



Title	Extended-Depth-of-Field Lensless Imaging via Optical Design and Physics-Constrained Reconstruction
Author(s)	Cunha Santos A. V. Silva Neto, Jose Reinaldo
Citation	大阪大学, 2026, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/105826
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

Extended-Depth-of-Field Lensless Imaging via Optical Design and Physics-Constrained Reconstruction

Submitted to
Graduate School of Information Science and Technology
The University of Osaka

January, 2026

Jose Reinaldo CUNHA SANTOS A V SILVA NETO

Abstract

Lensless cameras replace conventional lenses with thin coded masks placed near the sensor, enabling compact imaging systems. However, the resulting measurements arise from ill-posed operators whose point spread functions vary with depth and lateral translations, making reconstruction highly sensitive to model mismatch. Existing methods typically address this either through (1) explicit depth modeling, using multi-plane 3D formulations, dense calibration, and slow iterative solvers or through (2) supervised deep networks trained on lensed ground-truth data or simulated volumetric scenes. While effective in controlled settings, these strategies impose heavy calibration or computational demands and often generalize poorly to real-world scenes, limiting the practicality of current lensless systems.

This thesis aims to move lensless cameras toward practical all-in-focus 2D imaging, by jointly designing depth-robust coded masks and self-supervised reconstruction algorithms. The contributions span three main components.

First, we investigate restricted-aperture radial coded masks as a means to extend DOF while preserving a simple 2D shift-invariant convolution model. By comparing radial and non-radial masks in simulation and prototype-camera experiments with planar, dual-depth, and transparent-object scenes, we show that *radial* patterns in general reduce PSF depth dependence and already provide a clear EDoF advantage over conventional non-radial designs. Within this radial family, we introduce a radial shape parameterization and optimize the mask pattern to achieve a more broadband modulation response than handcrafted or randomly generated radial masks. As a result, the optimized radial mask simultaneously offers extended DOF and higher reconstruction quality across spatial frequencies, while remaining compatible with standard 2D

deconvolution.

Second, we generalize to *full-aperture radial masks* that extend beyond the sensor area to improve light efficiency and support wider depth of field. Using a new printed-mask prototype, we experimentally characterize PSF depth dependence and spatial shift-variance for both restricted- and full-aperture designs. This analysis reveals a trade-off: full-aperture masks improve depth robustness but introduce strong shift-variance, as different regions of the mask are projected onto the sensor. To address this, we propose an *artificially extended PSF* model, constructed by upsampling a centrally calibrated PSF to approximate the uncropped optical PSF. This enables an approximate shift-invariant forward model for full-aperture masks, implemented as a large-PSF convolution followed by cropping. Experiments with MNIST digits, QR codes, planar face images, and continuous-depth scenes demonstrate that this approximation supports all-in-focus lensless imaging while benefiting from increased light throughput.

Third, we explore *self-supervised neural reconstruction* as an alternative to supervised training. Instead of training on paired lensless–lensed images or on simulated data, we train reconstruction networks using only lensless measurements and a known forward model, by minimizing the discrepancy between measured sensor data and simulated measurements generated from the network output. We instantiate this idea with a standalone UNet and two convex–neural hybrid architectures that combine a Wiener or few-step ADMM deconvolution with a denoising UNet. Experiments on simulated and prototype-camera FFHQ datasets show that these self-supervised models can match or approach the perceptual quality of more expensive iterative methods, while reducing inference time to a few milliseconds per image and avoiding brightness/color biases potentially induced by lensed targets.

Collectively, these results demonstrate that depth-robust radial mask designs, extended-PSF modeling, and self-supervised neural priors form a coherent pathway toward practical, all-in-focus lensless imaging systems that do not rely on restrictive supervised approaches.

Contents

Abstract	i
1 Introduction	1
1.1 Background and motivation	1
1.2 Three obstacles to practical lensless cameras	3
1.3 Approach: Optics and algorithms for depth-robust lensless imaging	4
1.4 Contributions	6
1.5 Thesis organization	8
2 Theoretical background	10
2.1 Lensless imaging forward models	10
2.1.1 Full-matrix models	11
2.1.2 Separable operator models	12
2.1.3 Local convolution models	12
2.1.4 Shift-invariant convolutional models	13
2.2 Depth of field and depth-dependent forward models	14
2.2.1 PSF depth dependence	14
2.2.2 Depth-aware lensless forward model for 3D scenes	16
2.2.3 Extended depth of field and 3D modeling in lensless imaging	18
2.3 Reconstruction algorithms for noisy linear inverse imaging	19
2.3.1 Classical model-based methods	22

2.3.2	Supervised learning methods	25
2.3.3	Unsupervised and self-supervised neural reconstruction	27
3	Optimized radial mask for extended-depth-of-field lensless imaging	30
3.1	Problem statement and related work	30
3.2	Radial mask optimization	32
3.2.1	Mask parameterization	32
3.2.2	Optimization objective	34
3.2.3	Optimization analysis	35
3.3	Extended depth of field evaluation in simulations	37
3.3.1	Planar scene simulation	39
3.3.2	Dual depth scene simulation	40
3.4	Prototype lensless camera experiments	42
3.4.1	Continuous-depth experiment with a prototype camera	42
3.4.2	Reconstruction through transparent objects	45
3.5	Discussion and summary	46
4	Full-aperture radial masks for all-in-focus lensless imaging	49
4.1	Motivation and overview	49
4.2	Prototype camera and PSF characterization	51
4.2.1	Prototype camera with printed radial masks	51
4.2.2	PSF depth dependence analysis	53
4.2.3	PSF shift-variance observation	54
4.3	Shift-invariant deconvolution for small on-axis objects	56
4.3.1	MNIST simulation experiment	56
4.3.2	MNIST prototype camera experiment	59
4.4	Data-driven shift-variant deconvolution for large objects	60
4.4.1	QR code simulation experiment	60
4.4.2	QR code prototype camera experiment	62

4.5	Approximate shift-invariant deconvolution with extended PSFs	64
4.5.1	Extended-PSF forward-model approximation	64
4.5.2	Validating the extended-PSF shift-invariant model	66
4.5.3	In-focus FFHQ reconstruction: simulation and prototype	69
4.5.4	Continuous-depth scene reconstruction	71
4.6	Discussion and summary	74
5	Self-supervised lensless imaging	78
5.1	Motivation	78
5.2	Methodology	80
5.3	Experiments	83
5.3.1	Simulation experiments	83
5.3.2	Prototype camera experiments	86
5.4	Discussion and summary	88
6	Conclusion	91
	Acknowledgements	95
	Reference	97
	List of Publications	106
	Awards	107
A	Additional experiments	108
A.1	Quantization analysis via simulation	108

List of Figures

2.1	Lensless measurement formation diagram	11
2.2	PSF depth dependence illustration	15
2.3	Conceptual view of a noisy, ill-posed linear inverse imaging problem	20
3.1	Conceptual difference between radial and non-radial masks with respect to depth of field	32
3.2	Radial shape parameterization overview	33
3.3	MTF comparison between optimized radial mask and baselines (radial and non- radial)	36
3.4	Planar scene simulation to evaluate tradeoff between reconstruction quality and DOF between optimized radial and random masks	38
3.5	Dual depth scene simulated experiment setup	40
3.6	Dual depth scene simulated experiment results	41
3.7	Continuous scene experiment with a prototype camera	43
3.8	Transparent object scene experiment with a prototype camera	45
4.1	Diagram comparing depth-invariance of restricted- and full-aperture radial masks	50
4.2	PSF depth dependence analysis	52
4.3	Diagram illustrating the shift-variance of full-aperture radial PSF	55
4.4	Shift-invariant model approximation for small objects experiment	58
4.5	Data-driven shift-variant deconvolution experiment	61

4.6	Diagram describing the artificially extended PSF forward model	64
4.7	PSF size sweep analysis for the artificially extended PSF approximation	67
4.8	Planar scene experiment with extended PSF approximation	72
4.9	Continuous depth scene experiment	74
5.1	Comparison between supervised and self-supervised training approaches for lensless imaging	79
5.2	Reconstruction results for the self-supervised models for simulation and proto- type camera experiments	84
A.1	Quantization experiment diagram.	109
A.2	Quantization experiment results.	110

List of Tables

4.1	Optical parameters used to simulate PSF scaling with depth for the MNIST and QR-code experiments. Reproduced from [1].	57
4.2	FFHQ simulation analysis of in-focus reconstruction quality for restricted- and full-aperture radial masks. Reproduced from [1].	70
5.1	Quantitative analysis for simulated results on 20 reconstructions from the test set. For UDN method we show the metrics for the best reconstruction and the last (in parentheses) reconstruction over 20,000 iterations. Bold and underlined indicate best and second best methods, respectively. <i>Reproduced with permission of Springer Nature</i> [2].	85

Chapter 1

Introduction

1.1 Background and motivation

Conventional cameras rely on lenses to focus light from a three-dimensional (3D) scene onto a two-dimensional (2D) sensor. The lens both collects light and maps rays from different directions to distinct sensor locations, enabling a single snapshot to be inverted into a recognizable image with minimal computation. In contrast, lensless cameras [3, 4] replace the refractive lens with a thin coded mask placed directly in front of the sensor. The resulting measurement is a coded superposition of scene irradiance, and image formation becomes a noisy linear inverse problem in which computational reconstruction is essential for recovering a human-interpretable image.

Lensless architectures offer compelling advantages. Eliminating the lens enables extremely thin form factors [5], compatibility with non-visible spectral bands [6–9], and optical designs that are difficult or impossible to realize using conventional optics [10, 11]. These benefits are most compelling for space-constrained, close-up imaging where object distance can vary over centimeters and mechanical focusing is impractical. This includes applications in biomedical imaging, such as distal computational microendoscopy [12] and in-vivo lensless microscopy [13], where working distances are of centimeters and scene depth varies rapidly. Similar constraints appear in close-range visible-light sensing, including non-contact scanning with millimeter-

scale standoff variation [14] and QR-code recognition under uncertain target distance [15]. Beyond the visible spectrum, lensless designs are also being explored for long-wave infrared sensing, where refractive optics can be bulky or costly [16]. At the same time, removing the focusing element (i.e., lens) makes image formation highly geometry-dependent: the PSF can vary with depth and position, and reconstruction quality hinges on controlling and modeling these variations precisely. This is one of the motivations for the depth-robust optical design and physics-constrained reconstruction developed in this thesis.

Despite rapid progress in coded-mask design and reconstruction algorithms, most existing lensless systems remain tailored to constrained laboratory scenarios. Typical 2D deconvolution demonstrations assume a planar scene at a fixed distance and rely on supervised models trained with paired lensless–lensed captures, often acquired through beam-splitter setups [17, 18]. In contrast, 3D methods jointly estimate depth and intensity using multi-plane forward models, but these approaches generally require dense depth-wise PSF calibration, strong sparsity assumptions, and slow iterative solvers [19, 20]. Recent learning-based 3D techniques alleviate some of these burdens by training on synthetic volumetric datasets, yet they depend heavily on forward-model accuracy and often exhibit a performance gap when applied to real scenes [21–23]. Taken together, current pipelines either operate under restrictive conditions or incur substantial calibration and computational overhead, limiting their practicality in real-world settings where scenes span a range of depths and imaging conditions vary over time.

The goal of this thesis is to move lensless imaging toward a more practical regime where a single mask and reconstruction pipeline can operate robustly across depth variation without relying on lensed supervision. Rather than estimating depth explicitly or depending on extensive simulated training data, we pursue an alternative strategy: shaping the optical design and reconstruction model so that depth-dependent effects are suppressed, enabling all-in-focus 2D imaging for real scenes. This perspective motivates the three obstacles and corresponding solutions developed throughout this work.

1.2 Three obstacles to practical lensless cameras

Achieving this goal is challenging because lensless cameras face three tightly coupled obstacles:

- **Obstacle A: Limited depth of field.** In a naive lensless system, a single PSF is calibrated at one depth, and reconstruction assumes that the entire scene lies on that plane. In reality, scene points at different distances produce different PSFs; using a mismatched PSF leads to defocus-like artifacts and rapidly degrading image quality. Thus, the effective depth of field (DOF) of a standard lensless camera is limited, especially for close-range scenes where PSF scaling with depth is most pronounced.
- **Obstacle B: Model breakdown for large apertures and extended scenes.** Many reconstruction methods rely on a 2D *shift-invariant* convolution model, where the measurement is modeled as a convolution between the scene and a single PSF. This approximation can hold reasonably well when the coded mask occupies a restricted region near the sensor center and the field of view is modest. However, when the mask extends over the full aperture, as desired to collect more light, the PSF becomes strongly *shift-variant*: point sources at different locations see different portions of the mask pattern, and the measured PSFs change shape across the sensor. For large objects and wide fields of view, this invalidates the simple convolutional forward model and undermines standard deconvolution strategies.
- **Obstacle C: Slow and supervision-dependent reconstruction.** Classical model-based methods (e.g., ADMM with total variation regularization) provide physically faithful reconstructions but require many iterations per image and careful tuning, leading to high computational cost. Recent learning-based methods achieve impressive quality and speed by training deep networks to map lensless measurements to sharp images, but almost all rely on paired supervision: they either assume access to lensed reference images captured in parallel with lensless measurements or rely on simulation-approaches to generate the training data. This supervision can be expensive to acquire, may not reflect the long-term

deployment environment, and can introduce biases (e.g., brightness or color mismatch) that are hard to control.

We argue that these three obstacles are not independent. Depth-dependent PSFs (Obstacle A) motivate complex 3D forward models, whose size and shift-variance (Obstacle B) make reconstruction harder and more ill-conditioned, increasing the reliance on powerful priors or supervised training (Obstacle C). This thesis argues that significant progress can be made by jointly shaping both the *optical* design (the coded mask and camera geometry) and the *reconstruction* strategy, with a particular emphasis on:

- designing mask patterns that are inherently more robust to depth variation;
- developing approximations that recover a usable 2D shift-invariant model even for full-aperture masks; and
- leveraging self-supervised neural priors that avoid lensed ground-truth images while enabling fast inference.

1.3 Approach: Optics and algorithms for depth-robust lensless imaging

To address the above obstacles, this thesis takes a two-step approach, alternating between optical design and computational reconstruction.

From depth modeling to depth-robust masks. Chapter 2 develops the forward model perspective used throughout the thesis. We start by formulating lensless imaging as a linear operator mapping the scene to the sensor, both for planar scenes and for 3D scenes modeled as stacks of depth planes. A simple geometric model clarifies how the PSF scales with depth and why DOF collapses for near-field objects. We then review existing extended-DOF and 3D lensless approaches, distinguishing between methods that explicitly reconstruct multi-plane intensity and depth, and those that aim to engineer the PSF so that a single 2D model remains effective over a

wider depth range. This background motivates the search for mask designs that are intrinsically more depth-robust, and for reconstruction algorithms that can exploit such designs.

Obstacle A: Extending DOF with restricted-aperture radial masks. Chapter 3 tackles the limited-DOF problem under a restricted-aperture configuration, where the coded mask occupies a central region of the sensor. We investigate *radial* coded masks whose pattern is invariant under uniform scaling. The key idea is that if the PSF is invariant to uniform scaling, then variations in object distance do not affect the lensless measurement as severely. We introduce a radial mask parameterization, optimize the mask pattern within this parameterized framework, and compare the resulting design against random and other radial/non-radial baselines. Through simulated and prototype-camera experiments (including planar scenes, dual-depth scenes, and transparent-object setups) we show that the optimized radial mask offers significantly improved DOF while still allowing reconstruction via a simple 2D shift-invariant deconvolution model.

Obstacle B: Full-aperture radial masks and extended PSFs. Chapter 4 extends this idea to full-aperture radial masks that cover the entire sensor area. Such masks are desirable because they collect more light and allow larger fields of view, but they also introduce strong PSF shift-variance as different scene points illuminate different mask regions. We first build an improved prototype camera with lithographically printed masks and characterize its PSFs across depth and across the sensor. Comparing restricted- and full-aperture masks, we show experimentally that full-aperture radial masks yield PSFs that change more slowly with depth but more strongly with lateral translations.

To reconcile this with shift-invariant deconvolution, we then propose a novel approximation: an *artificially extended PSF*. By upsampling a centrally calibrated PSF beyond the sensor area, we approximate the “optical” PSF — the full shadow of the mask — and model shift-variance as a combination of shift-invariant convolution followed by cropping. Crucially, the scale-invariant geometry of radial masks makes this extension physically meaningful, which is not the case for generic mask patterns. We validate this approximation via PSF size sweeps, planar face-image experiments (FFHQ) and continuous-depth scenes, demonstrating that full-aperture radial masks can achieve all-in-focus behavior comparable to restricted masks while offering

improved light efficiency.

Obstacle C: Fast, self-supervised reconstruction with neural priors. Chapter 5 addresses the remaining obstacle: how to reconstruct lensless images quickly and robustly without relying on supervised training. We frame lensless reconstruction as a noisy linear inverse problem and review three families of reconstruction methods: classical model-based, supervised learning-based, and unsupervised/self-supervised approaches. Building on this, we propose self-supervised training schemes in which a reconstruction network is optimized only using lensless measurements and the known forward model. The loss compares measured sensor data to synthetic measurements obtained by passing the network output through the lensless forward operator, ensuring physics consistency without requiring ground-truth lensed images.

We instantiate this paradigm using three architectures: a standalone UNet, and two hybrid models that compose classical methods (Wiener deconvolution or a small number of ADMM iterations) with a learned UNet denoiser. Using simulated and prototype-camera FFHQ datasets captured with the same radial mask design as in previous chapters, we compare these self-supervised models against purely convex optimization baselines. The results show that hybrid self-supervised models can approach or surpass the perceptual quality of more expensive iterative schemes while reducing inference time to a few milliseconds per image, and they avoid the brightness and color biases that can arise when training on lensed supervision.

1.4 Contributions

This thesis advances lensless imaging toward practical all-in-focus 2D operation through joint optics-algorithm design. Its primary contributions span optical design, forward-model approximation, and self-supervised reconstruction.

1. Depth-robust restricted-aperture radial masks. We investigate radial coded masks whose patterns are invariant under uniform scaling and show that such designs intrinsically reduce PSF sensitivity to depth. Building on this insight, we propose a radial mask parameterization and

develop an optimization framework that discovers radial patterns with broadband modulation response. Simulated and prototype-camera experiments demonstrate that the optimized radial mask extends depth of field and improves reconstruction quality over both non-radial designs and handcrafted radial patterns, while remaining compatible with fast 2D shift-invariant deconvolution.

2. Full-aperture radial masks and an extended-PSF forward-model approximation. We generalize radial-mask design to full-aperture configurations that increase light throughput and support wider depth ranges, but introduce strong PSF shift-variance. Through experimental characterization, we analyze the depth and spatial variability of these PSFs and identify a fundamental trade-off between depth robustness and shift-invariance. To recover a tractable 2D model, we introduce an *artificially extended PSF*, obtained by upsampling a centrally calibrated PSF to approximate the uncropped optical footprint of the mask. This leads to a simple large-PSF convolution followed by cropping formulation that approximates shift-variant imaging with a shift-invariant operator. Experiments with QR codes, MNIST digits, planar face images, and continuous-depth scenes validate the feasibility of all-in-focus lensless imaging with full-aperture radial masks.

3. Self-supervised neural reconstruction for lensless imaging. To avoid the limitations of paired supervision and simulation-based training, we develop self-supervised reconstruction methods that rely solely on measured lensless images and the known forward model. We instantiate this approach using (i) a standalone UNet and (ii) two physics-informed hybrid architectures combining Wiener or few-step ADMM deconvolution with a denoising UNet. These models are trained by enforcing measurement consistency in the image formation process, without requiring lensed ground truth. Experiments on simulated and prototype-camera FFHQ datasets show that the resulting reconstructions match or approach the quality of slower iterative optimization methods while achieving millisecond-scale inference and avoiding potential brightness or color biases induced by supervised training.

Overall. Collectively, these contributions demonstrate that depth-robust optical design, tractable forward modeling for full-aperture masks, and self-supervised neural priors form a coherent pathway toward practical, all-in-focus lensless cameras that operate without restrictive supervised pipelines or explicit depth estimation.

1.5 Thesis organization

The remainder of this thesis is organized as follows.

- **Chapter 2** introduces the forward models and reconstruction algorithms used throughout the thesis. We derive 2D and 3D lensless forward models, illustrate PSF depth dependence, review existing extended-DOF and 3D lensless techniques, and discuss classical, supervised, and unsupervised reconstruction methods in the general linear inverse imaging framework.
- **Chapter 3** focuses on restricted-aperture radial coded masks as a means of extending DOF while retaining a simple 2D shift-invariant convolution model. We present the radial parameterization and optimization, analyze PSF depth robustness, and validate the design through simulations and prototype-camera experiments, including continuous-depth and transparent-object scenes.
- **Chapter 4** extends the radial mask concept to full-aperture designs and examines their depth and shift-variance properties using a new printed-mask prototype camera. We introduce the extended-PSF approximation that enables shift-invariant deconvolution with full-aperture masks and demonstrate all-in-focus reconstruction for planar and continuous-depth scenes.
- **Chapter 5** investigates self-supervised reconstruction for lensless imaging. We formulate self-supervised training objectives based on measurement-domain consistency, instantiate them with UNet and convex–neural hybrid architectures, and compare their performance

against classical optimization-based methods on both simulated and prototype-camera FFHQ datasets.

- **Chapter 6** concludes the thesis by summarizing the main findings, discussing limitations, and outlining future research directions toward practical lensless cameras that can operate robustly in complex real-world environments.

Chapter 2

Theoretical background

2.1 Lensless imaging forward models

In this thesis, we first consider lensless cameras in a two-dimensional, single-plane setting. Specifically, we model the object as a single irradiance distribution on a plane at a fixed distance from the mask, and study how this plane is mapped to the sensor. As illustrated in Fig. 2.1, the resulting measurement can be interpreted as a superposition of position-dependent point spread functions (PSFs) generated by individual bright points on this object plane. More specifically, let $\mathbf{x} \in \mathbb{R}^N$ denote a discretized representation of this object-plane irradiance, $\mathbf{y} \in \mathbb{R}^M$ the corresponding sensor measurements, and $\mathbf{n} \in \mathbb{R}^M$ an additive noise term; the lensless imaging process under incoherent lighting conditions can then be written in the general linear form

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \quad (2.1)$$

where $\mathbf{A} \in \mathbb{R}^{M \times N}$ is the forward operator that encodes how each object-plane element contributes to each sensor pixel, including the effects of the coded mask, sampling, cropping, and any shift- or position-dependent variation in the PSF. In the remainder of this section, we review several common structural assumptions on $\mathbf{A}$, from full matrices to convolutional forms, that simplify the model to different degrees and have direct implications for memory usage, computational cost, and reconstruction algorithm design. All of these formulations assume a two-

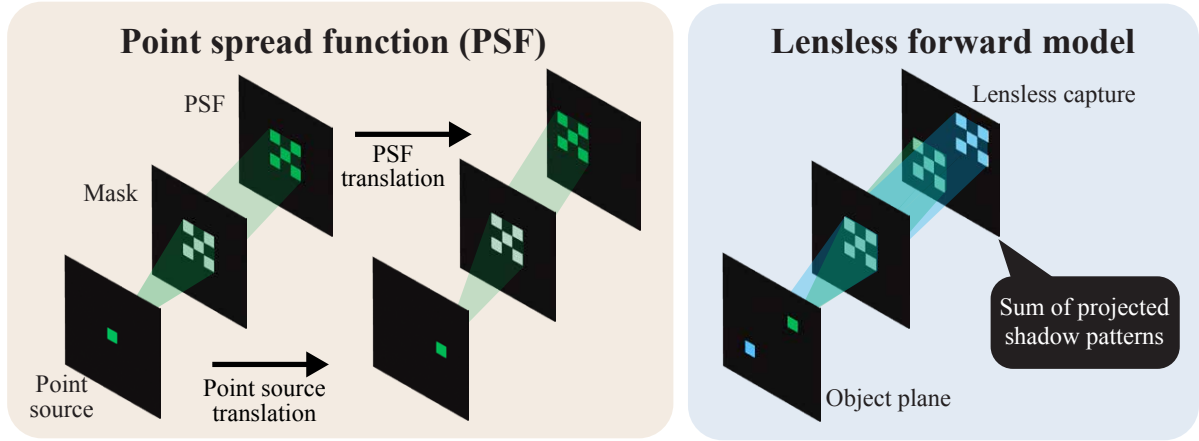


Figure 2.1: Lensless measurement formation as a superposition of shifted point spread functions (PSFs). Left: a single bright point source on the object plane casts a coded shadow of the mask onto the sensor; translating the point source laterally translates the PSF. Right: with multiple bright sources, the sensor measurement is the sum of their projected PSFs, forming the basic linear forward model used in this thesis.

dimensional, single-plane scene; in Sec. 2.2, we extend this framework to three-dimensional settings with multiple depth planes and depth-dependent PSFs.

2.1.1 Full-matrix models

The most general discrete representation of (2.1) treats $\mathbf{A}$ as a generic matrix. For a fixed object plane at a given distance, the full-matrix model can, in principle, capture arbitrary space-variant behavior, including diffraction patterns, and vignetting effects. This makes it an attractive conceptual starting point for analyzing the conditioning of the 2D inverse problem [5].

However, full-matrix models become quickly impractical for realistic sensor resolutions. For an object plane and measurement with resolution of 1024×1024 pixels, a fairly low resolution as per current standards, $\mathbf{A}$ would have on the order of 10^{12} entries, far beyond the memory capacity of typical embedded systems. Moreover, explicitly storing and multiplying by a dense $\mathbf{A}$ makes fast algorithms difficult to exploit, and increase the complexity of PSF-based cali-

bration. For these reasons, full-matrix models are rarely used directly for reconstruction, and are instead replaced by more structured representations that trade some modeling flexibility for computational tractability in the 2D plane. When moving to 3D, one would conceptually stack such operators across depth planes, further increasing memory and computational loads, as discussed in Sec. 2.2.

2.1.2 Separable operator models

A simplification of the full-matrix modeling is to assume that the forward operator is separable across spatial dimensions. A seminal work on this line of research is that of the doubly Toeplitz separable mask [24]. This approach leverages a coded mask with a separable pattern along vertical and horizontal directions, so that $\mathbf{A}$ can be approximated by separate convolutions as

$$\mathbf{y} = \mathbf{A}_L \mathbf{x} \mathbf{A}_R^T + \mathbf{n} \quad (2.2)$$

where $\mathbf{A}_L$ and $\mathbf{A}_R$ act along rows and columns of $\mathbf{x}$, respectively. Separable models significantly reduce storage requirements, enabling both iterative optimization methods [5, 25] and single-step inversion [18] methods.

Despite these advantages, strict separability remains an approximation for practical lensless cameras, and severely restricts the design space of the coded pattern of the mask, not being useful for general non-rectangular mask patterns (e.g., random or radial designs). Furthermore, despite its improvement in memory usage, its application to 3D scenes would require a stacking of these matrices along depth dimension further increasing computational cost. Consequently, we regard separable models as useful conceptual tools, but not as the primary modeling choice for the systems considered in this thesis.

2.1.3 Local convolution models

A more flexible and physically grounded alternative for a single object plane is to model lensless imaging as a space-variant convolution. In this approach, each point on the plane produces a

localized point spread function (PSF) on the sensor that varies minimally for adjacent regions of the object plane. Following this idea, the forward model can be approximated as a sum of patch-wise convolutions as

$$y = \sum_p \mathbf{h}_p * \mathbf{x}_p + \mathbf{n}, \quad (2.3)$$

where $\mathbf{x}_p$ denotes the p -th patch in the object-plane and $\mathbf{h}_p$ is the approximate PSF observed for the patch p of adjacent elements in the object plane.

Local convolution models strike a balance between expressiveness and efficiency. They can model space-variance due to off-axis aberrations or sensor nonuniformities, while still allowing reconstruction via convolution-like operations within each patch that can be performed efficiently in the frequency domain. These models have been used in traditional reconstruction settings [26] and single-step inversion [27]. However, their use comes with limitations. First, they require the calibration and storage of multiple PSFs, increasing both experimental and computational complexity [26, 28]. Second, patch boundaries can introduce stitching artifacts due to PSF discontinuity between adjacent patches [29, 30]. Third, extending this framework to 3D would also require a separate bank of local PSFs for each depth plane, further increasing calibration and storage demands.

2.1.4 Shift-invariant convolutional models

A particularly appealing special case, and the one adopted in this thesis as the baseline 2D model, is the shift-invariant convolutional formulation [31]. Here, the forward operator $\mathbf{A}$ corresponds to a discrete convolution with a single PSF $\mathbf{h}$ associated with a fixed object plane,

$$\mathbf{y} = C[\mathbf{h} * \mathbf{x}] + \mathbf{n}, \quad (2.4)$$

where $*$ denotes two-dimensional convolution, and $C(\cdot)$ is a cropping function that models the finite size of the image sensor. In practice, (2.4) can be efficiently computed via FFT-based circular convolution with appropriate padding and cropping [32]. The shift-invariant model assumes shift invariance of the PSF, meaning that the PSF does not depend on the spatial

position of the scene point, only on its relative displacement with respect to the sensor. This approximation on a single plane is justified under the following conditions: (i) Small distance between coded mask and image sensor, thus minimizing off-axis aberrations; (ii) restricted-aperture mask allowing the full mask pattern to be observed from an on-axis calibrated PSF.

The shift-invariant convolutional model offers several advantages in the 2D setting. It allows the forward operator to be represented compactly by a single PSF, drastically reducing memory requirements and enables efficient computation via FFTs, which is crucial for both iterative reconstruction and recent learning-based methods [17, 18, 27, 33]. In Sec. 2.2, we extend this formulation to three-dimensional scenes, discuss how depth-information on the measurements limits the DOF of general lensless cameras, and how previous methods in the literature addressed this challenge to achieve extended depth of field.

2.2 Depth of field and depth-dependent forward models

The 2D single-plane models in Sec. 2.1 assume that the scene lies on a plane at a fixed distance from the coded mask, so that the measurements are described by a forward operator $\mathbf{A}$ that models exclusively translations and shift-variant effects on a single depth plane. Real scenes, however, occupy a volume, and objects at different depths produce different patterns (i.e., PSF) on the sensor [19, 20]. In this section, we first illustrate the geometric origin of PSF change with depth in lensless cameras, then extend the single-plane formulation in (2.4) to three-dimensional scenes. This depth-aware viewpoint allows us to define depth of field (DoF) in lensless imaging in terms of PSF-model mismatch, and to review existing strategies for extending DoF, with particular emphasis on depth-robust mask designs that motivate the contributions of this thesis.

2.2.1 PSF depth dependence

The effective depth of field (DoF) of a lensless camera is closely tied to how its point spread function (PSF) scales with scene depth. To make this relationship explicit, we consider a simple geometric ray model of PSF formation for a coded mask system, illustrated in Fig. 2.2(a). Let

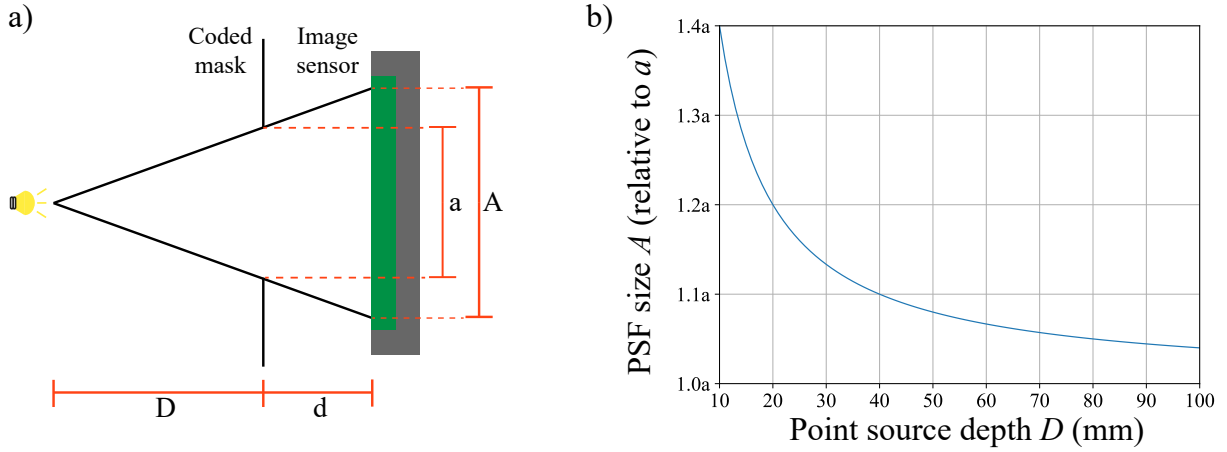


Figure 2.2: (a) Diagram illustrating the relationship between the light source depth D and the projected PSF size A . (b) Graph showing how the PSF size A scales depending on the point source depth D .

a denote the lateral size of the mask aperture, d the distance between mask and sensor, D the distance between a point source and the mask, and A the lateral size of the PSF observed on the sensor. We assume a square sensor and mask aperture for simplicity.

Geometric PSF scaling model. Under this geometric model, rays passing through the edges of the aperture form similar triangles between the point source and the sensor plane. From this similarity, the PSF size on the sensor can be approximated as

$$A = a \left(1 + \frac{d}{D} \right), \quad (2.5)$$

which captures the inverse dependence of PSF scaling on the object distance D . As the point source moves closer to the mask (smaller D), the PSF size A grows rapidly; conversely, for distant points (large D), changes in depth produce only modest variations in PSF size.

Figure 2.2(b) illustrates this relationship for typical mask and sensor parameters. The curves are computed from Eq. (2.5) using a mask–sensor distance of 4 mm. The nonlinearity is evident: small depth changes near the mask cause large changes in A , whereas the PSF becomes relatively stable for larger values of D .

Depth sensitivity and aperture size. The same model also clarifies how depth sensitivity

depends on the optical geometry. Considering two depths D and D' , the ratio of PSF sizes is

$$\frac{A}{A'} = \frac{a \left(1 + \frac{d}{D}\right)}{a \left(1 + \frac{d}{D'}\right)} = \frac{1 + \frac{d}{D}}{1 + \frac{d}{D'}}. \quad (2.6)$$

The ratio of change becomes independent of the aperture term a , showing that the *relative* change in PSF size between depths is governed only by the distances d and D . For a fixed mask–sensor spacing d , the system is therefore most sensitive to depth variations in the near field (small D), while far-field depth changes (large D) induce only mild PSF scaling. This geometric depth sensitivity underlies the strong DoF limitations often observed for close-range objects in lensless systems.

Implications for general coded-mask design. Although Fig. 2.2(a) is drawn for a simple restricted aperture, the same scaling principle applies to general coded masks (e.g., full-aperture masks) whose patterns do not enforce any depth-invariant structure. In such designs, the spatial variation of the pattern can be viewed as an ensemble of local apertures, each contributing to depth-dependent changes in the PSF. In the following subsections, we embed this per-point behavior into depth-aware forward models for 3D scenes and use that viewpoint to formalize depth of field in lensless imaging and motivate depth-robust mask designs.

2.2.2 Depth-aware lensless forward model for 3D scenes

We now embed the depth-dependent PSF behavior from Sec. 2.2.1 into a forward model for three-dimensional scenes. We model the scene as a stack of K object planes at discrete depths $\{z_k\}_{k=1}^K$ relative to the coded mask. Let $\mathbf{x}^{(k)} \in \mathbb{R}^N$ denote the discretized irradiance on the k -th plane, and let $\mathbf{A}^{(k)}$ denote the linear forward operator that maps $\mathbf{x}^{(k)}$ to the sensor for that depth.

General 3D linear model. Under incoherent imaging and a linear sensor response, the contributions of different depth planes add linearly. The total measurement can therefore be written as

$$\mathbf{y} = \sum_{k=1}^K \mathbf{A}^{(k)} \mathbf{x}^{(k)} + \mathbf{n}, \quad (2.7)$$

where $\mathbf{n}$ represents measurement noise and model mismatch. Each operator $\mathbf{A}^{(k)}$ may be fully

shift-variant and can, in principle, capture diffraction, occlusion, mask–sensor geometry, and any additional optical behaviour from the mask and image sensor [19].

By stacking all planes into a single vector

$$\tilde{\mathbf{x}} = [(\mathbf{x}^{(1)})^\top, \dots, (\mathbf{x}^{(K)})^\top]^\top,$$

(2.7) can be written compactly as

$$\mathbf{y} = \mathbf{A}_{3D}\tilde{\mathbf{x}} + \mathbf{n}, \quad \mathbf{A}_{3D} = [\mathbf{A}^{(1)} \ \dots \ \mathbf{A}^{(K)}], \quad (2.8)$$

where $\mathbf{A}_{3D}$ concatenates the depth-specific operators along columns. The overall 3D operator is substantially larger than its 2D counterpart, which directly impacts memory usage and the cost of applying $\mathbf{A}_{3D}$ in most reconstruction settings.

Shift-invariant convolutional approximation. In many lensless imaging systems, a common simplification is to assume that, for each depth plane z_k , the system is approximately shift invariant over the field of view. In this case, the operator $\mathbf{A}^{(k)}$ reduces to a 2D convolution with a depth-dependent PSF $\mathbf{h}^{(k)}$, yielding

$$\mathbf{y} = C \left[\sum_{k=1}^K \mathbf{h}^{(k)} * \mathbf{x}^{(k)} \right] + \mathbf{n}, \quad (2.9)$$

where $*$ denotes 2D convolution. This multilayer convolutional model can be viewed as a direct extension of the single-plane formulation in (2.4) to a stack of depth slices, each with its own PSF.

The depth dependence encoded in $\{\mathbf{h}^{(k)}\}$ reflects how the PSF shape change with z_k , due to geometric scaling (Sec. 2.2.1), diffraction effects, or additional optical components related to depth variation. For generic mask patterns, these PSFs can vary significantly with depth, leading to strongly depth-dependent blur and making reconstruction of extended 3D scenes ill-conditioned unless strong priors, multiple measurements, or task-specific constraints are imposed. In the following subsection, we use this multilayer model to formalize depth of field in lensless cameras in terms of the mismatch between the true depth-dependent operator and the single-PSF models commonly used in practice.

2.2.3 Extended depth of field and 3D modeling in lensless imaging

The depth-aware model in Sec. 2.2.2 shows that a general 3D scene is mapped to the sensor through a superposition of depth-dependent PSFs. In principle, one could follow this full 3D structure and attempt to recover all planes $\{\mathbf{x}^{(k)}\}$, together with their associated PSFs. In practice, however, such 3D lensless reconstructions are costly: they require depth-wise calibration, larger forward operators, and stronger priors, and are often unnecessary when the goal is simply to keep objects visually sharp over a range of depths. In this context, EDoF techniques for lensless imaging are naturally divided into two broad families: (1) *3D-based* methods that explicitly model and jointly recover intensity and depth, and (2) *2D-based* methods that focus on designing the PSF, so that a single 2D forward model remains valid over an extended depth range.

(3D) Joint intensity–depth reconstruction. A first class of methods embraces the full 3D structure of the forward model and aims to reconstruct both intensity and depth (or multiple depth slices) from the lensless measurement. Under the multi-plane formulation of Eq. (2.9), these approaches seek estimates of $\{\mathbf{x}^{(k)}\}$, sometimes together with a continuous depth map, in a depth-from-defocus (DFD) or light-field-like fashion. To mitigate the severe ill-posedness of this high-dimensional inverse problem, they typically increase the number of measurements or the expressivity of the forward operator, for example by:

- time-multiplexing coded patterns or variable-illumination conditions to enrich lensless measurements [34–36],
- calibrating PSFs at multiple depths and using them within an explicit 3D reconstruction algorithm [19, 20, 37],
- Data-driven optimization approaches, including end-to-end deep-optics frameworks [22, 23, 38–40].

These methods can achieve impressive 3D reconstructions or strong EDoF, but at the cost of increased system complexity: they often require multiple measurements, dense calibration

across depth, strong priors, and heavy computation. Moreover, the resulting pipelines may be tightly coupled to specific training datasets and hardware configurations, which can limit their deployability in unconstrained, general-purpose settings.

(2D) PSF design for depth-robust 2D models. A different strategy is to retain the simplicity of a single 2D forward model and instead engineer masks and optics so that the effective PSF changes more slowly with depth over a target range. In this view, the goal is not to reconstruct $\{\mathbf{x}^{(k)}\}$ plane by plane, but to design a coded mask for which a single calibrated PSF $\mathbf{h}$ remains a good approximation for all relevant depths $z \in [z_{\min}, z_{\max}]$. When successful, this allows the use of standard 2D deconvolution or learning-based reconstruction methods while still covering a substantial depth range.

This approach focuses on designing coded apertures whose patterns are more robust to depth-induced scaling or distortion [10, 22, 41–44]. In these systems, the effective EDoF is determined by how slowly the PSFs $\mathbf{h}^{(k)}$ deviate from one another as z_k varies, so that residual mismatch can be absorbed by the reconstruction algorithm without severe artifacts.

Positioning of this thesis. The work in this thesis is primarily aligned with the second strategy, but informed by the 3D perspective above. In Chapters 3–4, we explore *radial* mask designs that aim to suppress explicit depth cues in the measurements, so that a single 2D forward model can be used over an extended depth range. We show that such masks can effectively trade some in-focus sharpness for greatly improved robustness to PSF mismatch, enabling wide-DoF reconstructions even with simple 2D algorithms.

2.3 Reconstruction algorithms for noisy linear inverse imaging

Figure 2.3 illustrates the core difficulty of reconstruction in lensless imaging and related inverse imaging problems. A single noisy measurement $\mathbf{y}$ typically corresponds to many candidate solutions $\mathbf{x}$ that are consistent with the forward model, so the data alone cannot uniquely deter-

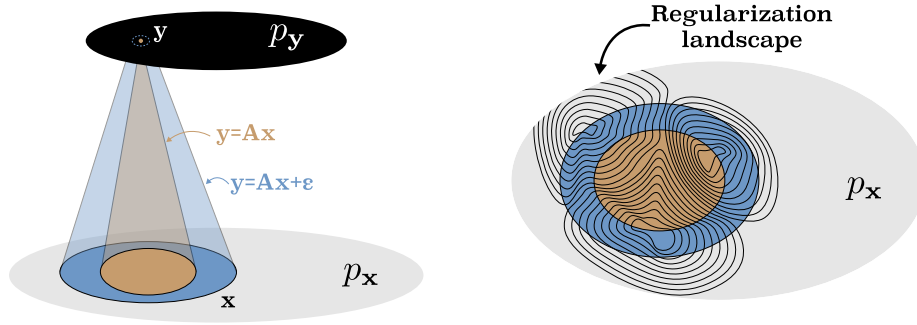


Figure 2.3: Conceptual view of a noisy, ill-posed linear inverse imaging problem. Left: many candidate images x can produce the same measurement y through the forward model. Right: adding a regularizer (prior) introduces additional information to restrict the search for solutions with desirable properties.

mine the scene. Regularization (or priors) effectively shrinks this set of feasible solutions to a smaller subset that reflects desirable image structure, such as smoothness, sparsity, or realism.

Lensless reconstruction can be viewed as one instance of a broader class of noisy linear inverse imaging problems. In this setting, a general measurement process is modeled as described in Eq. 2.1:

$$y = Ax + n, \quad (2.10)$$

where x denotes the information to be reconstructed (e.g., a 2D image or a stack of depth planes), A is a general linear forward operator, and n represents measurement noise and model mismatch. In the context of lensless imaging, A encodes the effect of the coded mask, propagation to the sensor, and any additional optics as shown in Sec. 2.1 and 2.2. The same formulation, however, also covers other thoroughly well-researched inverse imaging tasks such as motion deblurring, super-resolution and inpainting.

Ill-posed tasks. Even in the idealized noiseless setting $y = Ax$, many imaging operators A are rank-deficient. In this case, the measurement y does not uniquely determine x . Let x_p be any particular solution satisfying $y = Ax_p$, and let $\text{null}(A)$ denote the null space of A . As such, given a null-space element $v \in \text{null}(A)$ with $Av = 0$, all possible solutions to the ill-posed

inverse imaging can be written as

$$\mathbf{x} = \mathbf{x}_p + \mathbf{v}. \quad (2.11)$$

Equivalently, leveraging range-null decomposition [45, 46], if we decompose $\mathbf{x}$ into a component in the row space of $\mathbf{A}$ and a component in its null space,

$$\mathbf{x} = \mathbf{x}_R + \mathbf{x}_N, \quad \mathbf{x}_R \in \text{range}(\mathbf{A}^T), \mathbf{x}_N \in \text{null}(\mathbf{A}),$$

then the measurement depends only on $\mathbf{x}_R$:

$$\mathbf{y} = \mathbf{A}\mathbf{x}_R + \mathbf{A}\mathbf{x}_N = \mathbf{A}\mathbf{x}_R.$$

The null-space component $\mathbf{x}_N$ is completely invisible to the sensor. Thus, even with perfect, noiseless measurements, there are infinitely many scenes $\mathbf{x}$ consistent with the same $\mathbf{y}$ whenever $\text{null}(\mathbf{A}) \neq \{\mathbf{0}\}$.

Noise amplification and instability. In realistic settings, the measurements are corrupted by noise,

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \quad (2.12)$$

which further complicates reconstruction. Formally, even the row-space component $\mathbf{x}_R$ becomes difficult to recover stably: small perturbations in $\mathbf{y}$ can lead to large errors in any naive inversion of $\mathbf{A}$, especially along directions associated with small singular values. For example, using the Moore–Penrose pseudoinverse $\mathbf{A}^\dagger$ yields

$$\hat{\mathbf{x}} = \mathbf{A}^\dagger \mathbf{y} = \mathbf{A}^\dagger \mathbf{A}\mathbf{x} + \mathbf{A}^\dagger \mathbf{n},$$

so that noise is amplified in proportion to the inverse singular values of $\mathbf{A}$ [46]. As a result, unregularized solutions tend to be both non-unique (due to the null space) and numerically unstable (due to noise amplification), leading to poor reconstructions.

Role of priors. To obtain useful reconstructions, one must therefore combine the physics encoded in $\mathbf{A}$ with some form of prior information on $\mathbf{x}$. Intuitively, the data term constrains the component of $\mathbf{x}$ that is visible and stable through $\mathbf{A}$, while the prior controls how the remaining

degrees of freedom are filled in, especially those in or near the null space and along poorly conditioned directions. Different reconstruction paradigms embody this principle in different ways: classical optimization methods impose hand-crafted regularizers (e.g., 2D total variation [47]); supervised deep networks learn a direct mapping from paired examples and implicitly encode a prior through their training data; and more recent unsupervised approaches use the forward model itself (i.e., $\mathbf{A}$), together with learned priors, to guide reconstruction without requiring paired ground-truth images.

Three families of reconstruction methods. In the remainder of this section, we group reconstruction algorithms for (2.10) into three broad families:

- *Classical model-based methods*, which explicitly minimize a regularized objective that balances measurement consistency with hand-crafted priors;
- *Supervised learning methods*, which learn a parametric mapping from measurements to target images using paired training data with implicit regularization;
- *Self-supervised and unsupervised methods*, which exploit the known forward model, often together with powerful neural priors, to reconstruct images without explicit ground-truth supervision.

All three families are applicable to both 2D and 3D linear models, provided that the forward operator $\mathbf{A}$ is known or can be approximated. In the following subsections, we review each class.

2.3.1 Classical model-based methods

Classical reconstruction methods approach (2.10) by explicitly trading off measurement consistency and prior information through an optimization problem. A common starting point is the least-squares data term

$$\|\mathbf{y} - \mathbf{Ax}\|_2^2,$$

which enforces consistency between $\mathbf{Ax}$ and the measurement $\mathbf{y}$, combined with a regularizer $\Psi(\mathbf{x})$ that encodes prior assumptions on the image (e.g., smoothness or sparsity). The reconstruction is then obtained as

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \left\{ \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \tau \Psi(\mathbf{x}) \right\}, \quad (2.13)$$

where the regularization coefficient $\tau > 0$ balances data fidelity against prior strength. In lensless imaging, $\mathbf{A}$ is the forward operator derived from the coded-mask model (e.g., a 2D convolution with a PSF or a depth-dependent operator), so (2.13) directly couples the optical design with the reconstruction algorithm.

Linear inverse filtering. In the special case where $\mathbf{A}$ is a shift-invariant convolution with a known PSF $\mathbf{h}$ (see Eq. (2.4)), and the regularizer in (2.13) is a quadratic Tikhonov penalty $\Psi(\mathbf{x}) = \|\mathbf{x}\|_2^2$, the problem admits a closed-form solution as [48]

$$\hat{\mathbf{x}} = \mathcal{F}^{-1} \left(\frac{\mathcal{F}(\mathbf{h})^*}{|\mathcal{F}(\mathbf{h})|^2 + \lambda} \mathcal{F}(\mathbf{y}) \right), \quad (2.14)$$

where $\mathcal{F}(\cdot)$ and $\mathcal{F}^{-1}(\cdot)$ denote the discrete Fourier transform and its inverse, $(\cdot)^*$ denotes complex conjugate, and λ is a (scalar) regularization parameter that damps the reconstruction to avoid noise amplification. Intuitively, the filter in (2.14) inverts the blur where $|\mathcal{F}(\mathbf{h})|$ is large, and suppresses frequencies where $|\mathcal{F}(\mathbf{h})|$ is small and inversion would otherwise strongly amplify noise.

In lensless imaging, this Wiener-like deconvolution is often used as a fast baseline: given a calibrated PSF for a particular depth, one can compute an approximate reconstruction via a single forward–inverse FFT pair. However, because the filter is linear and tied to a fixed PSF, it struggles with strong depth dependence and shift dependence, as well as large field-of-view scenes or large-aperture masks that are more susceptible to cropping effects, often yielding noisy or over-smoothed results. These limitations motivate more flexible iterative formulations.

Iterative regularized reconstruction. When reconstruction conditions are not sufficient for linear inverse filtering, one solves (2.13) by iterative optimization. The basic idea is to start from an initial guess $\mathbf{x}^{(0)}$ and refine it by repeatedly applying updates that reduce the objective while respecting both the data-fidelity and the regularization prior terms.

A simple example is gradient descent on the smooth part of the objective. If we discard the regularization, we have

$$f(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2,$$

then the gradient of f is

$$\nabla f(\mathbf{x}) = \mathbf{A}^\top (\mathbf{A}\mathbf{x} - \mathbf{y}),$$

which measures the mismatch between simulated and observed measurements, back-projected through $\mathbf{A}^\top$. A gradient step

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - \eta \nabla f(\mathbf{x}^{(t)})$$

reduces this mismatch by moving $\mathbf{x}$ in the direction that best aligns $\mathbf{A}\mathbf{x}$ with $\mathbf{y}$, with the step size η controlling the rate of change. To incorporate priors, classical methods augment this basic scheme with regularization-aware updates:

- *Proximal and accelerated methods* such as ISTA and FISTA [49] handle objectives of the form $f(\mathbf{x}) + \lambda\Psi(\mathbf{x})$, where f is smooth and Ψ may be non-smooth (e.g., 2D total variation). Each iteration alternates between a gradient step on the data term and a proximal step that enforces the prior, often interpreted as a shrinkage or denoising operation. FISTA further adds a momentum-like term that combines the current and previous iterates before the proximal step, accelerating convergence for convex problems.
- *TwIST* [50] (Two-step IST) is another accelerated proximal scheme designed with sparse and TV-regularized imaging problems in mind. Like FISTA, it is based on proximal gradient updates, but it uses a different two-step recurrence that explicitly combines $\mathbf{x}^{(t)}$ and $\mathbf{x}^{(t-1)}$ before applying the shrinkage step. This modified recurrence can provide more stable and faster convergence than standard ISTA/FISTA in highly ill-conditioned deconvolution problems, while still alternating between a data-consistency update and a prior-enforcing update.
- *Splitting methods* such as ADMM [51] rewrite the reconstruction problem by introducing auxiliary variables that decouple the data term, the prior, and simple constraints. For lens-

less imaging, DiffuserCam [19] adopts this strategy by introducing separate variables for (i) the convolved image, (ii) a TV-regularization constraint, and (iii) a non-negative image estimate. ADMM then alternates between a least-squares update enforcing (i) measurement consistency, (ii) a proximal step that applies TV denoising, (iii) a projection enforcing non-negativity, and finally the dual-variable updates that keep the representations (i – iii) consistent. This splitting keeps each subproblem simple while still combining the physics operator, prior, and constraints into a single convergent optimization scheme.

3D extension for iterative optimization methods replace the 2D forward model (Sec. 2.1) by 3D forward models (Sec. 2.2) [19, 20, 34, 37]. This approach requires more careful depth-based PSF calibration, stronger sparsity assumptions, and longer compute times for convergence as they recover a 3D voxel of the captured scene.

From an intuitive standpoint, all these iterative methods iterate between two roles:

1. enforcing *physics consistency*, by pulling $\mathbf{Ax}$ towards the observed measurement $\mathbf{y}$;
2. enforcing *prior consistency*, by nudging $\mathbf{x}$ towards images that conform to assumed structure (smoothness, sparsity, non-negativity, etc.).

In lensless reconstruction, these classical model-based methods provide a transparent, interpretable baseline: the influence of the mask design and the forward model appears explicitly through $\mathbf{A}$ and $\mathbf{A}^\top$, and the effect of the prior can be controlled via the choice of $\Psi(\cdot)$ and its weight λ . However, their expressiveness is ultimately bounded by the chosen regularizer, which motivates the learning-based approaches discussed in the following subsections.

2.3.2 Supervised learning methods

Supervised learning approaches treat reconstruction as a data-driven mapping from measurements to target images. In lensless imaging, this typically involves capturing paired datasets $\{(\mathbf{y}_i, \mathbf{x}_i)\}$ where $\mathbf{y}_i$ are lensless measurements and $\mathbf{x}_i$ are "ground-truth" images acquired with a lensed camera, often using a beam-splitter setup and planar targets such as LCD screens [17,

25, 52]. A neural network f_θ is then trained to minimize a loss between $f_\theta(\mathbf{y}_i)$ and $\mathbf{x}_i$ (e.g., ℓ_2 , perceptual, or adversarial losses). Pure black-box architectures that do not explicitly include information of the optical system generally struggle even in this simplified setting [17], producing color-mismatched and lower quality reconstructions.

Physics-informed supervised models. Recent works therefore embed the physics into the network architecture. leADMM [17] unrolls an ADMM solver coupled with a UNet denoiser, enforcing data consistency while learning powerful priors. The learned primal–dual network of [33] follows a similar unrolling strategy in a primal–dual framework. FlatNet [18] couples an analytic inversion layer followed by a CNN that refines artifacts, and MWDN [48] extends this idea with multi-scale Wiener-like blocks. PhoCoLens [27] follow a similar logic of learnable single-inversion layer but applying it on a shift-variant modeling and extending previous approaches with a strong diffusion-based prior. In all these cases, knowledge of the optical system (PSF, cropping, etc.) is explicitly incorporated into the architecture, yielding reconstructions that are more robust than purely black-box networks trained on the same data.

Limitations for depth-varying scenes. Despite their success, supervised models inherit the assumptions of their training data. Most lensless datasets are acquired from planar scenes at a fixed distance, so the learned reconstructions are implicitly tied to a 2D setting at a single depth, and usually do not explicitly model additional PSF-varying effects such as shift-variance. Mismatch-aware architectures [53, 54] attempt to improve robustness by explicitly modeling or compensating for model errors, but still rely on paired supervision. More recent 3D extensions [22, 23, 38, 39] jointly design optics and reconstruction networks for depth-from-defocus or extended-DoF tasks, yet require multi-depth or volumetric training data that is challenging to obtain. As a result, current methods heavily rely on simulated data, which can create a mismatch between training (synthetic 3D scenes) and testing (in-the-wild real, continuous-depth scenes).

In this thesis, we focus instead on classical methods and *self-supervised/unsupervised* reconstruction frameworks that we introduce next.

2.3.3 Unsupervised and self-supervised neural reconstruction

The reconstruction algorithms in Sec. 2.3 can all be seen as ways of solving the same noisy linear inverse problem in (2.10), differing mainly in how they inject prior information on $\mathbf{x}$. Classical methods (e.g., ISTA, FISTA, ADMM) use *analytic* priors, while supervised learning methods train a reconstruction network directly from paired measurement–target data. Next, we define methods that rely on neural priors in scenarios without ground-truth images. We separate these methods into two classes: we use *unsupervised* for methods that do *not* train a network specifically for the given inverse problem, and use *self-supervised* for methods that *do* train a task-specific reconstruction network, but without access to ground-truth targets.

Unsupervised methods. Unsupervised methods in our terminology retain the same reconstruction objective as classical approaches,

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \Psi(\mathbf{x}),$$

but replace the hand-crafted regularizer $\Psi(\mathbf{x})$ with a neural prior. Most importantly, these methods *do not* train the network using measurement–ground-truth pairs for the target task.

Two approaches are most commonly used:

- *Untrained neural priors*, where the network is optimized only for the measurement at hand. Deep Image Prior (DIP) [55] is a seminal example: a randomly initialized CNN f_θ maps a fixed noise vector $\mathbf{z}$ to an image $\hat{\mathbf{x}} = f_\theta(\mathbf{z})$, and the parameters θ are optimized to minimize a data-consistency loss as

$$\mathcal{L}(\theta) = \frac{1}{2} \|\mathbf{y} - \mathbf{A}f_\theta(\mathbf{z})\|_2^2.$$

No external training images or targets are required: the convolutional network itself provides an *implicit prior*. Empirically, randomly initialized CNNs tend to reconstruct smooth, low-frequency structures first and only gradually fit high-frequency details and noise as optimization proceeds. As a result, if the optimization is stopped early, the network output typically captures the main edges and textures of a natural image while

suppressing pixel-wise noise; only with prolonged training does it begin to overfit the measurement noise. This frequency-biased behaviour of convolutions effectively acts as a built-in regularizer. This approach was extended to lensless imaging by Untrained Deep Networks (UDN) [56], which follow the same DIP strategy but applied to compressive lensless imaging tasks.

- *Pretrained generic priors*, where the network is trained for a different task (typically natural-image synthesis or denoising), and then reused as a prior for many inverse problems without further supervision. Examples include GAN- [57] and diffusion-model-based priors [46, 58–64] that are trained on large image collections and later used in plug-and-play, score-based, or diffusion-guidance frameworks for deblurring, inpainting, and super-resolution. Seminal methods such as diffusion posterior sampling (DPS) [58] and denoising diffusion null-space models (DDNM) [46] fall into this category: the diffusion model is trained for unconditional image generation, and at reconstruction time it is combined with the known forward operator $\mathbf{A}$ to solve new inverse problems without paired training data. One drawback of such methods is the inefficient multi-step sampling of diffusion models when compared to single-pass models (e.g., GAN), so recent works have focused on speeding up sampling using techniques such as distillation [65, 66].

In both cases, the inverse problem remains “unsupervised” in the sense that no ground-truth targets for the *specific* imaging task (e.g., lensless measurement $\rightarrow$ sharp scene) are used during training. The forward model enters only at inference time, through the data-consistency term.

Self-supervised methods. Compared to unsupervised methods, self-supervised reconstruction has been less explored in the inverse imaging literature. It occupies a different point in the design space: these methods *do* train a reconstruction network specifically for the target inverse problem, but use only unpaired or measurement-only data. Instead of minimizing a loss between the network output and ground-truth images, they construct surrogate objectives derived from the physics model, noise statistics, or redundancy in the measurements themselves. In the broader imaging literature, representative examples are the Noise2Noise [67], Noise2Self [68],

and Noise2Fast [69] families of denoising methods, which learn to remove noise without ever observing clean targets.

Formally, a typical self-supervised approach learns parameters θ of a reconstructor f_θ by minimizing a measurement-domain loss of the form

$$\min_{\theta} \mathbb{E}_{\mathbf{y}} [\mathcal{L}(\mathbf{y}, \mathbf{A}f_\theta(\mathbf{y}))],$$

where $\mathcal{L}$ enforces consistency with the forward model. The key point is that the network learns to approximate (or invert) $\mathbf{A}$ using only the observed measurements and the known physics, without access to paired ground-truth images.

Status in lensless imaging. While unsupervised and self-supervised neural-prior methods have been actively developed for generic inverse problems, their adoption in lensless imaging remains relatively limited. Most existing lensless pipelines either rely on classical model-based methods with analytic priors (e.g., TV) or on supervised networks trained from beam-splitter-based paired datasets. This gap is precisely where the later chapters of this thesis contribute: we design reconstruction frameworks for lensless cameras that (i) avoid supervision from lensed “ground-truth” targets, and (ii) leverage unsupervised/self-supervised training strategies, to investigate the impact of combining explicit physics-based forward models with neural priors on lensless imaging reconstruction quality and inference speed.

Chapter 3

Optimized radial mask for extended-depth-of-field lensless imaging

3.1 Problem statement and related work

In this chapter, we study how to design a lensless imaging system whose depth-dependent behavior can be well approximated by the 2D convolutional model in (2.4) over an extended range of scene depths. As discussed in Sec. 2.2, a general 3D scene is mapped to the sensor via a sum of depth-dependent PSFs,

$$\mathbf{y} = \sum_z \mathbf{h}_z * \mathbf{x}_z,$$

so that a reconstruction based on a single PSF calibrated at depth z^* can only recover objects lying near z^* . The gap between an object's true depth and the assumed depth manifests as blur or artifacts, and the range of depths for which this single-PSF reconstruction remains acceptable defines the depth of field (DOF) of the lensless system.

Existing approaches to extend DOF or recover 3D structure from lensless measurements can be roughly divided into two categories.

(I) Joint estimation of intensity and depth. A first line of work performs joint estimation of scene intensity and depth in a depth-from-defocus fashion. These EDoF and 3D imaging

methods aim to model the full multi-plane forward operator and reconstruct $\{\mathbf{x}_z\}$ for a set of depth planes z based on focus cues in the measurements. To ease this high-dimensional inverse problem, they typically increase the number of effective measurements using time-multiplexed coded patterns [34, 38] or multi-depth calibrated PSFs [19, 20]. Another direction jointly optimizes the optical system and reconstruction network in a data-driven, “deep optics” framework [22, 23, 39]. While powerful, these joint depth–intensity approaches lead to ill-posed, high-dimensional inverse problems that tend to require strong priors, multiple measurements, substantial computational cost, or learned components that can be difficult to interpret and analyze.

(II) PSF design for depth-robust measurements. A less explored strategy is that of designing coded masks with more depth-invariant PSFs. Instead of explicitly recovering depth, these methods design a coded mask that produces PSFs that are intrinsically more robust to depth, i.e., $\mathbf{h}_z$ changes as little as possible over a target depth range. By making the depth dependence of the PSF weak, one can recover scenes across a wider DOF while still using efficient 2D forward models and standard reconstruction algorithms.

Our approach. We investigate *radial coded masks* as a way to engineer depth-robust PSFs. As illustrated in Fig. 3.1, a random mask produces PSFs whose pattern changes noticeably under depth-induced scaling and sensor cropping, leading a 2D deconvolution calibrated at a single depth to reconstruct one object sharply while blurring another at a different depth. In contrast, radial patterns are approximately invariant under uniform scaling, so the central structure of the PSF remains similar across depths and can, in principle, support extended depth of field within a single 2D reconstruction model. In this chapter, we test this hypothesis by introducing a shape-preserving optimization procedure for the radial pattern and comparing the resulting optimized mask against handcrafted radial, randomly generated radial, and non-radial baseline masks, probing the limits of both reconstruction quality and DOF enabled by this design choice.

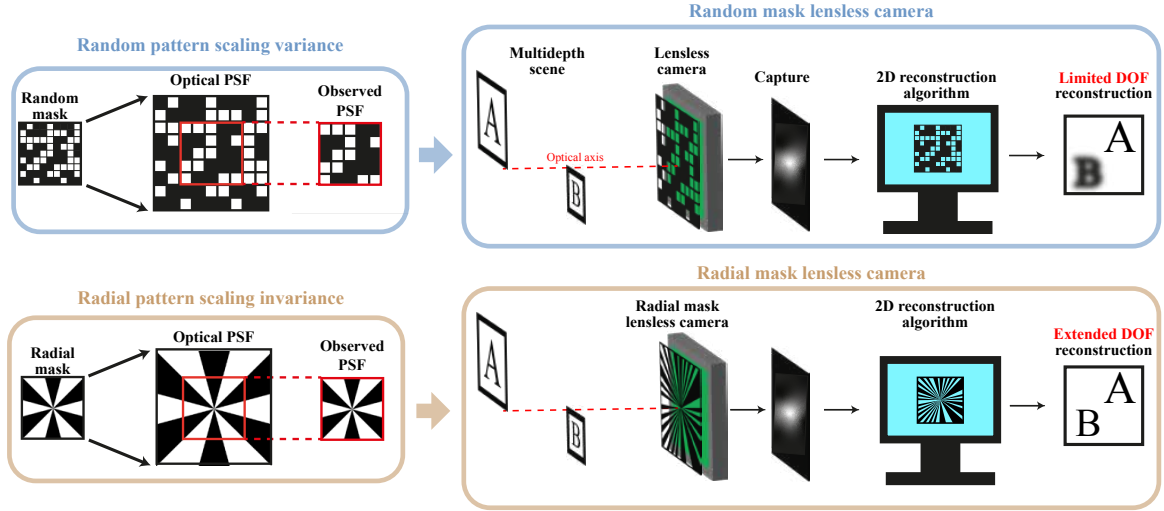


Figure 3.1: Conceptual illustration of depth behavior for different coded masks. (Top) A random mask yields depth-dependent PSFs and limited DOF when using a single-depth 2D reconstruction. (Bottom) A radial mask exhibits approximate scaling invariance, suggesting the potential for extended DOF within the same 2D reconstruction model. Adapted from [70](© 2023 IEEE).

3.2 Radial mask optimization

In this section we introduce our radial-shape-constrained optimization process. We start by presenting a radial mask parameterization, which constrains the radial shape of the mask throughout the optimization process. We then show the optimization objective, and explain how that is tied to the reconstruction quality achievable by the coded mask.

3.2.1 Mask parameterization

Past deep-optics studies typically optimize the mask by treating each element $m(x, y)$ independently, where $m(x, y)$ denotes the light transmittance at the discretized Cartesian coordinates (x, y) on the mask plane. This offers high design flexibility but yields masks with unconstrained global structure. Here, we introduce an additional constraint by enforcing a radial structure as illustrated in Fig. 3.2.

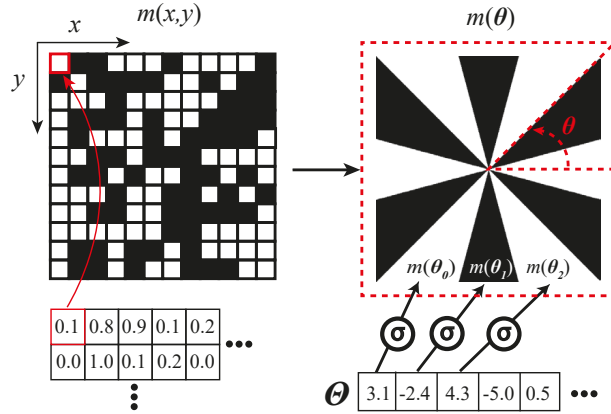


Figure 3.2: Overview of the radial coded mask parameterization. (Left) Conventional modeling with per-pixel transmittance $m(x, y)$ in Cartesian coordinates. (Right) Proposed radial parameterization, where the mask is divided into angular sections indexed by θ , each with transmittance $m(\theta)$ obtained by applying a sigmoid $\sigma(\cdot)$ to the corresponding entry of the parameter vector Θ . Reproduced from [70] (© 2023 IEEE).

We split the coded mask area into N_t radial sections

$$\Theta = [\theta_0, \theta_1, \dots, \theta_{N_t-1}]$$

along the angular coordinate, with each section originating at the center of the mask and extending toward the edge. All mask elements inside a given radial section share the same light transmittance value, i.e.,

$$m(x, y) = m(\theta_k) \quad \forall (x, y) \in \theta_k, \quad (3.1)$$

where k is the index of the k -th radial section and $(x, y) \in \theta_k$ denotes that the element $m(x, y)$ lies inside that section. We model the light transmittance of the k -th section as

$$m(\theta_k) = \sigma(m'(\theta_k)), \quad (3.2)$$

where $\sigma(\cdot)$ is a sigmoid function that constrains the value to the interval $[0, 1]$, and $m'(\theta_k)$ is a scalar parameter associated with section k . This parameterization allows us to represent the coded mask using a dimension-reduced vector

$$\mathbf{m}_\theta \in \mathbb{R}^{N_t},$$

and we define an operation $p(\cdot)$ that maps these angular-domain parameters to a coded mask in Cartesian coordinates. Ignoring additional effects such as diffraction, we approximate the PSF in (2.4) at the design distance (or at infinity) as

$$\mathbf{h} = p(\mathbf{m}_\theta). \quad (3.3)$$

3.2.2 Optimization objective

Our goal is to obtain a radial coded mask with a good broadband frequency response (i.e., MTF), so that information from the scene across a wide range of spatial frequencies (from smooth regions to sharp edges) is effectively transmitted to the lensless measurements. To that end, we leverage the modulation transfer function (MTF) of the PSF as the optimization objective.

Given a PSF $\mathbf{h}$, its normalized MTF is computed as

$$\text{MTF}(\mathbf{h}) = \text{normalize}(|\mathcal{F}[\mathbf{h}]|), \quad (3.4)$$

where $\mathcal{F}(\cdot)$ denotes the 2D discrete Fourier transform, $|\cdot|$ is the magnitude, and $\text{normalize}(\cdot)$ divides the spectrum by its DC value so that the MTF equals one at the origin. We then define the loss as

$$\mathcal{L}(\mathbf{m}_\theta) = -\text{mean}(\text{MTF}(\mathbf{h})) \quad (3.5)$$

$$= -\text{mean}(\text{MTF}(p(\mathbf{m}_\theta))), \quad (3.6)$$

where $\text{mean}(\cdot)$ computes the average over all spatial frequencies up to the cutoff. Since we wish to *maximize* the average MTF, we introduce the negative sign so that the problem can be expressed as the minimization

$$\hat{\mathbf{m}}_\theta = \arg \min_{\mathbf{m}_\theta} \mathcal{L}(\mathbf{m}_\theta). \quad (3.7)$$

We solve (3.7) using gradient descent with the Adam optimizer [71], with a learning rate of 0.01 and 2,000 iterations. We optimize a PSF with resolution of 140×140 . The mask parameters $\mathbf{m}_\theta$ are initialized by sampling uniformly in the interval $[-0.5, 0.5]$. We also employ gradient accumulation during optimization to speed up convergence.

3.2.3 Optimization analysis

We next analyze the masks obtained with the proposed optimization and compare their frequency responses against several radial and non-radial baselines.

Experiment 1: Comparison against handcrafted radial baselines. From the optimization experiments, we first select a representative optimized radial mask and compare its MTF against handcrafted radial baselines. Specifically, we consider cyclic (periodic) radial coded masks with 20, 40, and 60 radial sections, as well as a randomly generated radial mask. The cyclic masks are constructed by splitting the aperture into R radial sections of equal angular width and alternating zeros and ones between adjacent sections. The random radial mask uses the same number of sections as the optimized mask (i.e., 70), with each section's transmittance independently set to zero or one with equal probability.

The line MTFs of these handcrafted baselines and of the optimized radial mask are shown in Fig. 3.3 (top-left), and the corresponding patterns are shown in Fig. 3.3 (top-right). From the line MTFs, we observe that the optimized pattern achieves the desired broadband response: it competes with the sparser masks (e.g., Cyclic-20 and Radial-random) at lower frequencies and with the denser masks (e.g., Cyclic-40 and Cyclic-60) at higher frequencies.

Experiment 2: Comparison against randomly generated radial masks. To assess the effect of optimization statistically, we repeat the procedure 100 times and compare the resulting masks with 100 randomly generated radial masks. Both sets use the same number of radial sections (i.e., 70), and the random radial masks are created by assigning each section a binary transmittance (zero or one) with equal probability. For each type of mask, we compute the line MTF and report the mean and standard deviation across the 100 experiments.

As shown in Fig. 3.3 (bottom-left), the optimized radial masks consistently exhibit a stronger overall frequency response than the randomly sampled radial baselines, indicating that the MTF-based optimization provides a systematic advantage within the radial design class.

Experiment 3: Comparison against random (non-radial) masks. Finally, we compare the optimized radial mask with fully random (non-radial) coded masks. Random masks are known

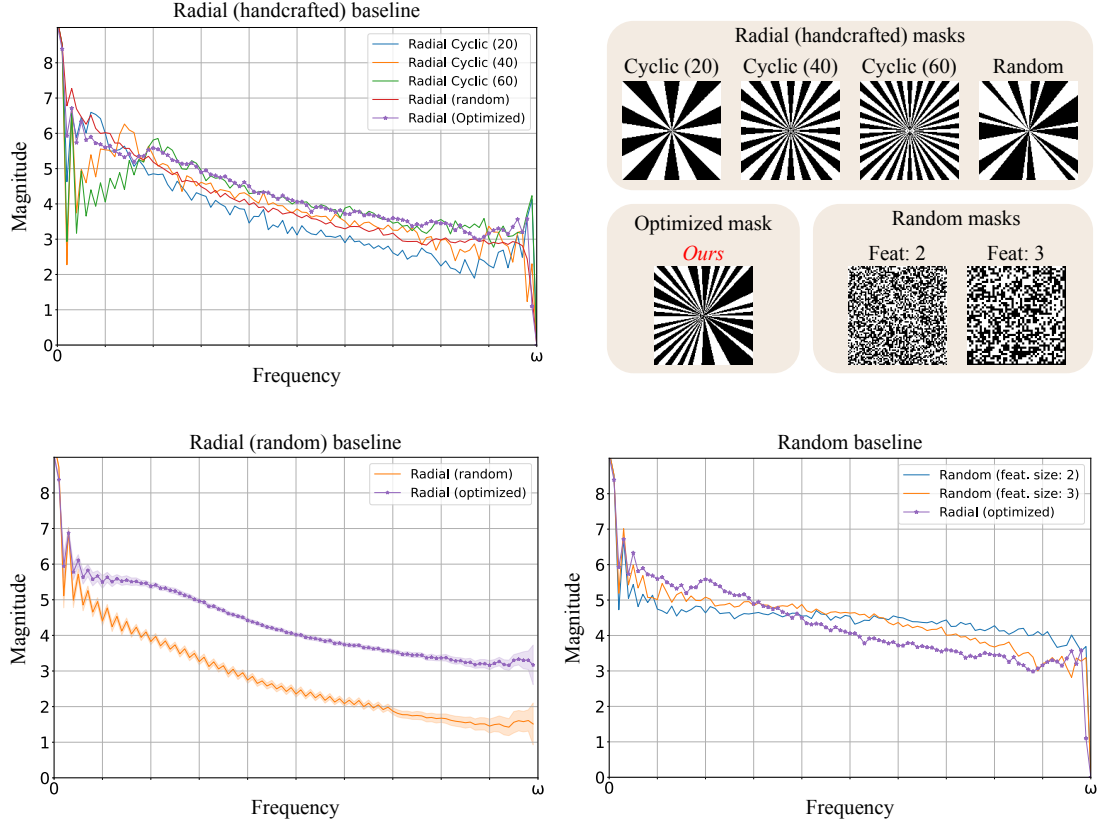


Figure 3.3: (Top-left) Line MTF comparison between our selected optimized radial mask and handcrafted radial masks with cyclic and random patterns. (Top-right) Mask patterns for handcrafted and random radial baselines, as well as our selected optimized radial mask. (Bottom-left) Comparison between the average line MTF obtained from 100 optimized radial masks and 100 randomly generated radial masks. (Bottom-right) Comparison between our selected optimized radial mask and random (non-radial) coded masks with feature sizes of 2 and 3 pixels. Adapted from [70] (© 2023 IEEE).

to have a well-balanced broadband response up to their cutoff frequency and thus serve as strong baselines for in-focus reconstruction quality. We generate two such masks with minimum feature sizes of 2 and 3 pixels; their patterns are shown in Fig. 3.3 (top-right). The corresponding line MTFs, together with that of our selected optimized radial mask, are shown in Fig. 3.3 (bottom-right).

As expected, the random masks provide stronger high-frequency response than the radial mask. This reflects the fact that the radial constraint significantly restricts the mask design space: some in-focus resolution is traded for the potential to enable extended depth of field. We revisit this trade-off between in-focus reconstruction quality (where random masks are advantageous) and out-of-focus robustness (where the radial mask is expected to excel) in the simulation study presented in Sec. 3.3.

3.3 Extended depth of field evaluation in simulations

In Sec. 3.2, we analyzed the frequency response of our optimized radial mask and compared it with several radial and non-radial baselines. That analysis provides insight into the information content of the corresponding lensless measurements and into the upper bound of reconstruction quality under a perfectly calibrated PSF and a matching scene. In this section, we extend the evaluation to three-dimensional settings by studying how object distance variations affect reconstruction quality for the same set of masks. The lensless measurements are generated using the multi-plane convolutional model in (2.9) under a noiseless setting, so that we can isolate the effect of PSF depth dependence and mask design on the effective DOF. Through these simulated experiments, we investigate whether radial masks can extend the DOF of a lensless camera and whether the optimized radial mask can achieve high-quality reconstructions under extended depth-of-field operation.

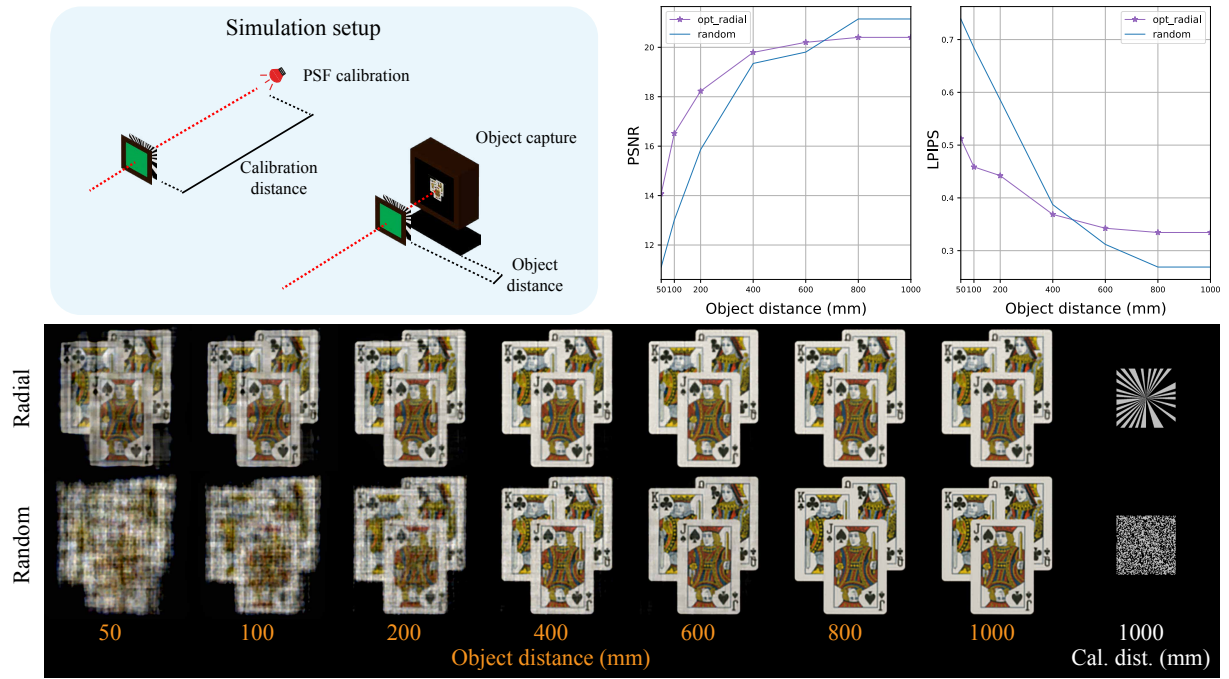


Figure 3.4: Planar scene simulations to evaluate the impact of mismatch between object and PSF depths on reconstruction quality. (Top-left) Experimental setup for simulated PSF calibration and object capture. (Top-right) Reconstruction metrics (PSNR and LPIPS) as a function of object distance when using a fixed-depth PSF. (Bottom) Reconstructions for objects placed at various distances from the coded mask for the optimized radial mask and the random mask.

3.3.1 Planar scene simulation

We first consider a single-plane reconstruction experiment. To evaluate the DOF of each mask, we vary the depth of a planar scene with respect to the lensless camera while keeping the PSF calibration depth fixed. This allows us to measure how reconstruction quality degrades as the mismatch between the object distance and the assumed PSF distance increases.

Experimental setup. The experimental setup is shown in Fig. 3.4(top-left). The scene consists of three cards arranged on a planar surface, whose distance from the camera is varied from 1,000 mm down to 50 mm. For each evaluated depth, we simulate the measurement using the PSF corresponding to that depth, but reconstruct the scene using only the PSF calibrated at the reference distance. To isolate the effect of depth-dependent PSF mismatch, we simulate a noiseless setting and compare the optimized radial mask with a random (non-radial) mask. In this experiment, the calibration PSF is measured at 1,000 mm, so objects placed at or near that depth should be reconstructed accurately, while objects farther away will exhibit increasing model mismatch.

Reconstruction and evaluation. We use the ADMM algorithm [19] for reconstruction using the same hyperparameters for all coded masks. We evaluate the results using Peak Signal-to-Noise Ratio (PSNR) and Learned Perceptual Image Patch Similarity (LPIPS) [72]. The reconstruction metrics as a function of depth are shown in Fig. 3.4(top-right), and example reconstructions for different object distances are shown in Fig. 3.4(bottom).

Results. As expected, the random mask achieves higher reconstruction quality for objects placed near the calibration depth, consistent with its stronger broadband MTF enabled by an unconstrained design space. In contrast, the optimized radial mask yields higher reconstruction quality at depths that deviate from the calibration distance, highlighting its advantage in extending the effective depth of field of the lensless imaging system.

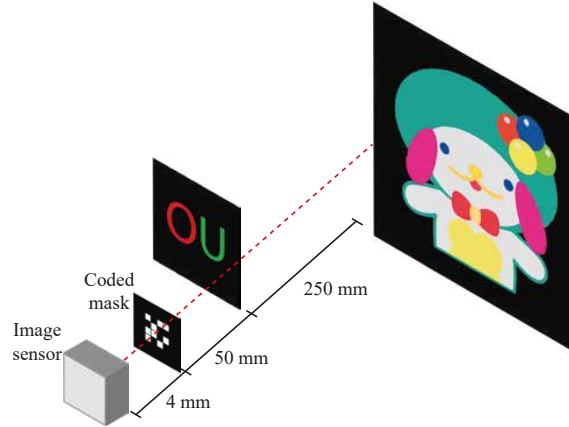


Figure 3.5: Diagram of the dual-depth scene simulated experiment. A near object is placed at 50 mm and a far object at 300 mm from the coded mask. PSFs are calibrated at a far depth, making the near object more challenging to reconstruct with a single fixed PSF. Reproduced from [70] (© 2023 IEEE).

3.3.2 Dual depth scene simulation

We next consider a dual-depth scene, where one object is placed close to the lensless camera and another is placed farther away. This setting allows us to compare refocusing capabilities of non-radial masks (e.g., random and Fresnel zone aperture (FZA) patterns), as demonstrated in [20], with the extended depth-of-field behavior of radial masks, which ideally do not require refocusing.

Experimental setup. The simulated setup is illustrated in Fig. 3.5: a near object is placed at 50 mm from the coded mask and a far object at 300 mm. As in Sec. 3.3.1, we calibrate the PSFs at a far depth (i.e., 300 mm) so the object closer to the mask is more challenging to reconstruct when using a single, fixed PSF. We compare our optimized radial mask against cyclic-20, cyclic-60, and randomly generated radial masks, as well as against FZA and random (non-radial) masks

Reconstruction and evaluation. We reconstruct the dual-depth scene using the ADMM [51] and the UDN algorithms [56], and evaluate PSNR separately for the in-focus (far) object and the out-of-focus (near) object. Reconstructions are shown in Fig. 3.6.

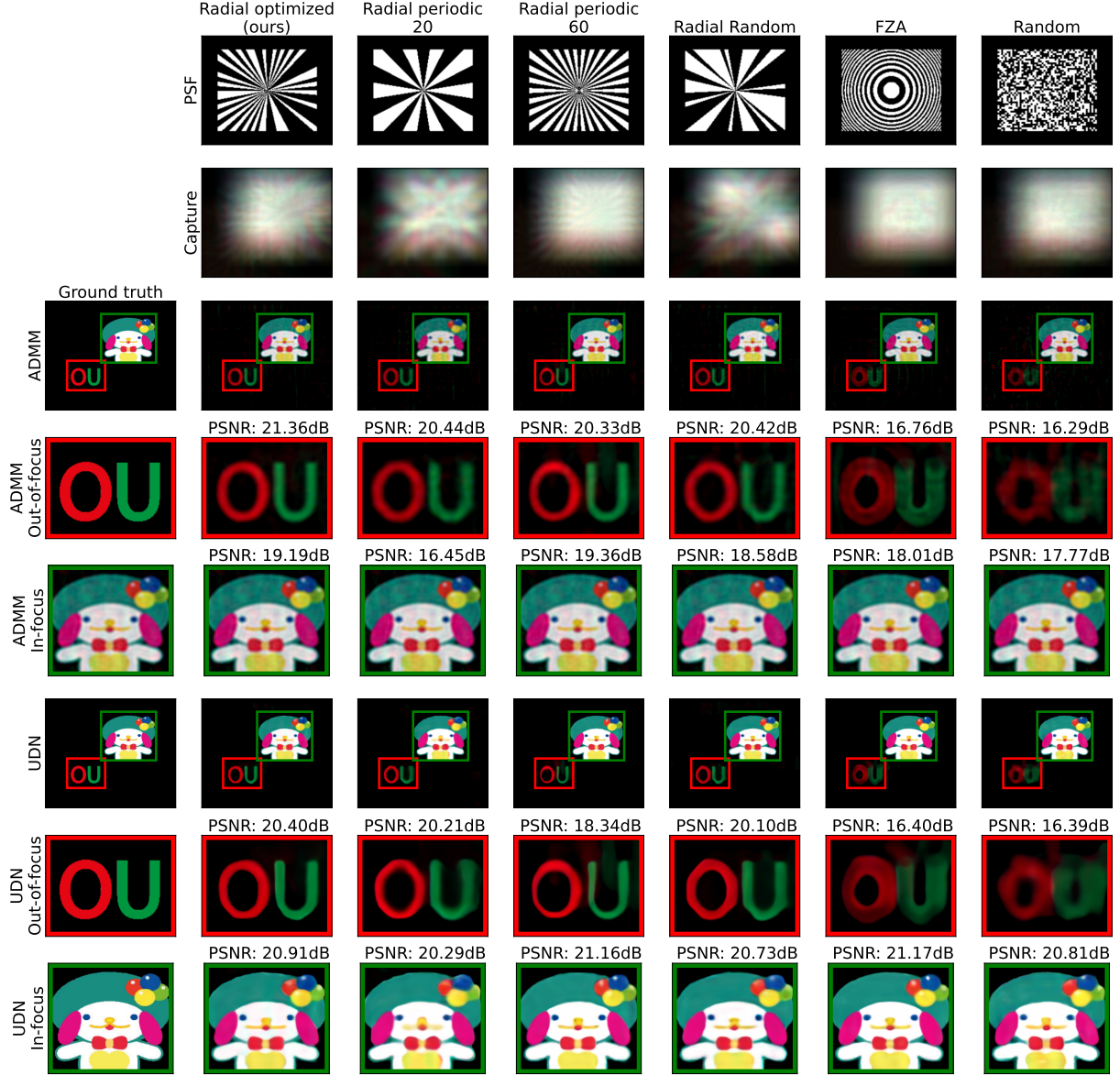


Figure 3.6: Reconstructions for the dual-depth scene simulation using different coded masks and reconstruction algorithms. Radial masks, and in particular the optimized radial mask, better preserve the near (out-of-focus) object while maintaining competitive quality on the far (in-focus) object. Reproduced from [70] (© 2023 IEEE).

Results. Overall, radial masks tend to achieve higher PSNR for the out-of-focus object (i.e., the OU pattern), and the optimized radial mask provides the highest PSNR for the near object for all tested reconstruction algorithms, while maintaining competitive performance on the far object. These results support the use of optimized radial masks as a means to extend the effective depth of field in lensless imaging.

3.4 Prototype lensless camera experiments

In Sec. 3.3, we used the multi-plane forward model to study how different coded masks behave under depth variations in an idealized setting. The simulations indicated that, while unconstrained random masks provide superior in-focus resolution near the calibration depth, radial masks—and in particular the optimized radial pattern—yield more stable reconstructions as the object moves away from the calibrated plane, effectively extending the depth of field.

In this section, we investigate the practical feasibility of our proposed radial mask by building a prototype lensless camera. Using the same family of coded masks, we conduct two optical experiments. The first involves a continuous-depth scene composed of a tilted chessboard and colored pawns placed at different distances from the mask, providing a direct counterpart to the simulated experiments. The second employs transparent objects in a challenging setup for depth estimation-based EDoF methods. We begin by describing the prototype setup and the continuous-depth scene experiment.

3.4.1 Continuous-depth experiment with a prototype camera

Prototype camera setup. To validate the extended DOF behavior of the radial masks in a real system, we built a prototype lensless camera consisting of an axial stack of a coded mask and an image sensor (Fig. 3.7(a)). The coded mask is implemented using a transmissive liquid-crystal spatial light modulator (SLM; LC2012 by HOLOEYE Photonics) and two polarization plates arranged in a crossed-Nicols configuration. Since all coded masks considered in this work are binary, the SLM transmittance is configured to approximate a binary response, and the central

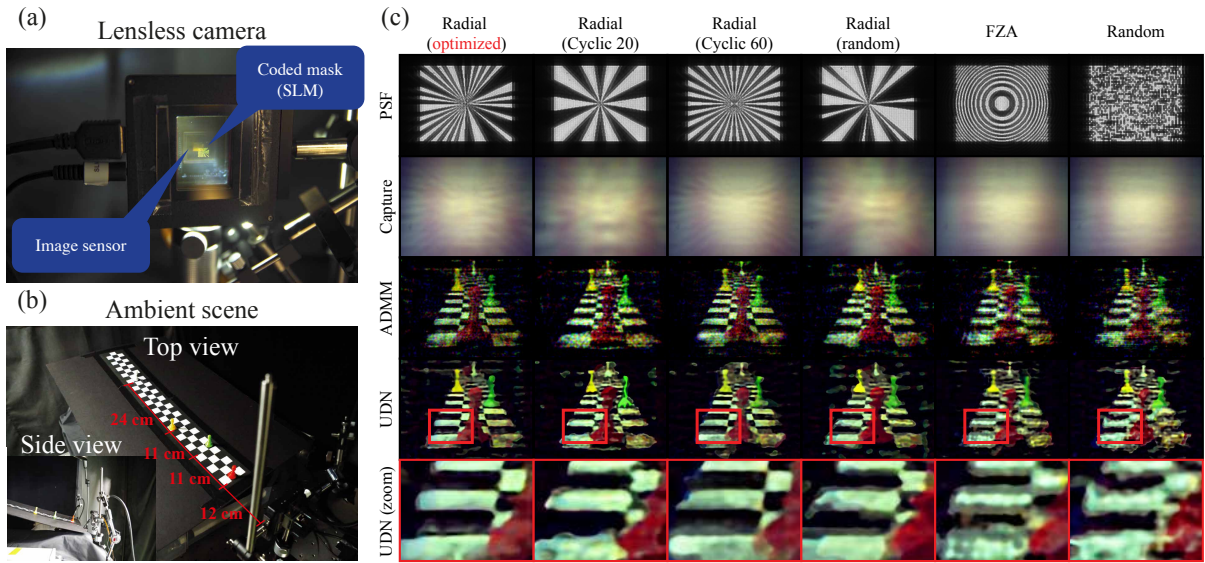


Figure 3.7: (a) Our prototype camera built using a CMOS image sensor and a spatial light modulator. (b) Setup for the continuous scene experiment, composed of a slanted plane with a checkerboard pattern on top and 4 pawns placed at multiple distances from the lensless camera. (c) Experimental reconstructions for the continuous depth scene. First row shows calibrated PSFs. The second shows the captured measurements, normalized for visualization. The third and fourth present the ADMM and UDN reconstructions, respectively. Reproduced from [70] (© 2023 IEEE).

188×228 pixels (with $36 \times 36 \mu\text{m}$ pitch) are used to display the mask patterns. The fill factor of the SLM is 58%.

Approximately 4.0 mm behind the SLM modulation plane, we place a color CMOS image sensor (BFS-U3-28S5C-BD, Teledyne FLIR) with 1464×1936 pixels and a pixel pitch of $4.5 \times 4.5 \mu\text{m}$. In the experiments, 8-bit RGB images are read out from the sensor and downsampled to 732×968 pixels for reconstruction. As in the simulations, the periphery of the mask is blocked to improve reconstruction stability, yielding an effective mask area of approximately 50%. The centers of the effective SLM area and the sensor are aligned using translation stages, and the planes of the two elements are adjusted to be parallel.

PSF calibration. The top row of Fig. 3.7(c) shows the calibrated PSFs. The PSF is measured experimentally by capturing the spherical wave emitted from a point light source placed approximately 60 cm away from the coded mask. The point source is constructed from a semiconductor laser with central wavelength 532 nm (Stradus 532 by Vortran Laser Technology) followed by a spatial filter (SFB-16DMRO-OBL40-25, SIGMA KOKI) containing a $25 \mu\text{m}$ pinhole, which together generate a spherical wave.

Experimental setup. Figure 3.7(b) shows the continuous-depth scene used in the experiments. A chessboard pattern is placed on an inclined plane in front of the prototype camera, and four colored pawns are positioned at different distances from the coded mask. The distances of the board and pawns are indicated in the figure. The scene is illuminated by a white LED light mounted above the lensless camera and directed toward the targets. The second row of Fig. 3.7(c) shows the captured lensless measurements for the different coded masks. Although these images are difficult to interpret visually, they contain multiplexed information from objects at multiple depths according to the multi-plane forward model in Eq. (2.9). The lensless captures are normalized independently for visualization. We use the ADMM (100 iterations) and UDN (150,000 iterations) methods for reconstructing the scene.

Results. Since the PSF is calibrated using a point source at 60 cm from the coded mask, objects located farther from the mask are closer to the calibrated depth and are expected to be easier to reconstruct, whereas objects closer to the camera should be more challenging. This trend can be

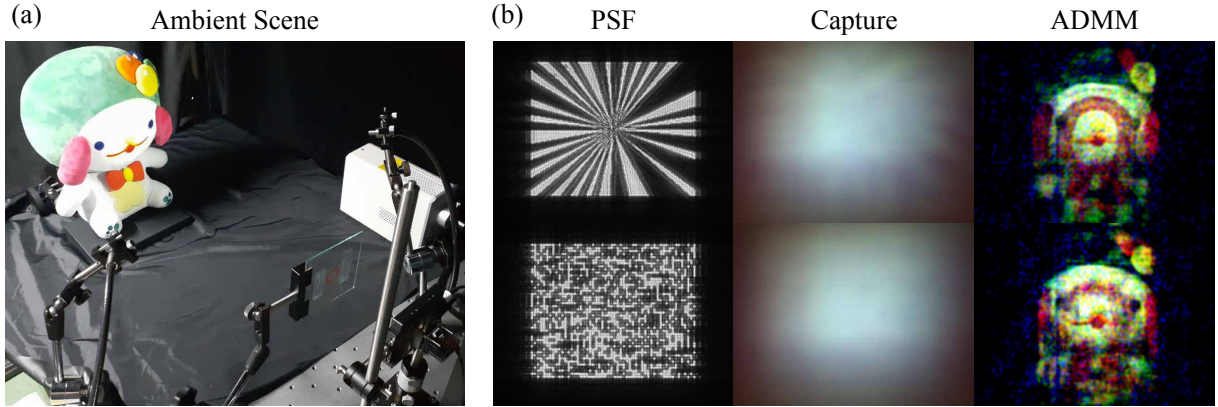


Figure 3.8: (a) Experimental setup for the transparent object reconstruction. (b) Calibrated PSFs, lensless captures and ADMM reconstructions for the transparent object scene. Reproduced from [70] (© 2023 IEEE).

observed in the ADMM and UDN reconstructions shown in Fig. 3.7(c). For non-radial masks, the chessboard squares nearer to the camera exhibit blurrier edges, and the red pawn—closest to the camera—is poorly reconstructed. In contrast, all radial masks provide extended depth of field, yielding sharper chessboard edges and more recognizable pawns across a wider range of depths.

Consistent with the simulation results in Sec. 3.3, a trade-off between high- and low-frequency behavior is observed for the periodic radial masks. The radial mask with 20 sections produces less sharp edges in the chessboard and pawns, whereas the mask with 60 sections yields sharper edges but slightly noisier reconstructions. The optimized radial mask achieves a balance between these extremes, reconstructing sharp chessboard edges while also preserving smoother, sparse regions such as the pawn bodies.

3.4.2 Reconstruction through transparent objects

A natural alternative approach to extended-DOF lensless imaging is to first estimate a depth map of the scene and then refocus the calibrated PSFs accordingly, so that each object is reconstructed using a PSF matched to its depth. In practice, however, depth-estimation-based

methods struggle in the presence of semi-transparent or transparent planes. In such scenarios, light from objects at multiple depths is mapped onto the same sensor pixels, and it is not obvious which depth should be assigned to those mixed measurements. As a result, depth estimates can become unreliable or ambiguous, limiting the effectiveness of subsequent refocusing.

Experimental setup. To gain further insight into the potential of radial masks in such cases, we conduct an experiment involving both a semi-transparent target and an object located behind a transparent plane. A red "O" pattern is printed on a transparent OHP film and mounted on a glass plate. The glass plate containing the "O" pattern is placed 12 cm in front of the prototype lensless camera, and a plush toy is positioned behind the glass at approximately 60 cm from the coded mask. The scene geometry is shown in Fig. 3.8(a). The PSFs are calibrated using a point source placed at 60 cm, so that, in principle, the reconstruction is biased in favor of the plush toy at the farther depth. Illumination is provided by two white LED lights placed on either side of the camera and directed toward the plush toy. We use the same prototype camera and calibration method described in Sec. 3.4.1, but now with only the random mask as a baseline.

Results. Figure 3.8(b) shows the calibrated PSFs, the captured lensless measurements, and the corresponding ADMM reconstructions for the different masks. The optimized radial mask is able to reconstruct a sharp semi-transparent "O" pattern on the glass plate while simultaneously recovering the plush toy behind it, despite the interference of the transparent plane. In contrast, the non-radial baseline (e.g., random masks) reconstruct the "O" poorly and exhibit reduced robustness to the layered depth structure. These results indicate that the proposed radial-mask-based system can maintain extended depth of field even in scenarios where depth-estimation-based approaches are expected to fail due to transparency-induced ambiguities.

3.5 Discussion and summary

In this chapter, we explored PSF engineering as a means to extend the depth of field of lensless cameras while retaining a simple 2D convolutional reconstruction model. We introduced a radial parameterization of the coded mask that enforces a shared transmittance along radial

sectors and enables a low-dimensional description of the mask. Within this constrained design space, we formulated an MTF-based optimization objective and showed that the resulting optimized radial pattern exhibits a broadband frequency response that systematically outperforms randomly generated radial masks. While fully random (non-radial) masks still offer superior high-frequency response at the calibrated depth, the radial constraint provides a principled way to trade some in-focus resolution for improved depth robustness.

Using multi-plane simulations, we then evaluated how these masks behave under object–PSF depth mismatch. For planar scenes, the random mask delivers the highest reconstruction quality near the calibration depth, but its performance degrades rapidly as the object moves away from that plane. In contrast, the optimized radial mask maintains higher reconstruction quality over a broader range of depths, indicating an extended effective DOF. A dual-depth experiment further highlighted this behavior: radial masks, and especially the optimized design, achieve better reconstruction of the near (nominally out-of-focus) object while remaining competitive on the far object for both ADMM and UDN reconstruction methods. These results support the view that radial masks can approximate the multi-depth forward model by a single-depth 2D convolution, enabling extended-DOF operation without explicit depth estimation.

We validated these findings on a prototype lensless camera. In the continuous-depth scene with a tilted chessboard and pawns at different distances, non-radial masks failed to reconstruct sharp details for objects far from calibration depth, whereas radial masks preserved sharper edges and recognizable details across a wider depth range. Among the radial designs, periodic masks with few sectors favored smoothness, while those with many sectors reconstructed high-frequency details. The optimized radial mask struck a balance between these extremes, providing sharp chessboard edges and smooth reconstruction of the pawns. In the transparent-object experiment, the optimized radial mask successfully reconstructed both a semi-transparent foreground pattern and a plush toy behind a glass plate, in a configuration where depth-estimation-based strategies are prone to fail due to transparency-induced ambiguities.

Overall, this chapter demonstrates that radial coded masks, optimized within a simple MTF-based framework, can extend the depth of field of lensless imaging systems while remaining

compatible with standard 2D reconstruction algorithms. At the same time, the use of an SLM and restricted aperture means that the radial structure is not fully exploited across the entire sensor area. In the next chapter, we build on these insights and investigate full-aperture radial mask designs aimed at achieving near depth-invariant PSFs and potentially paving the way toward all-in-focus lensless imaging.

Chapter 4

Full-aperture radial masks for all-in-focus lensless imaging

4.1 Motivation and overview

In Chapter 3, we showed that enforcing a radial structure on a coded mask and optimizing its MTF can extend the effective DOF of a lensless camera while still using a 2D shift-invariant convolution model for reconstruction. Using an SLM-based prototype camera with a restricted mask region, we demonstrated that radial masks yield more stable reconstructions across depth than non-radial patterns, and that the optimized radial design achieves a favourable trade-off between MTF response (i.e., compared to baseline radial patterns) and robustness to object–PSF depth mismatch (i.e., compared to baseline non-radial patterns).

Current limitations. This first-generation system, however, has two important limitations. First, from an *aperture* perspective, only a central portion of the SLM is used and the periphery is blocked to improve stability. Even if the displayed pattern is radial, this truncation reintroduces depth-dependent changes in the PSF, as shown in Fig. 4.1(Left), which can constrain the achievable DOF. Second, from an *implementation* perspective, the SLM’s relatively coarse pixel pitch and limited fill factor reduces the spatial resolution and contrast of the PSFs, which

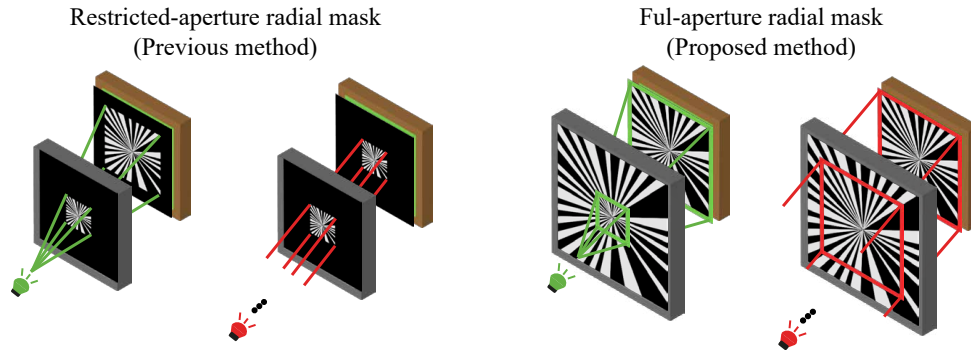


Figure 4.1: (Left) Restricted-aperture radial mask with a depth-induced PSF variation. (Right) Full-aperture radial mask with a depth-invariant PSF. Adapted from [1].

in turn constrains its achievable reconstruction quality.

Proposed solutions. In this chapter, we address both limitations from our previous research.

(1) On the *aperture* side, we investigate full-aperture radial masks designed to be larger than the sensor area, thus exploiting the radial structure across the camera’s entire FOV, as shown in Fig. 4.1(Right). The goal is to obtain PSFs that are more strongly correlated across depth than in the restricted-aperture case, pushing the multi-depth forward model closer to a depth-agnostic behavior and potentially paving the way to a true all-in-focus snapshot lensless camera. (2) On the *implementation* side, we replace the SLM with passive, printed amplitude masks. This change improves the optical quality of the PSFs and increases light efficiency. To isolate this implementation effect from aperture effects, we fabricate not only a full-aperture radial mask but also a restricted-aperture radial mask following the design used in Chapter 3. This allows us to separate the gains due to better mask realization from those due to increased aperture.

Chapter outline. To investigate whether the full-aperture radial mask achieves wider DOF than its restricted-aperture counterpart, we proceed in four steps. (1) We describe the printed-mask prototype camera, calibrate PSFs at multiple depths for both restricted- and full-aperture radial masks; we then quantify depth robustness via cross-correlation across depth planes. (2) We perform a proof-of-concept DOF evaluation using small objects aligned near the optical axis, where the impact of shift-variance is reduced and 2D shift-invariant deconvolution can

be applied, enabling a fair comparison between restricted- and full-aperture masks under the same reconstruction model. (3) We present a second proof-of-concept experiment in a data-driven reconstruction setting, where learned models can implicitly accommodate shift-variance, allowing us to compare restricted- and full-aperture masks on larger objects. (4) We propose an approximate shift-invariant deconvolution strategy based on an artificially extended PSF that enables the use of full-aperture radial masks within a 2D convolution framework, making it possible to reconstruct large planar targets and continuous-depth scenes.

4.2 Prototype camera and PSF characterization

In this section, we describe the new prototype lensless camera built around printed radial masks and analyze its depth-dependent PSFs. We consider two variants of the radial design: a *restricted-aperture* mask that follows the approach used in Chapter 3, and a *full-aperture* mask that extends the radial pattern to cover the full FOV of the sensor. Using this prototype, we first compare the depth robustness of these two masks via multi-depth PSF calibration and cross-correlation, and then assess the degree of spatial shift-variance in each design and its implications for 2D shift-invariant deconvolution that we used previously.

4.2.1 Prototype camera with printed radial masks

Mask fabrication. The new prototype lensless camera is designed to address two limitations of the lensless camera in Chapter 3: (1) the limited optical quality of the SLM-based mask and (2) the restricted effective aperture. we replace the previously used SLM with passive amplitude masks fabricated on a chromium-coated glass substrate. The optimized radial pattern is first designed in software and then transferred onto the chromium layer through standard mask fabrication procedures, yielding binary transmittance with well-defined edges and high fill factor. This fabrication procedure increases the mask resolution by one order of magnitude, from a feature size on the order of $36\ \mu\text{m}$ in the SLM implementation to approximately $1.62\ \mu\text{m}$ for the printed chromium mask. It also reduces aliasing and staircasing along radial directions

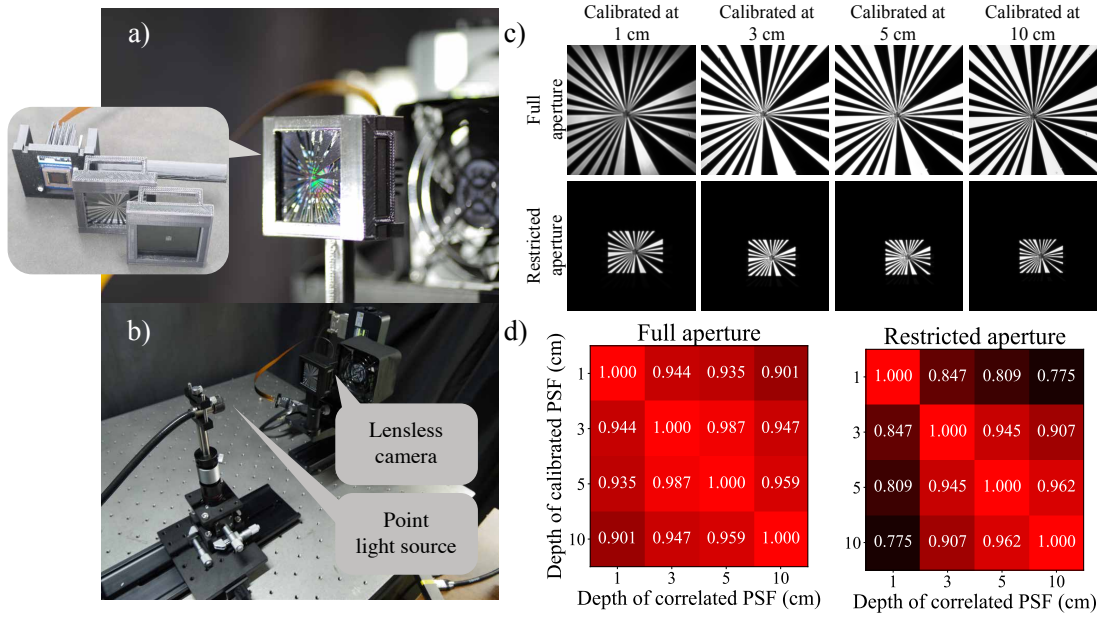


Figure 4.2: (a) Prototype camera built with restricted- and full-aperture printed radial masks and 3D printed holders for ease of mask swapping. (b) Multiple depth PSF calibration setup. (c) PSFs calibrated at multiple depths for restricted- and full-aperture masks. (d) Confusion maps with maximum cross-correlation values of PSFs at different depths. Reproduced from [1].

that arise when radial patterns are represented at low spatial resolution, potentially improving even further the depth-invariance of the radial PSF.

From this base pattern, we fabricate two masks: a *restricted-aperture* radial mask, in which only a central region is opened to reproduce the effective aperture used in the SLM prototype, and a *full-aperture* radial mask, in which the radial pattern occupies the entire clear area of the substrate so as to occupy the full FOV of the image sensor. This pair of masks printed masks provides a fair comparison between restricted- and full-aperture radial masks by allowing us to disentangle improvements due to a better physical realization (printed chromium mask versus SLM) from those due to increased aperture (full versus restricted).

Mechanical assembly. The masks are mounted in front of a color CMOS image sensor using custom 3D-printed holders. These holders maintain a fixed axial spacing between mask and sensor, align their optical axes, and block stray light entering from the periphery. As the sensor,

we use a Teledyne FLIR BFS-U3-28S5C-BD with a resolution of 1936×1464 pixels and a pixel pitch of $4.5 \mu\text{m}$. The restricted-aperture radial mask is fabricated so that its clear aperture covers approximately one quarter of the sensor extent along each axis. The full-aperture radial mask, in contrast, is printed larger than the sensor in order to cover its entire FOV. The assembled prototype camera is shown in Fig. 4.2(a).

Data acquisition and reconstruction. Raw RGB images are captured, optionally cropped and downsampled for reconstruction, and normalized for analysis. By swapping only the mask element while keeping the rest of the camera unchanged, we can directly compare the PSFs and reconstruction behavior of the restricted- and full-aperture radial designs.

4.2.2 PSF depth dependence analysis

Multi-depth PSF calibration. To quantify how strongly each mask design suppresses depth information in the PSF, we calibrate PSFs at multiple distances using the prototype camera. For a set of depths $\{z_k\}$ within the working range of the system, we place a point-like light source on the optical axis and record the corresponding lensless measurements. Each capture is treated as an estimate of the PSF $\mathbf{h}(z_k)$ at depth z_k for a given mask. The point light source is created using an LED source (Optosigma SLA-100B) with an attached pinhole aperture ($200 \mu\text{m}$ pinhole), mounted on a linear stage to facilitate its depth variation. The PSF calibration setup is shown in Fig. 4.2(b). The PSFs are calibrated for 4 depths (i.e., 1, 3, 5, and 10 cm). Results are shown in Fig. 4.2(c).

Cross-correlation across depth. For both restricted- and full-aperture masks, we evaluate PSF similarity for all pairs of depths $d_1, d_2 \in \{1, 3, 5, 10\}$. Since the calibrated PSFs are not perfectly aligned, we compute a *2D normalized cross-correlation map* by sliding one PSF over the other and then take the maximum correlation value as the similarity score between the two depths. Formally, for a pair of PSFs $\mathbf{h}(z_{d_1})$ and $\mathbf{h}(z_{d_2})$, we define

$$\rho(z_{d_1}, z_{d_2}) = \frac{\mathbf{h}(z_{d_1}) \star \mathbf{h}(z_{d_2})}{\|\mathbf{h}(z_{d_1})\|_2 \|\mathbf{h}(z_{d_2})\|_2}, \quad (4.1)$$

as the normalized correlation map, where $\star$ is the cross-correlation operation, and $\|\cdot\|_2$ is the L2 norm. We compute $\rho(z_{d_1}, z_{d_2})$ for all pairs $d_1, d_2 \in \{1, 3, 5, 10\}$, get the maximum value, and display the results as confusion maps in Fig. 4.2(d). The full-aperture radial mask consistently yields higher correlation values over a wider depth range, indicating that its PSFs change more slowly with depth variation. In other words, extending the radial pattern across the full aperture makes the forward model closer to the ideal depth-invariant behavior illustrated in Fig. 3.1, and suggests that a single PSF calibrated at a predefined depth may remain usable over a broader DOF. This analysis is performed with PSFs measured on-axis and does not yet account for spatial shift-variance across the field of view, which we analyze next.

4.2.3 PSF shift-variance observation

Restricted-aperture masks. Lensless cameras with restricted-aperture masks are often modeled with shift-invariant PSFs. Intuitively, even when the point source moves within the field of view, most of the coded mask pattern still projects onto the sensor, so the PSF shape changes are minimal. This is illustrated in Fig. 4.3(a): three point sources at different incidence angles but on the same depth plane project nearly the full mask pattern onto the sensor. On the right of Fig. 4.3(a), the two calibrated PSFs from the prototype camera (one on-axis and one near the edge) exhibit similar structure, differing primarily in intensity distribution but not on PSF shape.

Full-aperture masks. For full-aperture masks, this assumption breaks down. As shown in Fig. 4.3(b), point sources at different lateral positions on the same depth plane illuminate different regions of the coded mask, so distinct portions of the pattern are projected onto the sensor. This is reflected in the two calibrated PSFs in Fig. 4.3(b): unlike the restricted-aperture case, the on-axis and off-axis PSFs differ not only in overall intensity but also in shape, since different areas of the mask are illuminated. This position-dependent change in PSF structure indicates a stronger shift-variance of the full-aperture radial mask's PSF.

Challenges and solutions. The stronger shift-variance observed for full-aperture masks im-

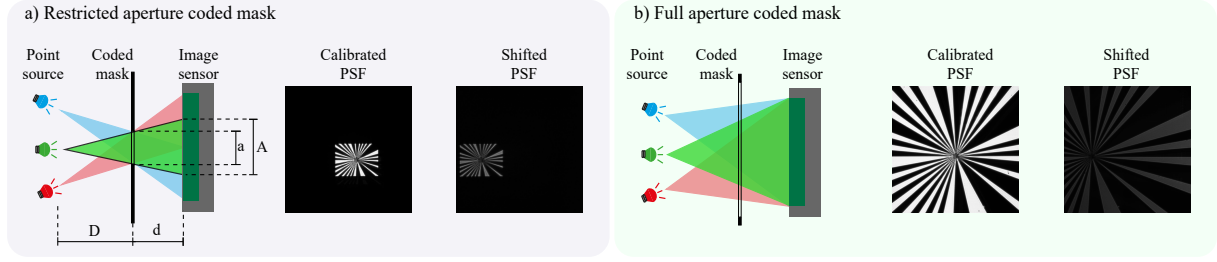


Figure 4.3: (a) Restricted-aperture radial mask, with entire mask pattern being projected onto the sensor from all directions, so that a centrally calibrated PSF and shifted PSF have the same information, albeit with different intensity values. (b) Full-aperture radial mask, with different angles projecting different mask regions and therefore producing different centrally calibrated and shifted PSFs. Adapted from [1].

plies that the simple shift-invariant model in Eq. (2.4) no longer provides a good approximation of the lensless measurements across the field of view. As a result, the deconvolution-based methods used in the previous chapter, which assume a single global PSF, are not immediately applicable in this setting. In the following sections, we explore three differing strategies that allows us to evaluate the full-aperture mask's DOF by circumventing this limitation. In Sec. 4.3, we target small objects placed near the optical axis of the camera, thus reducing the impact of shift-variance and allowing the 2D convolution model to remain a reasonable approximation; this allows a fair DOF comparison between restricted- and full-aperture masks under the same reconstruction methods as before, albeit in a limited scenario. In Sec. 4.4, we turn to data-driven reconstruction and use learned models to implicitly handle shift-variance, enabling DOF analysis for larger objects occupying larger amount of the FOV while acknowledging the black-box and training-dependent nature of this approach. Finally, in Sec. 4.5, we introduce a novel approximation based on artificially extended PSFs that allows us to apply shift-invariant-like deconvolution to full-aperture radial masks, making it possible to reconstruct large planar targets and continuous-depth scenes within a single, general framework.

4.3 Shift-invariant deconvolution for small on-axis objects

The goal of this section is to evaluate the effective DOF achieved by the full-aperture radial mask when reconstructing small objects at varying depths using a single calibrated PSF. As discussed in Sec. 4.2.3, the full-aperture mask exhibits strong shift-variance and is not well approximated by a 2D shift-invariant convolutional forward model. To make the problem tractable, we constrain our analysis to small, centered objects—specifically MNIST digits—so that the spatial variation of the PSF across the object support remains limited. In this localized setting, the convolutional forward model in Eq. (2.4) serves as a reasonable approximation, which allows us to apply efficient convex optimization-based reconstruction methods based on the 2D convolutional model.

Despite the model mismatch, we show that the reconstructed images are of sufficient quality to analyze the system’s depth invariance. We first present simulation results and then reconstructions from physical captures acquired with the prototype camera. The physical experiments also serve to validate that using the convolutional model as an approximation in simulation is fair, by comparing depth-sensitivity trends across both domains. Taken together, these experiments provide a consistent evaluation of the depth-robust behavior of the proposed full-aperture mask design relative to the restricted-aperture baseline.

4.3.1 MNIST simulation experiment

Experimental setup. This experiment evaluates the effective DOF of both the restricted- and full-aperture radial masks in simulation using small MNIST digits. The objective is to analyze how robust reconstruction is to PSF mismatch across varying object depths when using a single, fixed PSF for deconvolution. To isolate the effect of depth and avoid confounding factors such as sensor noise or quantization, all simulations are performed in a noiseless setting. The geometric parameters of the optical system used to simulate PSF scaling with depth are summarized in Table 4.1.

For each mask type, we simulate PSFs at four depths: 1, 3, 5, and 10 cm from the coded

Table 4.1: Optical parameters used to simulate PSF scaling with depth for the MNIST and QR-code experiments. Reproduced from [1].

	MNIST simulation	QR-code simulation
Mask aperture a	1.6 mm	1.6 mm
Mask–sensor distance d	4 mm	4 mm
Sensor width	6.5 mm	6.5 mm
Pixel pitch	$6.5/500 \approx 0.013$ mm	$6.5/360 \approx 0.018$ mm
Mask–object distance D	10, 30, 50, 100 mm	10, 100 mm

mask. MNIST digit images are placed at each of these depths, and synthetic sensor measurements are generated by convolving the object with the corresponding depth-specific PSF. The geometric parameters of the optical system and the PSF dimensions are summarized in Table 4.1.

Reconstruction and quantitative analysis. Reconstruction is performed using ADMM [51] under a shift-invariant convolutional model, with a single PSF calibrated at either 10 cm or 1 cm. This allows us to qualitatively assess how each calibrated PSF generalizes to other depths for each mask design.

To quantify DOF, we compute the average mean squared error (MSE) over 100 MNIST digit reconstructions for each pair of calibration depth and object depth. This yields a 4×4 “confusion matrix”, where each entry corresponds to the average reconstruction error when using a PSF calibrated at depth $z_{d_1} \in \{1, 3, 5, 10\}$ to reconstruct an object located at depth $z_{d_2} \in \{1, 3, 5, 10\}$.

Results. The qualitative simulation results in Fig. 4.4 (left) show that the restricted-aperture mask is more sensitive to PSF mismatch: a PSF calibrated at 10 cm yields poor reconstructions at 1 cm, and a PSF calibrated at 1 cm performs poorly at 10 cm. In contrast, the full-aperture radial mask produces visually consistent reconstructions across depths, even when there is a large mismatch between calibration and object depth. The MSE confusion matrices in Fig. 4.4 confirm this trend quantitatively: the full-aperture mask exhibits lower errors in the off-diagonal

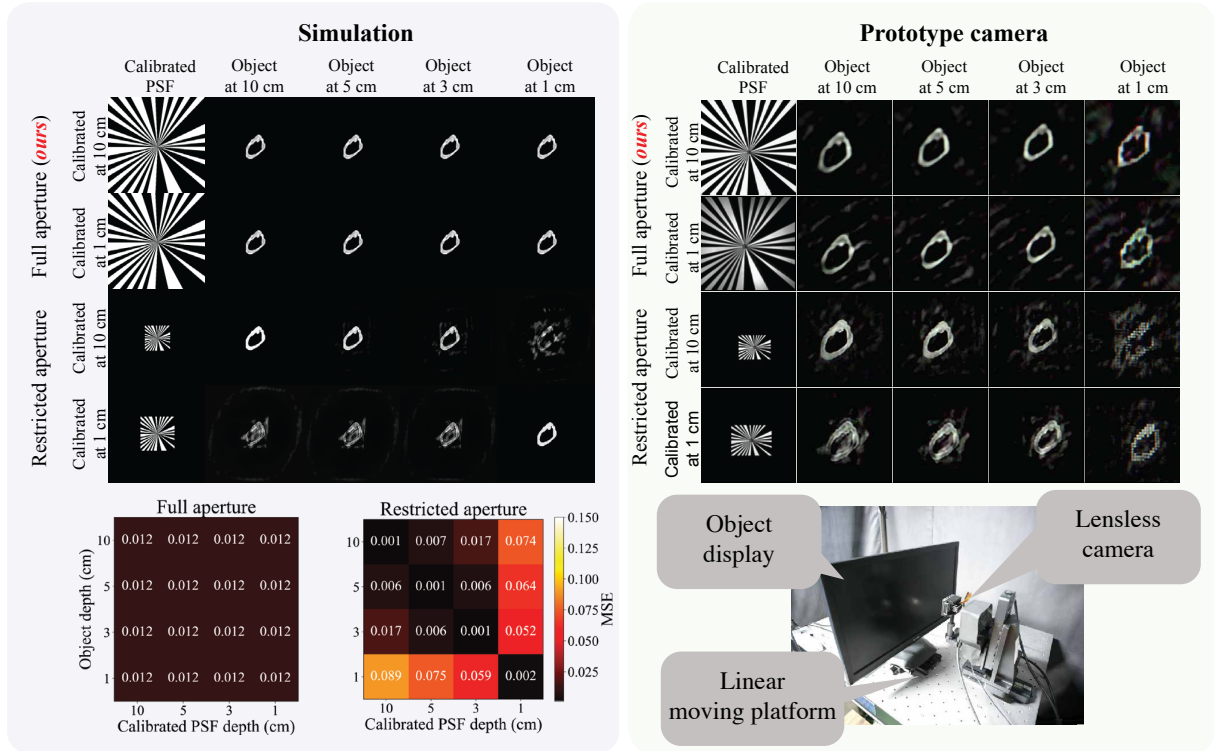


Figure 4.4: MNIST digit reconstruction for simulated and prototype-camera experiments. Left: simulation results for a representative digit and the corresponding MSE confusion matrices showing the impact of PSF depth mismatch for restricted- and full-aperture radial masks. Right: reconstructions of the same digit from real measurements captured with the prototype camera using a monitor-based display setup at multiple depths. Reproduced from [1].

entries, indicating a broader depth-invariant response than the restricted-aperture design.

4.3.2 MNIST prototype camera experiment

Experimental setup. We next evaluate the depth-dependent reconstruction performance of the restricted- and full-aperture radial masks using real measurements captured by our prototype lensless camera. The goals are twofold: (i) to validate, under real-world conditions, the DOF behavior observed in simulation, and (ii) to assess whether the convolutional model approximation provides a fair basis for analyzing DOF in the full-aperture setting.

MNIST digits are displayed on a monitor mounted on a platform with adjustable z -axis translation, allowing us to control the display depth, as illustrated in Fig. 4.4(right). We capture measurements at four object depths: 1, 3, 5, and 10 cm from the coded mask. The digit size on the monitor is chosen to be 20×20 pixels at 1 cm and to scale linearly with depth, resulting in sizes of 60×60 pixels at 3 cm, 100×100 pixels at 5 cm, and 200×200 pixels at 10 cm. Given a monitor pixel pitch of approximately $116 \mu\text{m}$, the smallest object has a size of roughly 2.3 mm and the largest of roughly 23 mm on the display. This setup yields small, localized objects near the optical axis, for which the shift-invariant approximation is most reasonable.

Reconstruction procedure. For each mask design, the objects are reconstructed using two fixed PSFs, one calibrated at 1 cm and the other at 10 cm, allowing us to evaluate how well each PSF generalizes across depth. Reconstructions are obtained using ADMM under a shift-invariant convolutional model with total variation regularization, mirroring the simulation setting in Sec. 4.3.1. This ensures that qualitative trends in depth sensitivity can be compared directly between simulated and real data.

Results. Figure 4.4 (right) shows representative reconstruction results for the prototype camera. For the restricted-aperture mask, reconstruction quality degrades markedly when the object is placed far from the calibration depth: digits at 1 cm become blurry and distorted when reconstructed with the 10 cm PSF, and the inverse mismatch exhibits similar artifacts. This behavior highlights the limited DOF of the restricted-aperture system. In contrast, the full-aperture radial

mask produces visually consistent reconstructions across all tested depths for both calibration PSFs, confirming its robustness to PSF mismatch in practice. Although minor discrepancies relative to simulation are expected due to factors such as noise, calibration errors, and monitor non-idealities, the overall trends align well with the results of Sec. 4.3.1 and support the conclusion that the full-aperture radial mask provides substantially improved depth invariance for localized objects.

4.4 Data-driven shift-variant deconvolution for large objects

In this section, we evaluate the performance of the proposed full-aperture radial mask on scenes with larger objects. Specifically, we consider QR codes placed at different distances from the camera. In contrast to the localized MNIST digits used in Sec. 4.3, these QR codes occupy a substantial portion of the field of view and thus experience strong spatial variation in the observed PSF. As illustrated in the PSF analysis (Sec. 4.2.3), the shift-invariant convolutional model becomes a poor approximation in this setting, and the deconvolution methods used up to this section in this thesis are no longer applicable.

To handle this shift-variance, we adopt a supervised learning approach: we train a deep neural network (U-Net) on lensless captures acquired at a single far depth (10 cm) and evaluate its generalization to a closer, unseen depth (2 cm). Since the network is trained on data from only one depth, successful reconstruction at 2 cm indicates that the *optical system*—rather than the network—supports depth-invariant reconstruction. We present experiments in both simulation (Sec. 4.4.1) and with real measurements from the prototype camera (Sec. 4.4.2), comparing the full-aperture and restricted-aperture radial masks under the same learning-based regime.

4.4.1 QR code simulation experiment

Motivation and setting. The MNIST experiments in Sec. 4.3 demonstrated extended DOF for small, centered digits, where the shift-invariant model remains a reasonable approximation. Many practical applications, however, involve larger objects that span a significant portion of

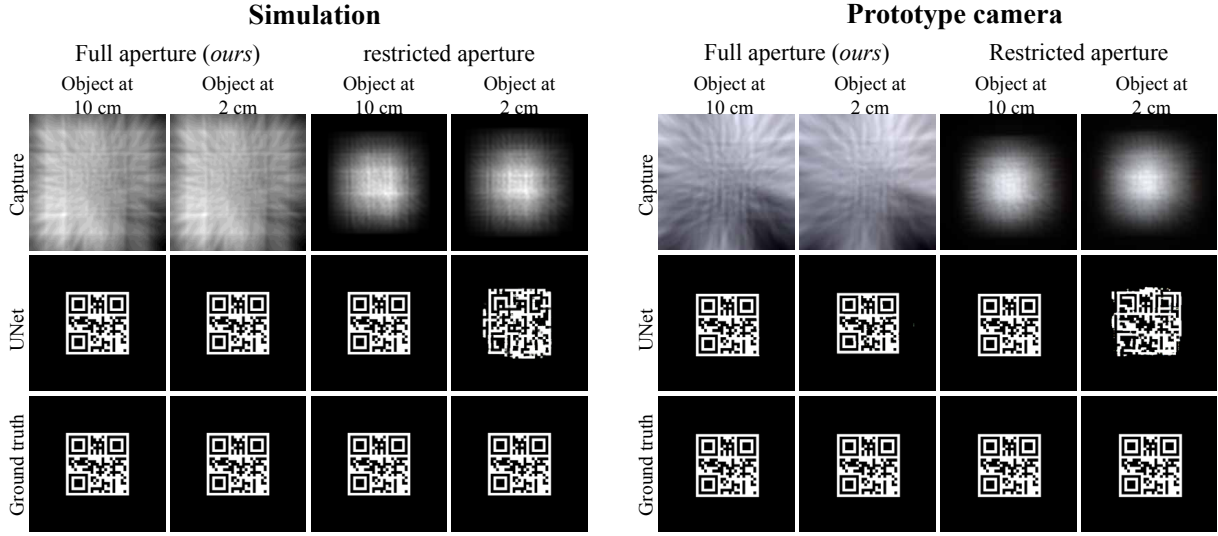


Figure 4.5: U-Net reconstruction of QR codes for near and far depths using a model trained only on far-distance captures. Left: simulation results for restricted- and full-aperture radial masks. Right: corresponding reconstructions from real measurements acquired with the prototype camera. Reproduced from [1].

the camera’s FOV. For such scenes, the PSF varies strongly across the field of view and the convolutional model breaks down. Here we investigate whether the proposed full-aperture radial mask can still support depth-robust reconstruction of large QR codes by using an implicitly learned shift-variant inverse mapping.

Experimental setup. We consider binary QR codes as test patterns and train a U-Net model [17] to map lensless measurements to the corresponding QR code images. The training set consists of 20,000 synthetically generated QR codes, created from random alphanumeric strings of up to 20 characters. For both full- and restricted-aperture masks, we simulate forward captures at two depths: 10 cm (used for training) and 2 cm (used only for testing). All simulations are performed at a resolution of 360×360 pixels, matching the prototype experiments in Sec. 4.4.2.

Neural reconstruction model. For training, we use a U-Net architecture following the implementation of [17], with a sigmoid activation in the final layer to produce outputs in $[0, 1]$ suitable for binary reconstruction. The network is trained for 300 epochs using the Adam opti-

mizer [71] with a learning rate of 1×10^{-4} and a batch size of 64. To encourage robustness to small misalignments and to better match the prototype setup, we apply random cropping from 360×360 to 300×300 pixels during training. We monitor validation loss during training and retain the model checkpoint that achieves the lowest validation error.

Results. Figure 4.5 (left) shows representative reconstruction results for a test QR code at both the training depth (10 cm) and the unseen test depth (2 cm). At 10 cm, both the restricted- and full-aperture masks produce accurate reconstructions, indicating that the learned model can successfully invert the lensless measurements at the training depth for either optical design. For the depth of 2 cm unseen during training, however, the behavior diverges: only the full-aperture radial mask continues to yield reliably decodable QR codes, whereas the restricted-aperture mask produces reconstructions that are severely degraded.

To quantify this effect, we evaluate reconstruction performance over 100 randomly generated QR codes at 2 cm. For each reconstruction, we decode the output image using a standard QR-code reader and compare the recovered string with the ground-truth content. The full-aperture mask achieves 100% reconstruction accuracy, recovering the correct encoded content in all 100 cases. In contrast, the restricted-aperture mask fails to produce a single correctly decoded QR code under the same conditions. These results indicate that, when paired with a learned model, the full-aperture radial mask supports significantly better depth generalization for large, shift-variant scenes than the restricted-aperture design.

4.4.2 QR code prototype camera experiment

Experimental setup and dataset. We now replicate the QR-code experiment using real measurements captured with the prototype camera, with the goal of validating the simulation findings in a physical setting. The overall setup mirrors that of the MNIST experiments in Fig. 4.4: a monitor displays QR codes at controlled depths in front of the lensless camera.

For training, we display QR codes at a fixed distance of 10 cm from the coded mask and capture lensless measurements for both the restricted- and full-aperture radial masks. The tar-

get images are the corresponding QR-code renderings shown on the monitor. The resulting dataset consists of three components: (1) the 20,000 synthetic QR-code targets, (2) the associated lensless measurements captured with the restricted-aperture mask, and (3) those captured with the full-aperture mask. The sensor images, originally 1464×1936 pixels, are cropped to 1440×1440 pixels and then downsampled to 360×360 to match the simulation resolution.

To test generalization, we display and capture 10 additional QR codes at a closer distance of approximately 2 cm from the coded mask, using the same monitor and camera configuration. These close-up captures are never shown during training and serve as a real-world counterpart to the 2 cm test depth in simulation.

Neural model and training protocol. The reconstruction network and training protocol are identical to those used in the simulation experiment (Sec. 4.4.1). For each mask type, we train a separate U-Net using the 10 cm training data and random cropping from 360×360 to 300×300 pixels, which helps remove alignment bias as a confounding factor. As before, we select the model checkpoint that achieves the lowest validation loss.

Results. Figure 4.5 (right) shows representative reconstructions from the prototype camera experiments. At the training depth of 10 cm, both the restricted- and full-aperture masks yield accurate and decodable QR codes, in line with the simulation results. When evaluated on the extreme close-up captures at 2 cm, however, only the full-aperture radial mask continues to produce reconstructions that can be reliably decoded. Although the close-up reconstructions for the full-aperture mask exhibit minor visual deviations from the target, the encoded content is recovered correctly in all tested cases. In contrast, the restricted-aperture mask fails to produce any correctly decoded QR codes at 2 cm; the reconstructed patterns are visibly distorted and cannot be decoded by a standard QR reader.

We repeated the 2 cm evaluation for 10 different QR codes. All reconstructions obtained with the full-aperture mask were successfully decoded, while none of the reconstructions from the restricted-aperture mask yielded the correct content. These findings reinforce the simulation results of Sec. 4.4.1 and further confirm that, when combined with a learned shift-variant deconvolution, the full-aperture radial mask provides wider depth generalization for large objects

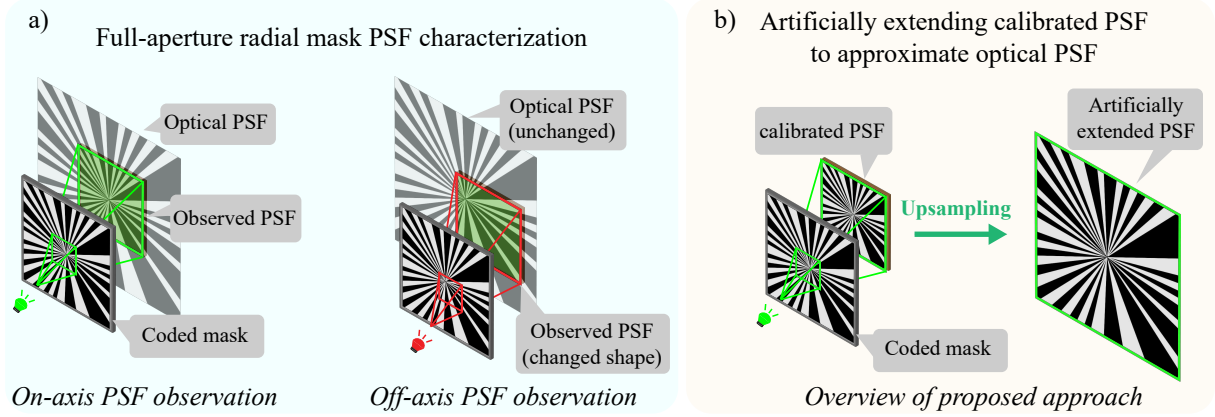


Figure 4.6: (a) Optical (unbounded) and observed (sensor-cropped) PSFs for on-axis and off-axis point sources: the on-axis case preserves shape up to scale, whereas off-axis shifts cause the observed PSF to change shape due to cropping. (b) Proposed extended-PSF approximation, where a centrally calibrated PSF is upsampled to approximate the optical PSF, enabling an approximately shift-invariant convolutional model followed by cropping. Reproduced from [1].

compared to its restricted-aperture counterpart.

4.5 Approximate shift-invariant deconvolution with extended PSFs

4.5.1 Extended-PSF forward-model approximation

Motivation. The experiments in Secs. 4.3 and 4.4 were intentionally designed to *work around* the increased shift-variance of the full-aperture radial mask. In the localized MNIST setting (Sec. 4.3), we restricted the object size and position so that the PSF could be approximated as shift-invariant. In the QR-code experiments (Sec. 4.4), we instead relied on a data-driven model to implicitly learn the spatially varying deconvolution. While both strategies were used to demonstrate the depth robustness of the full-aperture design, neither provides a general, physics-based forward model that can be used with standard deconvolution algorithms for large objects

and continuous-depth scenes.

In this section, we propose an alternative approach. Rather than constraining the scene or relying on learned inversion, we introduce a forward-model approximation that allows the full-aperture system to be treated as *approximately shift-invariant* by using an *extended PSF*. This approximation enables shift-invariant deconvolution for large planar targets and continuous-depth scenes, using the same class of convex optimization methods employed in the restricted-aperture case.

Optical versus observed PSF. To formalize this idea, we distinguish between the *optical PSF* and the *observed PSF*. The optical PSF corresponds to the full shadow of the coded mask that would be formed on an ideal, unbounded sensor plane: it extends beyond the physical sensor area and reflects only the geometry of the mask, the mask–sensor distance, and the object depth. In contrast, the observed PSF is the portion of this shadow that actually falls onto the finite sensor and is therefore captured in practice.

When a point source lies on the optical axis, the optical PSF is centered on the sensor. In this case, the observed PSF can be seen as a central crop of the optical PSF, and for a radial pattern specifically, their structure matches up to a scaling factor, so the two are effectively equivalent for modeling purposes. When the point source shifts laterally on the same depth plane, the optical PSF is simply translated across the (conceptual) infinite sensor plane, much like the shift-invariant PSFs considered in the previous chapter. However, because the physical sensor covers only a fixed window, the observed PSF becomes a crop of a *different* region of the same optical PSF. As a result, the *optical* PSF remains shift-invariant (it only translates), while the *observed* PSF becomes strongly shift-variant, since its shape changes with the cropping region selected by the sensor (see Fig. 4.6-a).

Extended PSF and shift-invariant modeling. The forward-model approximation we propose is to model the system at the level of the optical PSF rather than the observed PSF. Since the optical PSF is shift-invariant across lateral shifts at a fixed depth, we can write an approximate forward model as

$$\mathbf{y} \approx \mathcal{C}(\mathbf{h}_{\text{opt}} * \mathbf{x}),$$

where $\mathbf{h}_{\text{opt}}$ denotes an extended PSF representing the full mask shadow on an idealized plane, $*$ is 2D convolution, $\mathbf{x}$ is the scene, and $\mathcal{C}(\cdot)$ crops the convolved image to the sensor region. All the position dependence is concentrated in the cropping operator, while the convolution itself remains shift-invariant and can be implemented efficiently.

In practice, we do not directly observe $\mathbf{h}_{\text{opt}}$, since the sensor always crops the PSF. Our approach is therefore to construct an *artificially extended PSF* $\mathbf{h}_{\text{ext}}$ that approximates $\mathbf{h}_{\text{opt}}$ by synthetically enlarging a centrally calibrated observed PSF. For the full-aperture radial mask, this construction is reasonable because the mask pattern is approximately scale-invariant: upscaling the central observed PSF preserves its structural properties and provides a plausible estimate of the unseen outer regions of the coded mask. In other words, a single centrally calibrated PSF, together with an upsampling operation, is sufficient to approximate the full extended kernel, as shown in Fig 4.6(b).

By contrast, for generic full-aperture masks (e.g., random patterns or non-radial designs) simple upscaling of a central PSF would not faithfully represent the unobserved regions: different lateral positions would reveal genuinely different mask structures that are never seen by the central calibration. Accurately recovering the optical PSF in such cases would require explicit shift-variant calibration at multiple positions, which is considerably more complex. Thus, the full-aperture radial mask is attractive not only because it improves DOF, but also because its geometric structure makes the extended-PSF approximation both physically plausible and easy to calibrate from a single on-axis measurement.

4.5.2 Validating the extended-PSF shift-invariant model

Motivation. The extended-PSF approximation proposed in Sec. 4.5.1 models the full-aperture system as a shift-invariant convolution with a synthetically enlarged PSF, followed by cropping to the sensor region. Before relying on this model for large-scale reconstructions, we must verify two key aspects in practice: (1) that an extended PSF constructed from a single centrally calibrated measurement can indeed compensate for the strong shift-variance of the full-aperture

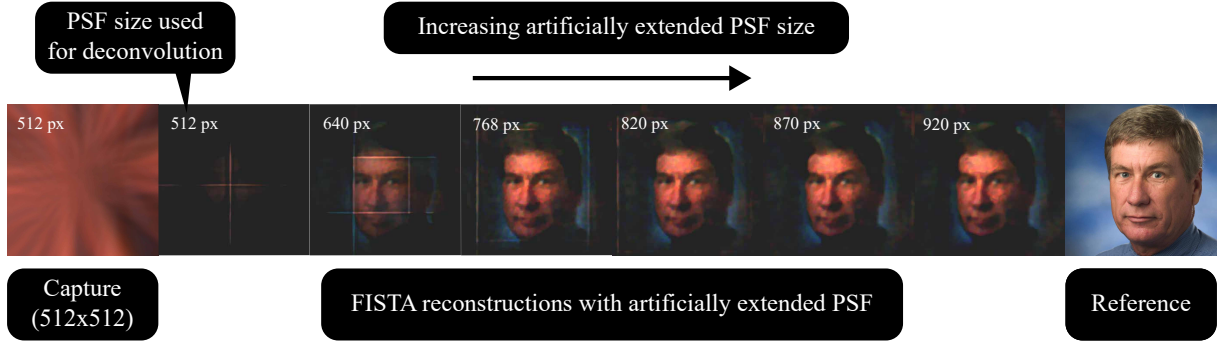


Figure 4.7: Results of the artificially extended PSF size sweep for an FFHQ capture. Adapted from [1].

mask, and (2) that there exists a practical kernel size (i.e., upsampling factor) for which reconstructions are stable and free of artifacts. To this end, we perform a PSF-size sweep using real FFHQ captures from the prototype camera, and empirically identify the smallest extended PSF that yields reliable reconstructions under the proposed shift-invariant framework.

Experimental setup. To determine a suitable size for the extended PSF, we perform a sweep experiment using real measurements from our prototype lensless camera and face images from the FFHQ dataset [73]. We adopt the same display configuration as in the MNIST experiments (Sec. 4.3): FFHQ images are shown on an LG Ultrafine 5K monitor and captured by the prototype camera equipped with both the restricted- and full-aperture radial masks. For this subsection, we focus on a representative capture obtained with the full-aperture mask.

The monitor is placed approximately 40 cm in front of the coded mask. We display 1800×1800 -pixel face images, which correspond to a physical object size of roughly 21×21 cm given the monitor pixel pitch of $116.7 \mu\text{m}$. During capture, the sensor gain is fixed at 0 dB and gamma correction is disabled. Exposure time is varied in later experiments to compare light efficiency, but for the PSF-size sweep we use a single exposure setting that yields a well-exposed measurement. The raw lensless image is cropped from the native sensor resolution of 1464×1936 pixels to a square region of 1464×1464 pixels and then downsampled to 512×512 pixels for reconstruction.

Extended-PSF candidates and reconstruction procedure. The full-aperture PSF is calibrated

using an on-axis point source placed at the same depth as the display (approximately 40 cm). From this centrally calibrated observed PSF, we construct a family of extended PSFs by synthetically enlarging its support. Specifically, we consider six candidate sizes for the extended kernel:

$$512 \times 512, \quad 640 \times 640, \quad 768 \times 768, \quad 820 \times 820, \quad 870 \times 870, \quad 920 \times 920$$

For each candidate size, we upsample the calibrated PSF to the desired resolution and use it as $\mathbf{h}_{\text{ext}}$ in the shift-invariant forward model

$$\mathbf{y} \approx \mathcal{C}(\mathbf{h}_{\text{ext}} * \mathbf{x}),$$

where $\mathbf{x}$ is the unknown scene, and $\mathcal{C}(\cdot)$ crops the convolved image back to the 512×512 measurement region.

Reconstruction is performed using FISTA [49] with the same regularization and optimization settings employed later in Sec. 4.5.3. For each PSF size, we reconstruct the image for 10,000 iterations. When $\mathbf{h}_{\text{ext}}$ is larger than the 512×512 measurement, the forward model naturally produces an estimated measurement with the same size as the PSF; this estimate is then center-cropped to 512×512 before computing the data-fidelity term. Representative reconstructions for each PSF size are shown in Fig. 4.7.

Results and choice of PSF size. The sweep reveals a clear dependence of reconstruction quality on the extended-PSF size. When the PSF support matches the measurement resolution (512×512), the reconstruction fails to recover a recognizable face: the model cannot adequately compensate for the strong shift-variance induced by sensor cropping of the full-aperture mask. Enlarging the PSF to 640×640 and 768×768 improves the result but still introduces visible artifacts, indicating that the extended kernel does not yet capture the effective spatial extent of the optical PSF.

At a size of 820×820 , these artifacts largely disappear and the reconstruction becomes visually stable and consistent, while further increases in kernel size do not produce noticeable improvements for this setup. We therefore adopt an extended PSF of 820×820 as the smallest

size that yields artifact-free reconstructions for large planar FFHQ targets. This choice strikes a practical balance between physical plausibility and computational cost, and is used in all subsequent experiments with the full-aperture radial mask in this chapter.

4.5.3 In-focus FFHQ reconstruction: simulation and prototype

Motivation. Having validated in Sec. 4.5.2 that an extended PSF of size 820×820 enables stable shift-invariant deconvolution for the full-aperture radial mask, we now investigate how this model performs relative to the restricted-aperture mask when reconstructing *in-focus* planar targets. In this section, the scene depth is fixed at the PSF calibration distance, so depth-of-field is not the limiting factor. Instead, we focus on the trade-off between sharpness, noise, and light efficiency for the two mask designs under a common shift-invariant reconstruction framework. We use both simulated and prototype-camera experiments based on FFHQ face images to assess reconstruction quality.

Simulation setup. We first perform a controlled simulation to compare reconstruction quality quantitatively. Using the optical geometry described in Table 4.1, we simulate lensless measurements for 100 FFHQ images resized to 256×256 pixels. For each mask type, we generate measurements by zero-padding the target images to the PSF size and convolving them with the corresponding kernel: a 512×512 PSF for the restricted-aperture mask and an 820×820 extended PSF for the full-aperture mask. Gaussian noise with standard deviation $\sigma = 0.005$ is added to the simulated measurements.

Reconstructions are performed with FISTA [49] under a shift-invariant forward model and identical regularization for both masks. For the full-aperture case, the estimated measurement is center-cropped to match the target resolution before computing the data-fidelity term, as in Sec. 4.5.2. We evaluate reconstruction quality using PSNR and LPIPS [72], reporting the mean and standard deviation over the 100 images.

Simulation results. Quantitative results are summarized in Table 4.2. The restricted-aperture mask achieves a PSNR of 24.17 ± 1.23 dB and an LPIPS score of 0.489 ± 0.039 , while the

full-aperture mask attains 24.11 ± 1.22 dB and 0.497 ± 0.035 , respectively. The differences are marginal, indicating that, at in-focus depths, the full-aperture radial mask combined with the extended-PSF model yields reconstruction quality that is essentially comparable to that of the restricted-aperture mask. Representative simulated reconstructions are shown in the bottom row of Fig. 4.8, and visually corroborate the quantitative metrics: both masks produce similar perceptual quality, with the restricted-aperture mask exhibiting slightly crisper high-frequency details.

Table 4.2: FFHQ simulation analysis of in-focus reconstruction quality for restricted- and full-aperture radial masks. Reproduced from [1].

Mask type	PSNR (dB) ($\uparrow$)	LPIPS ($\downarrow$)
Restricted-aperture mask	24.166 ± 1.23	0.489 ± 0.039
Full-aperture mask	24.110 ± 1.22	0.497 ± 0.035

Prototype-camera setup. We next validate these observations using real measurements from our prototype lensless camera. The hardware and display configuration follow the setup used in Sec. 4.5.2. FFHQ face images are shown on an LG Ultrafine 5K monitor placed approximately 40 cm from the coded mask. Each image is displayed at 1800×1800 pixels, corresponding to a physical size of roughly 21×21 cm for a pixel pitch of $116.7 \mu\text{m}$.

The prototype camera captures lensless images with both restricted- and full-aperture radial masks. During capture, the sensor gain is set to 0 dB and gamma correction is disabled. To evaluate light efficiency, we vary only the exposure time across four settings: 2000, 5000, 10,000, and $15,000 \mu\text{s}$. Raw measurements of size 1464×1936 pixels are cropped to a square region of 1464×1464 pixels and downsampled to 512×512 for reconstruction.

Prototype reconstruction model and results. Reconstruction from real measurements is performed with the same shift-invariant framework as in simulation. For the restricted-aperture mask, the calibrated PSF is resized to 512×512 and used directly as the convolution kernel. For the full-aperture mask, we employ the 820×820 extended PSF selected in Sec. 4.5.2: we

reconstruct an 820×820 image, generate an 820×820 estimated measurement, and then center-crop it to 512×512 for the data-fidelity term. FISTA is run for 10,000 iterations with identical regularization parameters for both masks. For visualization in Fig. 4.8 (top row), reconstructions are shown with enhanced brightness and contrast.

The prototype results align well with the simulation findings. At higher exposure times, both masks reconstruct visually convincing faces, with the restricted-aperture mask yielding slightly sharper fine detail. However, at the lowest exposure of $2000 \mu\text{s}$, the restricted-aperture reconstructions become significantly noisier, while the full-aperture mask still produces clean, stable reconstructions. This behaviour reflects the larger light-collecting area of the full-aperture design.

Discussion. Taken together, the simulation and prototype experiments indicate that, for in-focus scenes, the full-aperture radial mask combined with the extended-PSF shift-invariant model achieves reconstruction quality that is on par with the restricted-aperture mask, while offering markedly better light efficiency. This supports the use of full-aperture designs not only for their depth-of-field benefits (as shown in previous sections), but also as practical alternatives in low-light or short-exposure conditions where noise becomes a critical factor.

4.5.4 Continuous-depth scene reconstruction

Motivation. The FFHQ experiments in Sec. 4.5.3 evaluated the extended-PSF model in a planar, in-focus setting. To further assess the practical feasibility of the full-aperture radial mask and the proposed shift-invariant approximation, we now consider a scene with continuous depth variation. In this experiment, objects span a wide range of distances from the coded mask, so both depth-of-field and light efficiency play a central role. Our goal is to test whether the combination of full-aperture design and extended-PSF deconvolution can deliver robust reconstruction across depth, and how it compares to a conventional restricted-aperture mask under the same reconstruction framework.

Experimental setup. We use a small tabletop scene composed of two chess pawns (red and



Figure 4.8: Prototype and simulated results for the FFHQ reconstruction experiments of Sec. 4.5.3. Reproduced from [1].

green) and a plush toy, placed at different depths along the optical axis, from approximately 2 cm up to 50 cm from the coded mask. The approximate layout is illustrated in Fig. 4.9(left), where the plush toy is positioned near a reference depth of 40 cm and the red pawn is placed closer to the camera. For both restricted- and full-aperture radial masks, we calibrate the PSF using an on-axis point source at 40 cm, so that objects near this distance are nominally “in focus,” while those significantly closer or farther lie outside the calibration depth.

Lensless measurements are captured at four exposure times, $\{2000; 5000; 10,000; 15,000 \mu\text{s}\}$, with sensor gain fixed to 0 dB and gamma correction disabled, mirroring the settings used in Sec. 4.5.3. Raw images of size 1464×1936 pixels are cropped to a square region of 1464×1464 pixels and downsampled to 512×512 pixels prior to reconstruction.

Reconstruction model. Reconstruction follows the same shift-invariant deconvolution pipeline used for the FFHQ experiments. For the restricted-aperture mask, the calibrated PSF is resized to 512×512 and used directly as the convolution kernel. For the full-aperture mask, we employ the 820×820 extended PSF introduced in Sec. 4.5.2. As before, we reconstruct an 820×820 image, generate an estimated measurement of the same size via convolution, and then center-crop it to 512×512 for the data-fidelity term. In both cases, we use FISTA [49] with identical regularization and run 10,000 iterations. Reconstructions at different exposure times are displayed in Fig. 4.9(right).

Results and discussion. For the restricted-aperture mask, objects near the calibration depth (around 40 cm) are reconstructed with relatively sharp edges, especially at higher exposure times. However, the red pawn, which lies at a considerably different depth, appears noticeably degraded: its shape is blurred and corrupted by artifacts, and the reconstruction becomes increasingly noisy at lower exposures. This behaviour reflects the limited DOF of the restricted-aperture system and its lower light throughput.

In contrast, the full-aperture radial mask produces more balanced reconstructions across the entire depth range. While some fine details are slightly less sharp than in the best-case restricted-aperture reconstructions, the overall structure of the objects across all depths is preserved, and the reconstructions remain significantly less noisy at short exposure times. The red pawn, in

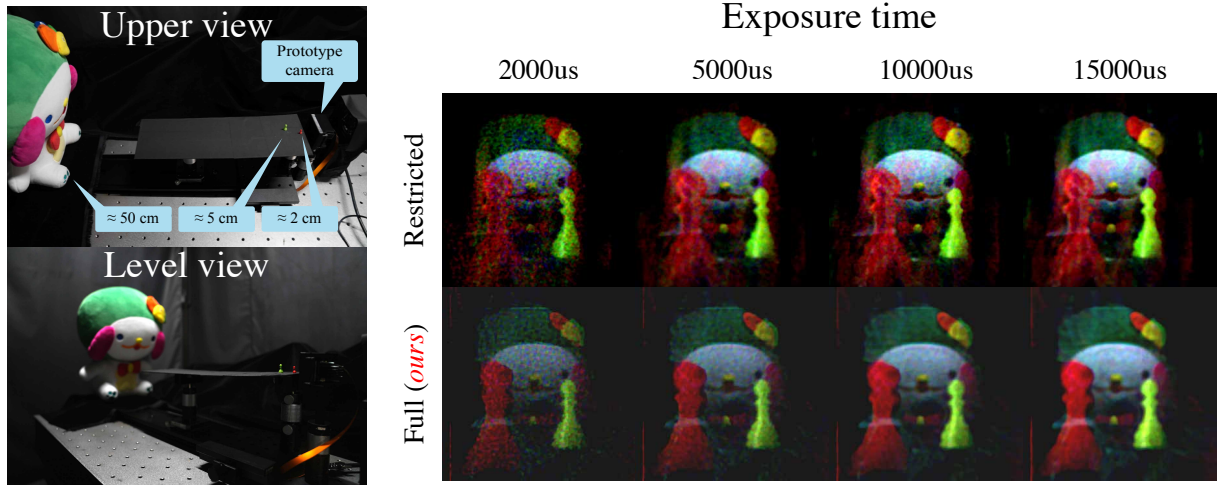


Figure 4.9: Experimental setup for the continuous depth scene reconstruction. (Left) Scene to be imaged, containing two pawns near the prototype camera (i.e., ≈ 2 and ≈ 5 cm, and a plushie toy placed 50 cm from the camera. (Right) Reconstructions of the captured scene for restricted- and full-aperture masks for multiple exposure times. Reproduced from [1].

particular, is much better resolved by the full-aperture mask under the extended-PSF model, demonstrating improved robustness to depth mismatch.

These continuous-depth results complement the planar FFHQ experiments: together, they show that the full-aperture radial mask, combined with the extended-PSF shift-invariant approximation, can deliver near in-focus quality, higher light efficiency, and more uniform reconstruction across depth than a restricted-aperture baseline, without resorting to depth estimation or multi-plane PSF calibration.

4.6 Discussion and summary

Summary of contributions. This chapter advances the radial-mask framework along two main axes. First, we move from an SLM-based prototype to a printed chromium mask, increasing mask resolution by an order of magnitude. With this higher-fidelity hardware, we revisited the depth-of-field behavior of the *restricted-aperture* radial mask and found that, in a continuous-

depth tabletop scene, its usable DOF is broader than suggested in the previous chapter: recognizable reconstructions extend from the far object at about 50 cm down to roughly 3 cm from the mask. Second, we extend the radial pattern to the *full aperture* and systematically compare restricted- and full-aperture designs. Across MNIST, QR-code, FFHQ, and continuous-depth experiments, we show that the full-aperture radial mask further widens the effective DOF, enabling reconstructions down to approximately 1 cm from the mask while maintaining in-focus quality near the restricted-aperture mask and offering significantly higher light throughput.

Depth robustness and shift-variance. The multi-depth PSF calibration experiments showed that extending the radial pattern across the full aperture significantly improves depth invariance: cross-correlation analysis across depth planes indicated that full-aperture PSFs change more slowly with depth than restricted-aperture PSFs, bringing the system closer to the ideal behavior illustrated in Chapter 3.1. At the same time, PSF measurements at different lateral positions revealed that full-aperture masks are more strongly shift-variant: different parts of the radial pattern are projected onto the sensor for off-axis point sources, leading to position-dependent PSF shapes. This trade-off of enhanced depth robustness at the cost of increased shift-variance motivated the subsequent experimental design and the development of the extended-PSF model.

Localized and learned shift-variant reconstruction. In the MNIST experiments, we restricted the scene to small digits centered near the optical axis so that the PSF variation across the object support remained limited. Under this localized regime, the full-aperture radial mask demonstrated a markedly broader effective depth of field than the restricted-aperture mask, even when using a single calibrated PSF and a standard ADMM-based deconvolution. The QR-code experiments relaxed the localization constraint by using large binary patterns that occupy a substantial portion of the field of view. Here, the shift-invariant convolutional model breaks down, and we instead employed a U-Net trained at a single depth. Both in simulation and with the prototype camera, the learned model generalizes across depth only when paired with the full-aperture radial mask: QR codes captured at unseen close-up distances are correctly decoded for the full aperture and fail for the restricted aperture. These results confirm that the optical design itself, rather than the network architecture, is responsible in enabling depth-generalization.

Extended-PSF approximation and shift-invariant deconvolution. To handle the stronger shift-variance of the full-aperture mask while retaining the efficiency of 2D deconvolution, we introduced an extended-PSF approximation: a centrally calibrated PSF is upsampled to a larger kernel that approximates the underlying optical PSF, and used in a shift-invariant convolution followed by cropping. A simple sweep over candidate PSF sizes on prototype FFHQ data showed that an 820×820 extended kernel is sufficient to obtain stable, artifact-free reconstructions. With this model in place, both simulated and real FFHQ experiments demonstrated that the full-aperture radial mask achieves in-focus reconstruction quality near the restricted-aperture baseline, while offering notably better performance under short exposure times and more wider DOF reconstruction in continuous-depth scenes.

Implications and limitations. Taken together, the results of this chapter paint a coherent picture: full-aperture radial masks combine (i) improved depth robustness, (ii) higher light throughput, and (iii) compatibility with efficient shift-invariant deconvolution via the extended-PSF approximation. At the same time, the approximation is not exact: residual shift-variance remains, and the method requires tuning of the extended PSF size and regularization parameters. Additionally, extremely wide fields of view or highly off-axis content may still challenge the model. Finally, the extended-PSF approximation critically relies on the scale-invariant geometry of the radial pattern and is not expected to transfer directly to arbitrary mask designs.

Connection to subsequent chapters. From the perspective of the overall thesis narrative, this chapter strengthens the case for radial coded masks as a practical optical front-end for lensless imaging. Chapter 3.1 showed that restricted-aperture radial masks can extend a lensless camera’s depth of field while remaining compatible with simple 2D deconvolution techniques. The present chapter demonstrates that a full-aperture version can achieve even wider DOF while having better light efficiency, and that its apparent shift-variance can be effectively managed through an extended-PSF forward model. In the subsequent chapters, we build on the results obtained so far and turn our attention to unsupervised reconstruction methods, an area of research under-explored in the recent literature. Our goal moving forward is to enable higher-quality lensless imaging without relying on beam-splitter-based supervision by relying on ground-truth-

less reconstruction methods.

Chapter 5

Self-supervised lensless imaging

In the previous chapters, we focused on the optical design of lensless cameras, showing how radial coded masks can substantially alleviate depth-of-field limitations and enable extended-DoF imaging with 2D reconstruction models. In this chapter, we turn to the *reconstruction side* of the problem. Our goal is to develop lensless reconstruction frameworks that (i) avoid reliance on paired lensed "ground-truth" images, and (ii) leverage neural priors to achieve fast, single-pass reconstructions once trained, while remaining tightly coupled to the underlying physics model.

5.1 Motivation

As discussed in Sec. 2.3, lensless reconstruction can be cast as a noisy linear inverse problem

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \quad (5.1)$$

where $\mathbf{A}$ encodes the physics of the lensless camera (coded mask, propagation, sensor response) and $\mathbf{n}$ models noise and mismatch. Classical model-based methods such as FISTA and ADMM reconstruct $\mathbf{x}$ by minimizing a regularized objective for each measurement, combining the data term with an explicit hand-crafted prior (Sec. 2.3). These approaches are interpretable and require no training data, but their expressiveness is limited and each new reconstruction may

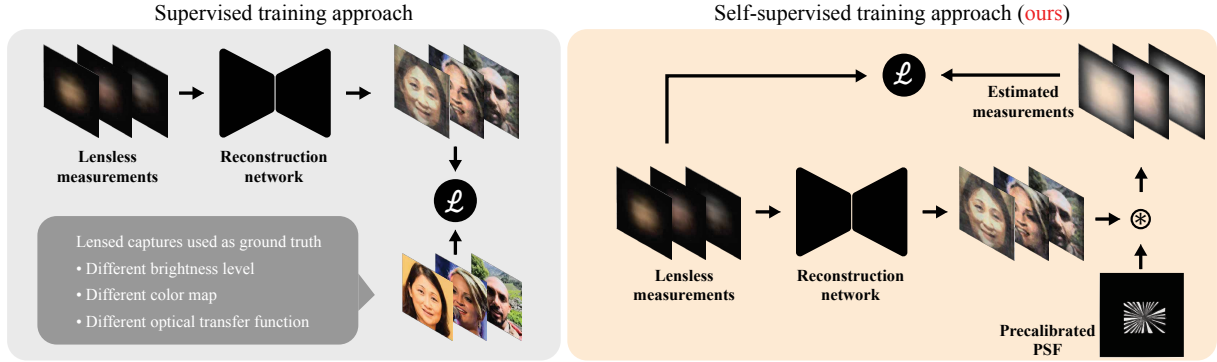


Figure 5.1: (Left) Supervised approach to train neural models for lensless imaging. (Right) Self-supervised approach for training neural models for lensless reconstruction. *Reproduced with permission of Springer Nature [2].*

require dozens or hundreds of iterations.

Supervised learning offers a complementary path: a reconstruction network f_θ is trained on paired lensless–lensed datasets, typically acquired via beam-splitter setups. Once trained, f_θ produces reconstructions in a single forward pass, dramatically reducing per-image compute. However, such models inherit the biases of the reference camera, depend on carefully aligned paired data, and are usually restricted to planar, depth-limited scenes.

In this chapter, we explore an alternative: *self-supervised* lensless reconstruction with neural priors. Instead of training on lensed ground truth, we train a reconstructor f_θ directly from lensless measurements and a known forward model, using a loss defined in the measurement domain. This keeps the reconstruction anchored to the physics of the lensless system, avoids cross-optics supervision bias, and—once training is complete—enables fast single-pass reconstructions comparable in speed to supervised networks but learned from measurement-only data. The following section formalizes this self-supervised formulation and describes the proposed architectures.

5.2 Methodology

Problem setting. Throughout this chapter we focus on the shift-invariant forward model introduced in Sec. 2.1.4, written as

$$\mathbf{y} = \mathbf{h} * \mathbf{x} + \mathbf{n},$$

where $\mathbf{x}$ is the unknown scene, $\mathbf{h}$ is a calibrated PSF for the lensless camera, and $\mathbf{n}$ models measurement noise and residual mismatch. Given only a set of lensless measurements $\{\mathbf{y}^i\}_{i=1}^N$ and the forward model in Eq. (5.2), our goal is to learn the mapping function

$$f_\theta : \mathbf{y} \mapsto \hat{\mathbf{x}}$$

parameterized by θ , that can be applied in a single forward pass at test time.

From one-image priors to pretrained network. A closely related idea to our work is the Deep Image Prior (DIP) [55], where a convolutional network is optimized *per image* so that its output, when passed through the forward model, matches the observed measurement. In the lensless setting, UDN [56] extends this idea by using a UNet whose weights are optimized from scratch for each measurement, leveraging the a convolutional network’s inductive bias as a powerful prior to filter the measurement noise.

While such DIP-style methods can produce high-quality reconstructions and require no training dataset, the neural model must be retrained from scratch for each image to be reconstructed. In contrast, our aim is to *pretrain* a single reconstruction network f_θ in a self-supervised fashion, so that after training, each new measurement can be reconstructed with a single forward pass while still being constrained by the physics model.

Self-supervised objective. Instead of minimizing a supervised loss between $f_\theta(\mathbf{y}^i)$ and paired lensed targets (as shown in Fig. 5.1-left), we define a self-supervised loss directly in the measurement domain (as shown in Fig. 5.1-right). We simulate the measurement corresponding to the reconstruction as $\tilde{\mathbf{y}}^i = f_\theta(\mathbf{y}^i) * \mathbf{h}$ and penalize its discrepancy from the observed measurement $\mathbf{y}^i$. Formally, we solve

$$\theta^* = \arg \min_{\theta} \sum_{i=1}^N \mathcal{L}(f_\theta(\mathbf{y}^i) * \mathbf{h}, \mathbf{y}^i), \quad (5.2)$$

where $\mathcal{L}$ is the mean squared error loss.

This objective has a structure similar to supervised training, but replaces ground-truth lensed images by a physics-consistency term. The reconstruction network f_θ is encouraged to produce reconstructions that, when mapped back to the measurement space through the known forward model, reproduce the actual lensless measurements.

Expected benefits. By minimizing (5.2), the learned reconstruction network combines three desirable properties:

1. *Fast inference.* After training, each new measurement is reconstructed by a single forward pass through f_θ , making inference as fast as in supervised pipelines and significantly faster than iterative optimization from scratch.
2. *Physics consistency.* The loss is defined through the forward model $\mathcal{F}$, ensuring that reconstructions remain consistent with the actual measurements and the calibrated PSF.
3. *Implicit neural prior.* The convolutional architecture and training dynamics of f_θ provide an implicit regularization effect, similar in spirit to DIP and UDN [55, 56], but amortized over a dataset rather than per-image optimization.

Hybrid physics–learned architectures. To assess how self-supervised training compares to classical convex optimization, we consider three architectures of increasing coupling between the physics-based and learned components:

- *naive UNet.* A UNet g_θ with the architecture and parameters defined in [17] is trained directly with the self-supervised loss (5.2). Here, the network is responsible for both inverting the forward model and enforcing prior structure. In this scenario, the estimated reconstruction is obtained as:

$$\hat{\mathbf{x}}_{UNet} = g_\theta(\mathbf{y}).$$

- *Wiener-U.* A hybrid model that first applies a fixed Wiener-like deconvolution (Sec. 2.3) to provide an initial estimate of the information to be reconstructed and then refines the

result with the denoising UNet g_θ :

$$\hat{\mathbf{x}}_{\text{Wiener-U}} = g_\theta \left(\text{Wiener}(\mathbf{y}; \mathbf{h}) \right)$$

The physics layer handles a large portion of the inverse problem in a linear, interpretable way, while the UNet focuses on removing artifacts. The entire pipeline is trained end-to-end using (5.2), but only the UNet parameters are learnable.

- *ADMM-U*. A second hybrid model that replaces the Wiener step with a small, fixed number of ADMM iterations based on the DiffuserCam formulation [19]. Starting from the measurement $\mathbf{y}$, we run a fixed number of ADMM steps (e.g., five) to obtain an initial estimate of the scene information that incorporates a TV prior and non-negativity constraints. A UNet g_θ then refines this estimate as:

$$\hat{\mathbf{x}}_{\text{ADMM-U}} = g_\theta \left(\text{ADMM}(\mathbf{y}; \mathbf{h}) \right)$$

This design leverages a strong, interpretable model-based reconstructor as the first stage, while the UNet acts as a learned denoiser. As in Wiener-U, only the UNet is trained; the ADMM steps are fixed and derived from the known forward model and regularizer.

All three architectures are trained under the same self-supervised objective (5.2), differing only in how much of the inverse problem is handled by explicit physics (Wiener or ADMM) versus the neural network. This allows us to systematically evaluate the trade-offs between pure learned reconstruction and hybrid physics–learned designs in terms of reconstruction quality, robustness, and computational cost.

Relation to classical and supervised methods. Conceptually, the proposed self-supervised models sit between classical convex optimization and fully supervised deep networks:

- Like classical methods, they are anchored to the physical forward model and require no lensed ground-truth images.
- Like supervised methods, they leverage an offline training phase that enables fast single-pass inference for new measurements.

In the subsequent sections of this chapter, we evaluate these self-supervised models on both simulated and prototype-camera datasets, comparing them against Wiener deconvolution, ADMM, and UDN baselines, and comparing these methods in terms of reconstruction quality and inference speed without lensed supervision.

5.3 Experiments

In this section, we evaluate the proposed self-supervised reconstruction framework on both simulated and real lensless data, and compare it against classical convex optimization baselines. Throughout, we use the optimized radial coded mask introduced in Chap. 3 and implemented as a printed mask in Chap. 4. For the hardware experiments, we reuse the prototype camera from Sec. 4.2, operated in its restricted-aperture configuration.

Our goals are: (i) to assess how close self-supervised neural reconstructions can come to well-tuned convex optimization in terms of PSNR and perceptual quality; and (ii) to quantify the corresponding reduction in inference time.

5.3.1 Simulation experiments

Experimental setup and implementation details. We first perform controlled simulations using the FFHQ [73] dataset as a representative natural-image prior. We select 20,000 images for training and 1,000 images for testing. Each image is resized to match the PSF support and is transformed into a synthetic lensless measurement via the convolutional forward model in Eq. (2.4), using the optimized radial mask from [70] and additive Gaussian noise with standard deviation $\sigma = 0.005$.

As convex baselines, we consider: UDN [56], Wiener filtering [48], and TV-regularized ADMM [51] with both 5 and 100 iterations. Hyperparameters for these methods (e.g., regularization weights, step sizes) are selected by random search on a validation subset.

For the self-supervised models, we instantiate three architectures: (i) a UNet similar to [17]; (ii) Wiener-U, i.e., Wiener deconvolution followed by a denoising UNet; and (iii) ADMM-U,

a) Simulation



b) Prototype camera



Figure 5.2: Reconstructions for simulated and prototype camera experiments comparing classical methods and self-supervised models on the FFHQ lensless dataset. *Reproduced with permission of Springer Nature [2].*

Table 5.1: Quantitative analysis for simulated results on 20 reconstructions from the test set. For UDN method we show the metrics for the best reconstruction and the last (in parentheses) reconstruction over 20,000 iterations. Bold and underlined indicate best and second best methods, respectively. *Reproduced with permission of Springer Nature [2].*

Convex Opt.	PSNR (dB)($\uparrow$)	LPIPS($\downarrow$)	Time (ms)	batch size($\downarrow$)
UDN (20k iter)	19.00(22.44)	0.475(0.348)	404×10^3	1
Wiener	10.03	0.723	0.179	64
ADMM (100 iter)	26.01	0.167	4×10^3	64
ADMM (5 iter)	24.81	0.246	3.703	64
Self-supervision (ours)				
UNet	20.61	0.483	<u>2.551</u>	64
Wiener + UNet	21.18	0.386	2.643	64
ADMM(5 iter) + UNet	<u>23.67</u>	<u>0.239</u>	6.355	64

i.e., ADMM with 5 iterations followed by a UNet denoiser. In Wiener-U and ADMM-U, the convex modules use exactly the same hyperparameters as in their convex-only counterparts. All networks are trained end-to-end with the self-supervised loss, using mean squared error in the measurement domain and the Adam optimizer [71] for 400 epochs.

Evaluation methodology. For quantitative comparison, we select 20 images from the 1,000-image test set and reconstruct them with classical and self-supervised methods. We report the average PSNR and LPIPS [72] as reconstruction quality metrics, and inference time (ms) in Tab. 5.1. For UDN, we show the metrics at the final iteration and, in parentheses, the best value attained within 20,000 iterations, to highlight its sensitivity to the stopping criterion.

Inference time is measured on an RTX 5090 GPU, using batch size 64 whenever possible. The average per-image runtimes are reported in Tab. 4.2. Qualitative examples of reconstructed FFHQ faces are shown in Fig. 5.2(a), where the first row contains convex baselines and the second row contains self-supervised models.

Results and analysis. The quantitative results in Tab. 5.1 reveal the trade-offs between reconstruction quality and computational cost:

- ADMM with 100 iterations achieves the highest PSNR and a strong LPIPS score, but requires on the order of 4×10^3 ms per image, making it impractical for real-time or high-throughput applications.
- UDN can reach better-quality reconstructions at its best iteration than at its final one, but its performance is highly sensitive to the number of iterations, with a total runtime exceeding 4×10^5 ms per image.
- The self-supervised UNet alone attains moderate PSNR and perceptual quality, but often produces overly smooth reconstructions, consistent with the known low-frequency bias of CNNs (see Fig. 5.2(a)).
- Hybrid self-supervised models (Wiener-U and ADMM-U) significantly improve both PSNR and LPIPS relative to their convex-only counterparts, while keeping inference times in the few-millisecond range. In particular, ADMM-U narrows the quality gap to ADMM-100 while being roughly three orders of magnitude faster.

Taken together, these results indicate that self-supervised hybrid models can approach the reconstruction quality of carefully tuned iterative solvers while providing orders-of-magnitude faster inference, making them attractive candidates for practical lensless systems.

5.3.2 Prototype camera experiments

Experimental setup and camera configuration. We next validate the self-supervised approach on real measurements acquired with the prototype lensless camera introduced in Chap. 4. The camera uses a CMOS image sensor (BFS-U3-28S5C-BD by Teledyne FLIR) with 1936×1464 pixels and $4.5 \mu\text{m}$ pitch, and a chromium-deposited radial coded mask on a $4 \text{ cm} \times 4 \text{ cm}$ glass substrate. In this chapter, we operate the system in a restricted-aperture configuration,

using only the central region of the mask; after cropping, the active mask area occupies approximately 17% of the effective sensor area.

We use the same acquisition setup used for the QR codes and the FFHQ experiments of Chap. 4, placing a display in front of the lensless camera, as shown in Fig. 4.4. FFHQ images are displayed on a monitor placed approximately 10 cm away from the coded mask, and 21,000 lensless measurements are captured. Raw sensor images are cropped from the native 1936×1464 resolution to 1000×1000 pixels and subsequently downsampled to 250×250 for reconstruction. Calibrated PSFs for this configuration are obtained following the procedure described in Sec. 4.2, and are used both in the convex baselines and in the self-supervised forward model.

For the prototype experiments, we reuse the same algorithmic configurations as in simulation: UDN, Wiener, and ADMM with 5 and 100 iterations as convex baselines, and UNet, Wiener-U, and ADMM-U as self-supervised models. Training is performed directly on the real lensless measurements, using the same self-supervised loss and optimization settings as in Sec. 5.3.1.

Evaluation methodology. Because the prototype captures involve real sensor noise, illumination variations, and calibration imperfections, we mainly focus on qualitative comparisons. Representative reconstructions are shown in Fig. 5.2(b) for all methods. For reference, we also display the corresponding PSF and captured measurement.

Although it is possible to compute PSNR and LPIPS against the displayed FFHQ images after suitable brightness and contrast normalization, such metrics are sensitive to global gains and offsets that may differ between simulation and hardware. We therefore treat them as secondary and emphasize visual fidelity, noise characteristics, and structural sharpness in our analysis.

Results and analysis. The trends observed in simulation largely carry over to the prototype camera:

- Wiener and short-iteration ADMM reconstructions tend to be either noisy or overly smoothed, depending on the chosen regularization level and iteration count.

- UDN frequently produces blurry outputs for the prototype data within the practical iteration limit, illustrating its sensitivity to the number of iterations and to model mismatch between simulation and hardware.
- The self-supervised UNet alone yields reconstructions that are globally plausible but somewhat soft, with occasional color or brightness shifts relative to the display images.
- Hybrid self-supervised models (Wiener-U and especially ADMM-U) produce sharper, more natural-looking reconstructions that better suppress noise and artifacts, while remaining consistent with the measurement data through the embedded physics-informed modules.

An additional observation from Fig. 5.2(b) is the mismatch in absolute brightness and contrast between some reconstructions and the displayed reference images. This highlights a potential source of bias in fully supervised pipelines: when trained directly against lensed ground-truth targets (e.g., beam-splitter pairs or display images), networks may implicitly absorb calibration errors and exposure differences, leading to unpredictable brightness corrections at test time. In contrast, the self-supervised models considered here are anchored in the measurement domain via the forward model, and do not rely on cross-optics supervision.

Overall, the prototype results support the main conclusions drawn from simulation: self-supervised hybrid reconstructions can deliver high-quality images from real lensless measurements, with substantially lower inference time than iterative convex solvers and without requiring lensed ground-truth supervision.

5.4 Discussion and summary

Reconstruction quality versus classical methods. The experiments in this chapter show that self-supervised neural reconstructions can approach the reconstruction quality of well-tuned classical methods while avoiding their high per-image computational cost. In simulation, ADMM with a large iteration budget still provides the strongest baseline in terms of PSNR

and LPIPS, as expected from a carefully regularized convex formulation. However, the self-supervised combinations of convex optimization and a UNet denoiser (Wiener-U and ADMM-U) consistently improve upon their purely convex counterparts with only a modest increase in inference time. A similar trend is observed for the prototype camera: the convex-only Wiener and shallow-iteration ADMM reconstructions tend to be noisier or over-smoothed, whereas the self-supervised variants recover sharper details and more visually pleasing textures.

Speed–quality trade-offs. A key motivation for this chapter was to decouple reconstruction latency from the number of optimization iterations. The comparison between UDN, iterative ADMM, and the self-supervised models highlights this point. Methods such as UDN or ADMM with many iterations can yield high-quality reconstructions but require seconds to minutes per image when run to convergence. In contrast, once trained, the self-supervised models perform a single feed-forward pass (optionally preceded by a small fixed number of convex steps), achieving inference times on the order of a few milliseconds for batched reconstructions. This suggests a practical regime in which lensless imaging benefits from the interpretability of classical physics-based methods during training, while leveraging neural networks for fast deployment-time inference.

Benefits of self-supervision for lensless imaging. Unlike supervised reconstructions, the proposed approach does not rely on paired lensed ground truths acquired via beam splitters or separate high-quality cameras. Instead, it uses only lensless measurements and a calibrated forward model to define the training loss. This has two advantages. First, it removes supervision bias: the network is not encouraged to match the brightness, contrast, or color balance of a separate reference imaging system, but is instead constrained to be consistent with the actual lensless measurements. Second, it simplifies data collection, since building a large-scale paired dataset is often more challenging than capturing unlabeled lensless measurements with a fixed prototype camera.

Limitations and scope. At the same time, the proposed self-supervised framework has limitations. It still assumes access to an accurate forward model (here, a shift-invariant convolution with a calibrated PSF), and the training distribution is tied to the scene statistics of the chosen

dataset (FFHQ faces in our case). Thus, generalization to very different scene classes or to strongly depth-varying settings is not guaranteed. In addition, while the self-supervised models reduce per-image inference time, they require an up-front training phase whose cost scales with dataset size and network capacity. Finally, our experiments in this chapter focus on restricted-aperture radial masks and planar scenes near a given calibration depth; extending the framework to fully exploit the depth-robust optics developed in previous chapters remains an open direction.

Summary of contributions in this chapter. This chapter has: (i) formulated lensless reconstruction as a self-supervised learning problem driven solely by measurement-domain consistency; (ii) proposed hybrid architectures that couple classical convex optimization (Wiener and ADMM) with a UNet denoiser in a self-supervised training loop; and (iii) demonstrated, through both simulations and prototype-camera experiments, that these models can achieve reconstruction quality comparable to iterative methods at a fraction of the inference time, without relying on lensed ground-truth supervision. These findings position self-supervised neural reconstruction as a possible building block for future lensless imaging systems, particularly when combined with the depth-robust optical designs developed in the earlier chapters of this thesis.

Chapter 6

Conclusion

Concluding thoughts. In summary, this thesis demonstrates that practical all-in-focus 2D lensless imaging can be achieved without explicit depth estimation or lensed supervision by jointly designing the optical system and the reconstruction model. Our contributions are threefold.

First, we established that radial mask structures provide a principled way to suppress depth cues on the lensless measurement. By introducing a radial parameterization and optimizing the mask pattern within this framework, we demonstrated both experimentally and in simulation that restricted-aperture radial masks achieve simultaneous extended depth of field and broadband MTF, while preserving compatibility with simple, fast, shift-invariant 2D reconstruction techniques.

Second, we extended the radial concept to full-aperture masks, revealing a core trade-off between increased depth robustness and increased shift-variance. We addressed this challenge through an artificially extended PSF approximation that recovers a usable 2D convolutional model from shift-variant measurements. This approximation enabled all-in-focus performance even for large-aperture masks that offer significantly improved light throughput.

Third, we showed that high-quality lensless reconstructions can be obtained without any lensed ground-truth images by leveraging self-supervised training based solely on measurement consistency. The proposed neural and hybrid architectures achieved reconstruction quality comparable to or better than traditional physics-based methods while operating at millisecond-scale

inference times and avoiding the potential biases inherent to supervised training.

Taken together, these results show that rather than treating depth dependence, shift-variance, and supervision as individual limitations, they can be mitigated or removed through carefully coordinated and complementary choices in mask shape/geometry, forward modeling, and reconstruction strategy. The resulting system moves lensless imaging closer to real-world viability, supporting robust and efficient 2D reconstruction across large depth ranges, scenes, and capture conditions.

Future directions. The contributions of this thesis also highlight several limitations and open questions to be addressed in future works.

Regarding the radial-mask design, our systems deliberately trade *depth information* for *extended depth of field*. By making the PSF more invariant to depth, we enable in-focus 2D reconstructions over a wide range of distances, but at the same time reduce the sensitivity needed for explicit 3D recovery. The proposed masks are therefore well-suited for all-in-focus 2D imaging, but not compatible with depth-estimation tasks. An important open question is whether it is possible to construct hybrid mask designs or multi-stage systems that retain the depth-robust behavior of radial patterns for 2D reconstruction, while still encoding sufficient depth-dependent structure for 3D recovery. This could involve combining radial structure with additional depth-encoding features or multi-mask configurations that switch between "EDoF mode" and "3D mode" depending on the task.

A second limitation of the radial mask lies in its restricted design space. As shown in Sec. 3.3.1, an unoptimized random coded mask can achieve a broader MTF response than our optimized radial pattern, which translates into higher in-focus reconstruction quality. This reveals an inherent trade-off between peak in-focus performance and extended depth of field. In this work, our optimization objective was tailored to obtain a broadband radial response under the radial-structure constraint. A promising direction for future research is to move beyond purely broadband MTF criteria and instead design objectives that emphasize natural image statistics, so that the radial pattern is optimized for the frequencies and structures that matter most for reconstructing real-world scenes.

A third limitation stems from the use of large aperture radial masks. While enabling larger DOF than restricted aperture masks, it was shown to produce lower reconstruction quality for objects within its DOF, as shown in Sec. 4.5.1. This is further highlighted by the quantization experiments in Sec. A.1, where we see that lost information in larger apertures due to cropping effect degrades reconstruction quality more severely than bit quantization. Additionally, the extended-PSF forward model proposed in Sec. 4.5.1, while physically motivated by the scale-invariant geometry of the radial mask and supported by current experiments, remains an approximation that may break down for very wide angles, or stronger off-axis aberrations. Developing more accurate yet tractable models for shift-variant lensless systems, and characterizing when extended-PSF approximations remain valid, are promising directions for future work, especially for applications requiring large fields of view or strict quantitative accuracy.

A fourth limitation stems from the use of relatively simple priors (i.e., UNet-based denoisers combined with Wiener or few-step ADMM) for our self-supervised approach in Ch. 5. This design, while competitive against traditional reconstruction methods (e.g., Wiener filter, ADMM), fall short of the photorealistic quality enabled by the current supervised training approaches in the literature. An exciting line of research is to integrate more powerful generative priors into the unsupervised/self-supervised lensless reconstruction pipeline. In particular, recent diffusion-based priors for inverse problems have shown state-of-the-art results in deblurring, inpainting, super-resolution, and compression artifact removal by guiding pretrained diffusion models with data-fidelity terms [46, 61, 63, 64]. However, these methods have rarely been evaluated in other challenging optical settings such as lensless imaging, where the forward operator can be highly ill-conditioned, depth-dependent, and shift-variant [66]. Systematically investigating diffusion-guided reconstruction for lensless cameras, understanding their failure modes under model mismatch and calibration errors, and adapting their guidance strategies to coded-mask physics are important open research directions.

Finally, this thesis has focused on controlled laboratory setups and scenes. Extending radial-mask designs, extended-PSF modeling, and self-supervised reconstruction to outdoor and high-dynamic-range scenes, and resource-constrained hardware platforms (e.g., embedded or mo-

ble devices) would move lensless cameras closer to real-world deployment. Addressing these challenges will likely require co-design across optics, reconstruction algorithms, and learning frameworks, continuing the theme that runs throughout this work: practical lensless imaging emerges not from specific choices of optics or computation alone, but from a carefully designed balance between the two.

Acknowledgements

I am deeply grateful to my supervisors, Professors Tomoya Nakamura and Yasushi Yagi, who have supported me from the very beginning of my journey at the University of Osaka. Their constant guidance, insightful feedback, and encouragement have been invaluable in helping me grow academically and as a researcher. I also wish to express my sincere thanks to Professor Yuta Nakashima for his support during the final stages of my doctoral studies.

I would like to express my appreciation to Professors Hajime Nagahara and Yuki Uranishi for kindly agreeing to serve on my thesis hearing committee and for taking the time to review and evaluate this work.

I am also grateful to everyone who has accompanied me throughout this journey, from my labmates to my personal friends, whose presence made stressful moments more bearable and good moments truly memorable.

Finally, I show my appreciation to my parents Joana Angélica and Marcos Cláudio and my sister Marina for their unending support of my ventures, and without who I would not be half the person I am today.

Copyrights note. In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of The University of Osaka's products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications/_standards/publications/rights/rights_link.html to learn how to obtain

a License from RightsLink. If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

Additionally, all figures shown on Chap. 5, when cited, are fully or partially reproduced with permission from Springer Nature.

Reference

- [1] Jose Reinaldo Cunha Santos A. V. Silva Neto, Hodaka Kawachi, Yasushi Yagi, and Tomoya Nakamura. Toward all-in-focus lensless imaging with full-aperture radial masks. *Opt. Express*, Vol. 33, No. 23, pp. 48112–48129, Nov 2025.
- [2] José Reinaldo Cunha Santos A. V. Silva Neto, Hodaka Kawachi, Yasushi Yagi, and Tomoya Nakamura. Self-supervised neural reconstructions for lensless imaging. *Opt. Review*, Vol. 32, p. 664–668, 2025.
- [3] Vivek Boominathan, Jacob T. Robinson, Laura Waller, and Ashok Veeraraghavan. Recent advances in lensless imaging. *Optica*, Vol. 9, No. 1, pp. 1–16, Jan 2022.
- [4] Shuowen Li, Yunhui Gao, Jiachen Wu, Mingjie Wang, Zhangcheng Huang, Shumei Chen, and Liangcai Cao. Lensless camera: Unraveling the breakthroughs and prospects. *Fundamental Research*, 2024.
- [5] M. Salman Asif, Ali Ayremlou, Aswin Sankaranarayanan, Ashok Veeraraghavan, and Richard G. Baraniuk. Flatcam: Thin, lensless cameras using coded aperture and computation. *IEEE Trans. Computational Imaging*, Vol. 3, No. 3, pp. 384–397, 2017.
- [6] Kristina Monakhova, Kyrollos Yanny, Neerja Aggarwal, and Laura Waller. Spectral diffusercam: lensless snapshot hyperspectral imaging with a spectral filter array. *Optica*, Vol. 7, No. 10, pp. 1298–1307, Oct 2020.

- [7] E. Caroli, J. B. Stephen, G. Di Cocco, L. Natalucci, and A. Spizzichino. Coded aperture imaging in x- and gamma-ray astronomy. *Space Sci. Rev.*, Vol. 45, pp. 349–403, 1987.
- [8] Scott D. Barthelmy, Louis M. Barbier, Jay R. Cummings, Ed E. Fenimore, Neil Gehrels, Derek Hullinger, Hans A. Krimm, Craig B. Markwardt, David M. Palmer, Ann Parsons, Goro Sato, Masaya Suzuki, Tadayuki Takahashi, Makoto Tashiro, and Jack Tueller. The burst alert telescope (bat) on the swift midex mission. *Space Sci. Rev.*, Vol. 120, pp. 143–164, 2005.
- [9] P. Ubertini, F. Lebrun, G. Di Cocco, A. Bazzano, A. J. Bird, K. Broenstad, A. Goldwurm, G. La Rosa, C. Labanti, P. Laurent, I. F. Mirabel, E. M. Quadrini, B. Ramsey, V. Reglero, L. Sabau, B. Sacco, R. Staubert, L. Vigroux, M. C. Weisskopf, and A. A. Zdziarski. Ibis: The imager on-board integral. *Astronomy and Astrophysics*, Vol. 411, pp. 131–139, 2003.
- [10] P. R. Gill and G. Stork. Lensless ultra-miniature imagers using odd-symmetry spiral phase gratings. *Imaging and Applied Optics*, p. CW4C.3, 2013.
- [11] Yi Hua, Yongyi Zhao, and Aswin C. Sankaranarayanan. Angle sensitive pixels for lensless imaging on spherical sensors. *arXiv preprint arXiv:2306.15953*, 2023.
- [12] Jaewook Shin, Dung N. Tran, Jasper R. Stroud, Sang Chin, Trac D. Tran, and Mark A. Foster. A minimally invasive lens-free computational microendoscope. *Science Advances*, Vol. 5, No. 12, p. eaaw5595, 2019.
- [13] Jesse K. Adams, Dong Yan, Jimin Wu, Vivek Boominathan, Sibao Gao, Alex V. Rodriguez, Soonyoung Kim, Jennifer Carns, Rebecca Richards-Kortum, Caleb Kemere, Ashok Veeraraghavan, and Jacob T. Robinson. In vivo lensless microscopy via a phase mask generating diffraction patterns with high-contrast contours. *Nature Biomedical Engineering*, Vol. 6, , 2022.
- [14] Hiroyuki Kawano. Lensless image scanner using multilayered aperture array for non-contact imaging. *Optical Engineering*, Vol. 55, No. 10, pp. 103–106, 2016.

- [15] Fangyu Liu, Jiachen Wu, and Liangcai Cao. Autofocusing of fresnel zone aperture lensless imaging for qr code recognition. *Opt. Express*, Vol. 31, No. 10, pp. 15889–15903, May 2023.
- [16] Ilya Reshetouski, Hideki Oyaizu, Kenichiro Nakamura, Ryuta Satoh, Suguru Ushiki, Ryuichi Tadano, Atsushi Ito, and Jun Murayama. Lensless imaging with focusing sparse ura masks in long-wave infrared and its application for human detection. In *Eur. Conf. Comput. Vis.*, pp. 237–253, 2020.
- [17] Kristina Monakhova, Joshua Yurtsever, Grace Kuo, Nick Antipa, Kyrollos Yanny, and Laura Waller. Learned reconstructions for practical mask-based lensless imaging. *Opt. Express*, Vol. 27, No. 20, pp. 28075–28090, Sep 2019.
- [18] Salman Siddique Khan, Varun Sundar, Vivek Boominathan, Ashok Veeraraghavan, and Kaushik Mitra. Flatnet: Towards photorealistic scene reconstruction from lensless measurements. *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 44, No. 4, pp. 1934–1948, 2022.
- [19] Nick Antipa, Grace Kuo, Heckel Reinhard, Ben Mildenhall, Emrah Bostan, Ren Ng, and Laura Waller. DiffuserCam: lensless single-exposure 3D imaging. *Optica*, Vol. 5, No. 1, pp. 1–9, 2018.
- [20] J. Tan, V. Boominathan, A. Veeraraghavan, and R. Baraniuk. Flat focus: depth of field analysis for the flatcam lensless imaging system. In *ICASSP*, pp. 6473–6477, 2017.
- [21] Nakkyu Baek, Donggeon Bae, Kyung Chul Lee, Namhoon Kim, Taeyoung Kim, Muhyeon Kang, and Seung Ah Lee. Lensless camera for single-shot depth estimation driven by synthetic dataset. In *Conference on Lasers and Electro-Optics Pacific Rim*, p. P2_135, 2024.
- [22] L. Jin, Y. Coole, J. B. Coole, M. T. Tan, X. Zhao, H. Badaoui, J. T. Robinson, M. D. Williams, A. M. G. Vigneswaran, R. R. Richards-kortum, and A. Veeraragha-

- van. Deepdof-se: affordable deep-learning microscopy platform for slide-free histology. *Nature Communications*, p. 2935, 2024.
- [23] D. Bagadthey, S. Prabhu, S. S. Khan, D. T. Frederick, V. Boominathan, Veeraraghavan A., and K. Mitra. Flatnet3d: intensity and absolute depth from single-shot lensless capture. *J. Opt. Soc. Am. A*, Vol. 39, No. 10, pp. 1903–1912, Oct 2022.
- [24] Michael J. DeWeert and Brian P. Farm. Lensless coded-aperture imaging with separable Doubly-Toeplitz masks. *Optical Engineering*, Vol. 54, No. 2, p. 023102, 2015.
- [25] Vivek Boominathan, Jesse K. Adams, Jacob T. Robinson, and Ashok Veeraraghavan. Phlatcam: Designed phase-mask based thin lensless camera. *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 42, No. 7, pp. 1618–1629, 2020.
- [26] Grace Kuo, Fanglin Linda Liu, Irene Grossrubatscher, Ren Ng, and Laura Waller. On-chip fluorescence microscopy with a random microlens diffuser. *Opt. Express*, Vol. 28, No. 6, pp. 8384–8399, Mar 2020.
- [27] Xin Cai, Zhiyuan You, Hailong Zhang, Jinwei Gu, Wentao Liu, and Tianfan Xue. Phocolens: Photorealistic and consistent reconstruction in lensless imaging. *Advances in Neural Information Processing Systems*, Vol. 37, pp. 12219–12242, 2024.
- [28] Ganghun Kim, Kyle Isaacson, Rachael Palmer, and Rajesh Menon. Lensless photography with only an image sensor. *Appl. Opt.*, Vol. 56, No. 23, pp. 6450–6456, Aug 2017.
- [29] Feng-Ping Li, Vemund Stenbekk Thorkildsen, Leiv-J Gelius, and Jian-Hua Yue. Space-varying iterative restoration of 2-d inversion models computed from marine csem data. *arXiv preprint arXiv:2210.09685*, 2022.
- [30] Yu Ren, Xiaoling Zhang, Xu Zhan, Xiangdong Ma, Yunqi Wang, Edmund Y. Lam, and Tianjiao Zeng. Large-field-of-view lensless imaging with miniaturized sensors. *arXiv preprint arXiv:2512.00488*, 2025.

- [31] E. E. Fenimore and T. M. Cannon. Coded aperture imaging with uniformly redundant arrays. *Appl. Opt.*, Vol. 17, No. 3, pp. 337–347, Feb 1978.
- [32] Zhencong Xiong, Wenjun He, Yijie Chen, Yifei Xu, Wenbo Wang, and Yuegang Fu. Lensless imaging through the fourier-admm network with a cost-effective and easily fabricated phase mask. *Opt. Lett.*, Vol. 50, No. 3, pp. 758–761, Feb 2025.
- [33] Oliver Kingshott, Nick Antipa, Emrah Bostan, and Kaan Akşit. Unrolled primal-dual networks for lensless cameras. *Opt. Express*, Vol. 30, No. 26, pp. 46324–46335, Dec 2022.
- [34] Y. Hua, S. Nakamura, M. S. Asif, and A. C. Sankaranarayanan. Sweepcam - depth-aware lensless imaging using programmable masks. *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 42, No. 7, pp. 1606–1617, 2020.
- [35] Yucheng Zheng and M. Salman Asif. Coded illumination for 3d lensless imaging. *IEEE Open Journal of Signal Processing*, Vol. 3, pp. 432–439, 2022.
- [36] Ariel Schwarz, Jingang Wang, Amir Shemer, Zeev Zalevsky, and Bahram Javidi. Lensless three-dimensional integral imaging using variable and time multiplexed pinhole array. *Opt. Letters*, Vol. 40, No. 8, pp. 1814–1817, Apr 2015.
- [37] M. Salman Asif. Toward depth estimation using mask-based lensless cameras. In *Asilomar Conference on Signals, Systems, and Computers*, pp. 1467–1470, 2017.
- [38] Yucheng Zheng, Yi Hua, Aswin Sankaranarayanan, and M. Salman Asif. A simple framework for 3d lensless imaging for programmable masks. In *Int. Conf. Comput. Vis.*, 2021.
- [39] Zhengdai Li, Ying Li, Wenyu Han, Kele Chen, Youming Guo, and Changhui Rao. End-to-end wiener-cnn optimization for lensless extended depth of field imaging with phase mask. *Opt. Communications*, Vol. 546, p. 129828, 2023.

- [40] Ying Li, Zhengdai Li, Kaiyu Chen, Youming Guo, and Changhui Rao. 3d reconstruction from focus for lensless imaging. *Appl. Opt.*, Vol. 63, No. 31, pp. 8212–8220, Nov 2024.
- [41] P. R. Gill. Odd-symmetry phase gratings produce optical nulls uniquely insensitive to wavelength and depth. *Opt. Letters*, pp. 2074–2076, 2013.
- [42] Shay Elmalem, Raja Giryes, and Emanuel Marom. Learned phase coded aperture for the benefit of depth of field extension. *Opt. Express*, Vol. 26, No. 12, pp. 15316–15331, Jun 2018.
- [43] Shiyu Tan, Yicheng Wu, Shou-I Yu, and Ashok Veeraraghavan. Codedstereo: Learned phase masks for large depth-of-field stereo. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 7170–7179, June 2021.
- [44] Ugur Akpinar, Erdem Sahin, Monjurul Meem, Rajesh Menon, and Atanas Gotchev. Learning wavefront coding for extended depth of field imaging. *Trans. Img. Proc.*, Vol. 30, p. 3307–3320, January 2021.
- [45] Johannes Schwab, Stephan Antholzer, and Markus Haltmeier. Deep null space learning for inverse problems: convergence analysis and rates. *Inverse Problems*, Vol. 35, No. 2, p. 025008, jan 2019.
- [46] Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. *Int. Conf. Learn. Represent.*, 2023.
- [47] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, Vol. 60, No. 1, pp. 259–268, 1992.
- [48] Ying Li, Zhengdai Li, Kaiyu Chen, Youming Guo, and Changhui Rao. Mwdns: reconstruction in multi-scale feature spaces for lensless imaging. *Opt. Express*, Vol. 31, No. 23, pp. 39088–39101, Nov 2023.

- [49] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm with application to wavelet-based image deblurring. In *ICASSP*, pp. 693–696, 2009.
- [50] José M. Bioucas-Dias and Mario A. T. Figueiredo. A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Trans. on Image Processing*, Vol. 16, No. 12, pp. 2992–3004, 2007.
- [51] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, Vol. 3, No. 1, p. 1–122, jan 2011.
- [52] Ashish Verma, Vivek Boominathan, Ashok Veeraraghavan, and Chandra Sekhar Seelamantula. Diffusion model based image reconstruction in lensless imaging. In *ICASSP*, pp. 1–5, 2025.
- [53] Tianjiao Zeng and Edmund Y. Lam. Robust reconstruction with deep learning to handle model mismatch in lensless imaging. *IEEE Trans. Computational Imaging*, Vol. 7, pp. 1080–1092, 2021.
- [54] Eric Bezzam, Yohann Perron, and Martin Vetterli. Towards robust and generalizable lensless imaging with modular learned reconstruction. *IEEE Trans. Computational Imaging*, Vol. 11, pp. 213–227, 2025.
- [55] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 9446–9454, June 2018.
- [56] Kristina Monakhova, Vi Tran, Grace Kuo, and Laura Waller. Untrained networks for compressive lensless photography. *Opt. Express*, Vol. 29, No. 13, pp. 20913–20929, Jun 2021.
- [57] Matheus Souza, Yidan Zheng, Kaizhang Kang, Yogeshwar Nath Mishra, Qiang Fu, and Wolfgang Heidrich. Latent space imaging. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2025.

- [58] Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *Int. Conf. Learn. Represent.*, 2023.
- [59] Hyungjin Chung, Jeongsol Kim, Sehui Kim, and Jong Chul Ye. Parallel diffusion models of operator and image for blind inverse problems. *IEEE Conf. Comput. Vis. Pattern Recog.*, 2023.
- [60] Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *Int. Conf. Learn. Represent.*, pp. 1–30, 2023.
- [61] Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *Int. Conf. Learn. Represent.*, 2023.
- [62] Litu Rout, Negin Raoof, Giannis Daras, Constantine Caramanis, Alex Dimakis, and Sanjay Shakkottai. Solving linear inverse problems provably via posterior sampling with latent diffusion models. In *Adv. Neural Inform. Process. Syst.*, 2023.
- [63] Bingliang Zhang, Wenda Chu, Julius Berner, Chenlin Meng, Anima Anandkumar, and Yang Song. Improving diffusion inverse problem solving with decoupled noise annealing. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 20895–20905, 2025.
- [64] Jeongsol Kim, Bryan Sangwoo Kim, and Jong Chul Ye. Flowdps : Flow-driven posterior sampling for inverse problems. In *Int. Conf. Comput. Vis.*, pp. 12328–12337, 2025.
- [65] Nimrod Berman, Ilan Naiman, Moshe Eliasof, Hedi Zisling, and Omri Azencot. One-step offline distillation of diffusion-based models via koopman modeling. In *Adv. Neural Inform. Process. Syst.*, 2025.
- [66] José Reinaldo Cunha Santos A. V. Silva Neto, Hodaka Kawachi, Yasushi Yagi, and Tomoya Nakamura. Null-space diffusion distillation for efficient photorealistic lensless imaging. *arXiv preprint arXiv:2511.12024*, 2025.

- [67] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. In *Int. Conf. Mach. learn.*, 2018.
- [68] Joshua Batson and Loic Royer. Noise2self: Blind denoising by self-supervision. In *Int. Conf. Mach. learn.*, 2019.
- [69] Jason Lequyer, Reuben Philip, Amit Sharma, Wen-Hsin Hsu, and Laurence Pelletier. A fast blind zero-shot denoiser. *Nature Machine Intelligence*, 2022.
- [70] José Reinaldo Cunha Santos A. V. Silva Neto, Tomoya Nakamura, Yasushi Makihara, and Yasushi Yagi. Extended depth-of-field lensless imaging using an optimized radial mask. *IEEE Trans. Computational Imaging*, Vol. 9, pp. 857–868, 2023.
- [71] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Int. Conf. Learn. Represent.*, pp. 1–15, 2015.
- [72] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 586–595, 2018.
- [73] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 43, No. 12, p. 4217–4228, dec 2021.

List of Publications

Journal Publications (related to this thesis)

1. José Reinaldo Cunha Santos A. V. Silva Neto, Tomoya Nakamura, Yasushi Makihara, and Yasushi Yagi, "Extended Depth-of-Field Lensless Imaging using an Optimized Radial Mask," *IEEE Trans. on Computational Imaging*, Vol. 9, pp. 857–868, Sep. 2023.
2. José Reinaldo Cunha Santos A. V. Silva Neto, Hodaka Kawachi, Yasushi Yagi, and Tomoya Nakamura, "Self-supervised neural reconstructions for lensless imaging," *Opt. Review*, Vol. 32, p. 664–668, 2025.
3. José Reinaldo Cunha Santos A. V. Silva Neto, Hodaka Kawachi, Yasushi Yagi, and Tomoya Nakamura, "Towards all-in-focus lensless imaging with full-aperture radial masks," *Opt. Express*, Vol. 33, No. 23, pp. 48112–48129, Nov. 2025.

Awards

1. Student presentation award at the 13th Japan-Korea Workshop on Digital Holography and Information Photonics (DHIP), 2023.

Appendix A

Additional experiments

A.1 Quantization analysis via simulation

Motivation. In this section, we analyze how analog–digital quantization interacts with coded-mask aperture size and how this interaction affects reconstruction quality. We focus on two questions: (1) Does bit quantization degrade reconstruction quality more severely for larger-aperture masks, and if so, by how much? (2) For large-aperture masks, is the impact of quantization more critical than the information loss caused by PSF cropping due to finite sensor size? This investigation provides additional insight onto the tradeoff between restricted- and full-aperture radial masks with respect to reconstruction quality, now considering quantization along with the increase in aperture size.

Methodology. We study the effect of quantization and aperture size using the following simulation pipeline: (1) simulate lensless measurements at high numerical precision (32-bit floating point) using the forward model; (2) apply an B -bit quantization model,

$$\mathbf{y}_{\text{quant}} = \text{round}(\mathbf{y} \times (2^B - 1)) / (2^B - 1), \quad (\text{A.1})$$

where B denotes the number of bits and $\mathbf{y} \in [0, 1]$ is the pre-quantization measurement; (3) reconstruct the scene from the quantized measurement $\mathbf{y}_{\text{quant}}$ using a fixed reconstruction algorithm.

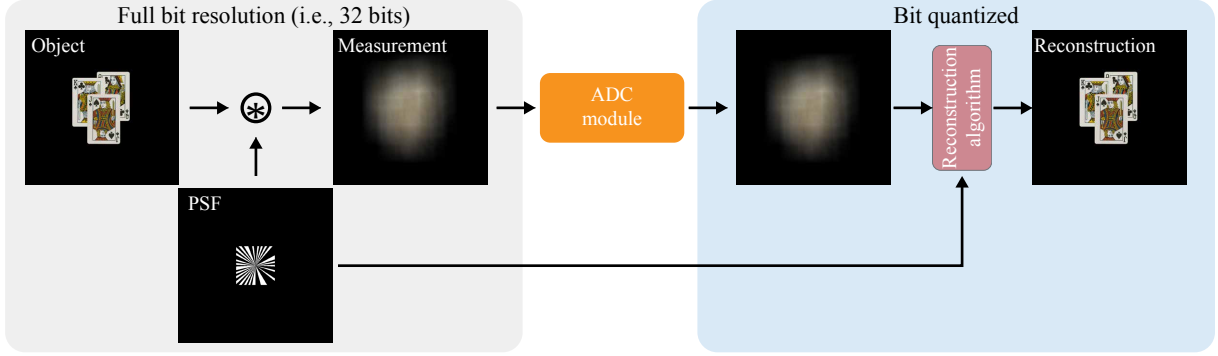


Figure A.1: Simulation pipeline for the quantization experiments. A high-precision (32-bit) lensless measurement is first generated by the forward model, then quantized by an B -bit ADC model, and finally used as input for reconstruction.

This procedure, illustrated in Fig. A.1, is repeated for multiple aperture sizes and bit depths. By evaluating reconstruction quality with metrics such as MSE and LPIPS, we can systematically assess how quantization and aperture size jointly influence the performance of lensless reconstruction.

Experiment. The simulation setup follows the planar-scene experiment in Sec. 3.3.1, using the same object (three cards arranged on a planar surface) and our optimized radial mask. The object plane is discretized on a 1024×1024 grid, with the target scene occupying the central 512×512 region. The sensor plane is likewise modeled as a 1024×1024 grid.

To study the interaction between aperture size and quantization, we vary the effective coded-mask aperture as a fraction of the sensor area. Specifically, we evaluate five aperture sizes corresponding to 1.6%, 6.3%, 25%, 56%, and 100% of the sensor area. For each aperture configuration, we simulate high-precision (32-bit) lensless measurements and then apply bit-depth quantization with

$$B \in \{8, 10, 12, 14, 16\}.$$

For every combination of aperture size and bit depth, we reconstruct the scene using the same deconvolution method (i.e., ADMM algorithm) and hyperparameters as in Sec. 3.3.1, and evaluate reconstruction quality with MSE and LPIPS metrics. The results are summarized as

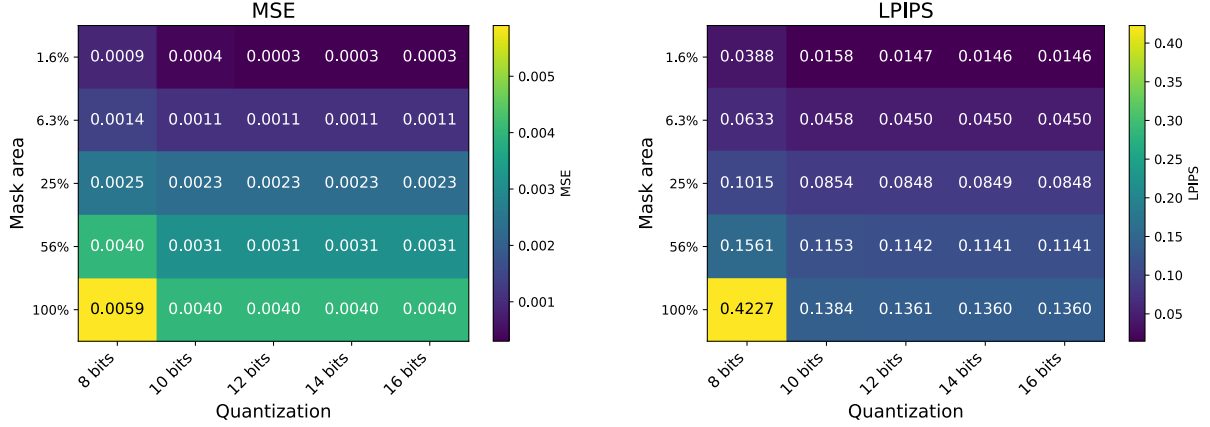


Figure A.2: Quantization analysis for the optimized radial mask. Left: MSE of reconstructed planar scenes as a function of aperture size (mask area with respect to sensor area) and ADC bit depth. Right: corresponding LPIPS values. In both metrics, reconstruction error increases primarily with aperture size, while bit-depth has a minor effect above 10–12 bits; a noticeable degradation appears only for the 8-bit setting, especially at large apertures.

heatmaps in Fig. A.2, enabling a direct comparison of how reconstruction quality varies jointly with aperture size and quantization level.

Results and discussion. The quantitative results reveal two trends. First, for any fixed bit depth, increasing the aperture size monotonically degrades reconstruction quality. At 16-bit quantization, the MSE increases from 3×10^{-4} to 4.0×10^{-3} when the aperture grows from 1.6% to 100% of the sensor area, while LPIPS rises from 0.0146 to 0.1360. Similar monotonic behavior is observed for all other bit depths. This indicates that, in our setting, the loss of information due to cropping associated with larger apertures has a stronger impact on reconstruction accuracy than quantization itself.

Second, the influence of quantization is relatively mild for bit depths of 10 bits and above. Across all aperture sizes, the MSE and LPIPS values for 12, 14, and 16 bits are nearly identical, and even 10-bit quantization introduces only mild degradation (e.g., for a 25% aperture, MSE remains constant while LPIPS slightly worsen from 0.0848 to 0.0854). A noticeable drop in quality appears only at 8 bits, particularly for the largest apertures: for a 100% aperture, LPIPS

increases from 0.1360 (16 bits) to 0.4227 (8 bits), indicating visible quantization-related artifacts. Nonetheless, even in this extreme case, the performance gap induced by changing the aperture from 1.6% to 100% is substantially larger than the gap between 16- and 8-bit quantization at a fixed aperture.

Overall, these results suggest that, for the optimized radial mask, quantization with 10–12 bits has a negligible effect on reconstruction quality compared to the dominant optical factors. Reconstruction degradations observed for large-aperture masks are primarily driven by cropping, rather than by ADC bit depth. In other words, within typical sensor bit depths, optical design choices (e.g., aperture size) remain the main determinant of reconstruction performance.