nature portfolio

Peer Review File

Bypassing the imaging thickness limit with a translationally-invariant nonlocal thin film

Corresponding Author: Professor Yuanmu Yang

This manuscript has been previously reviewed at another journal. This document only contains information relating to versions considered at Nature Communications.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors present a lensless imaging platform that utilizes an unstructured translationally-invariant dielectric multilayer film to achieve deterministic 2D imaging, depth-resolved 3D imaging, and optical diffraction tomography. The core concept to use the nonlocal Fabry-Pérot resonance for image formation is surprisingly simple yet elegant. The authors further elaborated how their approach can bypass the thickness bound of conventional, translationally-variant systems. The experimental realization is clean, and the demonstration of multifunctionality (2D, 3D, and tomography) within a single passive element is impressive. This work establishes a promising new direction for ultracompact optical systems. I believe this manuscript will be of broad interest to the photonics community. The authors have adequately addressed my previous concerns when the manuscript was submit to Nature Nanotechnology in their response. Only minor revisions are required.

1、The nonlocal thin film is attached to the protective spacer of the image sensor in this work. The imaging clipping
2、 What is the axial resolution of depth-resolved 3D imaging, and what are the primary factors that influence it in this ?
3、Although one might anticipate that the efficiency of this system would be very low, a direct comparison with that of conventional lens systems still needs to be calculated and provided.
4、This device works by restricting the natural spatial divergence of light to only a few degrees around the normal incidence direction, the resolution and the FOV will be low and the working wavelength will be very narrow. Given that this is an inherent limitation, in what specific practical scenarios might it still find applications and how to achieve it?

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
The authors have addressed at length my comments and the comments of the other reviewers, and have performed a considerable amount of additional work (mostly added to the Supplementary Information). They have moderated some of their claims and clarified the most confusing aspects of the manuscript (e.g., Fig. 1). I would have preferred a more direct acknowledgement of the limitations of the proposed nonlocal device (which are significant, especially in terms of resolution, bandwidth, efficiency) rather than dozens of pages arguing that these limitations are not a major concern. I'm still unconvinced by several statements made in the manuscript and response letter, some of which I found somewhat misleading or confusing. For example, the authors state in their response letter that "the current resolution is not a fundamental limit", and suggest practical strategies to overcome this issue, such as decreasing the object-to-sensor distance or shifting the operational wavelength or using curved detectors. While these approaches may be useful in practice, they do not address the fundamental issue of this device: by deliberately operating as a low-pass filter of spatial frequency, a significant amount of information is lost before reaching the sensor and cannot be recovered, and hence the resolution is inherently limited (and, no, low-pass filtering is not necessary for image formation; on the contrary, high-resolution imaging and especially super-resolution imaging aim to capture the widest possible range of spatial frequencies, rather than filter them out).

That said, I believe the paper is sufficiently novel, interesting, and thought-provoking to merit publication in a good journal. I agree that there are specific applications for which the limitations of this device may be acceptable given the advantage of reduced thickness. If the other reviewers and the editor are supportive of publication in Nature Communications, I would be comfortable with that decision, provided that the authors clearly and candidly discuss the inherent limitations of their device in the main text.

(Remarks on code availability)

Response to the Reviewers' Comments on the manuscript entitled "Bypassing the imaging thickness limit with a translationally-invariant nonlocal thin film" submitted to *Nature Communications*.

We would like to thank the editor and reviewers for their thorough assessment of our manuscript. We are very grateful for the reviewers' positive comments about the work as well as all the valuable suggestions. Below we provide a point-by-point response to all the questions and comments raised by each reviewer. Along with the response letter, we also resubmitted a manuscript that includes the updates to address the reviewers' concerns. With the improvements made, we hope the editor and reviewers may find our manuscript an exciting contribution to the broad optics community that may be suitable for publication in *Nature Communications*.

Response to Reviewer #1:

Reviewer #1: The authors present a lensless imaging platform that utilizes an unstructured translationally-invariant dielectric multilayer film to achieve deterministic 2D imaging, depth-resolved 3D imaging, and optical diffraction tomography. The core concept to use the nonlocal Fabry-Pérot resonance for image formation is surprisingly simple yet elegant. The authors further elaborated how their approach can bypass the thickness bound of conventional, translationally-variant systems. The experimental realization is clean, and the demonstration of multifunctionality (2D, 3D, and tomography) within a single passive element is impressive. This work establishes a promising new direction for ultracompact optical systems. I believe this manuscript will be of broad interest to the photonics community. The authors have adequately addressed my previous concerns when the manuscript was submitted to Nature Nanotechnology in their response. Only minor revisions are required.

***Response*:** We sincerely thank the reviewer for the thoughtful and positive evaluation of our work. We are also delighted that the reviewer finds our earlier responses to have adequately addressed the concerns raised during the previous review of the manuscript, and that only minor revisions are required. We have carefully considered the remaining minor points, and our detailed replies are provided in the point-by-point responses below.

Reviewer #1: 1. The nonlocal thin film is attached to the protective spacer of the image sensor in this work. The imaging clipping

***Response*:** We thank the reviewer for this insightful comment. In the current configuration, the nonlocal thin film is attached to the protective spacer of the image sensor, and the finite thickness of this spacer laterally displaces the diffraction pattern of obliquely incident light, so that at large illumination angles part of the pattern falls outside the finite sensor area and is clipped. This is a practical constraint of the present assembly rather than a fundamental limitation. As noted in the manuscript, directly integrating the thin film onto the sensor surface can eliminate image clipping caused by spacer thickness, thereby extending usable angular range. To demonstrate the feasibility of reducing the film-to-sensor distance, in our on-going work, we have successfully made a module in which, by redesigning the sensor structure and depositing the nonlocal thin film directly on the cover glass, the film is brought to within only about 100 μm of the sensor surface.

Accordingly, we have revised the corresponding sentence in Supplementary Section 16 as follows:

"Directly integrating the thin film onto the CMOS sensor surface or a thinner cover glass can eliminate image clipping caused by spacer thickness, thereby extending usable angular range."

Reviewer #1: 2. What is the axial resolution of depth-resolved 3D imaging, and what are the primary factors that influence it in this?

***Response*:** We thank the reviewer for this insightful question. We agree that analyzing the axial resolution of depth-resolved 3D imaging and its primary influencing factors is helpful for a more comprehensive characterization of the system.

In the original supplementary material, we characterized the sensitivity of the depth-dependent point spread functions (PSFs) through the Fisher information (FI) (Supplementary Section 12). Building on this analysis, the axial resolution of depth-resolved 3D imaging can be further quantified by the Cramér–Rao lower bound (CRLB)[1], which sets the theoretical lower bound of the depth-estimation variance over all unbiased estimators. At a 10 dB signal-to-noise ratio, the CRLB for the ring-mode PSF yields an axial resolution of 0.09–0.14 μm over the object distance range of ρ = 2–6 mm.

This axial resolution is primarily limited by truncation of the ring-mode PSF due to the sensor's finite active area, which caps the maximum captured ring diameter and hence the available Fisher information. Experimentally, the achieved axial resolution is further limited by temperature-induced drift of the film's resonance wavelength, which arises from the thermo-optic response of the high-Q Fabry–Pérot cavity and causes a mismatch between the calibrated and in-situ PSFs. Additionally, film-thickness non-uniformity introduces a position-dependent shift of the local resonance wavelength that deforms the PSF across the field. These effects can be mitigated by larger sensor sizes and shorter working distances, together with refined material selection, fabrication, and PSF calibration.

We have updated Supplementary Section 12 to report the axial resolution, as follows:

"Quantitatively, the CRLB derived from the axial FI yields an axial resolution of 0.09–0.14 μm for the ring-mode PSF over this distance range."

Reviewer #1: 3. Although one might anticipate that the efficiency of this system would be very low, a direct comparison with that of conventional lens systems still needs to be calculated and provided.

***Response*:** We appreciate the reviewer's insightful comment regarding the efficiency of our system relative to conventional lens systems. To properly benchmark the light throughput, we have introduced an "effective *f*-number" metric to quantify the system's efficiency. In the following, we first evaluate this metric for the present system, and then demonstrate an optimized configuration that improves the efficiency.

In a conventional imaging system, the image irradiance E is governed by the f-number (N_F), which is the standard metric for quantifying the light-gathering capability, following the relation $E \propto L/N_F^2$, where L is the source radiance[2]. For our translationally-invariant system, the light collection is governed by the squared angle-dependent transmission function. Unlike conventional lenses which possess a clear physical aperture, the transmission of our film varies continuously with the incident angle. To enable a direct comparison using the standard f-number metric, we normalize this continuous response to an equivalent top-hat function. We define an effective angular width $\Delta\theta_{eff}$ as the diameter of an equivalent top-hat angular response that yields the same total integrated energy as the transmission profile. The resulting image irradiance is given by,

$$E = L\pi\sin^2(\Delta\theta_{eff}/2). \quad \text{(R1)}$$

By equating these two expressions, the effective f-number is determined as,

$$N_F = \frac{1}{2\sin(\Delta\theta_{eff}/2)}. \quad \text{(R2)}$$

Substituting $\Delta\theta_{eff} = 1.22°$ for the nonlocal thin film, we obtain $N_F = 47.0$. Compared to a typical telecentric lens with an f-number of 8[3-5], the light collection efficiency of our system is approximately 1/35 of that of standard optics.

To further improve the efficiency, one potential strategy is to engineer a more complex PSF structure, rather than a single spot, at the cost of requiring deconvolution. Specifically, we optimized the system design by increasing the central layer thickness to 141.5 μm and setting the number of DBR pairs to $M = 7$. At an operating wavelength of 1550 nm, this configuration maintains a central spot FWHM of 87.7 μm while introducing a ring with a FWHM of 105.6 μm (Fig. R1a). Simulations indicate that this configuration expands the effective angular width $\Delta\theta_{eff}$ to 3.43° by collecting light through both the central lobe and the side ring. This enhancement reduces the effective f-number from 47.0 to 16.7. Although this configuration introduces a broader PSF with sidelobes, the spatial information is preserved and can be restored using standard linear deconvolution without requiring complex priors. The resulting image is shown in Fig. R1b. Applying Wiener deconvolution to the raw image with 20 dB added Gaussian noise successfully recovers the target structure (letter "E" with a line spacing of 200 μm) (Fig. R1c).

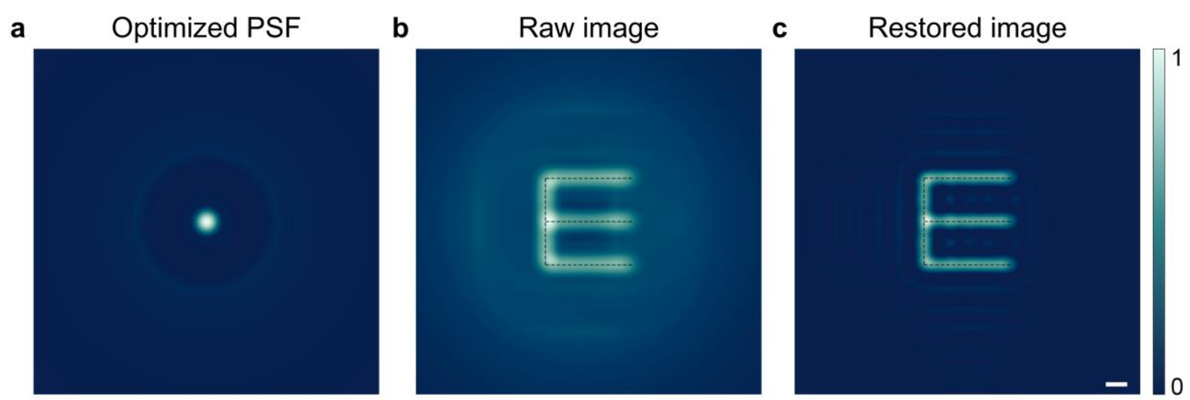

Figure R1 | Simulation of the efficiency-optimized system configuration with a central layer thickness of 141.5 μm and DBR pairs M =7. a, The calculated PSF exhibiting a central spot with a surrounding ring. **b,** The simulated image of the target object (letter "E" with a line spacing of 200 μm). **c,** The restored image after applying Wiener deconvolution. Scale bar: 100 μm.

We fully acknowledge that the proposed imaging platform is not a universal replacement for conventional optics in all scenarios. Due to the narrowband/dispersive nature of the system, it may be particularly suited for laser-based imaging applications. In Supplementary Section 17, we have quantified the light throughput and demonstrated the efficiency optimization strategy. Accordingly, we have updated the Discussion section in the Main Text:

"It has an effective *f*-number of 47.0, approximately six times the typical value of 8 for telecentric lenses[40-42] (Supplementary Section 17)."

Reviewer #1: 4. This device works by restricting the natural spatial divergence of light to only a few degrees around the normal incidence direction, the resolution and the FOV will be low and the working wavelength will be very narrow. Given that this is an inherent limitation, in what specific practical scenarios might it still find applications and how to achieve it?

***Response*:** We thank the reviewer for this important and constructive comment. We agree that the practical relevance of the proposed platform should be evaluated in light of its inherent trade-offs in spatial resolution, field of view (FOV), and spectral bandwidth. In brief, the current device is well suited to space-constrained sensing applications that employ active illumination, where eliminating lenses and free-space propagation is more critical than maximizing the numerical aperture or the spectral bandwidth.

We emphasize that the platform is not intended to compete with conventional high-NA microscopes, broadband cameras, or general-purpose passive imaging systems, nor is it a direct replacement for microscope objectives. Likewise, the present planar implementation should not be interpreted as a wide-angle far-field camera. Rather, it embodies a deliberate trade-off, prioritizing extreme compactness to enable image formation in a lensless architecture. Many imaging tasks do not require half-wavelength resolution as their primary objective, and high numerical aperture is therefore not a universal requirement. Similarly, although the current Bragg microcavity is narrowband, passive broadband operation is not essential for many high-impact applications based on active illumination. The device therefore targets a regime in which the dominant requirement is to enable image formation without requiring lenses or free-space propagation, with the thin film attached directly to an image sensor within an ultracompact, lensless architecture. Representative application domains in this regime include medical endoscopy[6], wearable sensors[7], robotic tactile sensing[8, 9], and compact localization metrology. In visual-tactile sensing, for example, acquiring accurate tactile information requires an extremely thin imaging backend. We have already presented a sufficient range of practical scenarios, together with concrete implementation schemes, in detail in our previous response.

Besides, the trade-offs noted by the reviewer can be systematically relaxed, and in part actively exploited, rather than merely tolerated. Regarding spatial resolution, although the demonstrated sub-100-μm resolution is modest relative to conventional microscopy, it is comparable to, and in some cases surpasses, that of other miniature imaging modalities benchmarked under identical sensor constraints, while the total track length of 12.2 μm is around two orders of magnitude thinner than competing lens-based miniature systems (Supplementary Section 8 and Table S1). The resolution is also not fixed. Reducing the object distance to 160 μm with 532 nm illumination improves it to 13.4 μm through proportional scaling of the film layer thicknesses, and cascading two tilted films narrows the angular linewidth by roughly an order of magnitude (Supplementary Section 10). Integrating the film onto a curved detector with the object near the center of curvature further provides geometric magnification, reaching an object-space resolution of approximately 3.91 μm with a hemispherical detector of radius R = 1 mm and M = 17 DBR pairs at a magnification of 100 (Supplementary Section 11).

Regarding the FOV, in the planar configuration the experimentally accessible FOV is set by the active area of the image sensor. Because the translationally-invariant film decouples thickness from pixel count, the directly imaged area can be expanded simply by adopting larger-format sensors, without increasing the film thickness. For long-distance observation, combining the film with a curved-sensor geometry enables ultracompact lensless telescopic designs in which magnification and FOV are tailored by the detector geometry rather than by a long focal length (Supplementary Section 11 and Fig. S13).

Regarding the working bandwidth, the narrowband and dispersive response is itself a useful degree of freedom rather than only a constraint. As demonstrated in the manuscript (Figs. 3 and 4), simply scanning the illumination wavelength unlocks depth-resolved 3D imaging and optical diffraction tomography. Where multiwavelength operation is needed, the film can be engineered for polychromatic response by supporting multiple Fabry–Pérot resonances, as in the RGB design of the tactile example above, while moderate-bandwidth sources such as LEDs suffice for centroid-based tasks, as in the localization example above. Broadband operation with preserved angular selectivity may be further explored by engineering the photonic band structure of the dielectric stack, for example by stacking multilayer films with varying periodicities[10, 11].

Finally, we respectfully clarify that the significance of bypassing the imaging thickness bound, and of achieving deterministic imaging with a stand-alone thin film, remains valid despite the associated trade-offs in resolution, FOV, and spectral bandwidth. Miller's thickness limit originates from the overlapping nonlocality required to resolve many distinct pixels through a shared aperture, which forces the thickness of a conventional translationally-variant system to scale linearly with pixel count, independent of numerical aperture. By contrast, our translationally-invariant architecture decouples system thickness from pixel count, thereby circumventing the scaling law inherent to conventional optics. Thus, our central claim is not that the present implementation eliminates all imaging trade-offs, but rather that it establishes a distinct imaging topology in which deterministic image formation is possible within an ultrathin, lensless architecture.

Accordingly, we have revised Supplementary Section 8 to state explicitly the specific practical scenarios in which the platform applies, as follows:

"For specific applications such as medical endoscopy[9], wearable sensors[10], and robotic tactile sensing[11, 12], the priority is often the extreme compactness required to integrate sensors into confined spaces rather than imaging resolution. In these specific practical scenarios, the thin film can be attached directly onto the image sensor and operated under active illumination, enabling lensless image formation with an ultracompact form factor unattainable with bulky conventional optics."

Response to Reviewer #4:

Reviewer #4: The authors have addressed at length my comments and the comments of the other reviewers, and have performed a considerable amount of additional work (mostly added to the Supplementary Information). They have moderated some of their claims and clarified the most confusing aspects of the manuscript (e.g., Fig. 1).

***Response*:** We sincerely thank the reviewer for carefully evaluating our revised manuscript and for recognizing the "considerable amount of additional work" we have carried out in response to the comments. We are also encouraged by the reviewer's observation that we have moderated our claims and clarified the previously confusing aspects of the manuscript, including Fig. 1.

Reviewer #4: I would have preferred a more direct acknowledgement of the limitations of the proposed nonlocal device (which are significant, especially in terms of resolution, bandwidth, efficiency) rather than dozens of pages arguing that these limitations are not a major concern.

***Response*:** We greatly appreciate the reviewer's insightful and constructive suggestion. We agree that the device's limitations in resolution, bandwidth, and efficiency should be more directly acknowledged in the main text. Accordingly, we have made the following revisions to the main text.

In the section presenting our 2D imaging results, we added a direct statement of the device's resolution:

"Our system achieves a TTL of just 12.2 µm, over two orders of magnitude thinner than conventional millimeter-scale miniature lens modules, while providing sub-100-µm resolution without the computational overhead required by other lensless imaging techniques (Supplementary Section 8). The system achieves a volumetric channel density[26] approximately 8.2 times greater than that of conventional lens-based systems (Supplementary Section 9). **Based on this performance profile, the current implementation is not intended to serve as a direct replacement for microscope objectives, but rather to extend image formation into a regime of extreme compactness inaccessible to conventional optics.** The imaging resolution can be further enhanced by reducing the object distance or by shifting to shorter operational wavelengths, which involves a straightforward scaling of the film's layer thicknesses. In addition, cascading two tilted thin films can reduce the angular linewidth of the transmission peak by roughly an order of magnitude compared to a single, upright film, leading to a substantial improvement in resolution without enhancing the cavity Q-factor (Supplementary Section 10)."

In the concluding Discussion, we added a direct statement of the device's narrowband nature and limited efficiency:

"The versatility of this platform points toward significant future advancements. The operating wavelength can be readily scaled from the ultraviolet to the infrared simply by adjusting the film's layer thicknesses. Performance can be further enhanced by directly depositing the dielectric stack onto the image sensor, a step that promises higher quality factors and streamlined integration. **In its present form, the nonlocal thin film is a narrowband device operating under active laser illumination. It has an effective *f*-number of 47.0, approximately six times the typical value of 8 for telecentric lenses[40-42] (Supplementary Section 17).** However, optimizing the film's spatial dispersion and angular selectivity could enable broadband operation or the engineering of more complex point spread functions for new imaging modes[43], a strategy that can also be leveraged to enhance the system's efficiency (Supplementary Section 17). Alternatively, the thin film can be engineered for polychromatic operation, enabling multiwavelength imaging within a single ultrathin element (Supplementary Section 18). Furthermore, moderate-bandwidth sources, such as LEDs, may suffice for applications like localization metrology that rely on PSF centroid estimation rather than size[44-46]."

We hope these revisions address the reviewer's concern by presenting the device candidly, with its resolution, bandwidth, and efficiency limitations stated directly in the main text, while clarifying the specific regimes in which its ultrathin, lensless architecture offers a distinct advantage.

Reviewer #4: I'm still unconvinced by several statements made in the manuscript and response letter, some of which I found, somewhat misleading or confusing. For example, the authors state in their response letter, that "the current resolution is not a fundamental limit", and suggest practical strategies to overcome this issue, such as decreasing the object-to-sensor distance or shifting the operational wavelength or using curved detectors. While these approaches may be useful in practice, they do not address the fundamental issue of this device: by deliberately operating as a low-pass filter of spatial frequency, a significant amount of information is lost before reaching the sensor and cannot be recovered, and hence the resolution is inherently limited…

***Response*:** We thank the reviewer for this comment. We agree that the practical routes we mentioned do not change the operating principle of the device. We wish to clarify that the resolution demonstrated in the present prototype reflects the specific configuration we adopted, namely the illumination wavelength, the object distance, and the film design. Accordingly, re-tuning these parameters or adopting a curved detector is, in practice, a useful way to improve the resolution for a given task.

Regarding the trade-off in resolution, we appreciate the reviewer's acknowledgement in a subsequent comment that "there are specific applications for which the limitations of this device may be acceptable given the advantage of reduced thickness". Indeed, the current implementation is not intended to serve as a direct replacement for microscope objectives, but rather to extend image formation into a regime of extreme compactness inaccessible to conventional optics, a capability required by space-constrained applications such as medical endoscopy, wearable sensors, and robotic tactile sensing. When evaluated under identical sensor constraints, our resolution is comparable to, and in some cases surpasses, that of other miniature imaging modalities (Supplementary Section 8). Crucially, our system achieves this with a total track length of just 12.2 μm, which is over two orders of magnitude thinner than competing lens-based miniature systems, and the field of view can be readily expanded without increasing the thickness. Moreover, the central point of this work, namely that the imaging thickness bound can be bypassed, remains valid regardless of the resolution attained.

Following the reviewer's comment, we have made the corresponding revisions in Supplementary Section 10:

"Taken together, these results indicate that the resolution reported in the main text corresponds to the specific implementation we adopted in the current prototype, and that the practically achievable resolution can be improved for a given task through appropriate design choices."

Reviewer #4: …(and, no, low-pass filtering is not necessary for image formation; on the contrary, high-resolution imaging and especially super-resolution imaging aim to capture the widest possible range of spatial frequencies, rather than filter them out).

Response**:** We thank the reviewer for this insightful comment. We agree with the reviewer that capturing a broader range of spatial-frequency components generally improves imaging resolution. Strictly speaking, however, a larger aperture does not necessarily mean higher resolution. Another condition is that the phase of the coherent transfer function (CTF) must maintain a small variation across the entire transmitted spectrum for all object positions (Supplementary Section 3). In many cases, enlarging the aperture increases the phase variation and degrades the in-phase interference, so the resolution becomes worse. Low-pass filtering is therefore sometimes applied deliberately to achieve better overall image quality.

For example, free-space propagation transmits a wide coherent band, but its rapidly oscillating chirp phase is left uncorrected, so the optical transfer function (OTF) collapses to a near-delta-like response. Conventional translationally-variant optics faces a field-dependent constraint because all object points share a single pupil. Enlarging the aperture of a given lens can even degrade the image quality across the field, as the CTF phase cannot maintain a small variation across the shared pupil for all object positions simultaneously. Our translationally-invariant thin film instead treats all object points identically, although its angular passband must still be chosen appropriately to balance the transmitted bandwidth against the phase variation.

This constraint is not peculiar to our device. In vision science, the human eye attains its best optical quality at an intermediate pupil diameter, with aberration limiting larger pupils and diffraction limiting smaller ones[12]. In photography, this trade-off produces the optimum aperture, commonly called the sweet spot, the moderate *f*-number at which a lens is sharpest, because opening the aperture increases aberration while stopping it down increases diffraction[13-15]. This behavior is illustrated in Fig. R2, in which we compute the MTF50 of a simulated commercial photographic lens, the Sigma 35mm F1.4 DG HSM Art, as a function of the *f*-number, where MTF50 is the spatial frequency at which the image-side modulation transfer function (MTF) falls to 0.5. The image quality peaks at a moderate *f*-number, while opening the aperture further instead degrades the image. Some form of angular spectrum filtering (low-pass filtering) therefore generally accompanies image formation, as it confines the transmitted spectrum to the range within which the phase variation can be kept small.

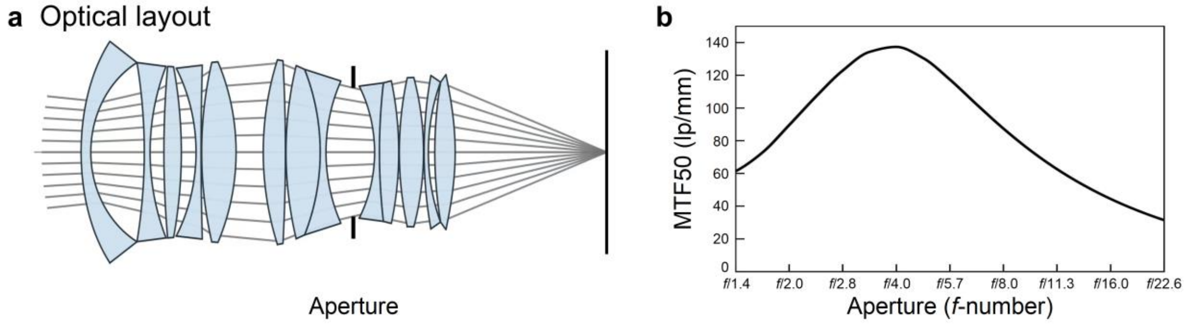

Figure R2 | The optimum aperture (sweet spot) of a commercial photographic lens. a, Optical layout of the Sigma 35mm F1.4 DG HSM Art lens. **b**, MTF50 of the lens at the image center versus *f*-number. MTF50 is the spatial frequency at which the image-side modulation transfer function (MTF) falls to 0.5.

Our nonlocal thin film operates at one extreme of this trade-off. It forms an image within an ultrathin, lensless architecture, and its translational invariance effectively assigns an individual aperture to each object point, allowing the CTF phase to maintain a small variation for all object positions without the constraints of a global pupil geometry. Its angular passband therefore reflects the finite-aperture low-pass behavior found in conventional optics.

Accordingly, we have slightly revised the corresponding sentence in Supplementary Section 3, which now reads:

"Consequently, some form of angular spectrum filtering (low-pass filtering) is generally needed for image formation, to prevent the CTF phase from undergoing large variations, thereby broadening the effective OTF."

We have also added the following sentence at the end of the discussion of conventional shared-aperture optics in the same section:

"This trade-off is manifested in the optimum aperture of conventional imaging systems, where the human eye attains its best optical quality at an intermediate pupil diameter[2] and a camera lens is sharpest at a moderate *f*-number[3-5]."

Reviewer #4: That said, I believe the paper is sufficiently novel, interesting, and thought-provoking to merit publication in a good journal. I agree that there are specific applications for which the limitations of this device may be acceptable given the advantage of reduced thickness. If the other reviewers and the editor are supportive of publication in Nature Communications, I would be comfortable with that decision, provided that the authors clearly and candidly discuss the inherent limitations of their device in the main text.

***Response*:** We sincerely thank the reviewer for this positive and generous assessment of our work. We are greatly encouraged that the reviewer finds the paper "sufficiently novel, interesting, and thought-provoking to merit publication in a good journal". We also appreciate the reviewer's recognition that "there are specific applications for which the limitations of this device may be acceptable given the advantage of reduced thickness", and we are grateful that the reviewer would be comfortable with the publication of this work in Nature Communications.

We fully agree with the reviewer that the inherent limitations of the device should be clearly and candidly discussed in the main text. As described in our earlier response, we have revised the main text to state these limitations explicitly. In the section presenting our 2D imaging results, we added a direct statement of the device's resolution:

"Our system achieves a TTL of just 12.2 µm, over two orders of magnitude thinner than conventional millimeter-scale miniature lens modules, while providing sub-100-µm resolution without the computational overhead required by other lensless imaging techniques (Supplementary Section 8). The system achieves a volumetric channel density[26] approximately 8.2 times greater than that of conventional lens-based systems (Supplementary Section 9). **Based on this performance profile, the current implementation is not intended to serve as a direct replacement for microscope objectives, but rather to extend image formation into a regime of extreme compactness inaccessible to conventional optics.** The imaging resolution can be further enhanced by reducing the object distance or by shifting to shorter operational wavelengths, which involves a straightforward scaling of the film's layer thicknesses. In addition, cascading two tilted thin films can reduce the angular linewidth of the transmission peak by roughly an order of magnitude compared to a single, upright film, leading to a substantial improvement in resolution without enhancing the cavity Q-factor (Supplementary Section 10)."

In the concluding Discussion, we added a direct statement of the device's narrowband nature and limited efficiency:

"The versatility of this platform points toward significant future advancements. The operating wavelength can be readily scaled from the ultraviolet to the infrared simply by adjusting the film's layer thicknesses. Performance can be further enhanced by directly depositing the dielectric stack onto the image sensor, a step that promises higher quality factors and streamlined integration. **In its present form, the nonlocal thin film is a narrowband device operating under active laser illumination. It has an effective *f*-number of 47.0, approximately six times the typical value of 8 for telecentric lenses[40-42] (Supplementary Section 17).** However, optimizing the film's spatial dispersion and angular selectivity could enable broadband operation or the engineering of more complex point spread functions for new imaging modes[43], a strategy that can also be leveraged to enhance the system's efficiency (Supplementary Section 17). Alternatively, the thin film can be engineered for polychromatic operation, enabling multiwavelength imaging within a single ultrathin element (Supplementary Section 18). Furthermore, moderate-bandwidth sources, such as LEDs, may suffice for applications like localization metrology that rely on PSF centroid estimation rather than size[44-46]."

With these statements in place, we believe the main text now discusses the inherent limitations of the device clearly and candidly, as the reviewer has requested. Once again, we thank the reviewer for the thorough and constructive comments throughout the review process, which have significantly improved the quality, clarity, and rigor of this work. We hope the revised manuscript is now suitable for publication in Nature Communications.

References

1. Shechtman, Y., Sahl, S. J., Backer, A. S. & Moerner, W. E. Optimal point spread function design for 3D imaging. *Phys. Rev. Lett.* **113**, 133902 (2014).
2. Kingslake, R. The Effective Aperture of a Photographic Objective. *J. Opt. Soc. Am.* **35**, 518-520 (1945).
3. Medley, S. S. et al. Periscope-camera system for visible and infrared imaging diagnostics on TFTR. *Rev. Sci. Instrum.* **56**, 1873-1884 (1985).
4. Zhang, S. et al. 4-domain twisted liquid crystal micropolarizer array for visible linear polarization imaging. *Opt. Express* **29**, 43226-43240 (2021).
5. Liu, C. et al. Correction of a digital micromirror device lithography system for fabrication of a pixelated liquid crystal micropolarizer array. *Opt. Express* **30**, 12014-12025 (2022).
6. Kiesslich, R., Goetz, M., Hoffman, A. & Galle, P. R. New imaging techniques and opportunities in endoscopy. *Nat. Rev. Gastroenterol. Hepatol.* **8**, 547-553 (2011).
7. Yokota, T., Fukuda, K. & Someya, T. Recent Progress of Flexible Image Sensors for Biomedical Applications. *Adv. Mater.* **33**, e2004416 (2021).
8. Dahiya, R. S., Metta, G., Valle, M. & Sandini, G. Tactile Sensing—From Humans to Humanoids. *IEEE Trans. Robot.* **26**, 1-20 (2010).
9. Abad, A. C. & Ranasinghe, A. Visuotactile Sensors With Emphasis on GelSight Sensor: A Review. *IEEE Sens. J.* **20**, 7628-7638 (2020).
10. Qu, Y. et al. Polarization-Independent Optical Broadband Angular Selectivity. *ACS Photonics* **5**, 4125-4131 (2018).
11. Yin, K. et al. Enabling Manufacturable Optical Broadband Angular-Range Selective Films. *ACS Nano* **15**, 19917-19923 (2021).
12. Campbell, F. W. & Gubisch, R. W. Optical quality of the human eye. *J. Physiol.* **186**, 558–578 (1966).
13. Kingslake, R. *Optics in Photography* (SPIE Optical Engineering Press, Bellingham, 1992).
14. Jacobson, R. E., Ray, S. F., Attridge, G. G. & Axford, N. R. *The Manual of Photography* 9th edn (Focal Press, Oxford, 2000).
15. Smith, W. J. *Modern Optical Engineering: The Design of Optical Systems* 4th edn (McGraw-Hill, New York, 2008).