

# Monocular metasurface camera for passive single-shot 4D imaging

Received: 7 March 2022

Accepted: 16 February 2023

Published online: 23 February 2023

 Check for updatesZicheng Shen<sup>1</sup>, Feng Zhao<sup>1</sup>, Chunqi Jin<sup>1</sup>, Shuai Wang<sup>1</sup> <sup>1</sup>, Liangcai Cao<sup>1</sup> <sup>1</sup> & Yuanmu Yang<sup>1</sup> <sup>1</sup> 

It is a grand challenge for an imaging system to simultaneously obtain multi-dimensional light field information, such as depth and polarization, of a scene for the accurate perception of the physical world. However, such a task would conventionally require bulky optical components, time-domain multiplexing, and active laser illumination. Here, we experimentally demonstrate a compact monocular camera equipped with a single-layer metalens that can capture a 4D image, including 2D all-in-focus intensity, depth, and polarization of a target scene in a single shot under ambient illumination conditions. The metalens is optimized to have a conjugate pair of polarization-decoupled rotating single-helix point-spread functions that are strongly dependent on the depth of the target object. Combined with a straightforward, physically interpretable image retrieval algorithm, the camera can simultaneously perform high-accuracy depth sensing and high-fidelity polarization imaging over an extended depth of field for both static and dynamic scenes in both indoor and outdoor environments. Such a compact multi-dimensional imaging system could enable new applications in diverse areas ranging from machine vision to microscopy.

Conventional cameras can only capture 2D images. In recent years, there has been a rapid development in 3D imaging techniques<sup>1–3</sup> towards emerging applications such as consumer electronics and autonomous driving. Moreover, a camera that can capture extra dimensions of the light field information, such as polarization and spectrum, may reveal even richer characteristics of a scene<sup>4–6</sup>, thus allowing the perception of a more “complete” physical world towards various tasks, such as object tracking and identification, with high precision.


However, capturing light field information beyond 2D intensity typically requires an optical system with a considerably increased size, weight, and power consumption. For instance, 3D imaging systems based on structured light<sup>7</sup> and time-of-flight<sup>8</sup> require active laser illumination. Binocular or multi-view cameras have a large form factor and a depth estimation accuracy and range constrained by their baseline length<sup>9</sup>. Polarization imaging systems are often based on amplitude or focal plane division or require time-domain multiplexing<sup>10</sup>. The

simultaneous measurement of multi-dimensional light field information can be even more challenging, often requiring an imaging system with a form factor and complexity far surpassing conventional cameras<sup>11–13</sup>.

It is highly desirable to have a compact monocular camera that can capture multi-dimensional light field information in a single shot under ambient illumination conditions. According to a generalized multi-dimensional image formation model<sup>13</sup>,

$$g(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{z,p,\dots}(x, y) \text{PSF}_{z,p,\dots}(x' - x, y' - y) dx dy + \eta, \quad (1)$$

where  $(x, y, z)$  and  $(x', y')$  represent the spatial coordinate in the object space and on the image plane, respectively.  $f$  is the system's input (target object) that contains multi-dimensional light field information,  $g$  is the system's output,  $p$  is the polarization state of light reflected from the target object, and  $\eta$  is a generic noise term. PSF is the point-spread function, which describes the imaging system's impulse

<sup>1</sup>State Key Laboratory of Precision Measurement Technology and Instruments, Department of Precision Instrument, Tsinghua University, Beijing 100084, China.  e-mail: [ymyang@tsinghua.edu.cn](mailto:ymyang@tsinghua.edu.cn)

response to a point light source. To obtain light field information beyond the 2D projection of intensity, such as depth  $z$  and polarization  $p$  of the target object, the PSF of the system should be strongly dependent on  $z$  and  $p$ . This degree of dependency is quantified by Fisher information<sup>14,15</sup> which determines the estimation accuracy of the corresponding parameter with a given system noise.

Taking depth estimation as an example, since any standard lens has a defocus that is dependent on the depth of the object, monocular cameras have already been used to estimate the depth of a scene by capturing multiple images under different defocus settings<sup>16–18</sup>. However, the depth-from-defocus method typically requires the physical movement of the imaging system and suffers from a low estimation accuracy due to the self-similarity of the system's PSF along the depth dimension.

More sophisticated depth-dependent PSFs, such as the double-helix PSF, have been proposed for depth estimation with a higher accuracy<sup>14,15,19,20</sup>. The double-helix PSF has two foci rotating around a central point, with the rotation angle dependent on the axial depth of the target object. A double-helix PSF can be generated using diffractive optical elements with a stringently tailored phase profile<sup>14,20</sup>. Subsequently, the depth of the target object can be retrieved by analyzing the power cepstrum of the acquired image<sup>20</sup>. However, the retrieval algorithm is computationally demanding and slow. Furthermore, due to the superposition of the twin-image generated by the two foci, the double-helix PSF method requires an additional reference image with an extended depth of field in order to reconstruct a high-fidelity 2D all-in-focus image of the scene. The reference image can be generated by an additional aperture or time-domain multiplexing but at the cost of significantly increasing the complexity of the imaging system<sup>21</sup>. Moreover, due to the  $C_2$  symmetry of the double-helix PSF, its depth measurement range is limited by the maximum rotation angle of  $180^\circ$ .

To construct a monocular camera that can efficiently retrieve multi-dimensional light field information of a scene in a single shot is an even more challenging task. Recently, an emerging class of sub-wavelength diffractive optical element, metasurface<sup>22–28</sup> has been found to be highly versatile to tailor the vectorial light field. Consequently, it opens up new avenues for various applications, including depth<sup>29–32</sup>, polarization<sup>33–36</sup>, and spectral<sup>37,38</sup> imaging. Very recently, Lin et al. theoretically proposed to leverage the end-to-end optimization

of a meta-optics frontend and an image-processing backend for single-shot imaging over a few discrete depth, spectrum, and polarization channels<sup>39</sup>. It remains a major challenge to build a compact multi-dimensional imaging system that allows high-accuracy depth sensing for arbitrary depth values in a wide range of indoor and outdoor scenes.

In this work, we experimentally demonstrate a monocular camera equipped with a single-layer metasurface that can capture 4D light field information, including 2D all-in-focus intensity, depth, and polarization, of a target scene in a single shot. Leveraging the versatility of metasurface to manipulate the vectorial field of light, we design and optimize a polarization-multiplexed metasurface with a decoupled pair of conjugate single-helix PSFs, forming a pair of spatially-separated twin-image of orthogonal polarization on the photosensor. The depth and polarization information of the target scene is simultaneously encoded in the decoupled twin-image pair. The PSF of the metasurface has a Fisher information two orders of magnitude higher than that of a standard lens. Combined with a straightforward image retrieval algorithm, we demonstrate high-accuracy depth estimation and high-fidelity polarization imaging over an extended depth of field for both static and dynamic scenes under ambient lighting conditions.

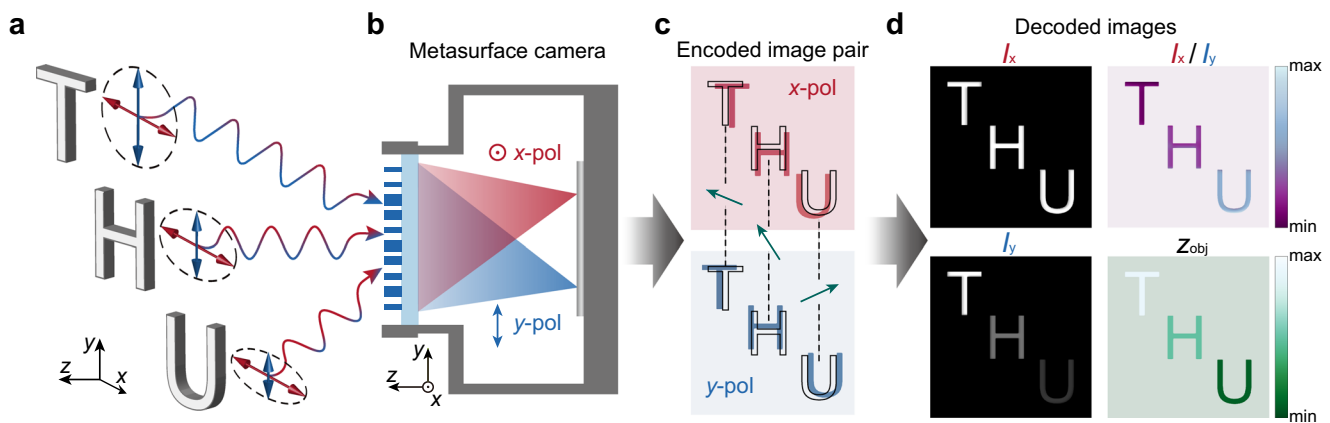
## Results

### Single-shot 4D imaging framework

The framework of the monocular metasurface camera for single-shot 4D imaging is schematically illustrated in Fig. 1. A single-layer metasurface is designed and optimized to generate a pair of conjugate single-helix PSFs that form a twin-image pair of the target object with orthogonal linear polarization laterally shifted on the photosensor (Fig. 1a). The depth of the scene is encoded in the local orientations of the translation vectors of the twin-image (Fig. 1b, c). Subsequently, one can computationally retrieve the all-in-focus 2D light intensity, depth, and polarization contrast of the scene (Fig. 1c, d).

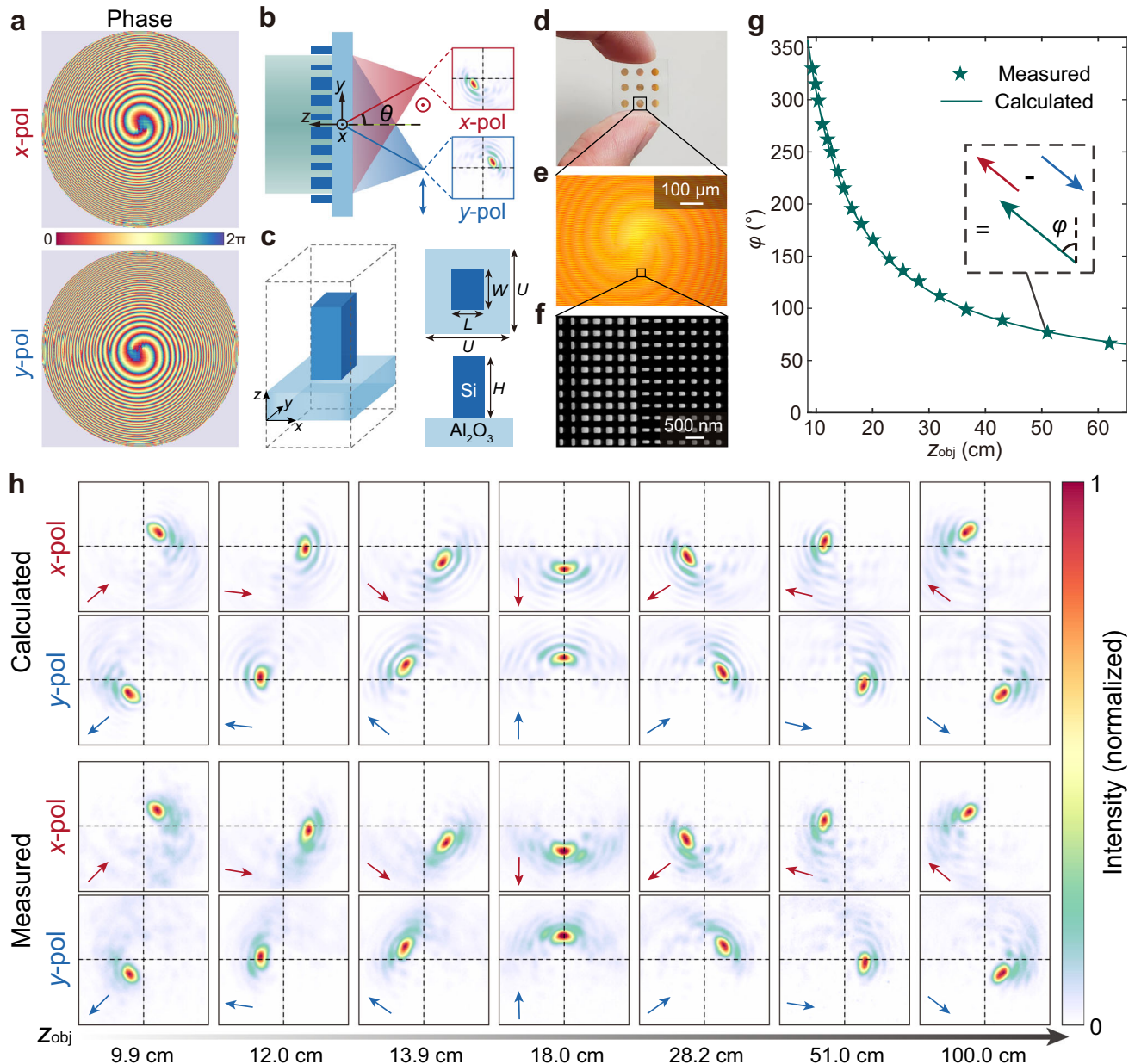
### Metasurface design and characterization

To generate a rotating single-helix PSF at a near-infrared operation wavelength  $\lambda = 800$  nm, we assume the placement of a metasurface at the entrance pupil of the imaging system, and initialize the transmission phase of the metasurface with Fresnel zones carrying spiral phase



**Fig. 1 | Framework of using a monocular metasurface camera for single-shot 4D imaging.** **a** The target scene consists of objects with different polarization states and depths. **b** Light emitted from the scene is split and focused on the photosensor, forming two images of orthogonal linear polarizations. The red and blue colors correspond to  $x$ - and  $y$ -polarized light, respectively. **c** The raw encoded image-pair recorded on the photosensor, with depth information of the scene encoded in the

local orientation of the translation vector (green arrows). The black outlines are the geometric image positions of the objects. **d** Decoded 4D image of the target scene, including all-in-focus 2D light intensity ( $I_x$  and  $I_y$ ), polarization contrast ( $I_x/I_y$ ), and depth ( $Z_{obj}$ ), can be retrieved from the raw image-pair by a straightforward, physically interpretable algorithm based on image segmentation and template matching of the object-pair.



**Fig. 2 | Metasurface design and characterization.** **a** Optimized metasurface phase profiles that can generate a decoupled pair of conjugate single-helix PSFs for x-polarized and y-polarized incident light, respectively. **b** Schematic of the metasurface that splits and focuses the light of orthogonal polarization on laterally shifted location of the photosensor with a decoupled pair of conjugate single-helix PSFs, where  $\theta = 8^\circ$  is the off-axis angle for polarization splitting. **c** The unit-cell of the metasurface is composed of silicon nanopillars of rectangular in-plane cross-sections on a sapphire substrate, with height  $H = 600$  nm, period  $U = 350$  nm, and width  $W$  and length  $L$  varying between 100 nm and 250 nm. **d–f** Photograph (**d**), optical microscopy image (**e**), and scanning electron microscopy image (**f**) of the fabricated metasurface sample,

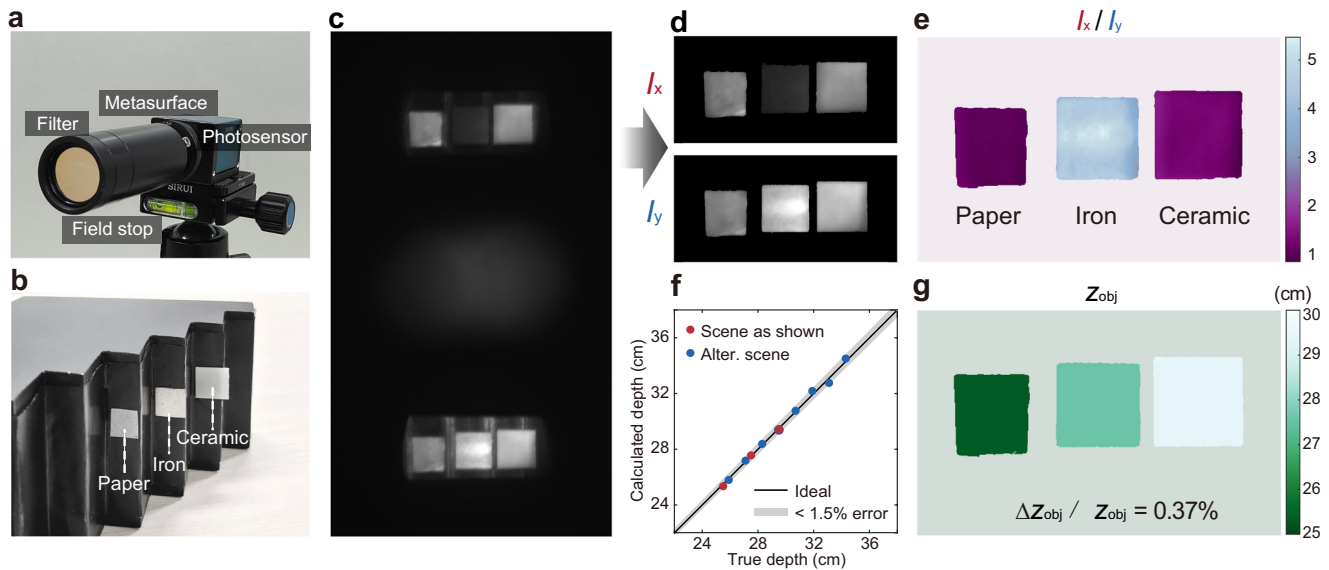
respectively. The other metalenses as shown in the photograph are fabricated for different purposes. **g** Calculated (green line) and measured (green star) orientation angle  $\varphi$  of the single-helix PSF pair as a function of the axial depth of a point object  $z_{obj}$ . The inset shows the method to extract  $\varphi$  from the translation relationship of the decoupled pair of conjugate single-helix PSFs. **h** Calculated and experimentally measured single-helix PSFs as a function of  $z_{obj}$ , for x-polarized and y-polarized incident light, respectively. The PSF pairs are conjugate with respect to the geometric image point (intersection point of black crosshairs). The arrows represent the vectors pointing from the geometric image point to the maximum of the PSFs.

profiles with gradually increasing topological quantum numbers towards outer rings of the zone plate<sup>40,41</sup> as,

$$\psi_r(u, \varphi_u) = \left\{ l\varphi_{ul} \left( \frac{l-1}{L} \right)^\varepsilon \leq u \leq \left( \frac{l}{L} \right)^\varepsilon, l=1, \dots, L \right\}, \quad (2)$$

where  $u$  is the normalized radial coordinate and  $\varphi_u$  is the azimuth angle in the entrance pupil plane.  $[L, \varepsilon]$  are adjustable design

parameters. Compared with Gauss–Laguerre mode-based approach widely used in the design of double-helix PSFs<sup>30,42,43</sup>, the adopted Fresnel zone approach can generate a more compact rotating PSF with the shape of the PSF kept almost invariant over an extended depth of field<sup>40</sup>. Subsequently, an iterative Fourier transform algorithm is used to maximize the energy in the main lobe of the rotating PSF within the 360° rotation range. The iterative optimization process further improves the peak intensity of the main lobe of the PSF by 36%. In



**Fig. 3 | Indoor imaging of a static scene.** **a** Photograph of the metasurface camera. **b** Photograph of the target scene consisting of three pieces of different materials (paper, iron, and ceramic) located at different axial depths. **c** Raw image-pair captured by the metasurface camera. **d** Decoded all-in-focus polarization image pair  $I_x$  and  $I_y$ . **e** Polarization contrast of the scene, defined as  $I_x/I_y$ , which clearly

distinct metallic and non-metallic materials. **f** Comparison of the calculated depth to the ground truth for the scene shown here and an alternative scene as shown in Supplementary Section 5. **g** Retrieved depth map of the scene, with a normalized mean absolute error ( $\Delta Z_{obj}/Z_{obj}$ ) of only 0.37%.

the final step of designing the transmission phase profile, a polarization splitting phase term,

$$\psi_{xf}(x,y) = -\frac{2\pi}{\lambda} \left[ \sqrt{x^2 + y^2 + f^2 + 2xf\sin\theta} - f \right], \quad (3)$$

$$\psi_{yf}(x,y) = -\frac{2\pi}{\lambda} \left[ \sqrt{x^2 + y^2 + f^2 - 2xf\sin\theta} - f \right], \quad (4)$$

is added to spatially decouple the conjugate single-helix PSFs, where  $f = 20$  mm is the focal length,  $\theta = 8^\circ$  is the off-axis angle for polarization splitting. The optimized phase profiles for both  $x$ - and  $y$ -polarized incident light are shown in Fig. 2a, along with the schematic of the metasurface that splits and focuses the light of orthogonal polarization, as shown in Fig. 2b. More detailed discussion of the transmission phase design is included in Supplementary Section 1.

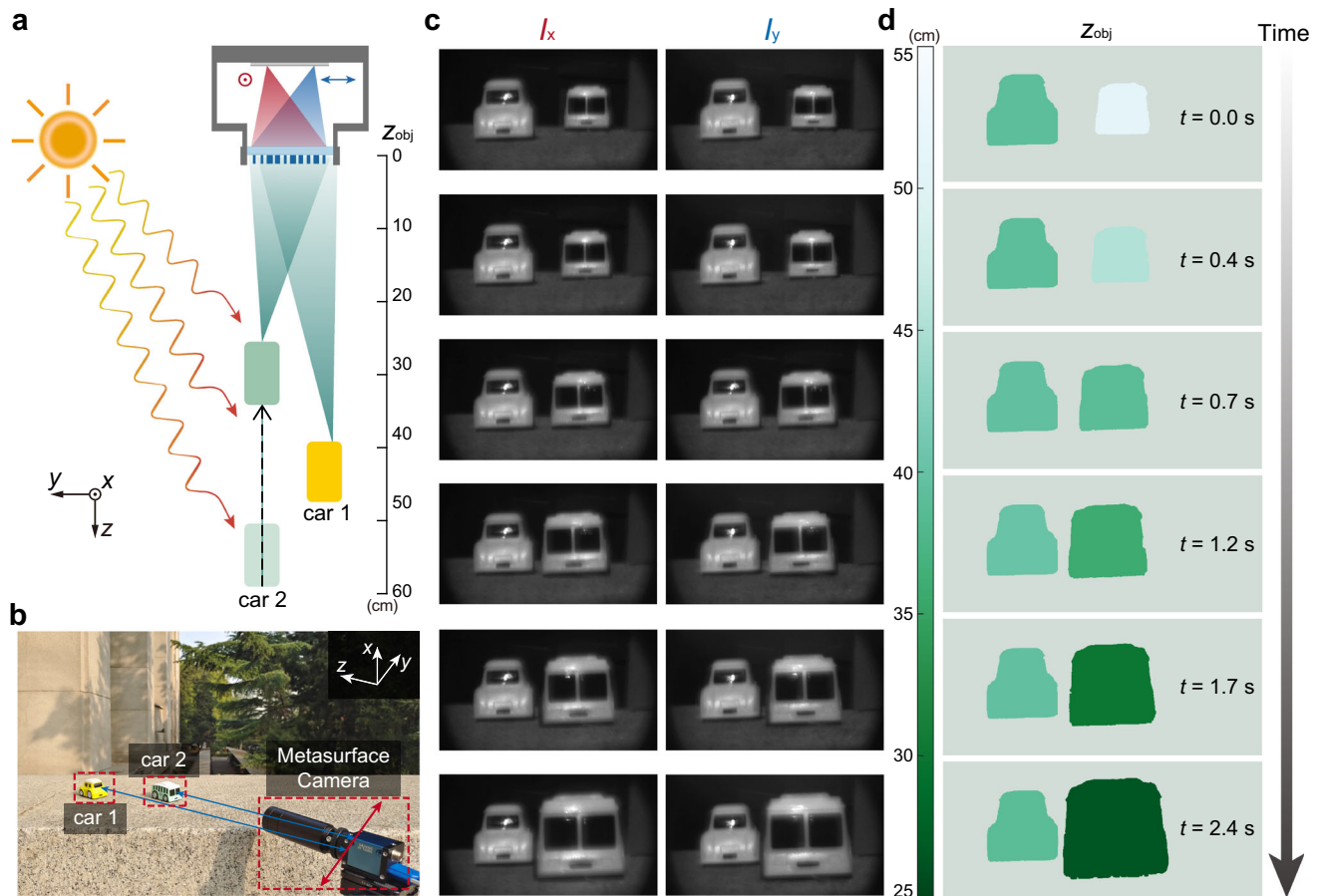
The unit-cell of the metasurface is composed of silicon nanopillars of rectangular in-plane cross-sections (Fig. 2c), thus allowing the independent control of the phase of  $x$ - and  $y$ -polarized incident light over a full  $2\pi$  range with near-unity transmittance (Supplementary Section 2). The smallest gaps between nanopillars are designed to be over 100 nm to reduce coupling between nanopillars<sup>44,45</sup>. The metasurface is fabricated using a complementary metal-oxide-semiconductor (CMOS)-compatible process (“Methods” and Supplementary Section 2), with an aperture diameter of 2 mm. The photograph, optical microscopy image, and scanning electron microscopy image of the fabricated metasurface are shown in Fig. 2d–f, respectively. We measure the polarization-dependent PSF pairs of the fabricated metasurface as a function of the axial depth of a point object ( $Z_{obj}$ ) (“Methods” and Supplementary Section 3) and confirm it is in close agreement with the calculation (Fig. 2g, h). The lens is measured to have a polarization extinction ratio of 35.6 (Supplementary Section 3) and a diffraction efficiency<sup>46</sup> of 44.54% and 43.93% for each input linear polarization (“Methods” and Supplementary Section 4). Since the single-helix PSF has a full 360° rotation range, its depth measurement range can be significantly extended compared to imaging systems with a double-helix PSF.

#### 4D imaging experiments

We assemble the metasurface with a CMOS-based photosensor with an active area of  $12.8 \times 12.8$  mm<sup>2</sup>, as shown in Fig. 3a. A bandpass filter with a central wavelength of 800 nm and a bandwidth of 10 nm and an aperture are installed in front of the camera to limit the spectral bandwidth and field of view (FOV), respectively. The assembled camera system has a size of  $3.1 \times 3.6 \times 13.5$  cm<sup>3</sup>, which may be further reduced with a customized photosensor and housing. To demonstrate single-shot 4D imaging, we set up a scene consisting of three pieces of different materials (paper, iron, and ceramic) located at different axial depths, as shown in Fig. 3b. When a partially polarized near-infrared light-emitting diode illuminates the scene, a raw image with depth and polarization information of the scene encoded can be captured (Fig. 3c).

The metasurface camera with a polarization-decoupled pair of conjugate single-helix PSFs forms two spatially separated images on the photosensor. Consequently, it avoids the superposition of twin-image presented in imaging systems with double-helix PSF. This allows high-fidelity all-in-focus 2D light intensity, depth, and polarization retrieval of the target scene using a straightforward, physically interpretable algorithm based on image segmentation and calculation of the local orientation of the translation vector of the object pair (Supplementary Section 5). It takes less than 0.4 s to reconstruct 4D images, each with up to  $500 \times 1000$  pixels, from a raw measurement on a laptop computer with Intel i7-10875H CPU and 16 GB RAM. This computation speed can be universally applied to 4D image reconstruction of any given scenes.

Figure 3d shows the retrieved all-in-focus 2D intensity images of the target scene for  $x$ - and  $y$ -polarized light, denoted as  $I_x$  and  $I_y$ , respectively. The polarization contrast can be calculated subsequently as  $I_x/I_y$ , which facilitates the clear distinction of metallic and non-metallic materials in the target scene (Fig. 3e). In comparison with the true depth, the retrieved depth map has a modest normalized mean absolute error (NMAE) value of 0.37%, defined as the mean value of the absolute error divided by its ground truth (Fig. 3f, g). We further verify that the NMAE of depth estimation can be below 1% for an alternative



**Fig. 4 | Outdoor imaging of a dynamic scene. a, b** Schematic (a) and photograph (b) of the outdoor scene for dynamic imaging experiment under sunlight illumination. Car 2 moves towards the metasurface camera at a nonuniform speed of

-10 cm/s, while car 1 stays still. **c** Decoded all-in-focus polarization image pairs  $I_x$  and  $I_y$  of the scene for selected frames of the recorded video. **d** Retrieved depth maps of the scene for selected frames of the recorded video.

scene with more objects over depths ranging from 25 cm to 34 cm (Supplementary Section 5).

A monocular camera is capable of capturing depth and intensity images for a dynamic scene under ambient lighting conditions in a single shot. To validate the concept, we record a video of a dynamic scene of moving toy cars (one toy car is kept still, the other moves at a nonuniform speed of  $\sim 10$  cm/s) using the monocular metasurface camera under sunlight illumination. The schematic and photograph of the scene are shown in Fig. 4a, b, respectively. Figures 4c and 4d show raw image pairs captured by the metasurface camera along with retrieved depth maps for selected frames of the recorded video (Supplementary Movie 1, played at  $0.3\times$  speed), which clearly reveals the absolute depth values and space-time relationship of the dynamic 3D scene with one of the toy cars moving over a distance of about 25 cm. The NMAE of depth estimation for the still and the moving toy car are 0.78% and 1.26%, respectively (Supplementary Section 5). The slightly higher depth estimation error for the outdoor dynamic scene, in comparison with indoor static scenes, may be due to the longer depth range, as well as the trade-off between signal-to-noise ratio and motion artifact of the captured images. Note that a longer integration time may lead to better signal-to-noise ratio but more image blur.

## Discussion

The prototype metasurface camera shown here can perform high-accuracy depth estimation in the centimeter range. For applications

aiming at a further distance, one can scale up the lens aperture. The depth estimation accuracy of the metasurface camera at a certain range is proportional to the rotation speed of the single-helix PSF, which is proportional to the square of the aperture size. For instance, an aperture size of 5 cm may allow depth estimation at 200-m-range with an estimated mean depth error still well below 1% (Supplementary Section 6).

The metasurface camera with the conjugate single-helix PSF has a Fisher information along the depth dimension over two orders of magnitude higher than that of a standard lens, resulting in much higher depth estimation accuracy (Supplementary Section 7). Furthermore, the single-helix PSF also allows the direct acquisition of high-quality 2D images with an extended depth of field, without the need for additional reference images (Supplementary Section 8).

The prototype metasurface camera works over a narrow spectral band and FOV. However, based on the multi-dimensional imaging framework, one may leverage the chromatic dispersion of a more sophisticated metasurface design to realize PSFs that are also strongly dependent on the wavelength of the incident light. In such a way, one may also achieve multispectral or even hyperspectral imaging using a monochromatic sensor. Using multiplexed meta-atoms, one may also achieve full-Stokes polarization imaging by directing the light of 3 pairs of orthogonal polarization states onto different areas of the photosensor plane (Supplementary Section 9). With the additional spectral or polarization channels, an important question is how to maximize the total

channel capacity without greatly sacrificing the spatial resolution or FOV of the imaging system, given the finite pixel number of the photosensor. Such a task may be partially fulfilled by compressive sensing<sup>47</sup>. To further increase the off-axis angle and the FOV of the camera system, one could use multi-level diffractive optics<sup>48</sup>, combine multiple layers of metasurfaces<sup>49,50</sup>, or by using proper aperture stops<sup>51</sup>. Although we only demonstrated depth estimation for segmented objects with uniform depth values here, ultimately, we expect that the complement of more advanced image retrieval algorithms, such as deep learning<sup>52,53</sup> and compressive sensing<sup>47,54</sup>, may facilitate accurate pixel-wise multi-dimensional image rendering for complex scenes. It is worth noticing that although deep learning has been proven to be a potent tool for numerous imaging tasks, including monocular depth estimation<sup>55–59</sup>, the method presented here is physics-driven and fully interpretable, without its operation relying on a prescribed training dataset that often has a bias towards certain scenario.

In summary, we have demonstrated a metasurface camera capable of capturing a high-fidelity 4D image, including 2D intensity, depth, and polarization in a single shot. It exploits the unprecedented ability of metasurface to manipulate the vectorial light field to allow multi-dimensional imaging using a single-piece planar optical component, with performance unattainable with conventional refractive or diffractive optics. A compact monocular camera system for multi-dimensional imaging may be useful in a myriad of application areas, including but not limited to augmented reality, robot vision, autonomous driving, remote sensing, and biomedical imaging.

## Methods

### Metasurface fabrication

The fabrication of the metasurface is done through a commercial service (Tianjin H-chip Technology Group). The process is schematically shown in Supplementary Fig. S6. The fabrication starts with a silicon-on-sapphire substrate with a thickness of the monocrystalline silicon film of 600 nm. Electron beam lithography (JEOL-6300FS) is employed to write the metasurface pattern using a negative tone resist Hydrogen silsesquioxane (HSQ). Next, the pattern is transferred to the silicon layer via dry etching directly using HSQ as the mask. Finally, the HSQ resist is removed by buffered oxide etchant.

### PSF and diffraction efficiency measurement setup

To measure the PSF of the fabricated metasurface, we construct a setup as shown in Supplementary Fig. S8. The illumination source consists of collimated light from a supercontinuum laser (YSL SC-PRO-15) and a bandpass filter (Thorlabs FB800-10) with a central wavelength of 800 nm and a bandwidth of 10 nm. The laser beam is expanded by a beam expander and focused by a convex lens with a focal length of 35 mm to generate the point light source. When measuring the PSF of the metasurface, the distance between the point light source and the metasurface is varied, with the distance between the metasurface and the photosensor kept fixed.

To estimate the polarization-dependent diffraction efficiency of the metasurface, the collimated laser beam is filtered by a linear polarizer (Thorlabs LPNIR100-MP2), with its spot size reduced using a convex lens with a focal length of 150 mm to match the aperture size of the metasurface. An optical power meter (Thorlabs PM122D) is first placed in front of the metasurface to measure the power of the incident light  $P_{\text{inc}}$ . Subsequently, to measure the power of the focused light  $P_f$ , a pinhole of 100- $\mu\text{m}$  diameter is placed in front of the power meter. The position of the power meter is spatially scanned and maximized near the designed focal

point of the metasurface. The diffraction efficiency of the metasurface is estimated as  $\eta = P_f / P_{\text{inc}}$ .

### Data availability

The data that support the plots within this paper and other findings of this study are available from the corresponding author on request.

### Code availability

The code that support the plots within this paper and other findings of this study are available from the corresponding author on request.

## References

- Cyganek, B. & Siebert, J. P. *An Introduction to 3D Computer Vision Techniques and Algorithms* (Wiley, 2011).
- Rogers, C. et al. A universal 3D imaging sensor on a silicon photonics platform. *Nature* **590**, 256–261 (2021).
- Kim, I. et al. Nanophotonics for light detection and ranging technology. *Nat. Nanotechnol.* **16**, 508–524 (2021).
- Shashar, N., Hanlon, R. T. & Petz, A. D. Polarization vision helps detect transparent prey. *Nature* **393**, 222–223 (1998).
- Grahn, H. & Geladi, P. *Techniques and Applications of Hyperspectral Image Analysis* (Wiley, 2007).
- Kadambi, A., Taamazyan, V., Shi, B. & Raskar, R. Depth sensing using geometrically constrained polarization normals. *Int. J. Comput. Vis.* **125**, 34–51 (2017).
- Geng, J. Structured-light 3D surface imaging: a tutorial. *Adv. Opt. Photonics* **3**, 128–160 (2011).
- McManamon, P. F. *Lidar Technologies and Systems* (SPIE, 2019).
- Brown, M. Z., Burschka, D. & Hager, G. D. Advances in computational stereo. *IEEE Trans. Pattern Anal. Mach. Intel.* **25**, 993–1008 (2003).
- Demos, S. G. & Alfano, R. R. Optical polarization imaging. *Appl. Opt.* **36**, 150–155 (1997).
- Oka, K. & Kato, T. Spectroscopic polarimetry with a channeled spectrum. *Opt. Lett.* **24**, 1475–1477 (1999).
- Meng, X., Li, J., Liu, D. & Zhu, R. Fourier transform imaging spectropolarimeter using simultaneous polarization modulation. *Opt. Lett.* **38**, 778–780 (2013).
- Gao, L. & Wang, L. V. A review of snapshot multidimensional optical imaging: measuring photon tags in parallel. *Phys. Rep.* **616**, 1–37 (2016).
- Greengard, A., Schechner, Y. Y. & Piestun, R. Depth from diffracted rotation. *Opt. Lett.* **31**, 181–183 (2006).
- Shechtman, Y., Sahl, S. J., Backer, A. S. & Moerner, W. E. Optimal point spread function design for 3D imaging. *Phys. Rev. Lett.* **113**, 133902–133902 (2014).
- Pentland, A. P. A new sense for depth of field. *IEEE Trans. Pattern Anal. Mach. Intel.* **9**, 523–531 (1987).
- Schechner, Y. Y. & Kiryati, N. Depth from defocus vs. stereo: how different really are they? *Int. J. Comput. Vis.* **39**, 141–162 (2000).
- Liu, S., Zhou, F. & Liao, Q. Defocus map estimation from a single image based on two-parameter defocus model. *IEEE Trans. Image Process.* **25**, 5943–5956 (2016).
- Pavani, S. R. P. et al. Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function. *Proc. Natl Acad. Sci. USA* **106**, 2995 (2009).
- Berlich, R., Brauer, A. & Stallinga, S. Single shot three-dimensional imaging using an engineered point spread function. *Opt. Express* **24**, 5946–5960 (2016).
- Quirin, S. & Piestun, R. Depth estimation and image recovery using broadband, incoherent illumination with engineered point spread functions. *Appl. Opt.* **52**, A367–A376 (2013).

22. Yu, N. et al. Light propagation with phase discontinuities: generalized laws of reflection and refraction. *Science* **334**, 333–337 (2011).
23. Yu, N. & Capasso, F. Flat optics with designer metasurfaces. *Nat. Mater.* **13**, 139–150 (2014).
24. Yang, Y. et al. Dielectric meta-reflectarray for broadband linear polarization conversion and optical vortex generation. *Nano Lett.* **14**, 1394–1399 (2014).
25. Yang, Y., Kravchenko, I. I., Briggs, D. P. & Valentine, J. All-dielectric metasurface analogue of electromagnetically induced transparency. *Nat. Commun.* **5**, 5753 (2014).
26. Arbabi, A., Horie, Y., Bagheri, M. & Faraon, A. Dielectric metasurfaces for complete control of phase and polarization with sub-wavelength spatial resolution and high transmission. *Nat. Nanotechnol.* **10**, 937–943 (2015).
27. Engelberg, J. & Levy, U. The advantages of metalenses over diffractive lenses. *Nat. Commun.* **11**, 1991 (2020).
28. Liu, M. et al. Multifunctional metasurfaces enabled by simultaneous and independent control of phase and amplitude for orthogonal polarization states. *Light Sci. Appl.* **10**, 107 (2021).
29. Guo, Q. et al. Compact single-shot metalens depth sensors inspired by eyes of jumping spiders. *Proc. Natl Acad. Sci. USA* **116**, 22959–22965 (2019).
30. Jin, C. et al. Dielectric metasurfaces for distance measurements and three-dimensional imaging. *Adv. Photonics* **1**, 036001 (2019).
31. Lin, R. J. et al. Achromatic metalens array for full-colour light-field imaging. *Nat. Nanotechnol.* **14**, 227–231 (2019).
32. Park, J. et al. All-solid-state spatial light modulator with independent phase and amplitude control for three-dimensional LiDAR applications. *Nat. Nanotechnol.* **16**, 69–76 (2021).
33. Arbabi, E., Kamali, S. M., Arbabi, A. & Faraon, A. Full-stokes imaging polarimetry using dielectric metasurfaces. *ACS Photonics* **5**, 3132–3140 (2018).
34. Rubin, N. A. et al. Matrix Fourier optics enables a compact full-Stokes polarization camera. *Science* **365**, 43 (2019).
35. Zhao, F. et al. Metalens-assisted system for underwater imaging. *Laser Photonics Rev.* **15**, 6 (2021).
36. Wei, J., Xu, C., Dong, B., Qiu, C.-W. & Lee, C. Mid-infrared semimetal polarization detectors with configurable polarity transition. *Nat. Photon.* **15**, 614–621 (2021).
37. Tittel, A. et al. Imaging-based molecular barcoding with pixelated dielectric metasurfaces. *Science* **360**, 1105–1109 (2018).
38. Wang, Z. et al. Single-shot on-chip spectral sensors based on photonic crystal slabs. *Nat. Commun.* **10**, 1020 (2019).
39. Lin, Z. et al. End-to-end metasurface inverse design for single-shot multi-channel imaging. *Opt. Express* **30**, 28358–28370 (2022).
40. Prasad, S. Rotating point spread function via pupil-phase engineering. *Opt. Lett.* **38**, 585–587 (2013).
41. Berlich, R. & Stallinga, S. High-order-helix point spread functions for monocular three-dimensional imaging with superior aberration robustness. *Opt. Express* **26**, 4873–4891 (2018).
42. Jin, C., Zhang, J. & Guo, C. Metasurface integrated with double-helix point spread function and metalens for three-dimensional imaging. *Nanophotonics* **8**, 451–458 (2019).
43. Colburn, S. & Majumdar, A. Metasurface generation of paired accelerating and rotating optical beams for passive ranging and scene reconstruction. *ACS Photonics* **7**, 1529–1536 (2020).
44. Kuznetsov, A. I., Miroshnichenko, A. E., Brongersma, M. L., Kivshar, Y. S. & Luk'yanchuk, B. Optically resonant dielectric nanostructures. *Science* **354**, aag2472 (2016).
45. Li, L., Zhao, H., Liu, C., Li, L. & Cui, T. J. Intelligent metasurfaces: control, communication and computing. *eLight* **2**, 7 (2022).
46. Engelberg, J. & Levy, U. Standardizing flat lens characterization. *Nat. Photon.* **16**, 171–173 (2022).
47. Donoho, D. L. Compressed sensing. *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006).
48. Hao, C. et al. Single-layer aberration-compensated flat lens for robust wide-angle imaging. *Laser Photonics Rev.* **14**, 2000017 (2020).
49. Arbabi, A. et al. Miniature optical planar camera based on a wide-angle metasurface doublet corrected for monochromatic aberrations. *Nat. Commun.* **7**, 13682 (2016).
50. Kim, C., Kim, S.-J. & Lee, B. Doublet metalens design for high numerical aperture and simultaneous correction of chromatic and monochromatic aberrations. *Opt. Express* **28**, 18059–18076 (2020).
51. Engelberg, J. et al. Near-IR wide-field-of-view Huygens metalens for outdoor imaging applications. *Nanophotonics* **9**, 361–370 (2020).
52. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
53. Wu, J., Cao, L. & Barbastathis, G. DNN-FZA camera: a deep learning approach toward broadband FZA lensless imaging. *Opt. Lett.* **46**, 130–133 (2021).
54. Wu, J. et al. Single-shot lensless imaging with fresnel zone aperture and incoherent illumination. *Light Sci. Appl.* **9**, 53 (2020).
55. Saxena, A., Sun, M. & Ng, A. Y. Make3d: Learning 3d scene structure from a single still image. *IEEE Trans. Pattern Anal. Mach. Intel.* **31**, 824–840 (2008).
56. Liu, F., Shen, C., Lin, G. & Reid, I. Learning depth from single monocular images using deep convolutional neural fields. *IEEE Trans. Pattern Anal. Mach. Intel.* **38**, 2024–2039 (2015).
57. Chang, J. & Wetzstein, G. Deep optics for monocular depth estimation and 3d object detection. in *Proc. International Conference on Computer Vision* 10193–10202 (2019).
58. Wu, Y., Boominathan, V., Chen, H., Sankaranarayanan, A. & Veeraraghavan, A. Phasecam3d—learning phase masks for passive single view depth estimation. in *IEEE International Conference on Computational Photography* 1–12 (2019).
59. Ikoma, H., Nguyen, C. M., Metzler, C. A., Peng, Y. & Wetzstein, G. Depth from defocus with learned optics for imaging and occlusion-aware depth estimation. in *IEEE International Conference on Computational Photography* 1–12 (2021).

## Acknowledgements

This work was supported by the National Natural Science Foundation of China (62135008, 61975251) and by the Guoqiang Institute, Tsinghua University.

## Author contributions

Z.S. designed and characterized the metasurface, developed the image retrieval algorithm with assistance from F.Z., C.J., and S.W.; Z.S., F.Z., C.J., L.C., and Y.Y. analyzed the data; Z.S. and Y.Y. prepared the manuscript with input from all authors; Y.Y. initialized and supervised the project.

## Competing interests

Y.Y. and Z.S. have submitted patent applications on technologies related to the device developed in this work. The remaining authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-36812-6>.

**Correspondence** and requests for materials should be addressed to Yuanmu Yang.

**Peer review information** *Nature Communications* thanks Cheng-Wei Qiu, Yurui Qu and Ting Xu for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023