content reconstructed by SVDeconv as input conditions. We train it for 200 epochs, following the training and inference settings used in StableSR[39].

In the first stage, SVDeconv is trained using range-space content derived from ground truth images. SVDeconv consists of two main parameter components: a learnable deconvolution kernel initialized

7

Table 1: Performance comparison of methods on DiffuserCam and PhlatCam

Dataset	PhlatCam						DiffuserCam					
Metrics	PSNR	SSIM	LPIPS↓	ManIQA	ClipIQA	MUSIQ	PSNR	SSIM	LPIPS↓	ManIQA	ClipIQA	MUSIQ
Ground Truth	*—*	*—*	*—*	*0.431*	*0.583*	*63.40*	*—*	*—*	*—*	*0.160*	*0.333*	*31.21*
WienerDeconv [44]	12.19	0.270	0.922	0.111	0.149	15.47	10.59	0.275	0.843	0.184	0.180	19.43
ADMM [7]	13.45	0.301	0.877	0.132	0.159	16.45	12.87	0.305	0.705	0.141	0.136	16.51
Le-ADMM-U [27]	20.12	0.515	0.405	0.138	0.180	24.64	22.35	0.668	0.253	0.110	0.185	20.16
MMCN [49]	20.39	0.524	0.346	0.145	0.206	27.65	24.09	0.744	0.238	0.121	0.183	21.47
UPDN [19]	20.48	0.533	0.352	0.158	0.215	29.32	24.67	0.747	0.256	0.139	0.178	22.56
FlatNet-gen [17]	20.53	0.549	0.375	0.203	0.287	44.94	21.43	0.696	0.254	0.134	0.244	23.15
DDNM+[41]	15.22	0.485	0.623	0.297	0.412	45.32	18.36	0.539	0.516	0.162	0.201	27.83
FlatNet+DiffBIR [23]	19.96	0.544	0.391	0.335	0.523	57.13	18.97	0.517	0.505	0.312	0.548	51.20
PhoCoLens(Ours)	22.07	0.601	0.215	0.357	0.565	62.20	24.12	0.748	0.161	0.172	0.339	32.84

with known PSFs, and a U-Net initialized with standard weights without pretraining. Once trained, SVDeconv processes input lensless measurements to estimate the range-space content of training samples. Subsequently, we use this estimated range-space content as input conditions for fine-tuning via null-space diffusion. During diffusion fine-tuning, we utilize a pre-trained diffusion model with frozen weights. We only train the supplementary conditioning modules like StableSR [35], to guide the reconstruction process effectively.

| WienerDeconv | Le-ADMM-U | FlatNet-gen | FlatNet+DiffBIR | Ours | Groundtruth |

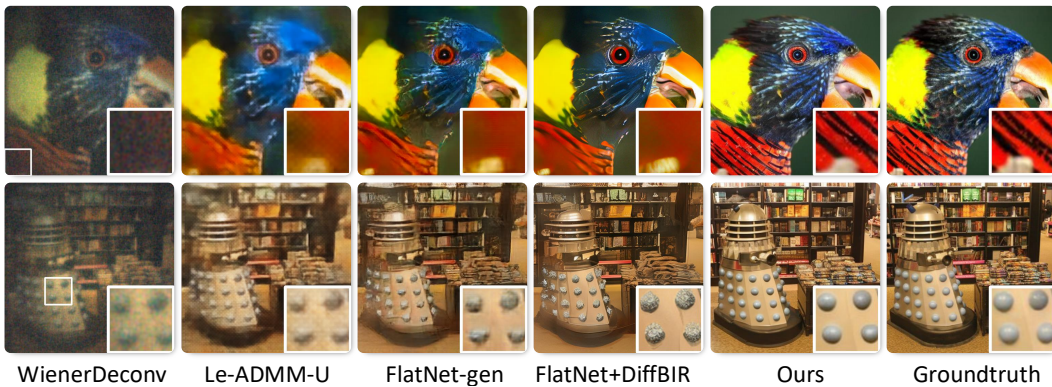



Figure 6: Qualitative comparison between our method and others on the PhlatCam dataset.

| Le-ADMM-U | UPDN | FlatNet+DiffBIR | Ours | Groundtruth |

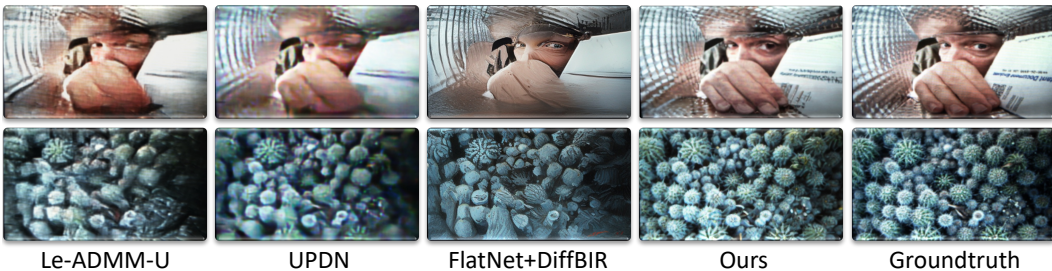



Figure 7: Qualitative comparison between our method and others on the DiffuserCam dataset.

5.3 Comparison with Other Approaches

We evaluate the performance of our proposed method by comparing it against both traditional and learning-based approaches, using qualitative and quantitative measures on the two datasets.

We compare traditional methods like Tikhonov regularized reconstruction in the Fourier domain (WienerDeconv [44]) and total variation regularization via ADMM [7]. Additionally, we evaluate learning-based methods like the unrolled network Le-ADMM-U [27], MMCN [49], and UPDN [19] and the feedforward deconvolution method FlatNet-gen. Additionally, we assess two diffusion-based methods: one uses a pre-trained Stable Diffusion [32] with a zero-shot inverse imaging sampling method DDNM+ [41], and another enhances outputs from FlatNet-gen using a pre-trained blind

8

Table 2: Comparison of deconvolution methods for range space reconstruction and original content reconstruction.

Reconstruction Target	Method	PhlatCam			DiffuserCam		
		PSNR	SSIM	LPIPS↓	PSNR	SSIM	LPIPS↓
Range Space	Le-ADMM-U	23.87	0.835	0.307	24.24	0.748	0.282
	SingleDeconv	25.61	0.874	0.263	24.94	0.799	0.230
	MultiWienerNet	26.43	0.880	0.256	26.32	0.823	0.189
	SVDeconv (Ours)	**26.97**	**0.894**	**0.253**	**28.84**	**0.856**	**0.144**
Original Content	Le-ADMM-U	20.12	0.515	0.405	22.35	0.668	0.253
	SingleDeconv	20.43	0.545	0.381	21.97	0.689	0.262
	MultiWienerNet	20.82	0.562	0.374	23.02	0.725	0.214
	SVDeconv (Ours)	**21.01**	**0.571**	**0.363**	**25.55**	**0.781**	**0.179**

FlatNet-gen   MultiWinenerNet

SVDenconv(Ours)   Groundtruth

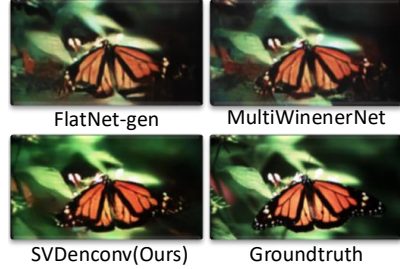



Figure 8: Comparison of deconvolution methods on a DiffuserCam example.

Table 3: Comparison of null space recovery methods

Method	PSNR	SSIM	LPIPS↓	ManIQA	ClipIQA	MUSIQ
StableSR [39]	14.93	0.446	0.624	0.64	0.235	33.84
DiffBIR [23]	16.21	0.432	0.502	0.353	0.512	56.65
DDNM [41]	17.62	0.539	0.517	0.148	0.201	27.82
Ours	**22.07**	**0.601**	**0.215**	**0.357**	**0.565**	**62.20**

Table 4: Comparison of different diffusion conditions

Condition	PhlatCam			DiffuserCam		
	PSNR	SSIM	LPIPS↓	PSNR	SSIM	LPIPS↓
WienerDeconv	20.38	0.551	0.301	20.42	0.632	0.282
FlatNet-gen	20.89	0.556	0.286	21.67	0.647	0.230
SVD-OC	21.16	0.572	0.276	22.43	0.679	0.185
Ours	**22.07**	**0.601**	**0.215**	**24.12**	**0.748**	**0.161**

DDNM   DiffBIR   Ours   Groundtruth

Figure 9: Comparison between our null space diffusion and other methods.

FlatNet-gen   SVD-OC   Ours   Groundtruth

Figure 10: Comparison between different diffusion conditions.

image restoration model DiffBIR [23]. Quantitative results are summarized in Tab. 1. Qualitative comparisons for PhlatCam and DiffuserCam are shown in Figures 6 and 7, respectively.

On the PhlatCam dataset, our method achieves the best performance on both fidelity (full-reference metrics) and visual quality (non-reference metrics) according to Tab. 1, This is further supported by the qualitative results in Fig. 6, where our reconstruction are the closest to the ground truth.

For the DiffuserCam dataset, our method achieves superior performance in consistency metrics (full-reference metrics, SSIM, and LPIPS). While UPDN achieves a slightly higher PSNR, the visual quality of its outputs suffers from artifacts, as shown in Fig. 7. This slight PSNR drop in our method is likely due to artifacts within the DiffuserCam ground truth itself, such as moiré patterns. These artifacts can penalize high-quality images with a lower PSNR, despite their good visual consistency. This is further evidenced by the visual comparison in Fig. 7. Conversely, FlatNet+DiffBIR, though achieving higher quality metrics (non-reference metrics, ManIQA, ClipIQA, and MUSIQ), fails to maintain data consistency (low full-reference metrics, PSNR, SSIM, and LPIPS). Also, as shown Fig. 7, FlatNet+DiffBIR sometimes badly distorts the original image content, like transferring a human hand into clothes in the top row. Overall, our method achieves the best trade-off between the data consistency and the visual quality. And thanks to the introduction of SVDencov, our method can reconstruct the field-of-view periphery better than other methods.

5.4 Analysis and Discussion

Effectiveness of Spatially-varying Deconvolution. To verify the effectiveness of the proposed SVDeconv, we compare it with other learnable deconvolution approaches. Baselines include Le-ADMM-U (a deep unrolled network implementing ADMM), SingleDeconv (using a single deconvolution kernel), and MultiWienerNet [48](employing multiple deconvolution kernels without weighted interpolation, initialized similarly to our approach with a single PSF). We focus on two types of content reconstruction: range space content reconstruction, where range space content is the target, and original content

9

reconstruction, using the original ground truth as the target. For a fair comparison, all methods utilize the same U-Net for denoising and perceptual enhancement and the same loss for training. Results in Tab. 2 show the efficacy of the proposed spatially-varying deconvolution in the faithful reconstruction of both range space content and original content. Further, Fig. 8 shows our method outperforms others, particularly in reconstructing fine details at the periphery (circled by white dashed boxes).

Effectiveness of Null-space Diffusion. Following the reconstruction of range space content, we perform null space content recovery conditioned on the previously reconstructed range space content. We compare our null-space diffusion model with other approaches, including conditional diffusion models like DiffBIR [23] and StableSR [39], as well as the zero-shot inverse imaging method DDNM [41]. Quantitative results in Tab. 3 and qualitative comparison in Fig. 9 demonstrate that our null-space diffusion performs the best in both consistency with the ground truth and visual quality.

Effectiveness of Range Content Conditions. One important design of our framework is to use the first stage reconstruction as the condition for the second stage diffusion. To evaluate the effectiveness of these conditions, we perform an ablation study where we replace our reconstructed range content with the outputs from WienerDeconv, FlatNet-gen, and the SVD-OC (SVDencov trained with original content), and re-train the conditional diffusion models using these conditions. As evident from Fig. 10, our range content condition significantly improves reconstruction consistency compared to others. This is because using output from models like FlatNet-gen introduces additional artifacts that hinder the second stage from recovering the original image. Furthermore, quantitative analysis in Tab. 4 confirms this observation. Our method outperforms these conditions in all fidelity metrics.

6 Conclusion and Limitation

We present PhoCoLens, an innovative two-stage approach for achieving photorealistic and consistent reconstructions in lensless imaging. The method leverages the strengths of two complementary approaches: spatially varying deconvolution for the range space ensures consistency, while null-space diffusion for the null space guarantees photorealism. Our experiments on two lensless cameras demonstrate PhoCoLens' effectiveness in reconstructing realistic scenes while preserving fidelity. One limitation of our approach is that the two-stage nature of PhoCoLens and the diffusion model's sampling time hinder its real-time applicability. Additionally, although PhoCoLens achieves the best fidelity, the diffusion model we used may still introduce high-frequency details that deviate from the original scene, particularly for smooth scenes lacking in detail. Future work will focus on two main aspects: accelerating the diffusion model sampling process to enable real-time photo and video capture with lensless cameras, and exploring 3D spatially varying point spread function (PSF) effects to capture 3D scene information.

Acknowledgment

We thank the reviewers for their valuable comments. This work is supported by the Shanghai Artificial Intelligence Laboratory, the RGC Early Career Scheme (ECS) No. 24209224, and CUHK Direct Grants (RCFUS) No. 4055189.

References

[1] Nick Antipa, Grace Kuo, Reinhard Heckel, Ben Mildenhall, Emrah Bostan, Ren Ng, and Laura Waller. Diffusercam: lensless single-exposure 3d imaging. *Optica*, 5(1):1–9, 2018.

[2] Muthuvel Arigovindan, Joshua Shaevitz, John McGowan, John W Sedat, and David A Agard. A parallel product-convolution approach for representing depth varying point spread functions in 3d widefield microscopy based on principal component analysis. *Optics express*, 18(7):6461–6476, 2010.

[3] M Salman Asif, Ali Ayremlou, Aswin Sankaranarayanan, Ashok Veeraraghavan, and Richard G Baraniuk. Flatcam: Thin, lensless cameras using coded aperture and computation. *IEEE Transactions on Computational Imaging*, 3(3):384–397, 2016.

[4] Eric Bezzam, Martin Vetterli, and Matthieu Simeoni. Privacy-enhancing optical embeddings for lensless classification. *arXiv preprint arXiv:2211.12864*, 2022.

10

[5] Vivek Boominathan, Jesse K Adams, Jacob T Robinson, and Ashok Veeraraghavan. Phlatcam: Designed phase-mask based thin lensless camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(7):1618–1629, 2020.

[6] Vivek Boominathan, Jacob T Robinson, Laura Waller, and Ashok Veeraraghavan. Recent advances in lensless imaging. *Optica*, 9(1):1–16, 2022.

[7] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

[8] Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. Ilvr: Conditioning method for denoising diffusion probabilistic models. *arXiv preprint arXiv:2108.02938*, 2021.

[9] Hyungjin Chung, Byeongsu Sim, Dohoon Ryu, and Jong Chul Ye. Improving diffusion models for inverse problems using manifold constraints. *Advances in Neural Information Processing Systems*, 35:25683–25696, 2022.

[10] Joseph W Goodman. *Introduction to Fourier optics*. Roberts and Company publishers, 2005.

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

[12] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[13] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

[14] Yi Hua, Yongyi Zhao, and Aswin C Sankaranarayanan. Angle sensitive pixels for lensless imaging on spherical sensors. *arXiv preprint arXiv:2306.15953*, 2023.

[15] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *Advances in Neural Information Processing Systems*, 35:23593–23606, 2022.

[16] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5148–5157, 2021.

[17] Salman Siddique Khan, Varun Sundar, Vivek Boominathan, Ashok Veeraraghavan, and Kaushik Mitra. Flatnet: Towards photorealistic scene reconstruction from lensless measurements. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(4):1934–1948, 2020.

[18] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[19] Oliver Kingshott, Nick Antipa, Emrah Bostan, and Kaan Akşit. Unrolled primal-dual networks for lensless cameras. *Optics Express*, 30(26):46324–46335, 2022.

[20] Grace Kuo, Fanglin Linda Liu, Irene Grossrubatscher, Ren Ng, and Laura Waller. On-chip fluorescence microscopy with a random microlens diffuser. *Optics Express*, 28(6):8384–8399, 2020.

[21] Chengbo Li, Wotao Yin, Hong Jiang, and Yin Zhang. An efficient augmented lagrangian method with applications to total variation minimization. *Computational Optimization and Applications*, 56:507–530, 2013.

[22] Ying Li, Zhengdai Li, Kaiyu Chen, Youming Guo, and Changhui Rao. Mwdns: reconstruction in multi-scale feature spaces for lensless imaging. *Optics Express*, 31(23):39088–39101, 2023.

[23] Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Ben Fei, Bo Dai, Wanli Ouyang, Yu Qiao, and Chao Dong. Diffbir: Towards blind image restoration with generative diffusion prior. *arXiv preprint arXiv:2308.15070*, 2023.

11

[24] Jun Luo, Yunfeng Nie, Wenqi Ren, Xiaochun Cao, and Ming-Hsuan Yang. Correcting optical aberration via depth-aware point spread functions. *IEEE transactions on pattern analysis and machine intelligence*, 2024.

[25] Elie Maalouf, Bruno Colicchio, and Alain Dieterlen. Fluorescence microscopy three-dimensional depth variant point spread function interpolation using zernike moments. *JOSA A*, 28(9):1864–1870, 2011.

[26] Kyoji Matsushima. Shifted angular spectrum method for off-axis numerical propagation. *Optics Express*, 18(17):18453–18463, 2010.

[27] Kristina Monakhova, Joshua Yurtsever, Grace Kuo, Nick Antipa, Kyrollos Yanny, and Laura Waller. Learned reconstructions for practical mask-based lensless imaging. *Optics express*, 27(20):28075–28090, 2019.

[28] Chong Mou, Xintao Wang, Liangbin Xie, Yanze Wu, Jian Zhang, Zhongang Qi, and Ying Shan. T2i-adapter: Learning adapters to dig out more controllable ability for text-to-image diffusion models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 4296–4304, 2024.

[29] Nurmohammed Patwary and Chrysanthe Preza. Image restoration for three-dimensional fluorescence microscopy using an orthonormal basis for efficient representation of depth-variant point-spread functions. *Biomedical optics express*, 6(10):3826–3841, 2015.

[30] Ken Perlin. Improving noise. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 681–682, 2002.

[31] Dikpal Reddy, Ashok Veeraraghavan, and Rama Chellappa. P2c2: Programmable pixel compressive camera for high speed imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 329–336. IEEE, 2011.

[32] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

[33] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[34] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 conference proceedings*, pages 1–10, 2022.

[35] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE transactions on pattern analysis and machine intelligence*, 45(4):4713–4726, 2022.

[36] Johannes Schwab, Stephan Antholzer, and Markus Haltmeier. Deep null space learning for inverse problems: convergence analysis and rates. *Inverse Problems*, 35(2):025008, 2019.

[37] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.

[38] Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. *arXiv preprint arXiv:2111.08005*, 2021.

[39] Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 2555–2563, 2023.

[40] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin CK Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. *arXiv preprint arXiv:2305.07015*, 2023.

12

[41] Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *International Conference on Learning Representations*, 2022.

[42] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[43] Haoyu Wei, Xin Liu, Xiang Hao, Edmund Y Lam, and Yifan Peng. Modeling off-axis diffraction with the least-sampling angular spectrum method. *Optica*, 10(7):959–962, 2023.

[44] Norbert Wiener. *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*. The MIT press, 1949.

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

[46] Xinge Yang, Qiang Fu, Mohamed Elhoseiny, and Wolfgang Heidrich. Aberration-aware depth-from-focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[47] Kyrollos Yanny, Nick Antipa, William Liberti, Sam Dehaeck, Kristina Monakhova, Fanglin Linda Liu, Konlin Shen, Ren Ng, and Laura Waller. Miniscope3d: optimized single-shot miniature 3d fluorescence microscopy. *Light: Science & Applications*, 9(1):171, 2020.

[48] Kyrollos Yanny, Kristina Monakhova, Richard W Shuai, and Laura Waller. Deep learning for fast spatially varying deconvolution. *Optica*, 9(1):96–99, 2022.

[49] Tianjiao Zeng and Edmund Y Lam. Robust reconstruction with deep learning to handle model mismatch in lensless imaging. *IEEE Transactions on Computational Imaging*, 7:1080–1092, 2021.

[50] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3836–3847, 2023.

[51] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.

[52] Wenhui Zhang, Hao Zhang, Kyoji Matsushima, and Guofan Jin. Shifted band-extended angular spectrum method for off-axis diffraction calculation. *Optics Express*, 29(7):10089–10103, 2021.

13

Appendix

A   PhlatCam Simluation

In section 3.2, we simulate a PhlatCam to interpret the detail of mismatch in a convolutional lensless imaging model. Here we explain how to simulate a lensless camera like PhlatCam.

Following previous work [5], our simulation begins by constructing a desired point spread function (PSF) based on the contours generated from random Perlin noise [30]. Given a specific distance between the mask and sensor, we optimize the phase mask to produce a PSF that matches the target. For this optimization, we employ the near-field phase retrieval algorithm (NfPR) [5]. This iterative algorithm alternates between the fields at the mask plane and the sensor plane, enforcing specific constraints at each location: the field amplitude at the mask plane is set to unity, and the intensity of the field at the sensor plane conforms to the target PSF. The algorithm utilizes forward Fresnel propagation for computing the field from the mask to the sensor plane and applies backward Fresnel propagation to compute from the sensor back to the mask. The inputs required for phase mask optimization include the target PSF, the mask-to-sensor distance (1mm in our setup), and the light wavelength, which in this case is set to 532 nm, a typical wavelength in the mid-visible spectrum.

| Target PSF | Optimized Phase Mask | Simulated PSF |

Figure A1: Visual examples for results of the range space content simulation.

Fig. A1 illustrates the target PSF, the optimized phase mask, and the simulated PSF resulting from the Fresnel propagation method applied to the designed phase mask. Given a desired phase mask, we can accurately simulate the lensless imaging process under various situations.

B   Range Space Content Approximation

In the first stage of our method (section 4.1), we want to reconstruct the range-space projection $\mathbf{A}^{\dagger}\mathbf{A}\mathbf{x}$ from the lensless measurement $\hat{\mathbf{y}}$. This reconstruction requires the known target $\mathbf{A}^{\dagger}\mathbf{A}\mathbf{x}$ corresponding to the given image $\mathbf{x}$. However, for real-world lensless imaging systems, directly obtaining the transfer matrix $\mathbf{A} \in \mathbb{R}^{N^2 \times M^2}$ is computationally infeasible due to its considerable size. Consequently, we resort to simulating the realistic lensless imaging process using an approximate form of $\mathbf{A}$ and emulate its pseudoinverse $\mathbf{A}^{\dagger}$ by inverting this approximation.

In cases where a single point spread function (PSF) $\mathbf{h}$ is calibrated, we approximate the forward imaging process as follows:

$$\mathbf{y}_{\text{sim}} = C(\mathbf{h} * \mathbf{x}) + \mathbf{n}, \tag{A1}$$

where C denotes the cropping operation due to the limited sensor size, $*$ represents the convolution operator, and $\mathbf{n}$ represents both sensor and quantization noise. In our practical implementation, the simulated $\mathbf{y}_{\text{sim}}$ is quantized to 12 bits, and sensor noise is introduced to achieve a signal-to-noise ratio (SNR) of 30.

The approximation for $\mathbf{A}^{\dagger}$ is derived by reversing the forward process, specifically by applying replicate padding to counteract the cropping and employing a Wiener deconvolution to reverse the convolution:

$$\mathbf{x}_{\text{range}} = D(P(\mathbf{y}_{\text{sim}}), \mathbf{h}), \tag{A2}$$

14

where the P represent replicate padding and the D represent Wiener deconvolution.

If multiple calibrated PSFs are available across different fields of view (FoV), the imaging forward process is better approximated by employing a low-rank model to mimic the spatially varying convolution. Correspondingly, the spatially varying deconvolution model discussed in section 4.1 could be utilized to simulate $\mathbf{A}^\dagger$.

In our PhlatCam and DiffuserCam setups, we only have a single calibrated PSF. Therefore, we constitute the range-space content of image $\mathbf{x}$ using the simple convolutions. It's important to note that even without multiple calibrated PSFs for a more accurate simulation, the generated range-space content remains a good approximation of the real counterpart. As shown in Section 5.4, our simulation still serves as better supervision compared to the original image content in the first stage.

Figure A2: Visual examples for results of the range space content simulation.

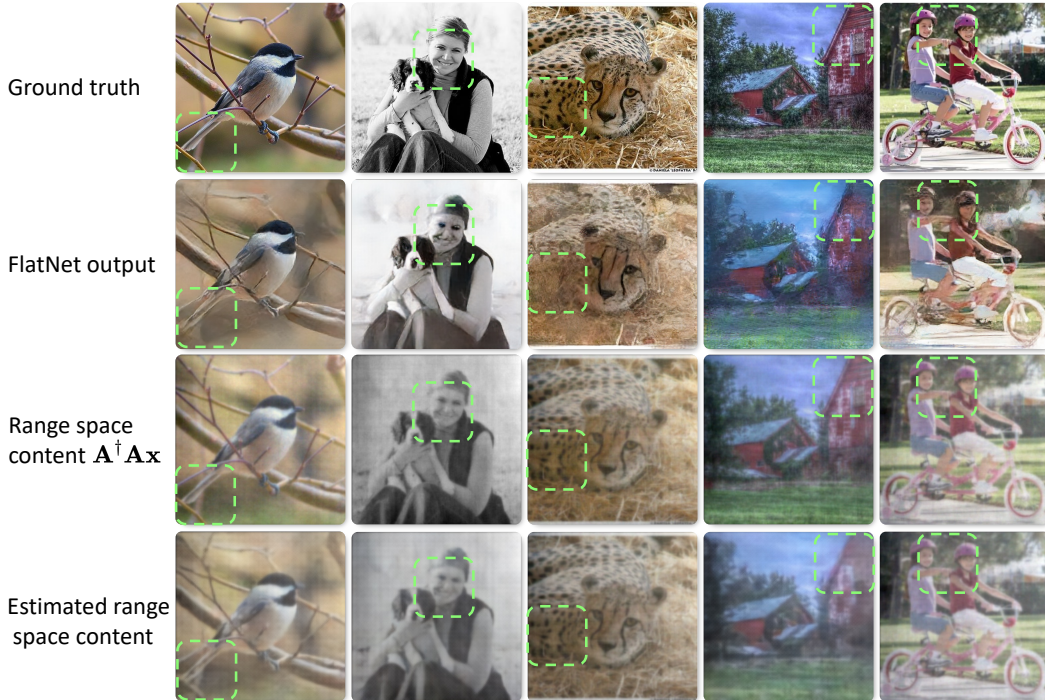



Figure A3: Examples of range space content estimated by our method.

Fig. A2 shows the visual examples of the original ground truth $\mathbf{x}$ and the corresponding simulated range space content $\mathbf{A}^\dagger\mathbf{A}\mathbf{x}$ of PhlatCam. It is shown that the range of space content $\mathbf{A}^\dagger\mathbf{A}\mathbf{x}$ keeps the low-frequency structure of the original scene and lost some high-frequency details from the lensless imaging process.

15

Fig. A3 shows the visual examples of range space content estimated by our method. The outputs from SVDeconv exhibit high visual consistency with the ground truth, as highlighted in the green boxes, in contrast to those from FlatNet. FlatNet alters the original scene's content and introduces incorrect high-frequency details. Conversely, our method preserves only the low-frequency content in the range space, aligning closely with the original ground truth.

C Training Target for the Null-space Diffusion Model

In section 6, we establish that the goal articulated in Eqn. (7) corresponds directly to the optimization objective described in Eqn. (8) used for diffusion training. We provide a detailed derivation to support this equivalence.

Diffusion models [32] can be characterized by a signal-to-noise ratio defined by two sequences $(\alpha_t)_{t=1}^{T}$ and $(\sigma_t)_{t=1}^{T}$. Given a data sample $\mathbf{x}_0$, the forward diffusion process q can be expressed as:

$$q(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t \mid \alpha_t \mathbf{x}_0, \sigma_t^2 \mathbb{I}), \tag{A3}$$

where $\mathcal{N}$ denotes the normal distribution.

The Markov property of the diffusion is given as:

$$q(\mathbf{x}_t \mid \mathbf{x}_s) = \mathcal{N}(\mathbf{x}_t \mid \alpha_{t|s}\mathbf{x}_s, \sigma_{t|s}^2 \mathbb{I}), \tag{A4}$$

for $s < t$, with:

$$\alpha_{t|s} = \frac{\alpha_t}{\alpha_s} \quad \text{and} \quad \sigma_{t|s}^2 = \sigma_t^2 - \alpha_{t|s}^2 \sigma_s^2. \tag{A5}$$

Given $\mathbf{x}_t = \alpha_t \mathbf{x}_0 + \sigma_t \epsilon$, and a range space condition $\mathbf{c} = \mathbf{A}^\dagger \mathbf{A}\mathbf{x}$, the diffusion output $\widetilde{\mathbf{x}}$ can be modeled by $\widetilde{\mathbf{x}} = x_\theta(\mathbf{x}_t, \mathbf{c}, t)$. To ensure reconstruction consistency such that $\mathbf{A}(\widetilde{\mathbf{x}} - \mathbf{x}_0) = \mathbf{0}$ in Eqn. (7), the optimization objective is defined as:

$$\min_\theta \mathbb{E}_{\mathbf{x}_0, t, \mathbf{c}, \epsilon \sim \mathcal{N}(0,1)} \left[\|\mathbf{A}(x_\theta(\mathbf{x}_t, \mathbf{c}, t) - \mathbf{x}_0)\|_2^2 \right]. \tag{A6}$$

Following previous work [12], we employ the reparameterization of the reconstruction term as a denoising objective:

$$\epsilon_\theta(\mathbf{x}_t, \mathbf{c}, t) = \frac{\mathbf{x}_t - \alpha_t x_\theta(\mathbf{x}_t, \mathbf{c}, t)}{\sigma_t}, \tag{A7}$$

to minimize:

$$\min_\theta \mathcal{L}_{\text{null}} = \mathbb{E}_{\mathbf{x}_0, t, \mathbf{c}, \epsilon \sim \mathcal{N}(0,1)} \left[\|\mathbf{A}(\epsilon_\theta(\mathbf{x}_t, \mathbf{c}, t) - \epsilon)\|_2^2 \right], \tag{A8}$$

which is exactly the Eqn. (8).

Due to the substantial computational cost associated with simulating the transfer matrix $\mathbf{A}$, an alternative optimization objective can be employed in practice. Specifically, we can approximate the goal described in Eqn. (8) using the following optimization formulation:

$$\min_\theta \mathcal{L}_{\text{null}_{\text{app}}} = \mathbb{E}_{\mathbf{x}_0, t, \mathbf{c}, \epsilon \sim \mathcal{N}(0,1)} \left[\|(\epsilon_\theta(\mathbf{x}_t, \mathbf{c}, t) - \epsilon)\|_2^2 \right]. \tag{A9}$$

D Qualitative Results

We present additional qualitative results demonstrating the superior performance of our methods in Fig. A4 for the PhlatCam dataset and in Fig. A5 for the DiffuserCam dataset.

16

| WienerDeconv | Le-ADMM-U | FlatNet-gen | FlatNet+DiffBIR | Ours | Groundtruth |

Figure A4: Qualitative comparison between our method and others on the PhlatCam dataset.

| Le-ADMM-U | UPDN | FlatNet+DiffBIR | Ours | Groundtruth |

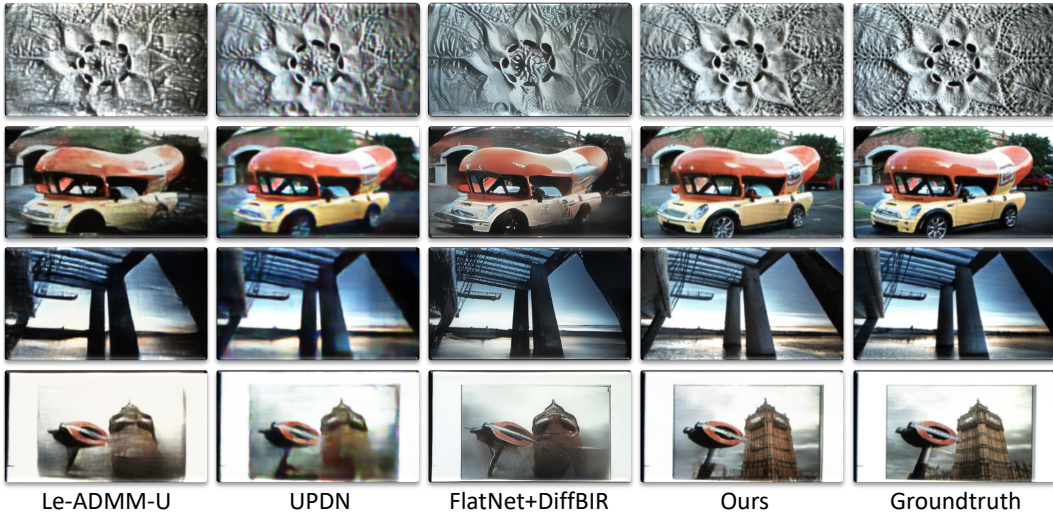



Figure A5: Qualitative comparison between our method and others on the DiffuserCam dataset.

17

E Broader Impacts

Lensless imaging algorithms represent a significant advancement in camera technology, offering both promising benefits and potential drawbacks for society. On the positive side, these algorithms facilitate the development of cameras that are smaller, lighter, and more energy-efficient. This can greatly expand their applicability in various fields, particularly in medical contexts where the miniaturization of imaging devices is crucial. For example, tiny lensless cameras could be integrated into endoscopic tools for less invasive procedures, improving patient outcomes and recovery times. Furthermore, such compact cameras could be used in wearable technologies to monitor health conditions continuously, providing real-time data to both patients and healthcare providers.

However, the proliferation of small, lensless cameras also introduces potential negative societal impacts, particularly concerning privacy. The ease with which these discreet cameras can be incorporated into everyday environments raises the possibility of misuse, such as unauthorized surveillance or covert photography. This concern is particularly acute in private spaces where individuals expect a high level of privacy. As such, while lensless imaging technology promises significant advancements in fields like healthcare, it also necessitates stringent regulations and ethical guidelines to prevent invasive use and protect individual privacy rights.

18

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction accurately reflect the contributions and scope.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Limitations are discussed in section 6.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

   Justification: The paper does not include theoretical results.

   Guidelines:

19

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: For details on reproducing the experimental results, please refer to section 5.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: Code and data will be available on the project website.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.

20

- Please see the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: The implementation details of the experiments are provided in section 5.2.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: We report the experiment results using the standard conventions in lensless imaging research, consistent with previous studies.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

21

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Evrey experiment can be completed within two days using a server equipped with eight NVIDIA RTX 3090 GPUs, a 64-core CPU, 256GB of RAM, and 1TB of storage.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: We strictly adhere to the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: The broader impacts of our work are explained in appendix E.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.

22

- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: We will not release data or models that have a high risk for misuse.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: We cite all original papers and make sure that our usage is legal.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: We do not introduce new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.

23

- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: Our work does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: Our work does not involve research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

24