Supplementary Information:

Beating bandwidth limits for large aperture broadband nano-optics

Johannes E. Fröch[1,2,,‖], Praneeth K. Chakravarthula[3,‖], Jipeng Sun[3], Ethan Tseng[3], Shane Colburn[2,4], Alan Zhan[4], Forrest Miller[2], Anna Wirth-Singh[1], Quentin A.A. Tanguy[2], Zheyi Han[2], Karl F. Böhringer[2,5,6], Felix Heide[3], Arka Majumdar[1,2,*]*

1: Department of Physics, University of Washington, Seattle, 98195, WA, USA

2: Department of Electrical and Computer Engineering, University of Washington, Seattle, 98195, WA, USA

3: Department of Computer Science, Princeton University, Princeton, NJ, 08544, USA

4: Tunoptix, 4000 Mason Road 300, Fluke Hall, Seattle, WA, 98195 USA

5: Department of Bioengineering, University of Washington, Seattle, WA, 98195, USA

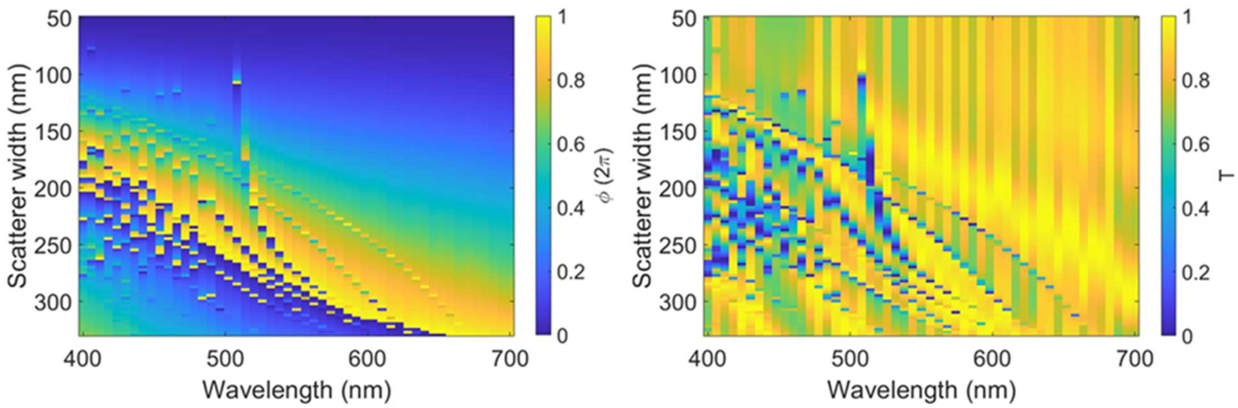
6: Institute for Nano-Engineered Systems, University of Washington, Seattle, WA, 98195, USA

[‖]: equal contribution

email: jfroech@uw.edu, arka@uw.edu

Supplementary Note 1. Details on meta-optic design

For all meta-optics designed in this work we used a square SiN post scatterer of ~ 800 nm height with a period of 350 nm. Phase delay and transmission as function of scatterer size were calculated using rigorous coupled wave analysis (RCWA), using S4.(*1*) The phase and transmission response of the scatterer is shown in Figure S1. For the meta-optics we considered only scatterer which achieved high transmission of > 90 % over a larger spectral range, while avoiding resonances.



Supplementary Figure 1*. Phase and Transmission response of scatterers for the wavelength range 400 nm – 700 nm.*

EDOF design

The metalens design of the main text was first based on an optimization of a symmetric phase profile using a dense sampling of wavelengths. In restricting the 1 centimeter diameter to a radially symmetric function, we dramatically reduce the memory requirements to simulate the design. Additionally, instead of computing the full point spread function for every iteration, we instead only compute the intensity at the focal spot, which acts as a useful proxy to concentrate power within a confined spatial location, and which we find is sufficient for enhancing MTF while mitigating computational requirements for the optimization. Our implementation uses the Rayleigh-Sommerfeld diffraction integral, exploiting the radial symmetry of the lens and the fact that the observation points in the integral are located only at the desired focal spot. This enables us to optimize for a dense sampling of 2000 wavelengths simultaneously without exceeding the memory requirements of our workstation, which used a V100 GPU.

To perform the optimization, we implemented the Rayleigh-Sommerfeld diffraction integral below in TensorFlow

$$E(x,y,z) = \iint_{-\infty}^{+\infty} E(x',y',0)\frac{e^{ikr}}{r}\frac{z}{r}\left(1+\frac{i}{kr}\right)dx'dy'$$

where the primed variables denote the source field coordinates (i.e., positions within the metasurface aperture), and the unprimed variables represent the positions of the observation points or on the destination plane. r is defined in the typical manner as below

$$r = \sqrt{(x - x')^2 + (y - y')^2 + z^2}$$

In our case, as $x = 0$ and $y = 0$ for the on axis focal spot, and as the source field is radially symmetric, we can simplify r as below:

$$r = \sqrt{r'^2 + z^2}$$

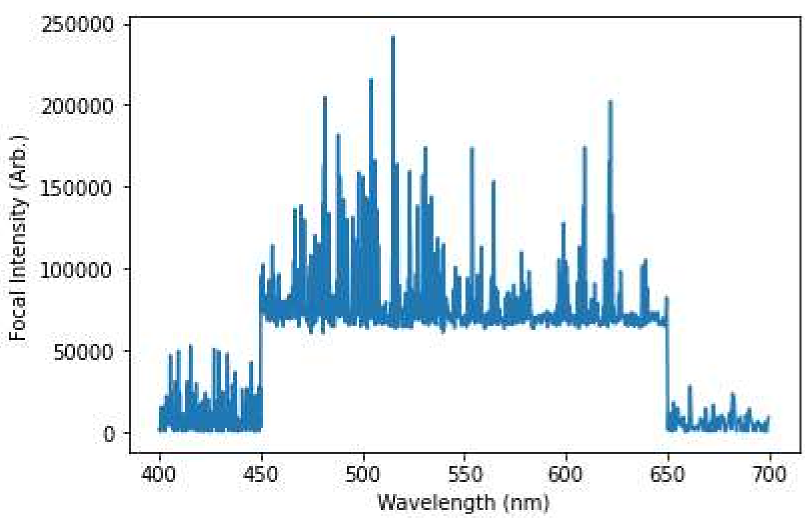
and rewrite the integral as

$$E_z = 2\pi \int_0^R E(r') \frac{e^{ikr}}{r} \frac{z}{r} \left(1 + \frac{i}{kr}\right) r' dr'$$

where E_z denotes the electric field on axis at a distance z from the origin along the z axis. Here, the integral becomes computationally far simpler as the 2-D integral is now a 1-D integral. We implement this in TensorFlow to enable gradient calculation via automatic differentiation, employing the Adam algorithm with a learning rate of 0.005. Our loss function is specified as

$$Loss = -\min_{\mathrm{i}} |E_i|^2$$

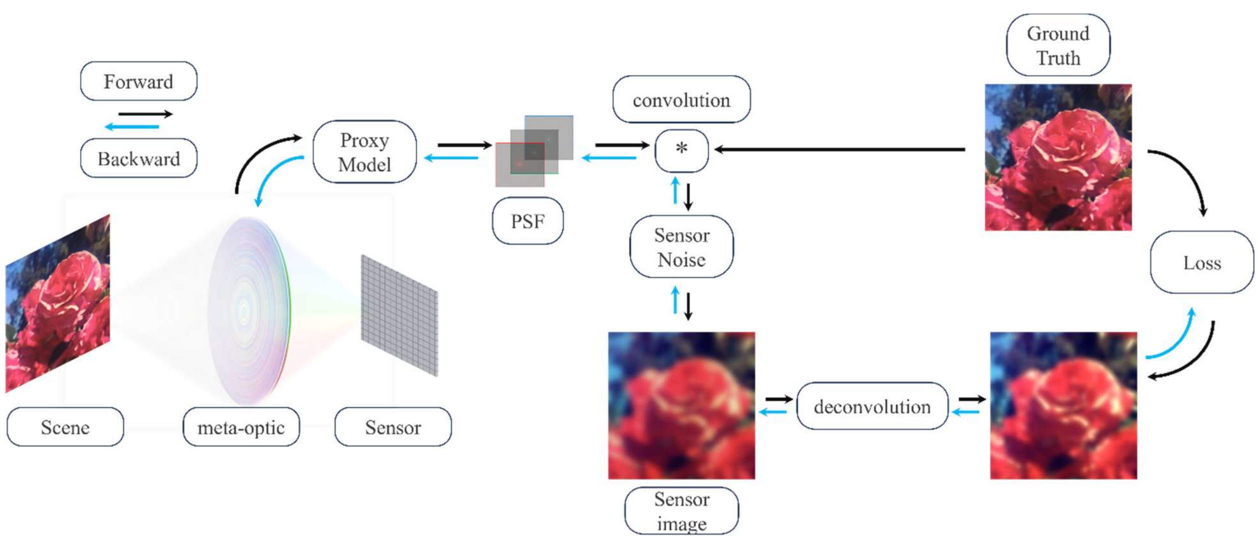
where E_i denotes the electric field at the focal point for the ith wavelength in the simulation. In our case, we optimized over 2000 wavelengths sampled between 450 nm and 650 nm. The loss function here serves to enhance the focal intensity of the least intense wavelength, having the effect of reducing the worst case performance across the wavelength band. The focal intensity in arbitrary intensity units after optimization of the 1 centimeter lens is shown in the Figure below.



Supplementary Figure 2: *The optimized focal intensity as a function of wavelength is shown. While there are several intensity peaks scattered across the 450-650 nm band, all the* *wavelengths in that range achieve a minimum threshold intensity as a result of the defined loss function.*

End-to-end designs

We require the end-to-end designed meta-optic to achieve the desired phase modulation at all visible wavelengths to design a broadband imaging lens. To this end, we build on the radially symmetric EDOF model described earlier and model the light propagation through metalenses with silicon nitride rectangular nanopillars and optimize the duty cycle (i.e., the width) of the nano-antennas. In a local neighborhood of these nano-antennas, we 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. We characterize metalenses with their local phase, which we tie to the structure parameters, i.e., the duty cycle, via a differentiable proxy model mapping the nanopillar structures to the resultant phase modulation. 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. The end-to-end design pipeline is illustrated in Figure S3.



Supplementary Figure 3. End-to-end design pipeline.

<u>*RCWA proxy for mapping phase and nano-scatterers*</u>

We first describe the scatterer geometry with the duty cycle of nano-antennas and analyze its modulation properties using rigorous coupled-wave analysis (RCWA). To achieve a differentiable mapping from phase to duty cycle, the phase as a function of duty cycle of the nano-antennas must be injective. Therefore, we fit the phase data of the metalens at a nominal design wavelength of 452nm to a polynomial proxy function of the form:

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

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 and the parameters a_i are fitted based on the RCWA analysis.

During the iterative optimization, we first apply the above phase-to-scatterer inverse mapping to determine the required duty cycle of the physical structure. Once the scatterer geometry is determined at the nominal wavelength, we then compute the resulting phase from the given scatterer geometry for other wavelengths using a second proxy function that maps scatterer geometry to phase. 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 to a radially symmetric second-order polynomial function of the form:

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

where λ is a non-nominal wavelength. Specifically, we compute the transmission coefficient data using RCWA and then fit the polynomial to the underlying RCWA-computed transmission coefficient data using linear least squares.

With the metalens phase and the inverse (phase to duty cycle) and forward (duty cycle to phase) mapping operators, we compute the phase modulation for broadband incident light. Using a fast Fourier transform (FFT) based band-limited angular spectrum method (ASM), we calculate the PSFs produced by the metalens as a function of wavelength to model full-color image formation. The PSF produced by the metalens for an incident beam of wavelength λ is computed as

$$\mathrm{PSF}_{\lambda} = f_{ASM}(\phi(r), \lambda, C_{meta})$$

where $\phi(r)$ 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 mapping proxy function coefficients, and $f_{ASM}(.)$ is the angular spectrum method implemented as a differentiable propagation function that generates the PSF for a given metasurface phase. Finally, the RGB image on the sensor plane is computed as

$$S = I \otimes PSF + \eta_{\mathrm{sensor}}$$

where $\otimes$ is a convolution operator, $\boldsymbol{I}$ is the groundtruth RGB image, and η_{sensor} is the sensor noise modeled as a per-pixel Gaussian-Poisson noise. Note that, for an input image $x \in [0,1]$ at a sensor pixel location, the measured noisy image on the sensor $f_{sensor}(x)$ is given by:

$$f_{sensor}(x) = \eta_g(0, \sigma_g) + \eta_p(x, a_p)$$

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.

With a measurement S as input, we recover the underlying image as

$$\tilde{I} = f_{deconv}(S, PSF, C_{deconv})$$

where C_{deconv} are the fixed parameters of our deconvolution method. To make the lens design process efficient both in terms of memory and compute, we employ a Wiener inverse filtering method in the design phase which is computed in one step and does not require any training like in neural network based methods.(*2*) However, note that after the meta-optic is designed and housed in the camera, we employ a computational image recovery backend for reconstructing high-fidelity images from the sensor measurements.

With the above synthetic metalens image formation model, we apply first-order stochastic gradient optimization to optimize for the metalens phases that minimize the error between the ground truth and recovered images. Specifically, we minimize the per-pixel mean squared error and maximize the perceptual image quality between the target image I and the recovered image $\tilde{I}$ as follows:

$$\widetilde{\phi}(r) = \underset{\phi}{\operatorname{argmin}} \sum_{i=1}^{T} \sum_{\lambda} \mathcal{L}\left(\widetilde{\boldsymbol{I}_{\lambda}^{(i)}}, \boldsymbol{I}_{\lambda}^{(i)}\right)$$

where T is the total number of training image samples used for the metalens phase optimization and $\mathcal{L}$ is the loss function used for the optimization given by

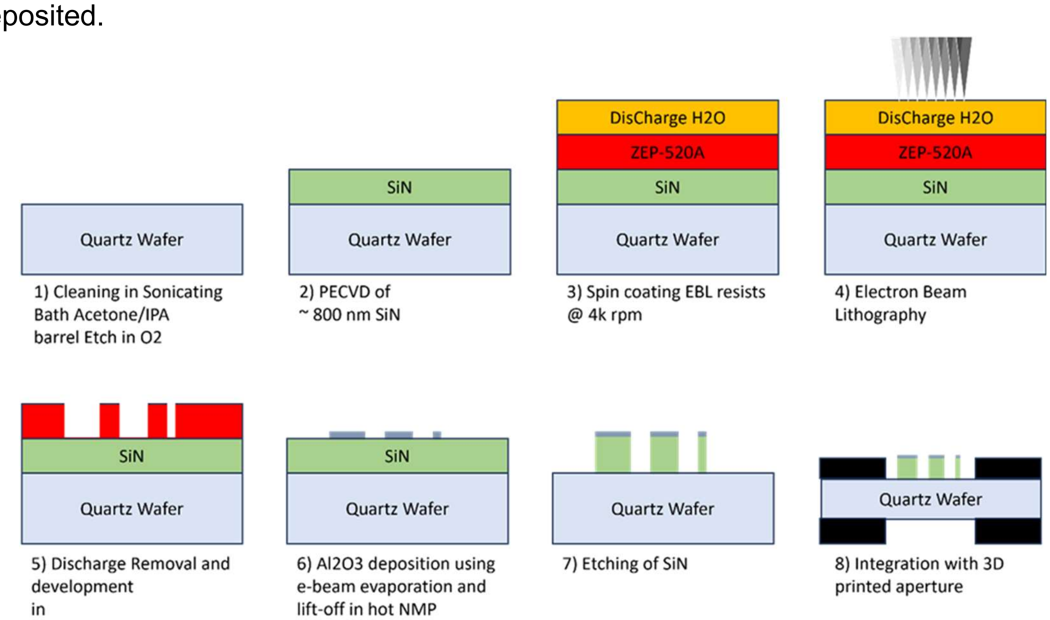
$$\mathcal{L} = \mathcal{L}_{MSE} + \mathcal{L}_{LPIPS}$$

which is a combination of per-pixel mean-squared error and learned perceptual image patch similarity (PIPS) metric(*3*).

We used the Adam optimizer with a learning rate of 0.001 running for 2 days to optimize for the meta-optic phase. Our optimizer was initialized with the EDOF metalens phase described in the previous section. Given the large compute overhead for simulating the meta-optic of 1cm diameter, instead of simulating the responses for the entire broadband spectrum at once for every iteration similar to the EDOF design approach, we sampled three wavelengths randomly from a pre-computed set of wavelengths discretized in intervals of 50 nm over the visible range for every 100 iterations.

Supplementary Note 2. Meta-optic fabrication

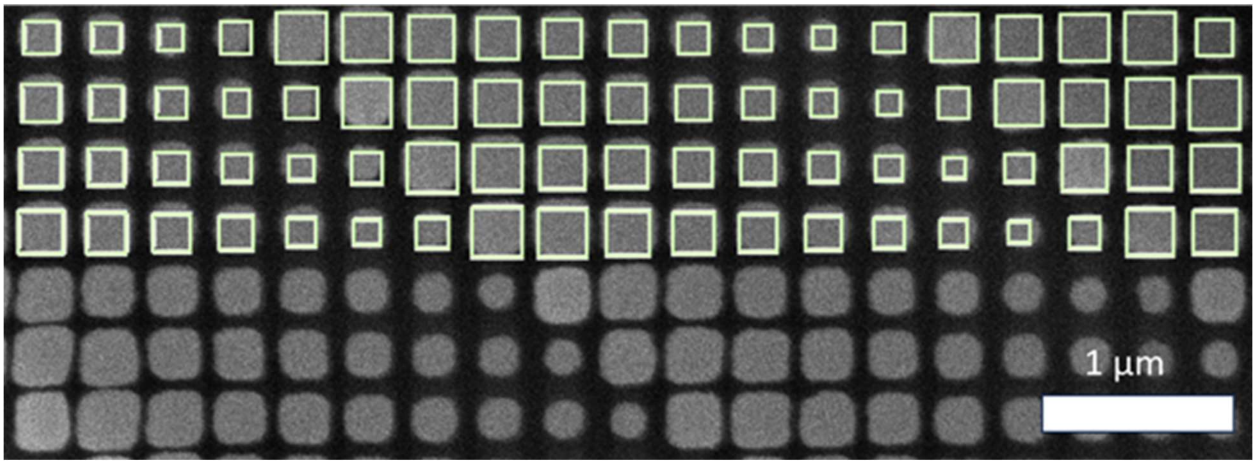
The fabrication process is schematically illustrated in Figure S4. All fabrications were completed in a clean room environment (Washington Nanofabrication Facility, ISO Class 5-7). (1) Quartz carrier wafers (with thickness of ~ 300 um) were purchased from University Wafer and cleaned in Acetone and IPA, as well as a short oxygen etching treatment in an Oxygen Barrel Etcher. (2) Then a ~ 800 nm thick SiN film was deposited on top of the wafer using plasma enhanced chemical vapor deposition (PECVD) in a SPTS PECVD chamber, with a mixture of Silane and Ammonia as the deposition precursors. After deposition the wafer was diced into 1.5 cm square pieces using a Disco Saw Dicer DAD123. (3) After brief cleaning (in Acetone and IPA in an ultrasonicating bath) and barrel etch step (O2, 100 W, 15s), a positive resist (ZEP 520 A) was spun onto the sample (4k rpm, thickness of ~ 400 nm), followed by baking at 180 °C for 3 min on a hot plate. A conductive polymer layer (DisCharge H2O) was subsequently spun on top at 4k rpm. (4) The resist layer was then patterned using a 8 nA, 100 keV electron beam (JEOL JBX6300FS) at a dose of ~ 300 μC cm−2. The writing time was about 4 1/2 hours. (5) After EBL, the conductive polymer layer was removed in a short IPA bath and subsequently the resist was developed at room temperature in Amyl Acetate for 2 min. Subsequently, the sample was descummed in a short barrel etch step (100 W, 15s). (6) Then using electron beam evaporation, a layer of ~ 75 nm AlOx was deposited. The mask was then lifted off overnight in an NMP bath at ~ 100 C on a hot plate. (7) Subsequently, the SiN layer was etched using a mixture of C4F8/SF6 in an inductively coupled reactive ion etcher (Oxford PlasmaLab System 100). We note that most of the AlOx layer is consumed during the process and only a negligible amount is left on the pillar. The remaining AlOx layer was not removed after etching. (8) Finally, the chip was integrated in a 3D printed holder and mounted with the sensor. For SEM imaging a thin conductive Au/Pd layer was deposited.



***Supplementary Figure 4**. Fabrication flow of the meta-optics.*

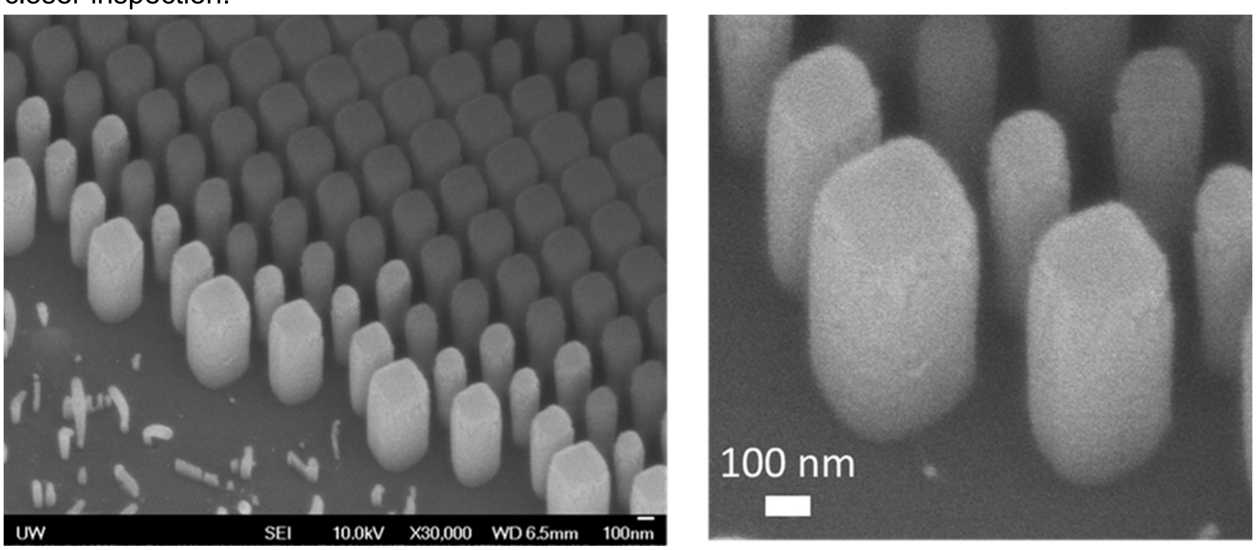
Supplementary Note 3. Morphological characterization of meta-optics

To characterize the meta-optics on the micro/nano scale we used a JEOL-JSM7400F Scanning Electron Microscope. To mitigate charging the sample was coated with a Au/Pd film. An image of the meta-optic compared to the intended GDS layout (Figure S5) highlights the accurate fabrication of the device on the nanometer scale with only minor deviations from the intended design.



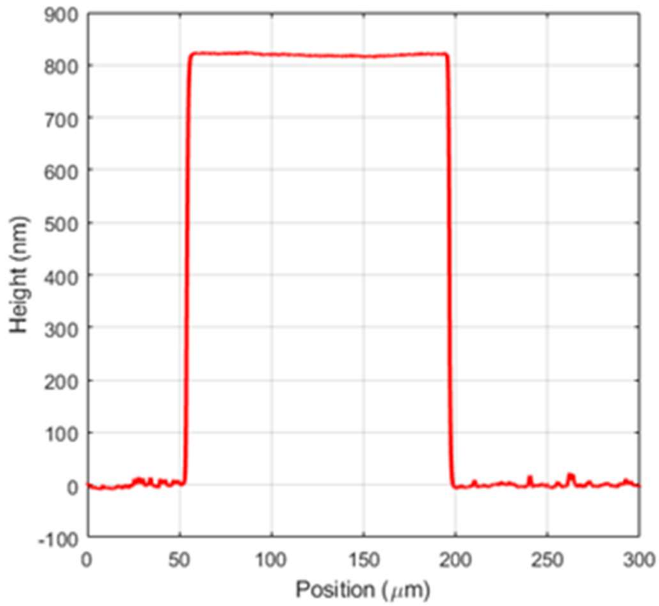
Supplementary Figure 5*. SEM image of the fabricated device, compared to the designed structure outline (overlayed in green).*

Further images (Fig. S6) from an oblique view show that the scatterer maintain a close to uniform footprint throughout their height with vertical sidewalls. We note that due to the processing conditions a certain edge roughness on the top remains, which however can only be seen at closer inspection.



Supplementary Figure 6*. SEM images at oblique view from the edge of the meta-optic aperture. A zoomed in image on the right, show some edge roughness close to the top, and otherwise smooth sidewalls of the individual scatterer.*

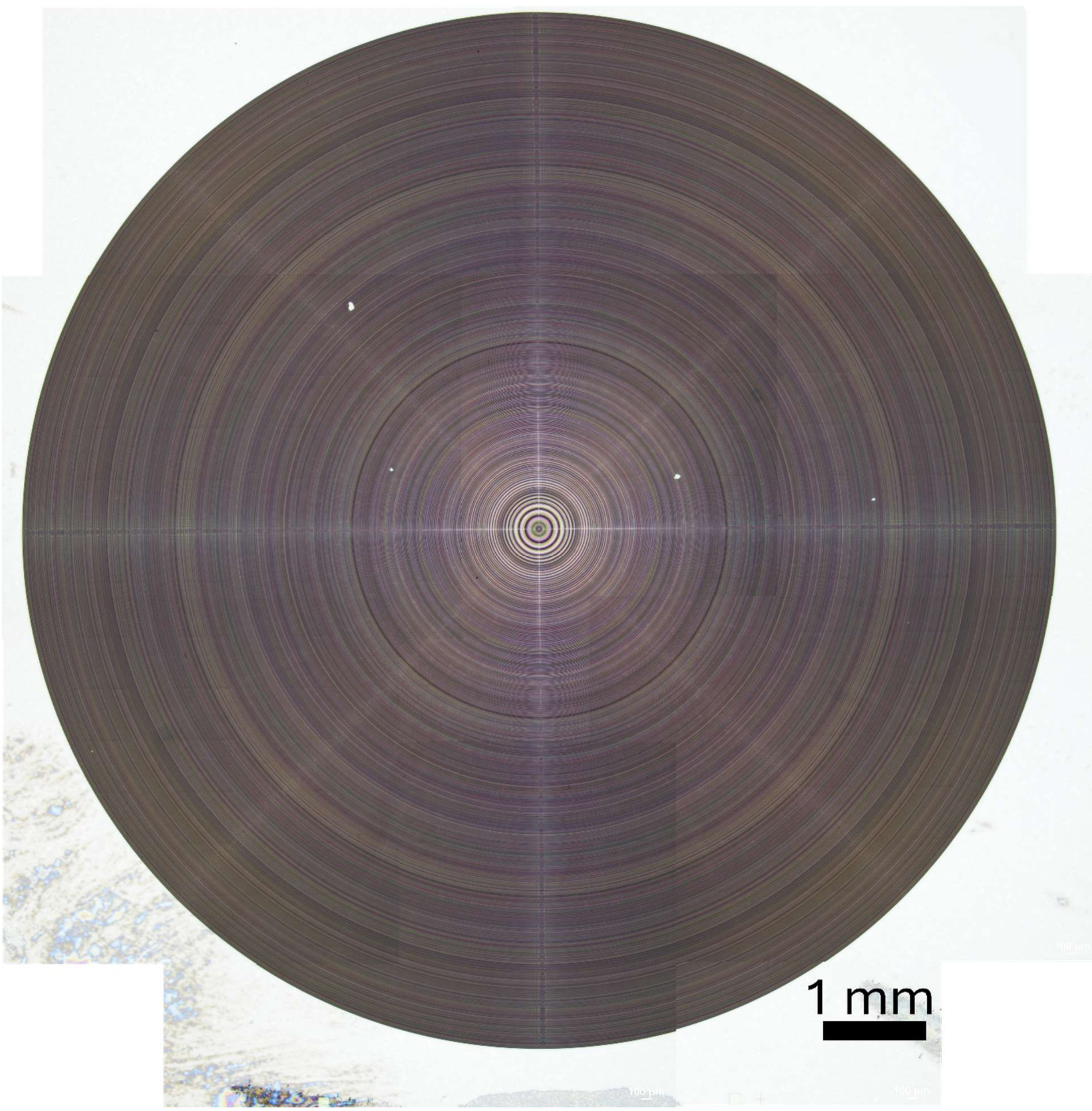
We further verified the height of the meta-optic using a profilometer (Fig. S7), which shows an approximate height of the structure of ~ 815 nm, measured at a reference marker. This value is close to the design height, the difference of ~15 nm is attributed to the residual AlOx mask.



Supplementary Figure 7. *Height profile of the device after etching on a reference marker area.*

Supplementary Note 4. Fabrication errors and tolerance of end-to-end designed meta-optics

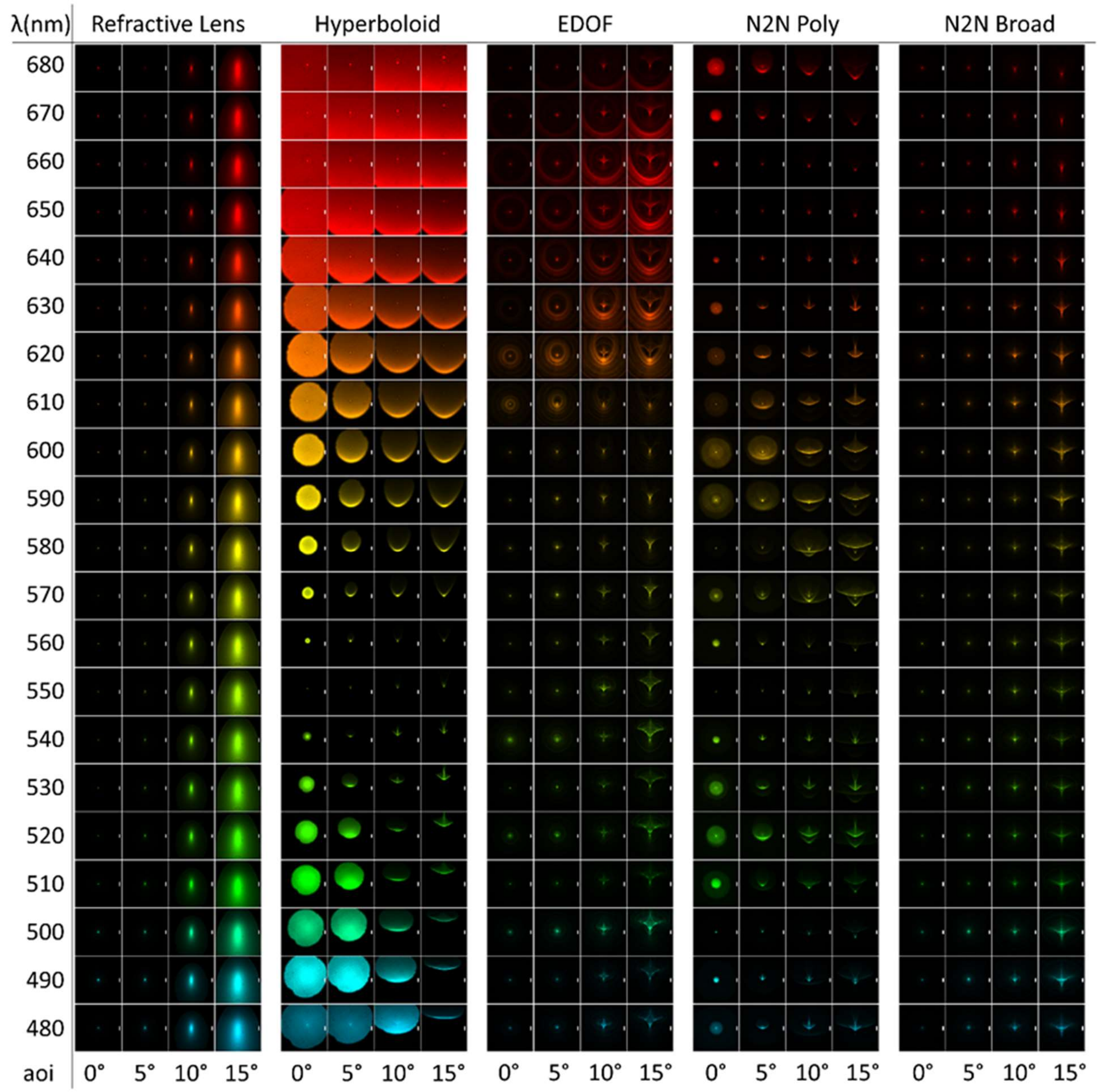
The fabrication accuracy of the 1cm large meta-optic is evident by several indicators. First, comparative SEM images overlayed with the intended structure outline show a very close overlap (Figure S5). Second, in microscope images we observed a uniform structural color across the same radial sectors for the entire aperture, illustrated in Figure S8. Third, the PSFs are radially uniform as shown in Figure S9, where the PSF intensity for RGB is plotted as function of radius with the mean square error overlayed.



***Supplementary Figure 8**: Stitched image of the entire aperture, showing uniform structural color.*

Supplementary Note 5. Meta-optic point spread function characterization

We characterized the point spread function as described in the Methods section. All PSF captures of the various meta-optic types are summarized in Figure S9, as function of wavelength and angle of incidence. It can be clearly seen that the hyperboloid only produces a narrow PSF for a small spectral range around the design wavelength of ~ 550 nm. The EDOF meta-optic exhibits a PSF which is more confined, yet certain wavelengths, such as ~ 620 nm or 540 nm, extend more. The poly chromatic end-to-end design is characterized by a very confined PSF for specific wavelengths (500 nm, 550nm, 580 nm, 650 nm), with a larger extends for in between wavelengths. The broadband design clearly shows the most balanced PSF, achieving a performance closest to the refractive lens in comparison.

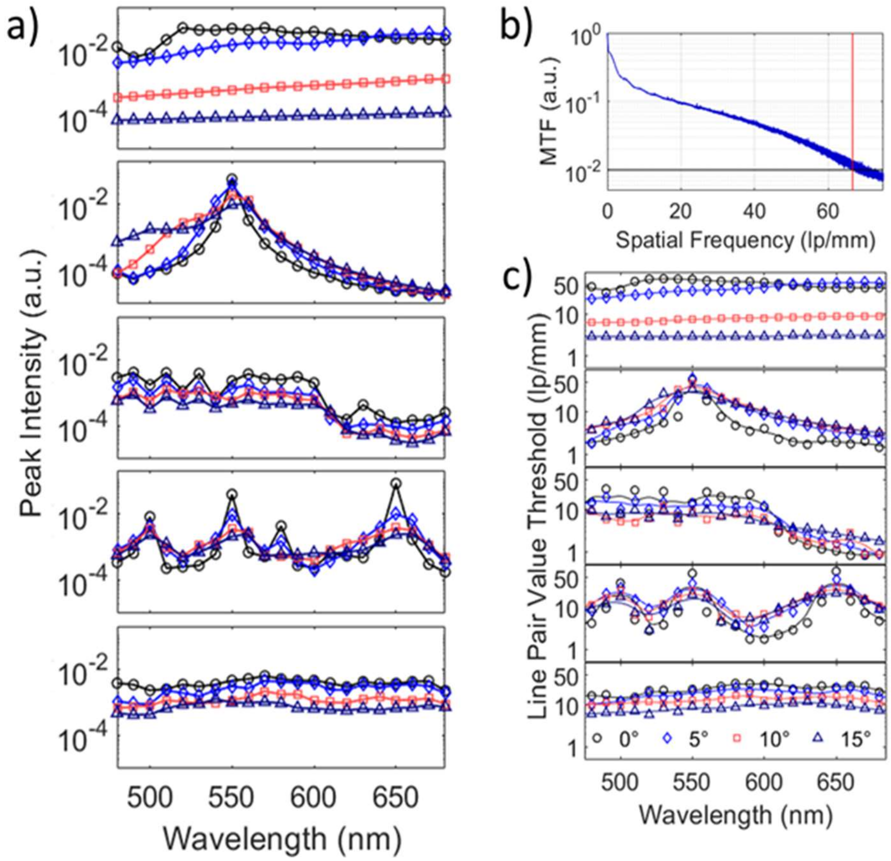


***Supplementary Figure 9**. PSF captures on sensor for different optics, as described in Figure 2a.*

To better compare the broadband performance of all lenses, we plotted the peak value of the center pixel normalized with respect to the total counts within a circular region of ~ 5 mm diameter around the center on the sensor. This allows us to qualitatively compare the focusing efficiency, as it considers the spatial extension of the PSF. Light that passes through the meta-optic directly without modification (i.e., unscattered light), decreases the contribution to the center pixel, and a tail appears or additional haze. As shown in Figure S10, we observe that overall, the refractive lens clearly performs best for small aoi, while it significantly degrades with increasing aoi. As is well known, a hyperboloid metalens achieves on-par performance with the refractive lens only for a narrow band of about ~ 10 nm. In contrast, meta-optics that are designed for broad wavelength range, show lower maximum peak intensity values, which however do not degrade as fast as the hyperboloid metalens. Even more, for larger aoi, we observe that the end-to-end designed meta-optics outperform the refractive lens.

To further compare the developed meta-optics, we used the modular transfer function (MTF), obtained as the absolute value of the Fourier transform of the PSF. The MTF curve depicts the achievable contrast value for a particular spatial frequency. However, from a system level perspective the MTF curve of the optic does not directly consider the capability of a computational backend to recover the image quality. Specifically, through applying a deconvolution step or more complex computational reconstruction methods, the image quality is significantly enhanced, thus augmenting the shortcomings of the optics and allowing one to circumvent the physical limitations in actual applications. To assess the suitability of the optics for a computational backend we therefore consider the line-pair value as the MTF decreases below a specific threshold value of 0.01.

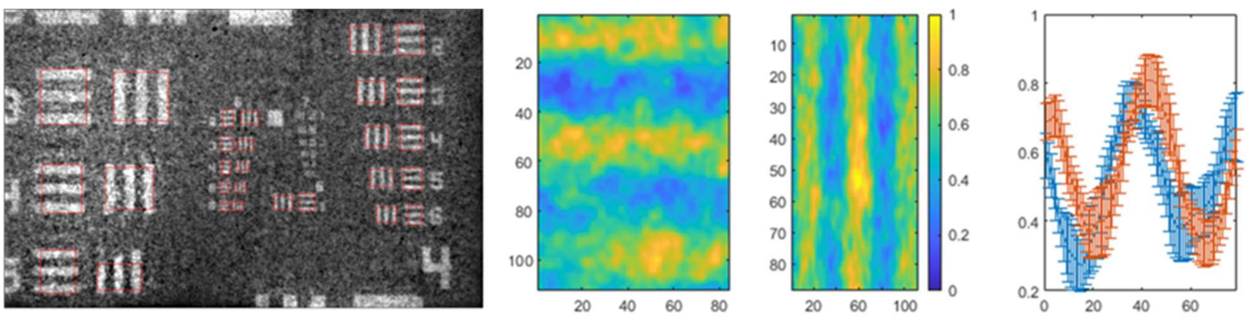
This value as a function of the wavelength and angle of incidence is plotted in Figure S10c for all considered optics. The refractive lens yields the highest performance throughout the spectral range for small aoi of 0° and 5°, however degrades quickly towards larger aoi of 10° and 15°. The hyperboloid lens exhibits excellent performance for a small spectral range but degrades outside that specific range. In comparison the computationally optimized EDOF design, provides good performance for a limited spectral range from 480 nm – 600 nm. The polychromatic end-to-end design achieves high performance for specific wavelengths but degrades outside. In comparison, the broadband end-to-end meta-optic provides mostly uniform performance across the broadband range, and further does not degrade as strongly for larger aoi. As shown later, this ensures uniform imaging capability across the entire field of view and color range.



Supplementary Figure 10. *a) Peak Intensity to integrated signal ratio, extracted for a wavelength range 480 nm – 680 nm. From top to bottom, we compare the refractive lens, the Hyperboloid metalens, the EDOF meta-optic, the polychromatic meta-optic, and the broadband meta-optic. b) example of finding the line-pair threshold for computational reconstruction. c) Plot of the Line-Pair value threshold for the different optics, listed in the same order as (a).*

Supplementary Note 6. System MTF measurement

As described, we directly determined the MTF line pair contrast by measuring the USAF 1951 target, using a broadband supercontinuum source (NKT FIR20) and tunable filter (NKT SUPERK SELECT) to select a specific wavelength for illumination of the target. For this measurement, we synchronized the wavelength filter sweep with the frame capture to obtain images of the target, and then extracted different line pair sections from the captures via an automated script. As shown in the Figure S11, different sections of the target were selected and averaged over the horizontal or vertical direction to obtain a mean intensity value along the line pair. The contrast was then calculated as the relative value of difference to sum of the max and min values. We note that the values plotted in Figure 2 of the manuscript were calculated as the mean value of horizontal and vertical line pairs. We captured the color image, while the white balance was set to zero, no AWB, no gain applied, and no sharpening.



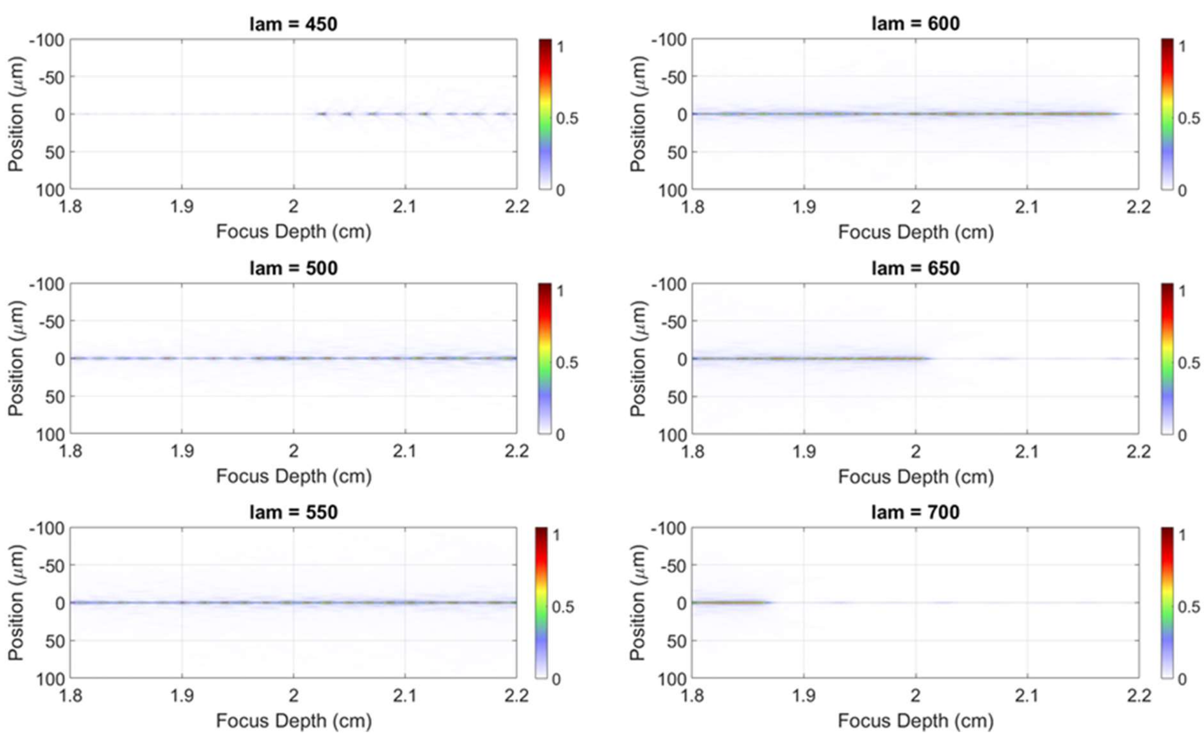
Supplementary Figure 11. *Example line pair value extraction from a captured frame of the USAF target. Areas that were used for the line pair contrast calculation are highlighted by red boxes. Also shown are the extracted average intensity across the line pair.*

Supplementary Note 7. Focusing profile of the meta-optics

As described in the main text, the broadband capability of the meta-optics arises from the extended depth of focus. To underpin this aspect, we measured the PSF of the broadband end-to-end design using a translational microscope setup, which enabled us to reconstruct the intensity profile along the optical axis for wavelengths of 450 nm to 700 nm in steps of 50 nm. Specifically, we used a 20 X Nikon Objective, mounted together with a Tube lens (Thorlabs TTL 180A) and a Sensor (Allied Vision ProSilica 1930 GT) on a programmable translation stage (Newport ILS100CC, Newport ESP301) with micrometer resolution.

The meta-optic was placed on an independent 3D translation stage, allowing alignment with respect to the optical axis. A collimated laser was transmitted through the meta-optic and the resulting image was captured with the microscope setup. After identifying the sample surface of the meta-optic, an automated script was used to move the stage by 7.5 μm at a time and then capture an image of the microscope setup.

After measurement, a further script was used to evaluate the data, cropping the image to a width of ~ 200 μm x 200 μm and setting the maximum as the center.

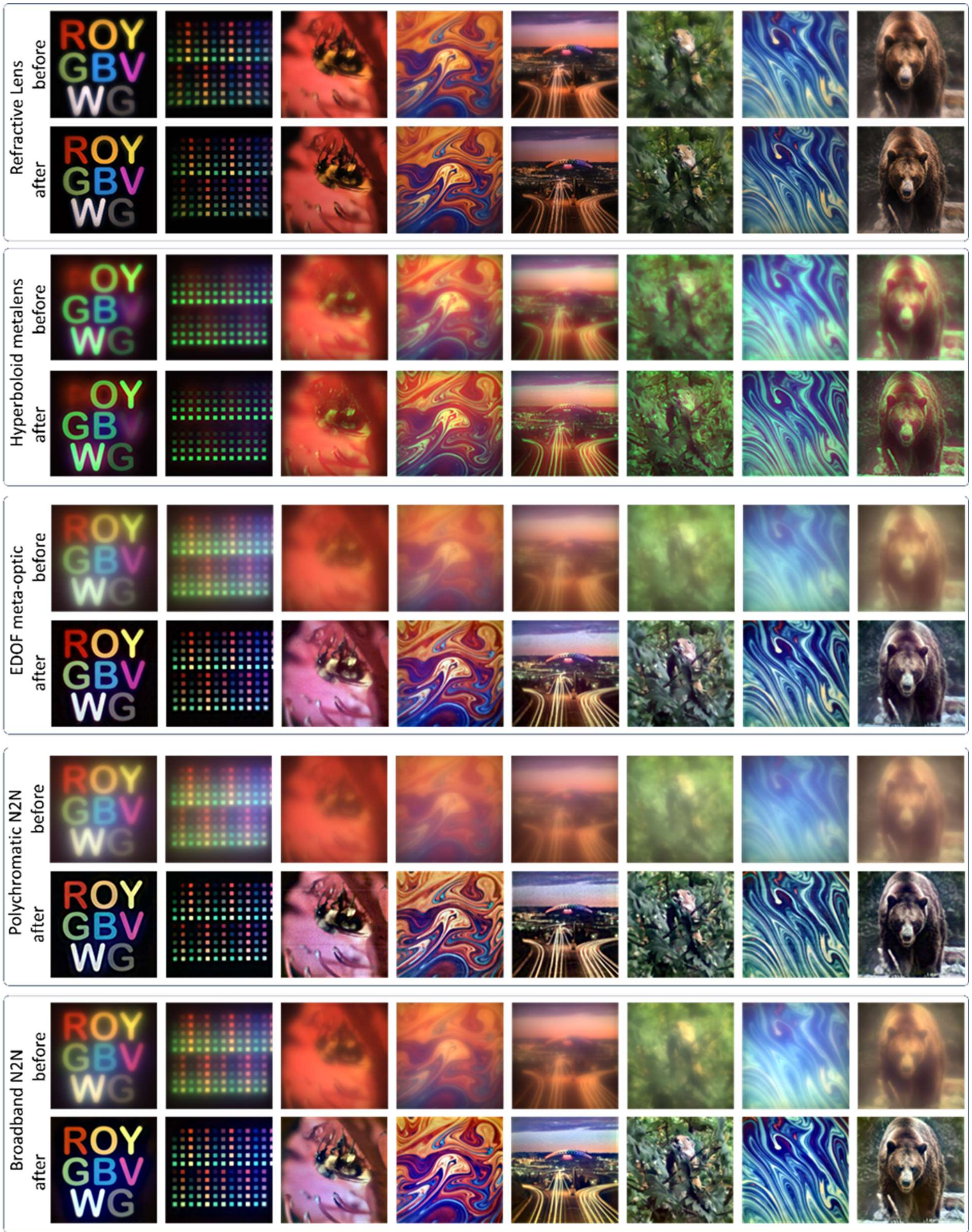


Supplementary Figure 12 *Focusing profile of the broad band meta-optic along the optical axis for different wavelengths (as given in the subplot titles).*

Supplementary Note 8. Further images of display captures

We have collated further image comparisons from the different optics in this section. Figure S13 shows images before and after Wiener Filtering for the different optics, including the refractive lens, the hyperboloid metalens, the EDOF meta-optic, the polychromatic end-to-end design, and the broadband end-to-end design. Specifically, we observe that in the captured images the refractive lens exhibits the clearest image. In comparison all images captured by different meta-optic types exhibit some degree of haze, due to an extended PSF. However, the broadband meta-optic is closest to the refractive lens with the clearest image, after image capture. Stemming from the more confined PSF across the spectral range, all wavelengths are equally focused on the sensor. In comparison images captured by the EDOF design and the polychromatic meta-optic retain a stronger haze in their respective images, due to the broader extension and unbalanced PSFs. Especially, as for the Wiener deconvolution, only one general PSF for each color channel can be considered, a more balanced PSF will ultimately yield more accurate computational reconstruction. Although this could be alleviated using some color filters, the total power on the sensor would then decrease.

Moreover, we can observe that some images captured via hyperboloid metalens appear to yield reconstructions close to the ground truth. However, the comparison in this Figure also shows that this is very scene specific and limited. For instance, images shown in the first two columns of different isolated colors clearly show that the color information is lost. Due to the limited dynamic range of the sensor, the more spread out and unfocused color information contained in the unfocused color channels, i.e., red and green channels are lost. In comparison the broadband meta-optic ensures that all colors are focused in a balanced manner.



450
451 ***Supplementary Figure 13****. Images captured from an OLED monitor for different types of optics.*
452 *Shown are images before and after computation.*
453

Supplementary Note 9. Probabilistic Diffusion Image Reconstruction

In this section, we describe how we recover the images from the metalens camera measurements using learned reconstruction. We formulate the image recovery as a model-based inverse optimization problem with a probabilistic sampling stage that samples a learned image prior. For learning the image prior, we use a probabilistic diffusion model that samples a distribution of plausible latent images for a given sensor measurement. Finally, the image reconstruction optimization problem is solved via splitting and unrolling the objective into a differentiable truncated solver.

To recover a latent image I from the sensor measurement S that relies on the physical forward model described above, we pose the deconvolution problem as a Bayesian estimation problem. In specific, we solve the following maximum-a-posteriori estimation problem with an abstract natural image prior $\Gamma(I)$:

$$\tilde{I} = argmin_I \underbrace{\frac{1}{2}||I \otimes k - S||^2}_{\text{Data Fidelity}} + \underbrace{\rho\Gamma(I)}_{\text{Prior Regularization}},$$

where $\rho > 0$ is a prior hyperparameter. The probabilistic natural image prior, in our case, allows for sampling the posterior of all plausible natural image priors instead of solving for a singular maximum of the posterior as a point estimate.

To solve the above equation, we split the non-linear and non-convex prior term from the linear data fidelity term via half-quadratic splitting to result in two simpler subproblems. To this end, we introduce an auxiliary variable *z* and pose the above minimization problem as

$$argmin_I \frac{1}{2}||I \otimes k - S||^2 + \rho\Gamma(z), \quad s.t.\ z = I.$$

which can be further reformulated as

$$argmin_{I,z} \frac{1}{2}\left|\left|I \otimes k - S\right|\right|^2 + \rho\Gamma(z) + \frac{\mu}{2}\left|\left|z - I\right|\right|^2, \mu \to \infty$$

where $\mu > 0$ is a penalty parameter, that $\mu \to \infty$ mandates equality $I = z$. We then relax μ and solve the problem iteratively by alternating between the following two steps,

$$I^{t+1} = argmin_I \frac{1}{2}||I \otimes k - S||^2 + \frac{\mu^t}{2}||I - z^t||^2,$$

$$z^{t+1} = argmin_z \frac{\mu^t}{2}||z - I^{t+1}||^2 + \rho\Gamma(z).$$

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 I given fixed values of z from the previous iteration and vice-versa.

Note that the first update from the above equations is a quadratic term that corresponds to the data fidelity term of the original objective. Assuming a circular convolution, this update can be solved in closed form with the following inverse filter update

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

where $\mathcal{F}(\ )$ is the Fourier transform, $\mathcal{F}^{*}$ is the complex conjugate of the Fourier transform, and $\mathcal{F}^{\dagger}$ is the inverse Fourier transform. The second update, however, includes the abstract image prior regularizer, which in general is non-linear and non-convex. The solution to this update is learned via a diffusion model that allows us to probabilistically sample the solution from a distribution Ω that is conditioned on the iterate I^{t+1} and the optimization penalty weights ρ, μ as inputs, the details of which we describe next.

Probabilistic Image Prior

We learn a diffusion model-based probabilistic image prior over a distribution Ω to handle the ambiguity in the deconvolution, wherein multiple clear latent images can be projected to the same noisy sensor measurement. Diffusion provided a probabilistic sampling approach to generate multiple samples, from which we can select the most suitable one. The forward process of diffusion entails progressively adding noise to a clean image, and learning to recover the underlying clean image from the noisy images. Our input x_0 to the diffusion model is a clean ground truth image I^{gt} with condition c defined as

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

where I^{gt} is the ground truth latent image, S is the sensor measurement, z^t is the auxiliary image coupling term and μ^t is the update weight term used for half-quadratic splitting (HQS), and $\gamma(T)$ is a positional encoding of T where $T \in [1,1000]$ is randomly sampled for each training iteration of the diffusion model. The symbol $\oplus$ denotes concatenation, as we condition the inputs by concatenating them along the channel dimension and employ self-attention to learn corresponding features.

The underlying neural network architecture of our diffusion model is a U-Net, and in each iteration while training our diffusion model, we add Gaussian noise to the clean image I^{gt} of the input x_0, proportional to T, to obtain x_t. The diffusion model is trained to recover a plausible image from the noisy x_t. We employ a least squares error metric for training the neural network. During the test time, our diffusion model recovers a plausible clean image iteratively from an input noisy image. In a traditional diffusion model, image generation is performed as

$$z' = (f \circ \ldots \circ f)(z_T, T),\ where\ f(x_t, t) = \Omega(x_t) + \sigma_t \epsilon,$$

where $z_T \sim \mathcal{N}(0, I), \sigma_t$ is the fixed standard deviation at the given step T, and $\epsilon \sim \mathcal{N}(0, I)$. This approach, however, results in long sampling times. To reduce the number of sampling steps, we adopt a non-Markovian diffusion process with the initial latent variable manipulated to guide the generated output as

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

In practice, we find that generation steps of 20 is sufficient for our experiments to conditionally recover the images from the noisy sensor measurements.

Supplementary Note 10. Further image comparisons for paired captures and neural backend

We present further images of the paired image capture system in Figure S14. We note that this set was not used to train the learned reconstruction method, but is unseen to assess the image quality. The left column are images captured with the compound refractive lens. The second column are captures with the broad band meta-optic without computational processing, the third column are images reconstructed with a Wiener deconvolution and block filtering, and the fourth column are images reconstructed with the neural backend. Throughout the various scenarios, the learned computational backend yields the highest image quality in all tested scenarios.

Supplementary Figure 14. *Additional examples and comparison of ground truth (compound optic capture), physics-based inverse filter, and learned reconstruction method. Captures of the compound camera are in the first column, raw broadband MO captures are in the second column, images reconstructed with the physics based inverse filter are in the third column, and images reconstructed with the learned reconstruction method are reported in the fourth column.*