Thin On-Sensor Nanophotonic Array Cameras

PRANEETH CHAKRAVARTHULA* and JIPENG SUN*, Princeton University, USA
XIAO LI, CHENYANG LEI, GENE CHOU, and MARIO BIJELIC, Princeton University, USA
JOHANNES FROESCH and ARKA MAJUMDAR, University of Washington, USA
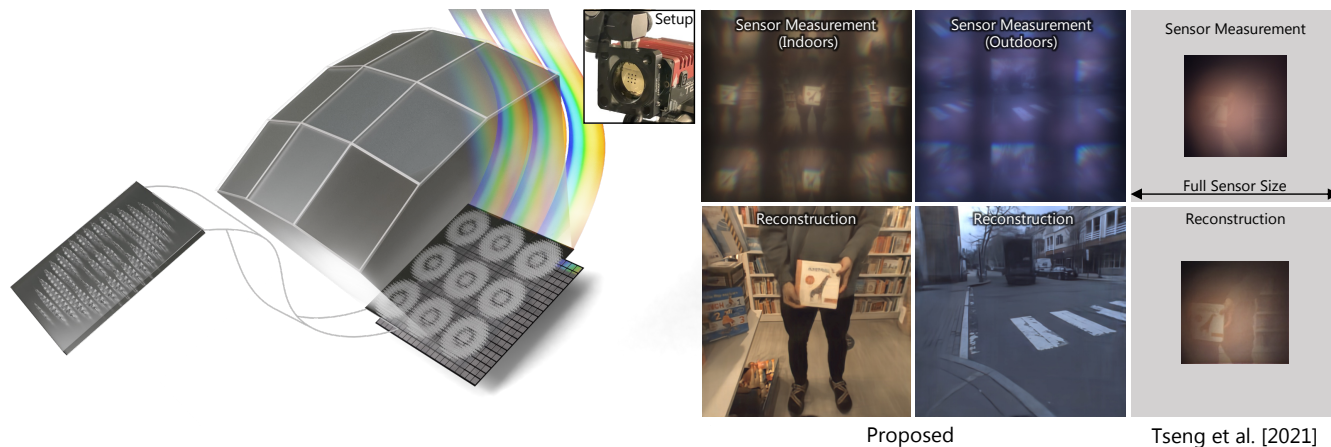FELIX HEIDE, Princeton University, USA

Fig. 1. We propose a thin nanophotonic imager that employs a learned array of metalenses to capture a scene in-the-wild. We devise an array of lenses mounted directly on top of the sensor cover glass (top center). Each lens in our array (illustrated on the left) is a flat (700 nm thick) metasurface area of nano-antennas which we design to focus light across the visible spectrum. The peripheral elements capture images at slanted field angles, making it possible to capture a wide field of view of 100°, more than twice as large as the most similar design from Tseng et al. [2021a] shown on the right side. The proposed computational camera is capable of recovering images outside of lab conditions under broadband illumination. We illustrate here the matching physical image size as measured on the sensor, with the gray area on right illustrating same sensor size.

Today's commodity camera systems rely on compound optics to map light originating from the scene to positions on the sensor where it gets recorded as an image. To record images without optical aberrations, i.e., deviations from Gauss' linear model of optics, typical lens systems introduce increasingly complex stacks of optical elements which are responsible for the height of existing commodity cameras. In this work, we investigate *flat nanophotonic computational cameras* as an alternative that employs an array of skewed lenslets and a learned reconstruction approach. The optical array is embedded on a metasurface that, at 700 nm height, is flat and sits on the sensor cover glass at 2.5 mm focal distance from the sensor. To tackle the highly chromatic response of a metasurface and design the array over the entire sensor, we propose a differentiable optimization method that continuously

samples over the visible spectrum and factorizes the optical modulation for different incident fields into individual lenses. We reconstruct a megapixel image from our flat imager with a *learned probabilistic reconstruction* method that employs a generative diffusion model to sample an implicit prior. To tackle *scene-dependent aberrations in broadband*, we propose a method for acquiring paired captured training data in varying illumination conditions. We assess the proposed flat camera design in simulation and with an experimental prototype, validating that the method is capable of recovering images from diverse scenes in broadband with a single nanophotonic layer.

Additional Key Words and Phrases: Computational Optics

*Indicates equal contribution.

Authors' addresses: Praneeth Chakravarthula, cpk@cs.unc.edu; Jipeng Sun, jipeng.sun@princeton.edu, Princeton University, USA; Xiao Li, xl2710@princeton.edu; Chenyang Lei, leichenyang7@gmail.com; Gene Chou, gchou@princeton.edu; Mario Bijelic, mario.bijelic@princeton.edu, Princeton University, USA; Johannes Froesch, jfroech@uw.edu; Arka Majumdar, arka@uw.edu, University of Washington, USA; Felix Heide, fheide@princeton.edu, Princeton University, USA.

1 INTRODUCTION

Cameras have become a ubiquitous interface between the real world and computers with applications across domains in fundamental science, robotics, health, and communication. Although their applications are diverse, *today's cameras acquire information in the same way they did in the 19th century*: they focus light on a sensing plane using a stack of lenses that minimize deviations from Gauss's linear model of optics [Gauss 1843]. In this paradigm, increasingly complex and growing sets of lenses are designed to record an image. Since the microfabrication revolution in the last century brought

us miniaturized sensors and electronic chips, it is now these optical systems that dictate a camera's size and weight and prohibit miniaturization without drastic loss of image quality [Asif et al. 2016; Peng et al. 2016a; Stork and Gill 2014]. For example, the optical stack of the iPhone 13 contains more than seven elements that make up the entire 8 mm of the camera length responsible for the camera bump. Unfortunately, attempts to use thinner single-element optics [Peng et al. 2016a; Stork and Gill 2014; Venkataraman et al. 2013], amplitude masks close to the sensor [Asif et al. 2016; Khan et al. 2020], or diffusers [Antipa et al. 2018; Kuo et al. 2017] instead of focusing optics have not been able to achieve the high image quality that conventional compound lens systems deliver.

The emerging field of nanophotonic metaoptics suggests an alternative. These optical devices rely on quasi-periodic arrays of sub-wavelength scatterers that are engineered to manipulate wavefronts. *In principle,* this approach promises new capabilities to drastically reduce the size and weight of these elements. The unprecedented ability to engineer each nanoscatterer enables optical functionality that is extremely difficult, if not impossible, to achieve using conventional optics: spectral and spatial filters [Camayd-Muñoz et al. 2020], polarization imagers [Arbabi et al. 2018], compact hyperspectral imagers [Faraji-Dana et al. 2019], depth sensors [Colburn and Majumdar 2020], and even image processors [Zhou et al. 2020]. Moreover, these flat optical elements are ultrathin, with a device thickness around an optical wavelength.

The imaging performance of existing meta-optics, however, is far from that of their counterpart refractive lenses [Colburn and Majumdar 2020; Tseng et al. 2021a]. While these lenses are corrected for wavelengths across the visible regime, the image quality is not on par with refractive lenses: third-order Seidel aberrations (e.g., coma, field curvature, and distortion) remain uncorrected as they are not even considered in the design procedure for these devices. Furthermore, the small apertures of the metalenses used ($\approx 50-100$ μm) severely limit the achievable angular resolution and total light collection (reducing the signal-to-noise ratio).

Increasing the field of view and aperture of these metalenses while simultaneously maintaining and improving aberration correction faces fundamental challenges: metasurfaces are inherently chromatic like any diffractive optics. For a metalens designed for a specific wavelength, the positions of the rings of constant phase decide the lens focusing behavior. When the incident wavelength changes, however, the imparted phase exhibits erroneous phase-wrapping discontinuities that vary significantly from the ideal response expected for the incident, non-design wavelengths [Arbabi et al. 2016]; this is the primary reason why metasurfaces exhibit chromatic aberrations. Recently, dispersion engineering has been investigated to design a metasurface to uniformly focus light across the full visible wavelength range [Chen et al. 2018; Wang et al. 2018a]. This technique relies on designing scatterers with not only a desired phase but also its higher-order response in the form of group delay and group delay dispersion. Recent work finds that there is a fundamental limit on the optical bandwidth for a dispersion-engineered metalens given a feasible aspect ratio and, therefore, small lens thickness – a limit that arises from the inherent time-bandwidth product [Presutti and Monticone 2020a]. The most successful approach to broadband imaging with metasurface optics from [Tseng et al. 2021a] relies on

end-to-end computational imaging and jointly designs lens parameters and computational reconstruction [Peng et al. 2019a; Sitzmann et al. 2018] with a differentiable forward model. Despite achieving increased image quality, the design limitations and computational and memory consumption of this approach is also fundamentally limited to designing *nanophotonic optics with a limited field of view of 40°, optimized for narrow wavelength bands, and low image resolutions of a few kilopixels* [Tseng et al. 2021a]. We aim to tackle this issue and design broadband computational nanophotonic cameras that can lift this limitation and make thin cameras possible, more than two orders of magnitude thinner and lighter than today.

In this work, we propose a flat camera that relies on an array of nanophotonic optics, which are learned for the broadband spectrum, and a computational reconstruction module that recovers a single megapixel image from the array measurements. The camera employs a *single flat optical layer sitting on top of the sensor cover glass* at 2.5 mm focal distance from the sensor. We introduce a differentiable forward model that approximates the highly chromatic wavefront response of a metasurface atom conditioned on the structure parameters in a local region. Instead of full-wave simulation methods that do not allow for simulating apertures larger than tens of microns across the visible band due to prohibitive memory and computational requirements, this differentiable model allows us to piggy-back on distributed machine learning methods and learn nanophotonic imaging across the entire band by stochastic gradient optimization over the continuous spectrum – in contrast to Tseng et al. [Tseng et al. 2021a] who optimize over the three fixed wavelengths of an OLED display. We achieve high-quality imaging performance across the entire visible band and more than double the field of view of existing approaches to 100° by separating the optical modulation for different optical fields into individual lenses in an array. We recover a latent image from our flat imager with a learned optimization method that relies on a diffusion model as a natural image prior. To tackle reconstruction in broadband illumination, we introduce a novel method to capture large datasets of paired ground-truth data in real-world illumination conditions.

Specifically, we make the following contributions:

- We introduce a flat on-sensor nanophotonic array lens that decomposes the joint optimization over field angle and broadband focusing into several subproblems with several smaller field of view. We propose a stochastic optimization method for designing the decomposed broadband array elements.
- We propose a novel learned probabilistic reconstruction method that relies on the physical forward model combined with a learned diffusion model as prior. To train the method, we propose an approach to capture paired real-world datasets.
- We analyze our method in simulation and compare the proposed method to alternative flat optical systems.
- We assess the method with a prototype camera system and compare it against existing metasurface designs. We confirm that the method achieves favorable image quality compared to existing metasurface optics across the entire spectrum and with a large field of view with a flat optical system on the sensor cover glass.

We will release all code, optical design files, and datasets.

Limitations. Compared to traditional cameras with larger optical systems, the proposed flat camera shares with existing computational flat cameras the need for GPU processing with high power consumption. Despite this limitation, the compute resources on modern smartphones present opportunities for the efficient implementation of the proposed reconstruction method on custom ASICs, potentially enabling fast inference on edge devices in the future. Our prototype does not use the full available optical aperture. To avoid optical baffles and overlap, we space out the sublenses in the array over non-contiguous regions, resulting in low total light efficiency. We also do not explicitly consider fabrication inaccuracies.

2 RELATED WORK

Flat Computational Cameras. Researchers have investigated several directions to reduce the height and complexity of existing compound camera optics. A line of work aims at reducing a complex optical stack of a handful to a dozen elements, to a single refractive element [Heide et al. 2013; Li et al. 2021; Schuler et al. 2013; Tanida et al. 2001] resulting in geometric and chromatic aberrations. Trading optical for computational complexity to address the introduced aberrations, these approaches have achieved impressive image quality comparable to a low-resolution point-and-shoot camera. Venkataraman et al. [2013] suppress chromatic aberrations by using an on-sensor array of color-filtered single lens elements, which turns the deconvolution problem into a chromatic light field reconstruction approach that is challenging to solve without artifacts. All proposed single-element refractive and diffractive cameras [Heide et al. 2016; Peng et al. 2015, 2016b] have in common that, although the optical stack itself decreases in height (less than a micron for diffractive elements), they require *long backfocal distances* of more than 10 mm prohibiting thin cameras. Lensless cameras[Antipa et al. 2018; Asif et al. 2016; Khan et al. 2020; Kuo et al. 2017; Liu et al. 2019; Monakhova et al. 2020; White et al. 2020] instead replace the entire optical stack with amplitude masks or diffusers that scramble the incoming wavefronts. Although this approach allows for thin cameras of a few millimeters in height, the information of a given scene point is distributed over the entire sensor. The light efficiency of these cameras is half of that of conventional lens systems, and recovering high-quality images from the coded measurements with large point spread functions of global support is challenging and, as such, the ill-posedness of the underlying reconstruction problem severely limits spatial resolution and requires long acquisition times. Using diffusers as caustic lenses has been investigated for 2D photography [Kuo et al. 2017], 3D imaging [Antipa et al. 2018] and microscopy [Kuo et al. 2020]. In addition to resulting in a challenging ill-posed reconstruction problem, the optimal distance from the diffuser to the sensor may vary from one diffuser to another [Boominathan et al. 2020]. In this work, we investigate an array of steered metasurface lenses as an alternative that allows for a short backfocal distance without mandating aberrations with global support or reducing light efficiency.

Metasurface Optics. Over the last few years, recent advancements in nanofabrication have made it possible for researchers to investigate optics by using quasi-periodic arrays of subwavelength scatterers to modify incident electromagnetic radiation. These ultra-thin

Table 1. Comparison of related work on thin cameras, where each criterion is fully ✓, partially (✓), or not ✗ met. See text for discussion.

	FlatCam [2016]	DiffuserCam [2017]	PiCam [2013]	Peng et al. [2016b]	Tseng et al. [2021a]	Ours
Camera Characteristics						
On-Sensor (< 2mm)	✓	(✓)	✓	✗	✓	✓
Light Efficient	✗	(✓)	(✓)	✓	✗	✓
Broadband	✓	✓	✓	(✓)	(✓)	✓
Wide Field of View	(✓)	(✓)	(✓)	✗	(✓)	✓
MTF	✗	✗	(✓)	✓	✗	✓
Fabrication	✓	(✓)	✗	✗	✓	✓
Well-posedness	✗	✗	(✓)	(✓)	(✓)	✓

metasurfaces allow the fabrication of freeform surfaces using single stage lithography. Specifically, meta-optics can be fabricated by piggy-backing on existing chip fabrication processes, such as deep ultraviolet lithography (DUV), without error-prone multiple etching steps required for conventional diffractive optical elements [Shi et al. 2022]. Each scatterer in a metasurface can be independently tailored to modify amplitude, phase, and polarization of wavefronts – light can be modulated with greater design freedom compared to conventional diffractive optical elements (DOEs) [Engelberg and Levy 2020; Lin et al. 2014; Mait et al. 2020; Peng et al. 2019b]. With these theoretical advantages in mind, researchers have investigated flat meta-optics for imaging [Aieta et al. 2012; Chen et al. 2022; Colburn et al. 2018; Lin et al. 2021; Yu and Capasso 2014], polarization control [Arbabi et al. 2015], and holography [Zheng et al. 2015]. However, existing meta-optics suffer from severe chromatic and geometric aberrations making *broadband imaging outside the lab infeasible with existing designs*. In contrast to diffractive optics, the wavelength-dependent aberrations are a direct result from non-linear imparted phase [Aieta et al. 2015; Lin et al. 2014; Wang et al. 2018b; Yu and Capasso 2014]. While methods using dispersion engineering [Arbabi et al. 2017; Khorasaninejad et al. 2017; Ndao et al. 2020; Shrestha et al. 2018; Wang et al. 2017] are successful in reducing chromatic aberrations, these methods are limited to aperture sizes of tens of microns [Presutti and Monticone 2020b]. Most recently, Tseng et al. [Tseng et al. 2021a] have proposed an end-to-end differentiable design approach for meta-optics that achieves full-color image quality with a large 0.5 mm aperture. However, while successful in imaging tri-chromatic bands of an OLED screen, their method does not perform well outside the lab and suffers from severe blur for fields beyond 40°. Recent advanced nano fabrication techniques have also made compact conventional cameras with wafer-level compound optics possible, e.g., OVT CameraCube [1] which, however, offers limited resolution and FoV. The proposed array design in this work optimizes image quality over the full broadband spectrum across the 100° deg FoV without increasing the backfocal length. Our method can potentially allow for one-step fabrication of the metalens directly on the camera sensor coverglass in the future, further shrinking existing wafer-level multi-element compound lens camera designs.

Differentiable Optics Design. Conventional imaging systems are typically designed in a sequential approach, where the lens and

---

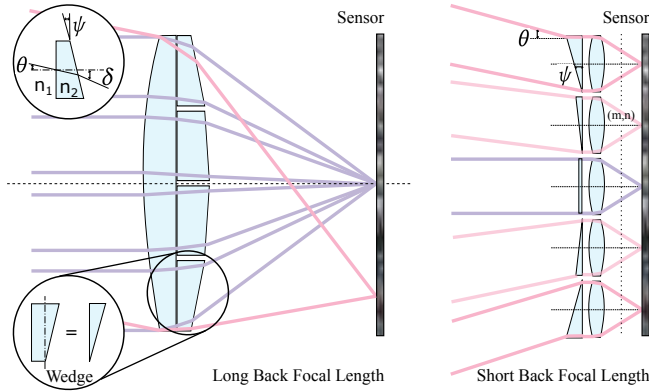[1]https://www.ovt.com/technologies/cameracubechip/

Fig. 2. Lens as a Combination of Prisms. A lens refracts an incoming set of parallel rays and bends them towards a focal point. With a piece-wise linear approximation, a lens can be thought of as a stack of infinitesimally small prisms. We depart from a single large lens with long back focal length, re-organize the prisms, and represent a large lens as a combination of appropriately chosen wedges and lenses with short focal length for high-quality camera imaging.
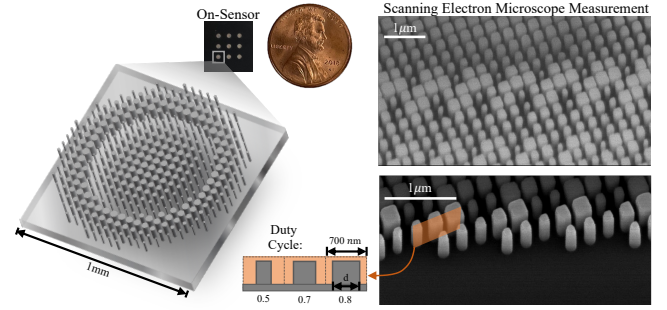

Fig. 3. Ultra-thin and Compact Metalens Array. Our on-sensor imaging optic, which is smaller than a penny, consists of an array of metalenses as shown above. Each metalens is made of optimized nano-antennas of size 350 nm, significantly smaller than the wavelength of visible light. Our optimized metalens nano-antenna structures scatter light from the entire visible spectrum.

sensors are hand-engineered concerning specific metrics such as RMS spot size or dynamic range, independently of the downstream camera task. Departing from this conventional design approach, a large body of work in computational imaging has explored jointly optimizing the optics and reconstruction algorithms, with successful applications in color image restoration [Chakrabarti 2016; Peng et al. 2019c], microscopy [Horstmeyer et al. 2017; Kellman et al. 2019; Nehme et al. 2020; Shechtman et al. 2016], monocular depth imaging [Chang and Wetzstein 2019; Haim et al. 2018; He et al. 2018; Wu et al. 2019], super-resolution and extended depth of field [Sitzmann et al. 2018; Sun et al. 2021], time-of-flight imaging [Chugunov et al. 2021; Marco et al. 2017; Su et al. 2018], high-dynamic range imaging [Metzler et al. 2020; Sun et al. 2020], active-stereo imaging [Baek and Heide 2021], hyperspectral imaging [Baek et al. 2021], and computer vision tasks [Tseng et al. 2021b]. In this work, we take a hybrid approach wherein we first optimize a nanophotonic lens array camera, designed with an inverse filter as an efficient proxy for the reconstruction method. We then devise a novel probabilistic deconvolution method conditioned on the measured signals for full-color image restoration, computationally compensating residual aberrations.

3 NANOPHOTONIC ARRAY CAMERA

In this section, we describe the nanophotonic array camera for thin on-sensor imaging. We design the imaging optic by learning an array of short back focal length metalenses with carefully designed phase profiles, enabling the camera to capture a scene with a large viewing angle. A learned image reconstruction method recovers the latent image from the nanophotonic array camera resulting in a thin on-chip imaging system. In the following, we first describe the nanophotonic array optic. In the remainder of this section, we then derive the differentiable image formation model for the metalens

array which we rely on to learn the phase profiles. In Sec. 4, we describe our reconstruction method.

3.1 Lens as a Combination of Prisms

A lens can be thought of as analogous to a series of *continuous* prisms as shown in Figure 2. A conventional prism refracts light causing its path to bend as

$$\delta = \theta - \psi + \sin^{-1}\left(\sqrt{n^2 - \sin^2\theta}\sin\psi - \sin\theta\cos\psi\right), \qquad (1)$$

where $n = n_2/n_1$ is the relative refractive index, ψ is the wedge angle of the prism, θ is the angle of incident light and δ is the angle of deviation, illustrated in Fig. 2. The greater the wedge angle, the greater the deviation of light path is. For smaller wedge and incident angles, the angle of deviation can be approximated as

$$\delta \approx (n-1)\psi. \qquad (2)$$

Therefore, narrow-angle prisms in an imaging setup merely result in a shift in image position. Specifically, light passing through the optical center of a lens where the surfaces are parallel to each other yields no prismatic effect in the center. However, light passing through the periphery of a lens experiences prismatic effects. The increasing angle between the opposing surfaces further from the lens center causes light to bend more and more, allowing the lens to focus light. Therefore, stacking a series of tiny prisms effectively makes a lens, notwithstanding the presence of aberrations. However, since different wavelengths are refracted differently, a large angle prism also causes different wavelengths to spread out in the image, causing chromatic aberrations due to the spectrum of colors from white light, coma, and astigmatism.

In this work, we design our optical layer as a combination of tiny co-optimized lens and prism phase elements, illustrated in Fig. 2, where the prisms help expand the field of view and the lenses reduce the back focal length. This allows us to devise a nanophotonic lens array to approximate the large lens, but trading off focal length with aberrations. Moreover, we also optimize the lens to correct some of the dispersion-based aberrations caused by the prism phases.

3.2 Radially Symmetric Nanophotonic Array

Nanophotonic meta-optics are ultrathin optical elements that utilize subwavelength nano-antenna scatterers to modulate incident light. Typically, these nano-antenna structures are designed for modulating the phase of incident light at a single nominal design wavelength, making meta-optics efficient for monochromatic light propagation. However, we require the meta-optic to achieve the desired phase modulation at all visible wavelengths to design a broadband imaging lens.

Metasurface Design Space. We design metasurfaces that consist of silicon nitride nanoposts with a height of 700 nm and a pitch of 350 nm on top of fused silica (n = 1.5), see Fig. 3. The rectangular pillars are made of a high refractive index material that is transparent. We keep these parameters fixed and then optimize the width (= length) of the nano-antennas between 100–300 nm, determining what we call the local "duty cycle", see again Fig. 3.

In a local neighborhood of these nano-antennas, we are able to simulate the phase for a given duty cycle using rigorous-coupled wave analysis (RCWA), which is a Fourier-domain method that solves Maxwell's equations efficiently for periodic dielectric structures. As such, in the following, we characterize metalenses with their local phase, which we tie to the structure parameters, i.e., the duty cycle, via a differentiable model.

Radially Symmetric Metalens Array. We model the metasurface phase ϕ, which we treat as a differentiable variable in our design, as a radially symmetric per-pixel basis function

$$\phi(x_i, y_j) = \phi(r), \quad r = \sqrt{x_i^2 + y_j^2}, \ i, j \in \{1, 2, ..., N\}, \quad (3)$$

where N is the total number of pixels along each axis, (x_i, y_j) denotes the nano-antenna position and r is its distance from the optical axis of the metalens. The per-pixel basis allows each nano-antenna along one radius of the metasurface to vary independently of the other nano-antennas without constraints. We constrain the metalens to be radially symmetric as opposed to optimizing the phase in a per-pixel manner to avoid local minima. Additionally, a spatially symmetric design imparts a spatially symmetric PSF which reduces the computational burden as it allows the simulation of the full field-of-view by only simulating PSFs along one axis.

We impose an additional wedge phase of varying wedge angles over each metalens element to achieve a wider field of view. Therefore, for an $M \times N$ nanophotonic array, the phase of each element is given by

$$\phi^{m,n}(x_i, y_j) = \phi^{m,n}(r) = \phi(r) + \frac{2\pi}{\lambda_w} \left(x_i \sin \psi_x^{m,n} + y_i \sin \psi_y^{m,n} \right), \quad (4)$$

where $\phi^{m,n}(x_i, y_j)$ is the phase modulation at the (x_i, y_j)-th nano-antenna of the metalens in m-th row and n-th column, λ_w is the wavelength for which the wedge phase was defined, and (ψ_x, ψ_y) are the selected wedge angles along each axis. Note that for given wedge angles of a metalens element in the array, the additional wedge phase is constant whereas the radially symmetric phase is optimizable.

Since the phase is defined only for a single nominal design wavelength, we apply two operations in sequence at each scatterer position in our metasurface: 1) a phase-to-structure inverse mapping to compute the scatterer geometry at the design wavelength for a given phase and 2) a structure-to-phase forward mapping to calculate the phase at other target wavelengths given a scatterer geometry. To allow for direct optimization of the metasurface phase, we model both the above operators as polynomials to ensure differentiability, which we describe below.

RCWA Proxy Mapping Operators. We describe the scatterer geometry with the duty cycle of nano-antennas and analyze its modulation properties using rigorous coupled-wave analysis (RCWA). The phase as a function of duty cycle of the nano-antennas must be injective to achieve a differentiable mapping from phase to duty cycle. To this end, we fit the phase data of the metalens at the nominal design wavelength to a polynomial proxy function of the form

$$d(r) = \sum_{i=0}^{N} a_i \left(\frac{\phi(r)}{2\pi} \right)^{2i}, \quad (5)$$

where $d(r)$ is the required duty cycle at a position r from the optical axis on the metasurface, $\phi(r)$ is the desired phase for the nominal wavelength λ_0, and the parameters a_i are fitted. We set the nominal wavelength $\lambda_0 = 452$ nm for all of our experiments.

After applying the above phase-to-scatterer inverse mapping to determine the required physical structure, we compute the resulting phase from the given scatterer geometry for other wavelengths using a second scatterer-to-phase proxy function. This forward mapping function maps a combination of the nano-antenna duty cycle and incident wavelength to an imparted phase delay. We model this proxy function by fitting the pre-computed transmission coefficient of scatterers under an effective index approximation [Tseng et al. 2021a] to a radially symmetric second-order polynomial function of the form

$$\widetilde{\phi}(r, \lambda) = \sum_{n=0}^{2} \sum_{m=0}^{2} b_{nm} d(r)^n \lambda^m, \quad n + m \leq 2, \quad (6)$$

where λ is a *non-nominal* wavelength. Specifically, we compute the transmission coefficient data C_{META} using RCWA and then fit the polynomial to the underlying RCWA-computed transmission coefficient data using linear least squares. For details on fitted polynomial coefficients for the inverse and forward mappings and additional details on proxy functions, please see Supplementary Material.

Single Lens Element Image Formation. With the metalens phase described by Eq. (4) and the mapping operators defined in Eq. (5) and Eq. (6), we compute the phase modulation for a broadband incident light. Using a fast Fourier transform (FFT) based band-limited angular spectrum method (ASM), we calculate the PSFs produced by each metalens in the array as a function of wavelength and field angle to model full-color image formation over the entire field of view. The spatially varying PSF as produced by each element in the nanophotonic array for an incident beam of wavelength λ at an angle θ is

$$\text{PSF}_{\theta, \lambda}^{m,n} = f_{\text{META}} \left(\phi^{m,n}(r), \theta, \lambda, C_{\text{META}} \right), \quad (7)$$

where $\phi^{m,n}(r)$ is the optimizable radially symmetric metasurface phase and C_{META} are the set of fixed parameters such as aperture and focal length of the metalens, and $f_{\text{META}}(\cdot)$ is the angular spectrum method as a propagation function that generates the PSF $\mathbf{k}$ for a given metasurface phase. Finally, the RGB image on the sensor plane is

$$\mathbf{S} = \mathbf{I} \otimes \mathbf{k} + \eta_{\text{SENSOR}}, \qquad (8)$$

where $\otimes$ is a convolution operator, $\mathbf{I}$ is the groundtruth RGB image, and η_{SENSOR} is the sensor noise modeled as per-pixel Gaussian-Poisson noise.

Specifically, for an input signal $x \in [0, 1]$ at a sensor pixel location, the measured noisy signal $f_{\text{SENSOR}}(x)$ is given by

$$f_{\text{SENSOR}}(x) = \eta_g(0, \sigma_g) + \eta_p(x, a_p), \qquad (9)$$

where $\eta_g(0, \sigma_g) \sim \mathcal{N}(0, \sigma_g^2)$ is the Gaussian noise component and $\eta_p(x, a_p) \sim \mathcal{P}(x/a_p)$ is the Poisson noise component.

Spatially Varying Array Image Formation. We simulate the spatially varying aberrations in a patch-wise manner. We first divide the overall FoV into an $M \times N$ grid of patches for a nanophotonic array with $M \times N$ metalens elements. For incident broadband light at field angle θ, we then compute $\text{PSF}_{\theta,\lambda}$ for each metalens element in the array with varying wedge angles, see Eq. 4. While we use $\text{PSF}_{\theta,\lambda}$ for the image formation forward model, we permute the PSFs for different wavelengths for deconvolution. This process acts as a regularization to the PSF design and avoids variance across the spectrum, essential for robust imaging in the wild. After design and fabrication, we account for mismatches between the simulated PSF by our proxy model and the experimentally measured PSF by performing a PSF calibration step.

Differentiable Nanophotonic Array Design. With a measurement $\mathbf{S}$ as input, we recover the latent image as

$$\tilde{\mathbf{I}} = f_{\text{DECONV}}(\mathbf{S}, \mathbf{k}, C_{\text{DECONV}}), \qquad (10)$$

where C_{DECONV} are the fixed parameters of the deconvolution method. To make our lens design process efficient, we employ an inverse filtering method in the design of our optic which does not require training and allows it to be computed in one step, as opposed to the proposed method in Sec. 4, see Supplemental Material for detail.

With this synthetic image formation model in hand, our nanophotonic array imaging pipeline allows us to apply first-order stochastic gradient optimization to optimize for the metalens phases that minimize the error between the ground truth and recovered images. In our case, given an input RGB image $\mathbf{I}$, we aim to find a metalens array that will recover $\mathbf{I}$ with high fidelity with short back focal length to achieve compact and ultra-thin imaging device with wide FoV. To design our optical system, we minimize the per-pixel mean squared error and maximize the perceptual image quality between the target image $\mathbf{I}$ and the recovered image $\tilde{\mathbf{I}}$. To this end, we use first-order stochastic gradient descent solvers to optimize for individual metalens elements in the nanophotonic array as follows

$$\tilde{\phi}(r)^{m,n} = \underset{\{\phi\}}{\arg\min} \sum_{i=1}^{T} \sum_{\theta,\lambda} \mathcal{L}\left(\tilde{\mathbf{I}}_{\theta,\lambda}^{(i)}, \mathbf{I}_{\theta,\lambda}^{(i)}\right), \qquad (11)$$

where T is the total number of training image samples and the images are measured by the (m, n)-th metalens in the array and the loss function

$$\mathcal{L} = \mathcal{L}_{\text{MSE}} + \mathcal{L}_{\text{LPIPS}}. \qquad (12)$$

Specifically, we design the metalens to work in the entire broadband visible wavelength range and modulate the incident wave fields over a 60° FoV. We notice that PSFs vary smoothly across the FoV and hence we sample it in regular intervals of 15° during optimization, whereas the wavelengths are sampled in intervals of 50 nm over the visible range. We use the Adam optimizer with a learning rate of 0.001 running for 15 hours over the dataset described in Sec. 5 to optimize for the meta-optic phase. We further fine-tune the metalens phase to suppress side lobes in the PSF to eliminate the haze that corrupts the sensor measurements, especially the ones captured in the wild. Once the optimization is complete, we use the optimized radially symmetric metalens with the appropriate wedge phases to manufacture our meta-optic.

Full Spectrum Phase Initialization. To aid the optimization from above, we propose a full spectrum metalens phase initialization wherein a rotationally symmetric metalens phase is optimized to maximize the focal intensity at the center of the field of view. Specifically, we initialize the optimization described in Eq. 11 with the solution to another optimization problem with the following objective

$$\tilde{\phi}(r) = \underset{\{\phi\}}{\arg\min} \sum_{\lambda=400nm}^{700nm} -f_{\text{META}}(\phi, \theta = 0, C_{\text{META}})\Big|_{(x_s, y_s)=(0,0)}. \qquad (13)$$

where (x_s, y_s) are the coordinates on the sensor plane. In other words, the solution to the above optimization problem finds a metalens phase that focuses all broadband light energy at the center of the sensor plane, thereby significantly reducing chromatic artifacts. We sample the wavelengths in steps of 10 nm and further use a per-pixel error function on the computed PSF in order to further improve the phase initialization. Note that similar to the phase described in Eq. (3), we use a per-pixel basis for solving the above metasurface phase which we later use to initialize Eq. (11).

Finally, the phase obtained by solving the optimization problem described in Eq. (11) is fabricated and installed on the sensor of the prototype camera, see Sec. 5.2. The measurements by this ultra-thin compact camera follow Eq. (8), and we next describe how the latent images are recovered.

4 PROBABILISTIC IMAGE RECOVERY

This section describes how we recover images from measurements of the on-sensor array camera. We first formulate the image recovery task as a model-based inverse optimization problem with a probabilistic sampling stage that samples a learned prior. We solve the optimization problem via splitting and unrolling into a differentiable truncated solver. To learn a natural image prior along with the unrolled solver, we propose a probabilistic diffusion model that samples a multi-modal distribution of plausible latent images. For ease of notation, we first describe the image recovery algorithm for a single lens element before describing the recovery method for the entire array.
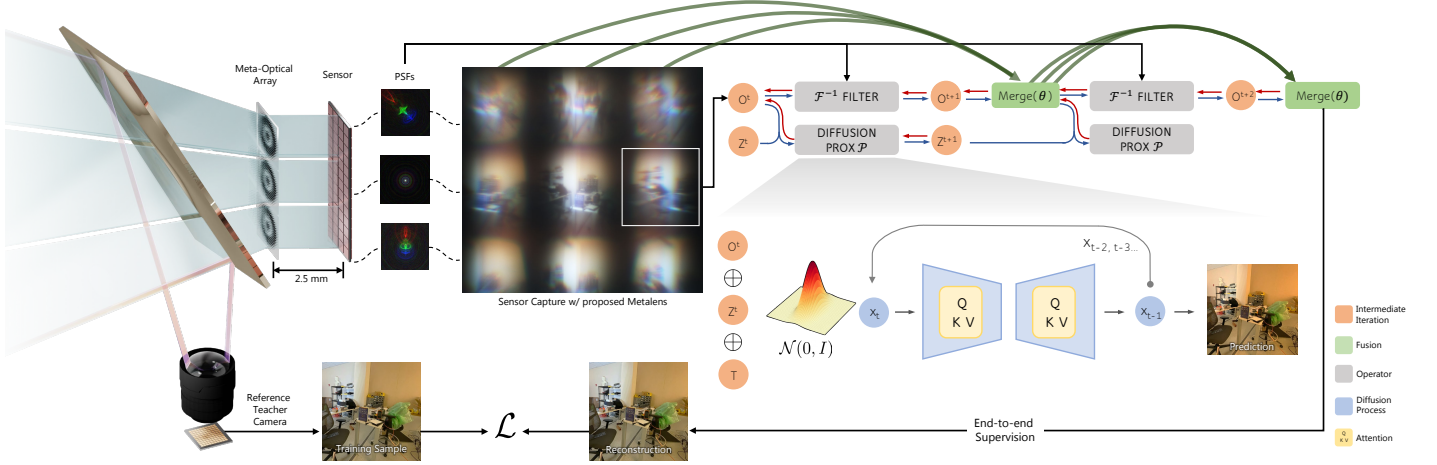
Fig. 4. Overview of Probabilistic Image Reconstruction. We propose a deconvolution method that relies on the physics-based forward model (PSFs illustrated on the left) along with a learned probabilistic prior (diffusion model on the bottom right). The proposed method is an unrolled optimization method that alternates between inverse filtering steps using the calibrated PSFs, diffusion steps that sample from the natural image manifold with conditioning on the current iteration, and a merging step that combines the image estimates from all sub-apertures (green). The unrolled optimization method is trained in an end-to-end fashion with paired training samples captured with a co-axial reference camera (bottom and left), see text for details.

4.1 Model-based Splitting Optimization for a Single Element

We propose a method to recover the latent image $\mathbf{I}$ from the sensor measurement $\mathbf{S}$ that relies on the physical forward model described in Eq. (8). We represent the spatially varying PSF of the array camera as $\mathbf{k}$ in the following for brevity. Following a large body of work on inverse problems in imaging [Bertero et al. 2021; Romano et al. 2017; Venkatakrishnan et al. 2013], we pose the deconvolution problem at hand as a Bayesian estimation problem. Specifically, we solve a maximum-a-posteriori estimation problem [Laumont et al. 2022] with an abstract natural image prior $\Gamma(\mathbf{I})$, that is

$$\tilde{\mathbf{I}} = \arg\min_{\{\mathbf{I}\}} \underbrace{\frac{1}{2}\left\|\mathbf{I} \otimes \mathbf{k} - \mathbf{S}\right\|^2}_{\text{Data Fidelity}} + \underbrace{\rho\Gamma(\mathbf{I})}_{\text{Prior Regularization}}, \qquad (14)$$

where $\rho > 0$ is a prior hyperparameter. However, instead of solving for the singular maximum of the posterior as a point estimate, we employ a *probabilistic prior that samples* the posterior of all plausible natural image priors. In other words, this will allow us to sample multiple plausible reconstructions near the maximum.

To solve Eq. (14), we split the non-linear and non-convex prior term from the linear data fidelity term to result in two simpler subproblems via half-quadratic splitting. To this end, we introduce an auxiliary variable $\mathbf{z}$, and pose the above minimization problem as

$$\arg\min_{\{\mathbf{I}\}} \frac{1}{2}\left\|\mathbf{I} \otimes \mathbf{k} - \mathbf{S}\right\|^2 + \rho\Gamma(\mathbf{z}) \quad s.t. \quad \mathbf{z} = \mathbf{I}. \qquad (15)$$

We further reformulate the above minimization problem then as

$$\arg\min_{\{\mathbf{I},\mathbf{z}\}} \frac{1}{2}\left\|\mathbf{I} \otimes \mathbf{k} - \mathbf{S}\right\|^2 + \rho\Gamma(\mathbf{z}) + \frac{\mu}{2}\left\|\mathbf{z} - \mathbf{I}\right\|^2, \quad \mu \to \infty, \qquad (16)$$

where $\mu > 0$ is a penalty parameter, that $\mu \to \infty$ mandates equality $\mathbf{I} = \mathbf{z}$. We relax μ and solve the above Eq. (16) iteratively by alternating between the following two steps,

$$\begin{cases} \mathbf{I}^{t+1} = \arg\min_{\{\mathbf{I}\}} \frac{1}{2}\left\|\mathbf{I} \otimes \mathbf{k} - \mathbf{S}\right\|^2 + \frac{\mu^t}{2}\left\|\mathbf{I} - \mathbf{z}^t\right\|^2, \\ \mathbf{z}^{t+1} = \arg\min_{\{\mathbf{z}\}} \frac{\mu^t}{2}\left\|\mathbf{z} - \mathbf{I}^{t+1}\right\|^2 + \rho\Gamma(\mathbf{z}). \end{cases} \qquad (17)$$

where t is the iteration index and μ^t is the updated weight in each iteration. We initialize our method with $\mu^0 = 0.1$ and exponentially increase its value for every iteration. Note that we solve for $\mathbf{I}$ given fixed values of $\mathbf{z}$ from the previous iteration and vice-versa.

The first update from the iteration (17) is a quadratic term that corresponds to the data term from Eq. (14). Assuming circular convolution, it can be solved in closed form with the following inverse filter update

$$\mathbf{I}^{t+1} = \mathcal{F}^{\dagger}\left(\frac{\mathcal{F}^*(\mathbf{k})\mathcal{F}(\mathbf{S}) + \mu^t\mathcal{F}(\mathbf{I}^t)}{\mathcal{F}^*(\mathbf{k})\mathcal{F}(\mathbf{k}) + \mu^t}\right), \qquad (18)$$

where $\mathcal{F}(\cdot)$ denotes the Fast Fourier Transform (FFT), $\mathcal{F}^*(\cdot)$ denotes the complex conjugate of FFT, and $\mathcal{F}^{\dagger}(\cdot)$ denotes the inverse FFT. Please refer to the Supplemental Material for the derivation.

However, the second update from iteration (17) includes the abstract regularizer and it is, in general, non-linear and non-convex. We learn the solution to this minimization problem with a diffusion model that allows us to probabilistically sample the solution space near the current iterate $\mathbf{I}^{t+1}$. Specifically we sample from a distribution Ω that is conditioned on the iterate $\mathbf{I}^{t+1}$ and the optimization penalty weights ρ, μ as inputs

$$\mathbf{z}^{t+1} \sim \Omega(\mathbf{I}^{t+1}, \rho, \mu^t). \qquad (19)$$

Next, we describe how we learn and sample from this prior in our method.

4.2 Diffusion-based Image Prior

We propose a diffusion-based prior Ω [Ho et al. 2020; Kawar et al. 2022; Sohl-Dickstein et al. 2015] to handle an ambiguity in deconvolution: multiple clear latent images can be projected to the same measurement $\mathbf{I}$. Diffusion provides a probabilistic approach to generate multiple samples, from which we can select the most suitable one. Figure 4 illustrates the proposed prior model.

We first devise the forward process of diffusion by adding noise and learning to recover the clean image. We denote our input x_0 as I^{gt}, and our condition c is defined as

$$c = I^{gt} \oplus \mathbf{S} \oplus z^t \oplus \mu^t \oplus \gamma(T), \qquad (20)$$

where I^{gt} is the ground truth latent image, $\mathbf{S}$ is the sensor measurement, z^t is the auxiliary image coupling term defined in Eq. (15), μ^t is an update weight defined in Eq. (16), and $\gamma(T)$ is a positional encoding of T where $T \in [1, 1000]$ is the timestep randomly sampled for each training iteration of the diffusion model. Note that the subscript t in z^t and μ^t refers to the HQS iteration from Eq. (17), separate from T which refers to the diffusion timestep.

Here, $\oplus$ is the concatenation symbol, as we condition the inputs by concatenating them along the channel dimension and employ self-attention [Vaswani et al. 2017] to learn corresponding features.

To train our diffusion model, in each iteration we add Gaussian noise to the I^{gt} of x_0 proportional to T to obtain x_t. Specifically, we train the model Ω to recover O^{gt} from x_t. Similar to [Chou et al. 2022], we recover O^{gt} rather than the added noise. To tackle moderate misalignment in our dataset, we employ a Contextual Bilateral loss (CoBi) which is robust to misalignment of image pairs in both RGB and VGG-19 feature space[Zhang et al. 2019]. Our overall training objective is

$$\mathcal{L}_{\text{diff}} = \mathcal{L}_{\text{CoBi}}(\Omega(x_t), I^{gt})_{RGB} + \lambda \mathcal{L}_{\text{CoBi}}(\Omega(x_t), I^{gt})_{VGG}, \qquad (21)$$

where λ is empirically selected via experimentation. The architecture of our diffusion model is a UNet [Ronneberger et al. 2015] following [Ho et al. 2020]. We provide further details in the Supplemental Document.

During test time, our diffusion model performs generation iteratively. In the vanilla DDPM [Ho et al. 2020], generation is performed as follows

$$z' = (f \circ ... \circ f)(z_T, T), \ f(x_t, t) = \Omega(x_t) + \sigma_t \epsilon, \qquad (22)$$

where $z_T \sim \mathcal{N}(0, I)$, σ_t is the fixed standard deviation at the given timestep, and $\epsilon \sim \mathcal{N}(0, I)$. However, this results in long sampling times. Instead, we follow DDIM [Song et al. 2021], and adopt a non-Markovian diffusion process to reduce the number of sampling steps. Furthermore, we use the "consistency" property that allows us to manipulate the initial latent variable to guide the generated output. As a result, $f(x_t, t)$ from Eq. (22) can be defined as

$$f(x_t, t) = \sqrt{\alpha_{t-1}} \left(\frac{x_t - \sqrt{1 - \alpha_t}\Omega(x_t)}{\sqrt{\alpha_t}} \right) + \sqrt{1 - \alpha_{t-1} - \sigma_t^2} \cdot \Omega(x_t) + \sigma_t \epsilon. \qquad (23)$$

In practice, we find generation timesteps of 20 sufficient for our experiments.

4.3 Learned Array Deconvolution and Blending

The nanophotonic array lens measures an array of images, each with a different FoV, which we deconvolve and merge together to form a wider FoV image. We employ the probabilistic image recovery approach described in Sec. 4.1 for deconvolving the array of images. Specifically, the individual latent images in the array are recovered by solving

$$\tilde{\mathbf{I}}_{m,n} = \arg \min_{\{\mathbf{I}_{m,n}\}} \frac{1}{2} \left\| \mathbf{I}_{m,n} \otimes \mathbf{k}_{m,n} - \mathbf{S}_{m,n} \right\|^2 + \rho\Gamma(\mathbf{I}_{m,n}), \qquad (24)$$

where (m, n) corresponds to the sub-image in the m-th row and n-th column of the sensor array measurement. For solving this, we first acquire real PSF measurements $\mathbf{k}_{m,n}$ for each element of the metalens array. An sensor measurements $\mathbf{S}_{m,n}$ are acquired as a dataset of images captured in various indoor and outdoor environments, as described next in Sec. 5, to allow for learning the probabilistic prior Γ over natural images.

The recovered array of latent images are finally blended together into a wider FoV super-resolved image to approximately match the sensor resolution. Given an (m, n) array of input images $\{\mathbf{I}_{m,n}\}$ where $(m, n) = \{0, 1, 2, ...\}$, our goal is to produce a wide-range image $\mathbf{I}_B$, which is obtained by appropriately correcting, stitching and blending the individual sub-images recovered from the metalens array measurement. To this end, we employ a modified UNet blending network to learn the blending function f_B which takes a blended homography-transformed stack of concatenated mn sub-images (see Sec. 5.1) and a coarse alpha-blended wide-range image $\mathbf{I}_B^\alpha$ as input, and produces the correctly blended image as the output,

$$\mathbf{I}_B = f_B(\{\mathbf{I}_{m,n}\}, \mathbf{I}_B^\alpha). \qquad (25)$$

To learn the function f_B, the blending network is supervised over groundtruth images acquired using an aberration corrected compound optic camera, see Sec. 5.2. The loss function $\mathcal{L}$ used is a combination of pixel-wise error and perceptual loss during the training

$$\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_{\text{LPIPS}}, \qquad (26)$$

to allow for accurate reproduction of color and features while also accounting for any misalignments in the in-the-wild captured data pairs as well as systemic errors in the prototype data acquisition setup. Moreover, supervising our blending network on the full sensor resolution groundtruth image measurements also allows for recovering a high-fidelity latent image from the $m \times n$ low resolution sub-images from the metalens array camera.

4.4 Implementation

The proposed deconvolution framework and the learned blending network are implemented in PyTorch. The training for the overall deconvolution approach is done iteratively and progressively to sample over a large plausible manifold of latent images from the sensor measurements. For all the training purposes, we use a dataset with groundtruth images of 800×800 pixels resolution and 9 patches of 420×420 pixels sub-images measured from individual metalenses in the nanophotonic array. Training was performed using the paired groundtruth and metalens array measurements acquired using the experimental paired-camera setup, see Sec. 5. During the deconvolution step, an initial filtered image obtained

according to Eq. (18) is passed through a probabilistic diffusion-prior model that progressively corrupts the filtered image with additive noise and recovers the latent image by sampling over the manifold of probability distribution of image priors. To preserve color fidelity, we normalize the image to range [0, 1]. We use the ADAM optimizer with $\beta_1 = 0.5$ and $\beta_2 = 0.999$, and $\lambda = 1.2$ for the training objective in Eq. (21).

5 EXPERIMENTAL PROTOTYPE AND DATASETS

This section describes the dataset and camera prototype we use to train the proposed reconstruction network. The training dataset consists of simulated data and captured paired image data. We first describe the synthetic dataset, then the capture setup and the acquisition of the proposed paired dataset. Finally, we describe the fabrication process of the proposed nano-optical array.

5.1 Synthetic Dataset

Training the probabilistic image recovery network described in Sec. 4 requires a large and diverse set of paired data, which is challenging to acquire in-the-wild. Therefore, we simulate the nanophotonic array camera with the corresponding metalens design parameters, to generate a large synthetic dataset of paired on-sensor and groundtruth measurements. We use this large synthetic dataset for training alongside a smaller real-world dataset for fine-tuning. Each metalens in the array camera has a focal length of 2 mm and covers an FoV of $60°$ for a broadband illumination, with the center-to-center distance between the metalenses on-chip being 2.42 mm. Due to the circular aperture of each metaoptic, the sensor measurements exhibit vignetting at higher eccentricities.

For a given groundtruth image, we first crop 9 images that correspond to the final 3×3 metalens array camera measurement, with each metalens measurement corresponding to $60°$ FoV and the groundtruth image corresponding to a total of $90°$ FoV. Each of the 9 images are subjected to vignetting where we model the vignetting mask as a fourth-order Butterworth filter with linear intensity fall-off, given by

$$\mathbf{V} = \left(1 + \left(\frac{||w||^2}{f_c^2} \right)^4 \right)^{-1} \tag{27}$$

where $|| \cdot ||^2$ denotes the squared magnitude, w is the spatial frequency and f_c is the cutoff frequency of the filter. All parameters are matched to the experimental setting. Note that we apply this filter on each individual metalens measurement only as an intensity mask to the sensor image and our cutoff frequency corresponded to $45°$ of the metalens FoV. The vignetted images are convolved with the simulated PSFs on the sensor as described in Eq. (7) and further corrupted by simulated sensor noise described in Eq. (8). The simulated individual metalens measurements are then resized and arranged in a 3×3 array to simulate the nanophotonic sensor capture. To this end, we first compute homographies between the 9 local image patches as measured by the real nanophotonic array camera and the ground truth compound optic camera, which is described next in Sec 5.2, to transform the ground truth image to map that of the sensor capture. We then utilize these homography transforms to project each of the 9 simulated metalens measurements onto the

appropriate local patch on the sensor

$$\hat{\mathbf{p}}_{mn}^{gt} \rightarrow \mathbf{H}_{mn}\mathbf{p}_{mn}^{s}, \tag{28}$$

where $\hat{\mathbf{p}}_{mn}^{gt}$ denotes the coordinates in the ground truth image corresponding to the FoV as captured by the $(m, n) - th$ metalens in the array camera, $\mathbf{p}_{mn}^{s}$ denotes the sensor coordinate corresponding to the (m, n)-th metalens measurement and $\mathbf{H}_{mn}$ denotes the corresponding homography. The final sensor measurement is simulated as

$$\mathbf{S}_{mn} = (\mathbf{H}_{mn}^{-1}\mathbf{I} * \mathbf{V}) \otimes \mathbf{k}_{mn} + \eta_{\text{SENSOR}}, \tag{29}$$

$$\mathbf{S} = \sum_{m,n} \mathbf{S}_{mn} \quad s.t. \quad (m, n) \in \{0, 1, 2\}; \quad m + n \leq 2, \tag{30}$$

where $\mathbf{S}_{mn}$ denotes the (m, n)-th array measurement on the sensor, $\mathbf{S}$ being the final sensor measurement, $\mathbf{H}_{mn}^{-1}$ and $\mathbf{k}_{mn}$ being the corresponding inverse homography and PSF, respectively. The sensor noise added is as determined by the parameters $C_{\text{SENSOR}} = \{\sigma_g, a_p\}$ which we determine to be $\sigma_g = 1 \times 10^{-5}$ and $a_p = 4 \times 10^{-5}$ using the calibration method as described in Foi et al. [Foi et al. 2008].

To generate the full synthetic dataset, we randomly sample 10,000 images from a combination of ImageNet [Deng et al. 2009] and MIT 5K [Bychkovsky et al. 2011] datasets for groundtruth images. Our training dataset contains 8000 images and the validation and test data splits contain 1000 each. The networks trained on our synthetic dataset are then further finetuned on in-the-wild real data. Specifically, we pre-train the diffusion model on our 8000-image synthetic training dataset by 200 epochs and fine-tune the model weights on the training set of the experimental dataset we discussed in the following section. We evaluate the fine-tuned model with SSIM, PSNR, and LIPIPS metrics and select the best-performing weights for final real-world testing.

5.2 Experimental Setup and Dataset

To acquire the paired experimental data, we developed a hardware setup shown in Figure 5, which can simultaneously capture real-world scenes from the metalens array camera and a reference camera that has a conventional off-the-shelf lens. To this end, we use a plate beam splitter, which splits world light into two optical paths by 70% transmission and 30% reflection. The positions of the plate beam splitter and the two cameras are precisely aligned such that the optical centers and the optical axes of the two cameras are as close as possible. The two cameras are synchronized to capture scenes with the same timestamps.

We employ an Allied Vision GT1930 C sensor of 5.86 micron pixel pitch and 1936x1216 resolution for the metalens array camera such that the effective FoV (Field-of-View) from all the metalens elements in the array can be captured in the same frame. The same sensor is used for the reference camera which has a 3.5 mm focal length, wide FoV lens from Edmund Optics such that we can achieve a FoV larger than the full FoV of the metalens array camera in the "ground truth" captures. A third Allied Vision GT1290C camera with 3.75 micron pixel pitch and 1280x960 resolution is used for mounting the metalens proposed by Tseng et al. [2021a], which we compare against in Section 6. We use Precision Time Protocol (PTP) to synchronize all the cameras such that the captures are taken at
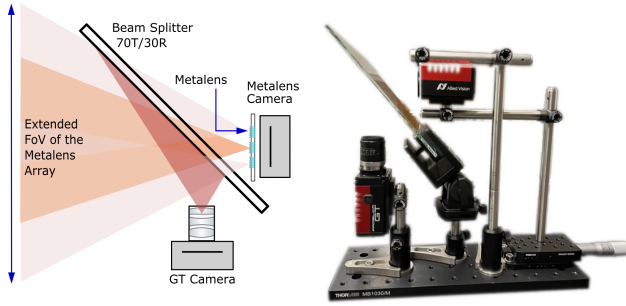
Fig. 5. Paired Data Acquisition. In our capture setup, we employ a plate beam splitter, which splits world light into two optical paths by 70% transmission and 30% reflection such that the setup can simultaneously capture real-world scenes with one camera in the transmission path that employs the designed metalens array and another camera in the reflection path that employs a conventional off-the-shelf lens (GT camera). The two cameras are aligned and calibrated to map one to the other captures, see text for details.

the same timestamps with sub-millisecond precision. After we align the sensor parallel to our fabricated metalens array, we perform fine alignment between the sensor and the metalens array with a 3D translation stage, where the sensor is mounted to. When the alignment is completed, the sensor captures the effective FoV of all the metalens array elements and the images are focused on the sensor plane. See Supplemental Material for details.

After the alignment, we conduct PSF measurements of the individual metalens elements in the array, which are used in the model-based part of the image reconstruction method. The light sources that we use are red, green, and blue fiber-coupled LEDs from Thorlabs (M455F3, M530F2, and M660FP1). The fiber has a core size of 800 microns diameter and the fiber tip is placed 340 mm away from the metalens array such that it can be approximated as a point source with an angular resolution that is the same as the angular resolution of one pixel in the captured metalens images (8̃ arc-min). The PSFs of all the metalens elements are captured in the same frame. By turning on and off each individual color LED, we can acquire the PSFs of different colors. When alternating between colors, we change the input of the fiber without introducing mechanical shifts to the output of the fiber such that the position of the point light source is fixed.

Next, we align the optical center and the optical axis of the central element from the metalens array camera to those of the reference camera. We use collimated laser and pinhole apertures to make sure the beam splitter is positioned at a 45° tilting angle. Then, we set up the position of the metalens array camera and adjust the laser beam height such that the transmission path is incident on the center metalens element. The center of the reference camera is positioned in the reflection beam path and the distance between the beam splitter and the reference camera sensor is adjusted to the same as that between the beam splitter and the metalens array camera. We achieve accurate alignment by observing a reference target with both cameras simultaneously until the two cameras are aligned.

After all the alignment is completed, the setup is mounted on a tripod with rollers, as shown in Figure 5, such that it can be moved around indoors and outdoors for acquiring a diverse dataset. In the capture process, the exposure time is chosen so that the photos from the two cameras are bright but unsaturated, and the frame rate is chosen to make sure that there is a sufficient difference in the scenes for a few consecutive frames. When we change the scenes, the exposure time of the two cameras is adjusted proportionally such that the image brightness is adapted to different scenes while the light throughput ratio between the two cameras is unchanged. The exposure time of the metalens array camera ranges between 50 ms to 150 ms and the exposure time of the reference camera is 1.2× that of the metalens camera.

Per-pixel Mapping between Two Cameras. To find the per-pixel mapping between the reference camera and the metalens array camera, we have the two cameras capture red, green and blue checkerboard patterns shown on a large LCD screen and then calibrate the distortion coefficients of the two cameras per color channel. After the image acquisition, we perform image rectification for the captures from both cameras. Then, to account for the difference in camera FoV and the difference in viewing perspectives between each metalens array element and the reference camera, we perform homography-based alignment to map the reference camera captures to the captures from all the metalens array elements.

5.3 Fabrication of the Meta-Optic

The optimized meta-optic design described in Sec. 3 was fabricated in a 700 nm SiN thin film on fused silica substrate. First a SiN thin film was deposited on a fused silica wafer via plasma-enhanced chemical vapor deposition. The meta-optic array was then written on a single chip via electron beam lithography (JEOL-JBX6300FS, 100 kV) using a resist layer (ZEP-520A) and discharging polymer layer (DisCharge H2O). After development, a hard mask of alumina (65 nm) was evaporated, and subsequently lift-off overnight in NMP at 110 °C. After a brief plasma clean to remove organic residues, we used inductively-coupled reactive ion etching (Oxford Instruments, PlasmaLab100) with a fluorine based etch chemistry to transfer the meta-optic layout from the hard mask into the underneath SiN thin film. Finally, we created apertures for the meta-optics to exclude unmodulated light that passed through non-patterned regions. These apertures were created through optical direct write lithography (Heidelberg-DWL66) and subsequent deposition of a 150 nm thick gold film. Our array has a total size of ~ $7mm^2$ with elements $1mm$ in diameter and $F\#2.4$. We avoid optical baffles in our prototype and, to ensure no overlap, instead space out the lenslets over the wafer with ~ 15% of the area being used as apertures. However, note that we do not use the peripheral regions of each sublens; hence, we use non-continuous regions of pixels totaling ~ 40% of the full sensor area. In the future, integrating optical baffles to separate array elements may eliminate the need for separation. However, fabricating and aligning baffles is not a simple feat and we prototype our camera without them. Please refer to the Supplemental Material for additional details.
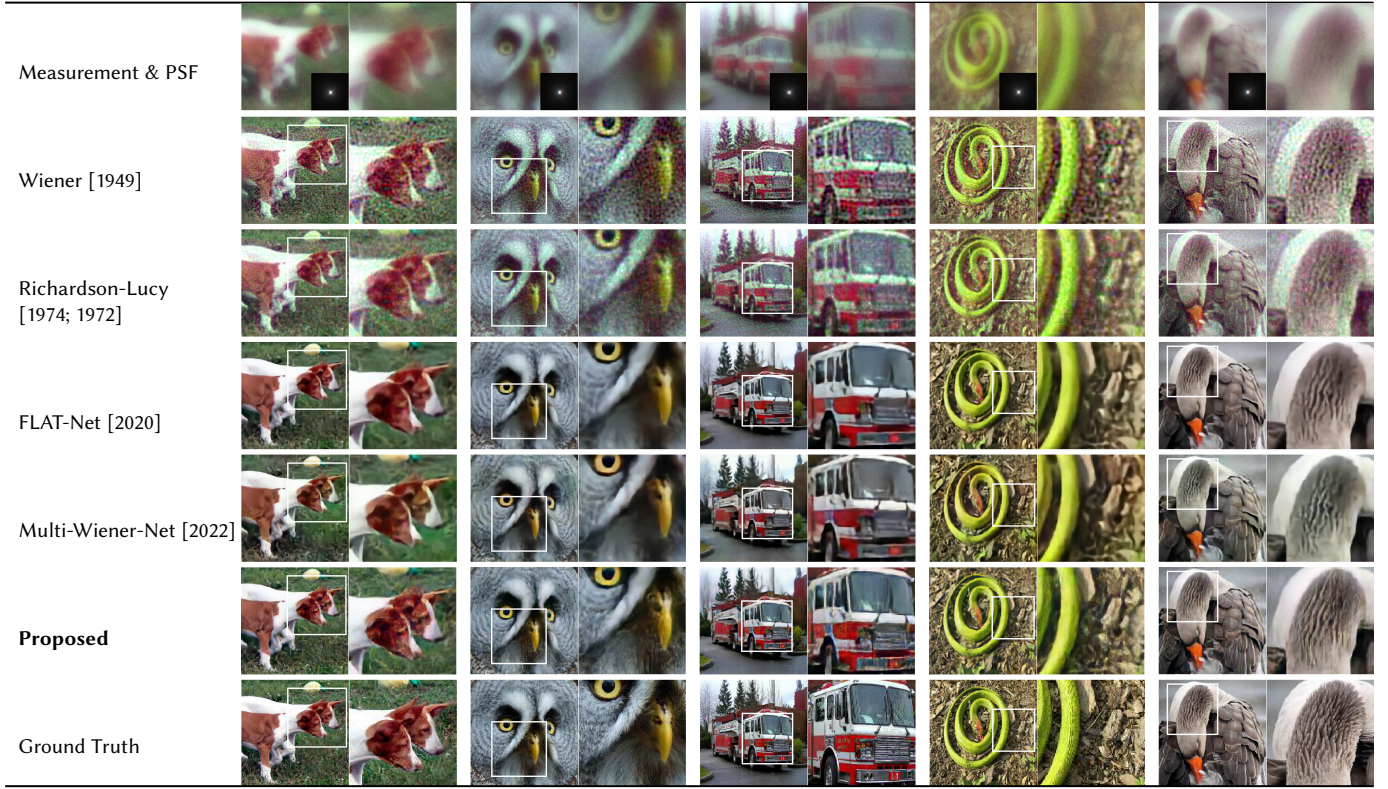
Fig. 6. Synthetic Qualitative Assessment of Diffusion-based Deconvolution. Results on *unseen* validation set. The two conventional deconvolution approaches (Wiener [Wiener et al. 1949] and Richardson-Lucy [Richardson 1972]) suffer from apparent reconstruction noise. The predictions from Flatnet [Khan et al. 2020] and Multi-Wiener-Net [Yanny et al. 2022] are overly smooth with high-frequency details missing. The proposed probabilistic reconstruction method is capable of recovering fine details, such as the grass (left), feathers (center left), or grill (center) without severe reconstruction noise.

6 ASSESSMENT

6.1 Synthetic Evaluation

Before validating the proposed method on experimental captures, we separately evaluate the probabilistic deconvolution method and the proposed thin camera in simulation. To this end, we use unseen test set (consisting of 1000 images) from our synthetic dataset described in Sec. 5.1 to assess the method with paired ground truth data.

Assessment of Probabilistic Deconvolution. Existing non-blind deconvolution methods do not operate on several sub-aperture images that are combined together to form a final image. To assess the proposed probabilistic deconvolution method in isolation, and allow for a fair comparison, instead of considering all nine sub-apertures of the proposed meta-optic, we consider only the central portion. Doing so allows us to compare the proposed reconstruction method with a single PSF and image – the setting that existing non-blind deconvolution methods are addressing. For this experiment, we drop the blending operator from the proposed method described in Sec. 4 and train the remainder of the method as described next.

We report qualitative and quantitative results in Table 2 and Figure 6 which both validate the proposed reconstruction method. Specifically, we compare our method to existing conventional and learned non-blind deconvolution methods, that is Wiener inverse

Table 2. Qualitative Assessment of Probabilistic Non-blind Deconvolution. To evaluate the proposed reconstruction method, we simulate aberrated and noisy images of the central lens in our optical design, see Sec. 5.1. We evaluate all methods on our synthetic validation set and find that the proposed method outperforms all baselines in SSIM, PSNR, and LPIPS [2018].

	Wiener	Richardson-Lucy	Flatnet [2][2020]	Multi-Wiener-Net [2022]	Proposed
SSIM ↑	0.452	0.486	0.679	0.648	**0.754**
PSNR [dB] ↑	19.38	19.97	24.63	22.90	**25.80**
1-LPIPS ↑	0.360	0.495	0.672	0.569	**0.773**

filtering [1949] and Richardson Lucy iterations [1974; 1972] as traditional methods, and the FLAT-Net [Zhou et al. 2020] and Multi-Wiener Net [Yanny et al. 2022] as recent learning-based approaches. We retrain the two learning-based approaches on our synthetic data for a fair comparison. Table 2 confirms that the proposed method outperforms all compared methods in all metrics, that is SSIM, PSNR, and LPIPS [Zhang et al. 2018]. Although all learned methods are trained on the same data, the proposed method improves on the existing learned baseline methods by a margin of more than 1 dB in PSNR. The qualitative results reported in Figure 6 confirm this trend. While the learned approaches, FLAT-Net and Multi-Wiener-Net are robust to noise and function in presence of large blur kernels such

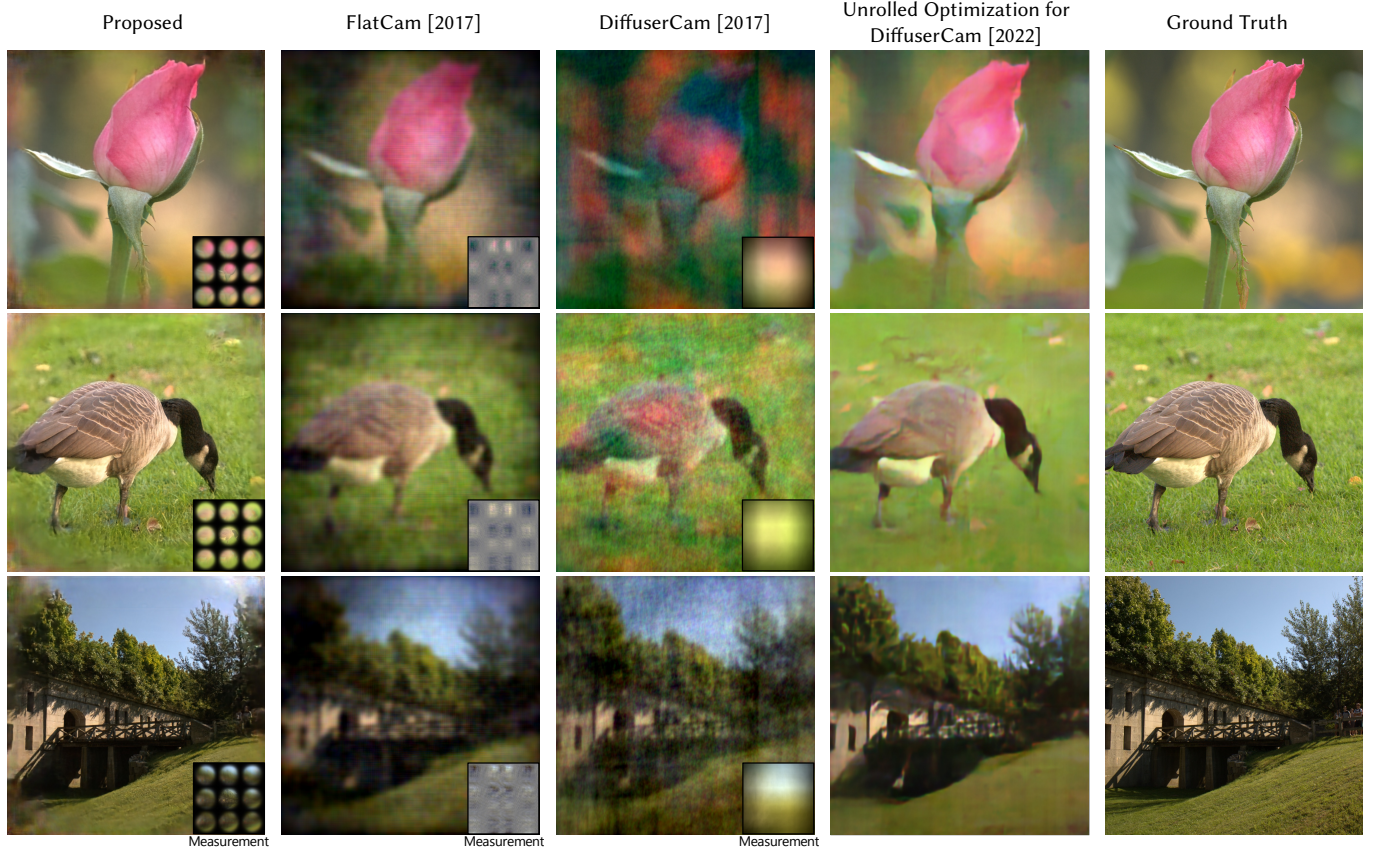| Proposed	FlatCam [2017]	DiffuserCam [2017]	Unrolled Optimization for DiffuserCam [2022]	Ground Truth

Fig. 7. Synthetic Assessment of Thin Cameras. Results on *unseen* validation set. Alternative thin sensing approaches in FlatCam [2017] and DiffuserCam [Kuo et al. 2017] allow for capturing rays from a large cone of angles, however, they mix spatial and color information in PSFs with support of the entire sensor, see insets. This makes the recovery of high-frequency details challenging, even for learning-based methods [Kingshott et al. 2022]. The proposed camera design captures high-quality information across almost the entire field of view.

as the one simulated in this experiment – in contrast to the conventional methods – their predictions tend to be oversmoothed. The proposed probabilistic prior is capable of recovering fine details, such as the swan feather structure (right), headlights of the fire truck (center) and details in grass and face of the dogs (left).

Validation of Thin Imager Design. Next, we validate the proposed thin camera design in simulation. We again rely on the unseen test set from our synthetic dataset described in Sec. 5.1 to evaluate the method with ground truth data available. We now consider all nine sub-apertures on the sensor that require employing the blending operator we dropped for the experiments described above. Figure 7 includes simulated sensor measurements as insets for a few scenes, see Supplemental Document for additional simulations.

The qualitative and quantitative evaluations in Table 3 and Figure 7 validate the proposed thin camera design. Here, we compare the proposed thin imager to successful imaging methods with a flat form factor: the FlatCam [2017] design, which employs an amplitude mask placed in the sensor cover glass region instead of a compound lens, and DiffuserCam [2017] relying on a caustic PSF resulting from a diffuser placed above the coverglass. In addition to

Table 3. Qualitative Assessment of Thin Camera Design. To evaluate the nanophotonic array camera design proposed in this work, we simulate aberrated and noisy images for our 3 × 3 array following Sec. 5.1, and recover images with the proposed probabilistic reconstruction method. We assess the image quality compared to FlatCam [Asif et al. 2017] and DiffuserCam [Antipa et al. 2018] as alternative thin camera design approaches. We evaluate all methods on our *unseen* synthetic validation set and find that the proposed design compares favorably in SSIM, PSNR, and LPIPS [2018].

	FlatCam [2017]	FlatNet [2020]	DiffuserCam [2017]	Kingshott et al. [2022]	Proposed w/ Tikhonov	Proposed
SSIM ↑	0.544	0.533	0.479	0.594	0.731	**0.892**
PSNR [dB] ↑	19.25	20.61	16.42	21.24	25.47	**32.66**
1-LPIPS ↑	0.292	0.426	0.255	0.348	0.71	**0.803**

evaluating the image formation and reconstruction methods proposed in the original works, we also evaluate recent learning-based reconstruction methods, including FlatNet [Khan et al. 2020] which is capable of learning from FlatCam observations, and the unrolled optimization method with neural network prior from Kingshott et al. [2022] that recovers images from diffuser measurements. We retrain the learning-based approaches on our synthetic data for a fair comparison. The proposed thin imager improves on all alternative
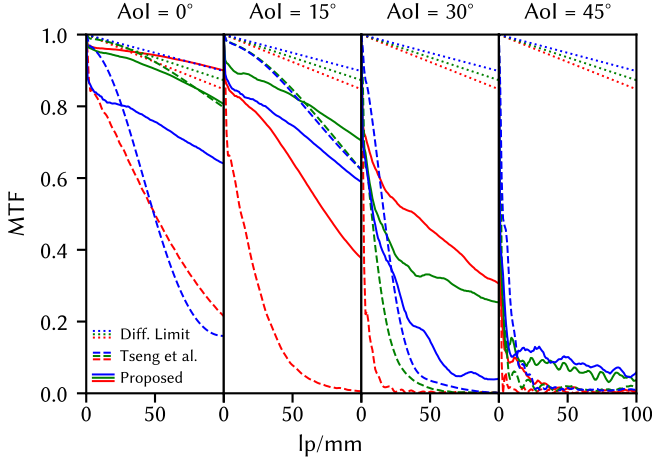
Fig. 8. MTF for different angles of incidence. The calculated MTF for various angle of incidence (AoI) for array MO with varying wedge phase profile. Thick solid lines correspond to the diffraction limit. Dashed lines correspond to calculated MTF curves for Tseng et al. [2021a]. Full thin lines correspond to the MO, presented in this work, with wedge phase as specified in the title. The x-axes represent line pairs per mm, within the specific resolution range for the used sensor, with a pixel size of 5.7 μm.
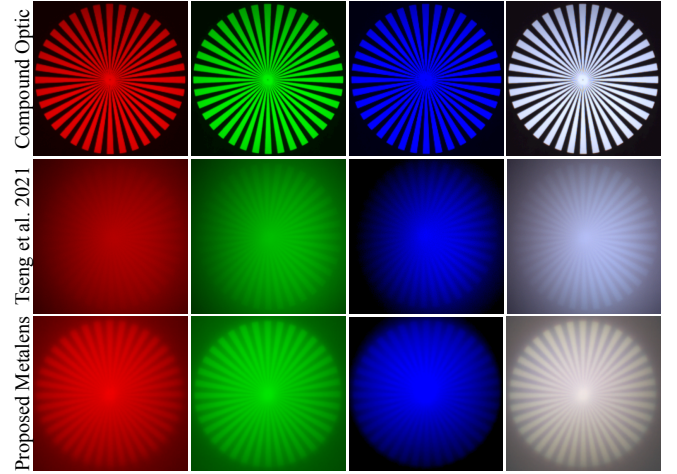

Fig. 9. Sensor measurement of the Siemens Star calibration pattern. Even without reconstruction, our metalens design significantly reduces scattering compared to the previous state-of-the-art design proposed by Tseng et al. [2021a] across all wavelengths, thereby allowing for broadband imaging in-the-wild. When combined with probablistic deconvolution method, the proposed nanophotonic array camera robustly recovers the latent image.

designs both quantitatively and qualitatively. While FlatCam and DiffuserCam sensing allow the capture of rays from a large cone of angles, the spatial and color information is entangled in PSFs with support of the entire sensor, making the recovery of high-frequency content challenging independent of the FoV. As such, all examples in Figure 7 confirm the trend from the quantitative Table 3 — the proposed metasurface array imager is able to image fine details across almost the entire field of view. The flower (top) and draw bridge (bottom) are reconstructed with high fidelity across the entire image. Only in the peripheral corners of the image, the proposed method is not capable of recovering details as image information is missing, which is filled in by the blending network.

The proposed camera design benefits from both the optical design and the probabilistic prior. To analyze the contribution of these two components, we conduct an ablation experiment by replacing the diffusion prior with a non-learned prior. Because spatial priors, including Total Variation (TV) regularization and neural network-based learned priors, both can "hallucinate" frequency content missing in the measurements (e.g., high-frequency edges in the case of TV), we compare our approach to Tikhonov regularization [Golub et al. 1999] as a traditional per-pixel prior. We observed an average PSNR of 25.5 dB, which still outperforms all alternative flat camera designs by more than 4 dB. The proposed diffusion prior further improves this by 7.2dB with the same input data used by all methods. These additional evaluations further validate both the optical design and effectiveness of the diffusion prior.

6.2 MTF and FoV Analysis

Next, we analyze the optical performance of the proposed nanophotonic array lens via its theoretical modulation transfer function

(MTF), *i.e.*, the ability of the array lens to transfer contrast at a given spatial frequency (resolution) from the object to the imaging sensor. As discussed in Sec. 3, our lens is optimized for broadband illumination across the visible spectrum and to span an effective FoV of 70° for a 3 × 3 and an FoV of 80° for a 5 × 5 metalens array, respectively, with each individual lens in the array capturing a total FoV of 45°. We calculate the MTF of our array designs and compare to the the recent design from Tseng et al. [2021a], which is reported to achieve a total FoV of 40°.

The analysis in Figure 8 validates that the proposed metalens exhibits a significantly improved MTF under different angle of incidence (AoI) compared to the existing designs, approaching the efficiency of diffraction limit for normal incidence of light (shown in black in Figure 8) and achieving higher MTF values for larger incidence angles. Importantly, even at large AoI, the MTF is sufficiently large to enable reconstruction through a computational backend. This allows for high-fidelity signal measurements and robust full-color deconvolution. With increasing angles of incident light, although the MTF performance drops, our metalens compares favorably to the design from Tseng et al. The MTF performance of our lens is also reflected in the raw sensor measurements. Figure 9 reports the raw sensor measurement using a compound optic, the metalens by Tseng et al. [2021a], and the measurement from our proposed metalens design. Although Tseng et al. [2021a] designed their metalens specifically for trichromatic red, green and blue wavelengths, the sensor measurement of the calibration Siemens star pattern show significant scattering resulting in loss of image quality. Note that the MTF considerations above are based on the direct sensor measurement without the algorithmic framework employed for latent image recovery. However, the effective MTF which also considers the image recovery algorithm [Fontbonne et al. 2022] which we review later in our experimental results.

---

[2]We compare here against the generalized version of Flatnet [2020] which does not require separable point-spread functions.
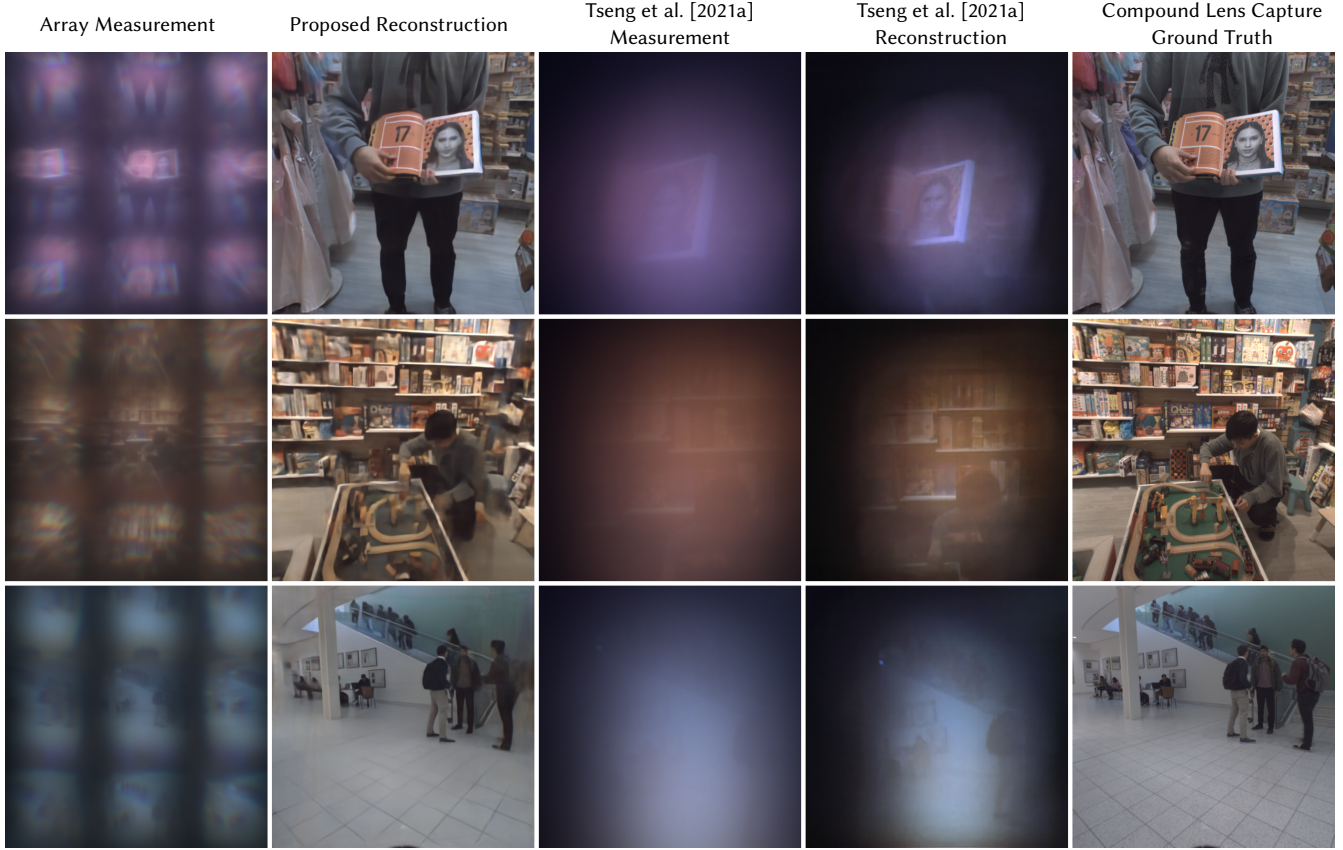
Fig. 10. Experimental evaluation of the proposed thin nanophotonic camera on broadband indoor scenes. The proposed nanophotonic array optic with the probabilistic deconvolution method reconstructs the underlying latent image robustly in broadband lit environments. As a comparison, Tseng et al. [2021a] cannot recover the color and spatial details of the scenes well. See Supplemental Material for lab captures acquired of a narrow-band screen where the proposed design and Tseng et. al both function adequately.

6.3 Experimental Assessment

In the following, we validate the proposed camera design with experimental reconstructions from captures acquired with the prototype system from Sec. 5.2. To this end, we aim to capture scenes that feature high-contrast detail, depth discontinuities, and color spanning a large gamut. To test the camera system in in-the-wild environments, we acquire scenes in typical broadband indoor and outdoor scenarios. As such, we note that, to the best of our knowledge, ours is the *first demonstration of broadband nanophotonic imaging outside the lab*. Figure 10 and Figure 11 shows images as captured by the previous state-of-the-art [Tseng et al. 2021a] and proposed thin-lens camera, and the corresponding reference images captured using a compound optical lens for a variety of indoor and outdoor scenes. The reconstructions from Tseng et al. were measured on a sensor of smaller size, see Sec 5.2, which we do not resize here, unlike in the Teaser 1.3 The images recovered using the proposed nano-optic and reconstruction algorithm outperform existing approaches. The proposed thin camera is capable of imaging the scene adequately with accurate color reproduction. While the peripheral regions, reconstructed from elements with strong wedge phase elements, contain less spatial detail than the center region, they still recover detail over the entire design field of view. The center region of the recovered

images has relatively high image quality and captures fine detail present in the reference images. The reconstructed images suffer from no apparent chromatic aberrations, which have been an open problem in the design of broadband metasurface optics.

Comparison to Neural Nano-Optics [Tseng et al. 2021a]. We compare the proposed design experimentally to the broadband design from Tseng et al. [2021a]. While their lens design is the most successful existing broadband metalens design, it is designed for a fixed set of three wavelength bands. As Tseng et al. [2021a] report, their design performs well for the narrow selective spectrum of an OLED display that is imaged with an optical relay system. We confirm this experimentally in the Supplemental Material. For the full broadband scenarios that we tackle in our work, their design comes with severe scattering that is not apparent when imaging a screen with black surrounding region, as shown in Figure 10 and Figure 11. For the experimental scenes captured in our work, the proposed method significantly outperforms this existing design in image quality and size of the field of view. This experiment validates the proposed broadband design methodology and the array sensing approach we investigate in this work.

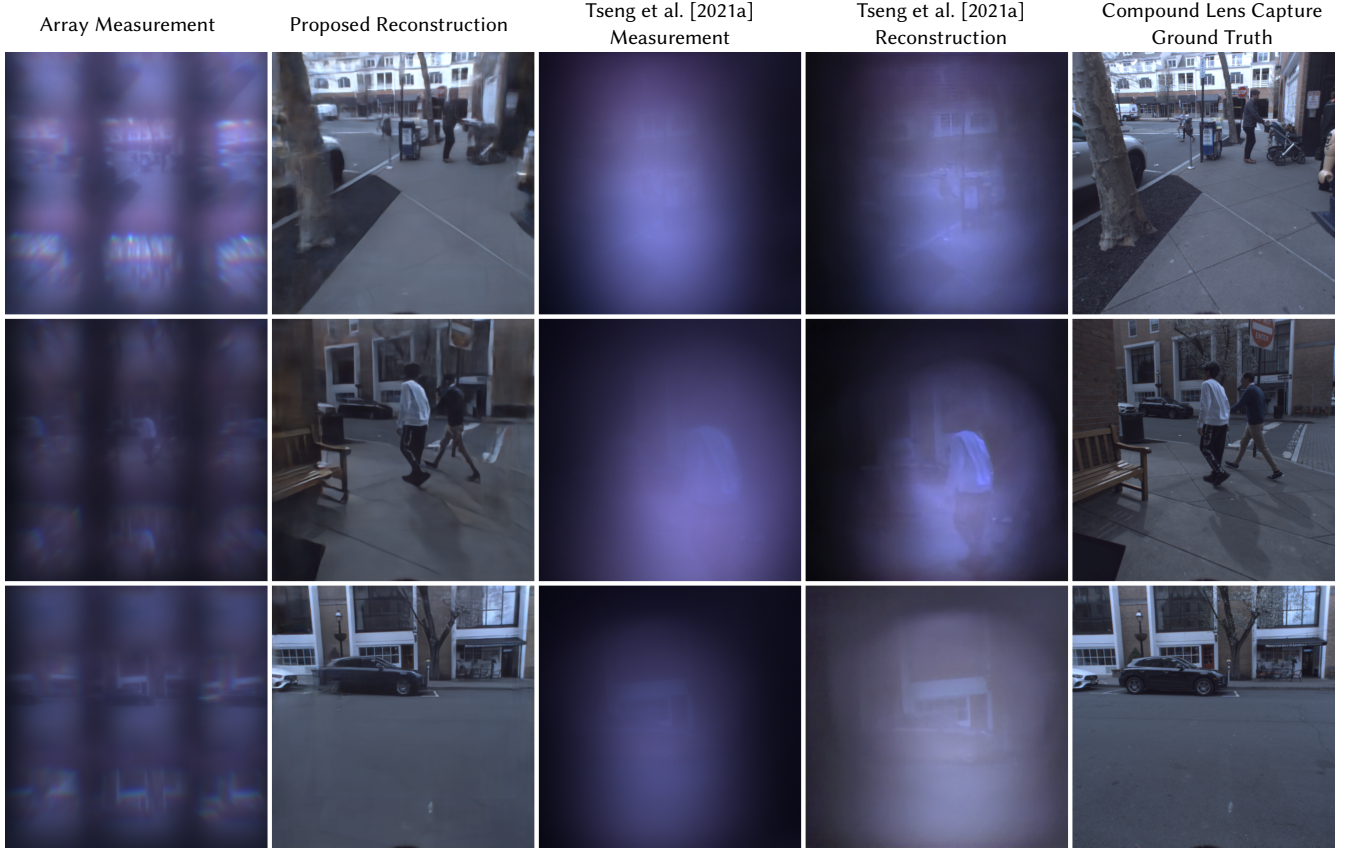| Array Measurement	Proposed Reconstruction	Tseng et al. [2021a] Measurement	Tseng et al. [2021a] Reconstruction	Compound Lens Capture Ground Truth

Fig. 11. Experimental evaluation of the proposed thin nanophotonic camera on broadband outdoor scenes. The proposed nanophotonic array optic with the probabilistic deconvolution method reconstructs the underlying latent image robustly in broadband lit environments, outperforming Tseng et al. [2021a] significantly in outdoor scenes. See Supplemental Material for lab captures acquired of a narrow-band screen where the proposed design and Tseng et. al both function adequately.

Experimental Validation of Denser 5 × 5 Design. In addition to the 3 × 3 array investigated above, we have also fabricated an additional 5 × 5 array with an additional peripheral set of nanophotonic lenslets to cover a larger field of view of 120°. Unfortunately, the sensors available to us (with sufficient lead time) were slightly too small to capture the entire array, and spacing the elements closer would have required baffles and the removal of the coverglass on the sensor [3]. Figure 12 reports the measurement of a still life scene captured under broadband quartz light *without* image reconstruction. Although the full capture of the array is cropped due to the available sensor size, we can observe the increasing field of view from the center element to the periphery. Validating the proposed design, as indicated in the figure, the red plush toy, completely outside the field of view of the center element, enters the field of view of the middle, and moves to the center in the left element.

7 CONCLUSION

We investigate a flat camera that employs a novel array of nanophotonic optics that are optimized for broadband spectrum and collaboratively capture a larger field of view than a single element. The

---

[3]The epoxy-glued cover glasses on commodity mass-market sensor packages cannot be removed without specialized tools or destroying the sensor.
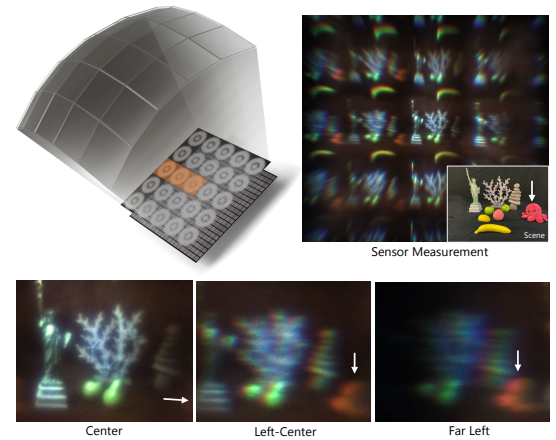

Fig. 12. Experimental Assessment of 5 × 5 Prototype Lens. Although the fabricated array does slightly exceed the available sensor area, see text, the capture of the horizontal set of elements (indicated in orange) illustrates that neighboring elements with increasing wedge angles capture successively oblique field angles. Moving towards the array periphery, the red plush toy (indicated with white arrow) is entering (center-left) and moving progressively towards the center of the field of view of each individual lens.

proposed nanophotonic array is embedded on a metasurface that sits on top of the sensor cover glass making the proposed imager thin and manufacturable with a single-element optical system. Although we devise a differentiable lens design method for the proposed array metasurface sensor – allowing us to suppress aberrations across the full visible spectrum that exist in today's heuristic and optimized metasurface optics – the proposed design is not without aberrations. We propose a probabilistic image reconstruction method that allows us to recover images in presence of scene-dependent aberrations in broadband – an open problem using metasurface optics. We validate the proposed nanophotonic array camera design experimentally and in simulation, confirming the effectiveness not only of the optical design, compared against existing broadband metasurface optics, but also the deconvolution method, compared in isolation or against alternative thin camera designs. In the future, we plan to explore integrating low-cost baffles and the co-design with sensor color-filter arrays into the proposed design which requires scale-able fabrication integrated into the sensor cover glass. We hope that the proposed camera can not only inspire novel designs, e.g., flexible sensor arrays, but also re-open an exciting design space computational photography community has explored in the past, that is light field arrays, color-multiplexed arrays, and task-specific array optics – all now directly on the sensor.

ACKNOWLEDGMENTS

REFERENCES

Francesco Aieta, Patrice Genevet, Mikhail A. Kats, Nanfang Yu, Romain Blanchard, Zeno Gaburro, and Federico Capasso. 2012. Aberration-Free Ultrathin Flat Lenses and Axicons at Telecom Wavelengths Based on Plasmonic Metasurfaces. *Nano Letters* 12, 9 (2012), 4932–4936. https://doi.org/10.1021/nl302516v

Francesco Aieta, Mikhail A. Kats, Patrice Genevet, and Federico Capasso. 2015. Multiwavelength achromatic metasurfaces by dispersive phase compensation. *Science* 347 (2015), 1342–1345.

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

Amir Arbabi, Yu Horie, Mahmood Bagheri, and Andrei Faraon. 2015. Dielectric metasurfaces for complete control of phase and polarization with subwavelength spatial resolution and high transmission. *Nature Nanotechnology* 10 (2015), 937–943.

Ehsan Arbabi, Amir Arbabi, Seyedeh Mahsa Kamali, Yu Horie, and Andrei Faraon. 2016. Multiwavelength polarization-insensitive lenses based on dielectric metasurfaces with meta-molecules. *Optica* 3, 6 (2016), 628–633.

Ehsan Arbabi, Amir Arbabi, Seyedeh Mahsa Kamali, Yu Horie, and Andrei Faraon. 2017. Controlling the sign of chromatic dispersion in diffractive optics with dielectric metasurfaces. *Optica* 4, 6 (Jun 2017), 625–632. https://doi.org/10.1364/OPTICA.4.000625

Ehsan Arbabi, Seyedeh Mahsa Kamali, Amir Arbabi, and Andrei Faraon. 2018. Full-Stokes imaging polarimetry using dielectric metasurfaces. *Acs Photonics* 5, 8 (2018), 3132–3140.

M. Salman Asif, Ali Ayremlou, Aswin C. Sankaranarayanan, Ashok Veeraraghavan, and Richard G. Baraniuk. 2017. FlatCam: Thin, Lensless Cameras Using Coded Aperture

and Computation. *IEEE Transactions on Computational Imaging* 3, 3 (2017), 384–397.

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

Seung-Hwan Baek and Felix Heide. 2021. Polka Lines: Learning Structured Illumination and Reconstruction for Active Stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Seung-Hwan Baek, Hayato Ikoma, Daniel S Jeon, Yuqi Li, Wolfgang Heidrich, Gordon Wetzstein, and Min H Kim. 2021. Single-shot Hyperspectral-Depth Imaging with Learned Diffractive Optics. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2651–2660.

Mario Bertero, Patrizia Boccacci, and Christine De Mol. 2021. *Introduction to inverse problems in imaging.* CRC press.

Vivek Boominathan, Jesse K Adams, Jacob T Robinson, and Ashok Veeraraghavan. 2020. Phlatcam: Designed phase-mask based thin lensless camera. *IEEE transactions on pattern analysis and machine intelligence* 42, 7 (2020), 1618–1629.

Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. 2011. Learning Photographic Global Tonal Adjustment with a Database of Input / Output Image Pairs. In *The Twenty-Fourth IEEE Conference on Computer Vision and Pattern Recognition*.

Philip Camayd-Muñoz, Conner Ballew, Gregory Roberts, and Andrei Faraon. 2020. Multifunctional volumetric meta-optics for color and polarization image sensors. *Optica* 7, 4 (2020), 280–283.

Ayan Chakrabarti. 2016. Learning sensor multiplexing design through back-propagation. *Advances in Neural Information Processing Systems* 29 (2016).

Julie Chang and Gordon Wetzstein. 2019. Deep Optics for Monocular Depth Estimation and 3D Object Detection. *ArXiv* abs/1904.08601 (2019).

Ji Chen, Xin Ye, Shenglun Gao, Yuxin Chen, Yunwei Zhao, Chunyu Huang, Kai Qiu, Shining Zhu, and Tao Li. 2022. Planar wide-angle-imaging camera enabled by metalens array. *Optica* 9, 4 (2022), 431–437.

Wei Ting Chen, Alexander Y Zhu, Vyshakh Sanjeev, Mohammadreza Khorasaninejad, Zhujun Shi, Eric Lee, and Federico Capasso. 2018. A broadband achromatic metalens for focusing and imaging in the visible. *Nature Nanotechnology* 13, 3 (2018), 220–226.

Gene Chou, Yuval Bahat, and Felix Heide. 2022. Diffusion-SDF: Conditional Generative Modeling of Signed Distance Functions. *arXiv preprint arXiv:2211.13757* (2022).

Ilya Chugunov, Seung-Hwan Baek, Qiang Fu, Wolfgang Heidrich, and Felix Heide. 2021. Mask-ToF: Learning Microlens Masks for Flying Pixel Correction in Time-of-Flight Imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9116–9126.

Shane Colburn and Arka Majumdar. 2020. Metasurface generation of paired accelerating and rotating optical beams for passive ranging and scene reconstruction. *ACS Photonics* 7, 6 (2020), 1529–1536.

Shane Colburn, Alan Zhan, and Arka Majumdar. 2018. Metasurface optics for full-color computational imaging. *Science Advances* 4 (2018), eaar2114.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 248–255.

Jacob Engelberg and Uriel Levy. 2020. The advantages of metalenses over diffractive lenses. *Nature Communications* 11 (2020), 1991.

MohammadSadegh Faraji-Dana, Ehsan Arbabi, Hyounghan Kwon, Seyedeh Mahsa Kamali, Amir Arbabi, John G Bartholomew, and Andrei Faraon. 2019. Hyperspectral imager with folded metasurface optics. *ACS Photonics* 6, 8 (2019), 2161–2167.

A. Foi, Mejdi Trimeche, V. Katkovnik, and K. Egiazarian. 2008. Practical Poissonian-Gaussian Noise Modeling and Fitting for Single-Image Raw-Data. *IEEE Transactions on Image Processing* 17 (2008), 1737–1754.

Alice Fontbonne, Hervé Sauer, and François Goudail. 2022. End-to-end optimization of optical systems with extended depth of field under wide spectrum illumination. *Applied Optics* 61, 18 (2022), 5358–5367.

Carl Friedrich Gauss. 1843. *Dioptrische Untersuchungen von CF Gauss.* in der Dieterich-schen Buchhandlung.

Gene H Golub, Per Christian Hansen, and Dianne P O'Leary. 1999. Tikhonov regularization and total least squares. *SIAM journal on matrix analysis and applications* 21, 1 (1999), 185–194.

Harel Haim, Shay Elmalem, Raja Giryes, Alex Bronstein, and Emanuel Marom. 2018. Depth Estimation From a Single Image Using Deep Learned Phase Coded Mask. *IEEE Transactions on Computational Imaging* 4 (2018), 298–310.

Lei He, Guanghui Wang, and Zhanyi Hu. 2018. Learning Depth From Single Images With Deep Neural Network Embedding Focal Length. *IEEE Transactions on Image Processing* 27 (2018), 4676–4689.

Felix Heide, Qiang Fu, Yifan Peng, and Wolfgang Heidrich. 2016. Encoded diffractive optics for full-spectrum computational imaging. *Scientific reports* 6 (2016), 33543.

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

Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems* 33 (2020), 6840–6851.

Roarke Horstmeyer, Richard Y. Chen, Barbara Kappes, and Benjamin Judkewitz. 2017. Convolutional neural networks that teach microscopes how to image. *ArXiv* abs/1709.07223 (2017).

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

Michael Kellman, Emrah Bostan, Michael Chen, and Laura Waller. 2019. Data-Driven Design for Fourier Ptychographic Microscopy. In *2019 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 1–8.

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

Mohammadreza Khorasaninejad, Zhujun Shi, Alexander Y Zhu, Wei-Ting Chen, Vyshakh Sanjeev, Aun Zaidi, and Federico Capasso. 2017. Achromatic metalens over 60 nm bandwidth in the visible and metalens with reverse chromatic dispersion. *Nano Letters* 17, 3 (2017), 1819–1824.

Oliver Kingshott, Nick Antipa, Emrah Bostan, and Kaan Akşit. 2022. Unrolled primal-dual networks for lensless cameras. *Opt. Express* 30, 26 (Dec 2022), 46324–46335. https://doi.org/10.1364/OE.475521

Grace Kuo, Nick Antipa, Ren Ng, and Laura Waller. 2017. DiffuserCam: diffuser-based lensless cameras. In *Computational Optical Sensing and Imaging*. Optical Society of America, CTu3B–2.

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

Rémi Laumont, Valentin De Bortoli, Andrés Almansa, Julie Delon, Alain Durmus, and Marcelo Pereyra. 2022. On Maximum-a-Posteriori estimation with Plug & Play priors and stochastic gradient descent. *arXiv preprint arXiv:2201.06133* (2022).

Xiu Li, Jinli Suo, Weihang Zhang, Xin Yuan, and Qionghai Dai. 2021. Universal and flexible optical aberration correction using deep-prior based deconvolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2613–2621.

Dianmin Lin, Pengyu Fan, Erez Hasman, and Mark L. Brongersma. 2014. Dielectric gradient metasurface optical elements. *Science* 345, 6194 (2014), 298–302. https://doi.org/10.1126/science.1253213

Zin Lin, Charles Roques-Carmes, Raphaël Pestourie, Marin Soljačić, Arka Majumdar, and Steven G. Johnson. 2021. End-to-end nanophotonic inverse design for imaging and polarimetry. *Nanophotonics* 10, 3 (2021), 1177–1187. https://doi.org/doi:10.1515/nanoph-2020-0579

Fanglin Linda Liu, Vaishnavi Madhavan, Nick Antipa, Grace Kuo, Saul Kato, and Laura Waller. 2019. Single-shot 3D fluorescence microscopy with Fourier DiffuserCam. In *Novel Techniques in Microscopy*. Optica Publishing Group, NS2B–3.

Leon B Lucy. 1974. An iterative technique for the rectification of observed distributions. *The astronomical journal* 79 (1974), 745.

Joseph N. Mait, Ravindra A. Athale, Joseph van der Gracht, and Gary W. Euliss. 2020. Potential Applications of Metamaterials to Computational Imaging, In Proc. Frontiers in Optics / Laser Science. *Frontiers in Optics / Laser Science*, FTu8B.1. https://doi.org/10.1364/FIO.2020.FTu8B.1

Julio Marco, Quercus Hernandez, Adolfo Muñoz, Yue Dong, Adrián Jarabo, Min H. Kim, Xin Tong, and Diego Gutierrez. 2017. DeepToF: off-the-shelf real-time correction of multipath interference in time-of-flight imaging. *ACM Trans. Graph.* 36 (2017), 219:1–219:12.

Christopher A Metzler, Hayato Ikoma, Yifan Peng, and Gordon Wetzstein. 2020. Deep optics for single-shot high-dynamic-range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1375–1385.

Kristina Monakhova, Kyrollos Yanny, Neerja Aggarwal, and Laura Waller. 2020. Spectral DiffuserCam: lensless snapshot hyperspectral imaging with a spectral filter array. *Optica* 7, 10 (2020), 1298–1307.

A. Ndao, Li-Yi Hsu, Jeongho Ha, Junhee Park, C. Chang-Hasnain, and B. Kanté. 2020. Octave bandwidth photonic fishnet-achromatic-metalens. *Nature Communications* 11 (2020), 3205.

Elias Nehme, Daniel Freedman, Racheli Gordon, Boris Ferdman, Lucien E Weiss, Onit Alalouf, Tal Naor, Reut Orange, Tomer Michaeli, and Yoav Shechtman. 2020. Deep-STORM3D: dense 3D localization microscopy and PSF design by deep learning. *Nature methods* 17, 7 (2020), 734–740.

Yifan Peng, Qiang Fu, Felix Heide, Shuochen Su, and Wolfgang Heidrich. 2015. Computational imaging using lightweight diffractive-refractive optics. *Optics Express* 23 (2015), 31393–31407.

Yifan Peng, Qiang Fu, Felix Heide, and Wolfgang Heidrich. 2016a. The Diffractive Achromat: Full Spectrum Computational Imaging with Diffractive Optics. *ACM Trans. Graph. (SIGGRAPH ASIA 2016)* 35, 4, Article 31 (July 2016), 11 pages.

Yifan Peng, Qiang Fu, Felix Heide, and Wolfgang Heidrich. 2016b. The Diffractive Achromat Full Spectrum Computational Imaging with Diffractive Optics. *ACM Transactions on Graphics (TOG)* 35 (2016), 1–11.

Yifan Peng, Qilin Sun, Xiong Dun, Gordon Wetzstein, Wolfgang Heidrich, and Felix Heide. 2019a. Learned Large Field-of-View Imaging With Thin-Plate Optics. *ACM Transactions on Graphics (SIGGRAPH Asia)* 38, 6 (2019).

Yifan Peng, Qilin Sun, Xiong Dun, Gordon Wetzstein, Wolfgang Heidrich, and Felix Heide. 2019b. Learned large field-of-view imaging with thin-plate optics. *ACM Transactions on Graphics (TOG)* 38 (2019), 1–14.

Yifan Peng, Qilin Sun, Xiong Dun, Gordon Wetzstein, Wolfgang Heidrich, and Felix Heide. 2019c. Learned large field-of-view imaging with thin-plate optics. *ACM Trans. Graph. (Proc. Siggraph Asia)* 38, 6 (2019), 219–1.

Federico Presutti and Francesco Monticone. 2020a. Focusing on bandwidth: achromatic metalens limits. *Optica* 7, 6 (2020), 624–631.

Federico Presutti and Francesco Monticone. 2020b. Focusing on bandwidth: achromatic metalens limits. *Optica* 7, 6 (Jun 2020), 624–631. https://doi.org/10.1364/OPTICA.389404

William Hadley Richardson. 1972. Bayesian-based iterative method of image restoration. *JoSA* 62, 1 (1972), 55–59.

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

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*.

C. Schuler, H. Burger, S. Harmeling, and B. Schölkopf. 2013. A Machine Learning Approach for Non-blind Image Deconvolution. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1067–1074.

Yoav Shechtman, Lucien E Weiss, Adam S. Backer, Maurice Y. Lee, and W E Moerner. 2016. Multicolour localization microscopy by point-spread-function engineering. *Nature photonics* 10 (2016), 590–594.

Zheng Shi, Yuval Bahat, Seung-Hwan Baek, Qiang Fu, Hadi Amata, Xiao Li, Praneeth Chakravarthula, Wolfgang Heidrich, and Felix Heide. 2022. Seeing through Obstructions with Diffractive Cloaking. *ACM Transactions on Graphics (SIGGRAPH)* 41, 4 (2022).

Sajan Shrestha, Adam C. Overvig, Ming Ying Lu, Aaron Stein, and Nanfang Yu. 2018. Broadband achromatic dielectric metalenses. *Light: Science & Applications* 7 (2018), 85.

Vincent Sitzmann, Steven Diamond, Yifan Peng, Xiong Dun, Stephen Boyd, Wolfgang Heidrich, Felix Heide, and Gordon Wetzstein. 2018. End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Trans. Graph. (TOG)* 37, 4 (2018), 114.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*. PMLR, 2256–2265.

Jiaming Song, Chenlin Meng, and Stefano Ermon. 2021. Denoising Diffusion Implicit Models. In *International Conference on Learning Representations*. https://openreview.net/forum?id=St1giarCHLP

David Stork and Patrick Gill. 2014. Optical, mathematical, and computational foundations of lensless ultra-miniature diffractive imagers and sensors. *International Journal on Advances in Systems and Measurements* 7, 3 (2014), 4.

Shuochen Su, Felix Heide, Gordon Wetzstein, and Wolfgang Heidrich. 2018. Deep End-to-End Time-of-Flight Imaging. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), 6383–6392.

Qilin Sun, Ethan Tseng, Qiang Fu, Wolfgang Heidrich, and Felix Heide. 2020. Learning Rank-1 Diffractive Optics for Single-shot High Dynamic Range Imaging. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Qilin Sun, Congli Wang, Fu Qiang, Dun Xiong, and Heidrich Wolfgang. 2021. End-to-end complex lens design with differentiable ray tracing. *ACM Transactions on Graphics (Proc. Siggraph)* 40, 4 (2021), 1–13.

Jun Tanida, Tomoya Kumagai, Kenji Yamada, Shigehiro Miyatake, Kouichi Ishida, Takashi Morimoto, Noriyuki Kondou, Daisuke Miyazaki, and Yoshiki Ichioka. 2001. Thin observation module by bound optics (TOMBO): concept and experimental verification. *Applied optics* 40, 11 (2001), 1806–1813.

Ethan Tseng, Shane Colburn, James Whitehead, Luocheng Huang, Seung-Hwan Baek, Arka Majumdar, and Felix Heide. 2021a. Neural Nano-Optics for High-quality Thin Lens Imaging. *Nature Communications* (December 2021).

Ethan Tseng, Ali Mosleh, Fahim Mannan, Karl St-Arnaud, Avinash Sharma, Yifan Peng, Alexander Braun, Derek Nowrouzezahrai, Jean-Francois Lalonde, and Felix Heide. 2021b. Differentiable Compound Optics and Processing Pipeline Optimization for End-to-end Camera Design. *ACM Transactions on Graphics (SIGGRAPH)* 40, 4 (2021).

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg. 2013. Plug-and-play priors for model based reconstruction. In *2013 IEEE Global Conference on Signal and Information Processing*. IEEE, 945–948.

Kartik Venkataraman, Dan Lelescu, Jacques Duparré, Andrew McMahon, Gabriel Molina, Priyam Chatterjee, Robert Mullis, and Shree Nayar. 2013. Picam: An ultra-thin high performance monolithic camera array. *ACM Trans. Graph. (TOG)* 32, 6 (2013), 166.

Shuming Wang, Pin Chieh Wu, Vin-Cent Su, Yi-Chieh Lai, Mu-Ku Chen, Hsin Yu Kuo, Bo Han Chen, Yu Han Chen, Tzu-Ting Huang, Jung-Hsi Wang, et al. 2018a. A broadband achromatic metalens in the visible. *Nature nanotechnology* 13, 3 (2018), 227–232.

Shuming Wang, Pin Chieh Wu, Vin-Cent Su, Yi-Chieh Lai, Mu-Ku Chen, Hsin Yu Kuo, Bo Han Chen, Yu Han Chen, Tzu-Ting Huang, Jung-Hsi Wang, Ray-Ming Lin, Chieh-Hsiung Kuan, Tao Li, Zhen lin Wang, Shining Zhu, and Din Ping Tsai. 2018b. A broadband achromatic metalens in the visible. *Nature Nanotechnology* 13 (2018), 227–232.

Shuming Wang, Pin Chieh Wu, Vin-Cent Su, Yi-Chieh Lai, Cheng Hung Chu, Jia-Wern Chen, Shen-Hung Lu, Ji Chen, Beibei Xu, Chieh-Hsiung Kuan, et al. 2017. Broadband achromatic optical metasurface devices. *Nature Communications* 8, 1 (2017), 187.

Alexander D. White, Parham Porsandeh Khial, Fariborz Salehi, Babak Hassibi, and Ali Hajimiri. 2020. A Silicon Photonics Computational Lensless Active-Flat-Optics Imaging System. *Scientific Reports* 10 (2020), 1689.

Norbert Wiener, Norbert Wiener, Cyberneticist Mathematician, Norbert Wiener, Norbert Wiener, and Cybernéticien Mathématicien. 1949. *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*. Vol. 113. MIT press Cambridge, MA.

Yicheng Wu, Vivek Boominathan, Huaijin Chen, Aswin Sankaranarayanan, and Ashok Veeraraghavan. 2019. PhaseCam3D — Learning Phase Masks for Passive Single View Depth Estimation. *2019 IEEE International Conference on Computational Photography (ICCP)* (2019), 1–12.

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

Nanfang Yu and Federico Capasso. 2014. Flat optics with designer metasurfaces. *Nature Materials* 13 (2014), 139–150.

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

Xuaner Zhang, Qifeng Chen, Ren Ng, and Vladlen Koltun. 2019. Zoom to learn, learn to zoom. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3762–3770.

Guoxing Zheng, Holger Mühlenbernd, Mitchell Kenney, Guixin Li, Thomas Zentgraf, and Shuang Zhang. 2015. Metasurface holograms reaching 80% efficiency. *Nature Nanotechnology* 10 (2015), 308–312.

You Zhou, Hanyu Zheng, Ivan I Kravchenko, and Jason Valentine. 2020. Flat optics for image differentiation. *Nature Photonics* 14, 5 (2020), 316–323.