Mengu et al. *Light: Science & Applications* (2023)12:86
https://doi.org/10.1038/s41377-023-01135-0

Official journal of the CIOMP 2047-7538
www.nature.com/lsa

ARTICLE Open Access

Snapshot multispectral imaging using a diffractive optical network

Deniz Mengu[1,2,3], Anika Tabassum[1,2,3], Mona Jarrahi[1,2,3] and Aydogan Ozcan[1,2,3]✉

Abstract

Multispectral imaging has been used for numerous applications in e.g., environmental monitoring, aerospace, defense, and biomedicine. Here, we present a diffractive optical network-based multispectral imaging system trained using deep learning to create a virtual spectral filter array at the output image field-of-view. This diffractive multispectral imager performs spatially-coherent imaging over a large spectrum, and at the same time, routes a pre-determined set of spectral channels onto an array of pixels at the output plane, converting a monochrome focal-plane array or image sensor into a multispectral imaging device without any spectral filters or image recovery algorithms. Furthermore, the spectral responsivity of this diffractive multispectral imager is not sensitive to input polarization states. Through numerical simulations, we present different diffractive network designs that achieve snapshot multispectral imaging with 4, 9 and 16 unique spectral bands within the visible spectrum, based on passive spatially-structured diffractive surfaces, with a compact design that axially spans ~$72\lambda_m$, where λ_m is the mean wavelength of the spectral band of interest. Moreover, we experimentally demonstrate a diffractive multispectral imager based on a 3D-printed diffractive network that creates at its output image plane a spatially repeating virtual spectral filter array with $2 \times 2 = 4$ unique bands at terahertz spectrum. Due to their compact form factor and computation-free, power-efficient and polarization-insensitive forward operation, diffractive multispectral imagers can be transformative for various imaging and sensing applications and be used at different parts of the electromagnetic spectrum where high-density and wide-area multispectral pixel arrays are not widely available.

Introduction

Multispectral imaging has been an instrumental tool for major advances in various fields, including environmental monitoring[1], astronomy[2–4], agricultural sciences[5,6], biological imaging[7–9], medical diagnostics[10,11], and food quality control[12,13] among many others[14–20]. One of the simplest ways to achieve multispectral imaging is to sacrifice the image acquisition time in favor of the spectral information by capturing multiple shots of a scene while changing the spectral filter in front of a monochrome camera[21]. Another traditional form of multispectral imaging relies on push-broom scanning of a one-dimensional detector array across the field-of-view (FOV)[22]. While these multispectral imaging techniques provide sufficient spectral and spatial resolution, they suffer from relatively long data acquisition times, hindering their use in real-time imaging applications[23]. An alternative solution that allows simultaneous collection of the spatial and spectral information is to split the optical waves emanating from the input FOV onto different optical paths each containing a different spectral filter, followed by a 2D monochrome image sensor array[24,25]. However, this approach often leads to more complex and bulky optical systems since it requires the use of multiple focal-plane arrays, one for each band, along with other optical components.

Modern-day snapshot spectral imaging systems often use coded apertures in conjunction with computational image recovery algorithms to digitally mitigate these shortcomings of traditional multispectral imaging systems.

Correspondence: Aydogan Ozcan (ozcan@ucla.edu)
[1]Electrical and Computer Engineering Department, University of California, Los Angeles, CA 90095, USA
[2]Bioengineering Department, University of California, Los Angeles, CA 90095, USA
Full list of author information is available at the end of the article

© The Author(s) 2023
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/.

One of the earliest forms of coded aperture snapshot spectral imaging used a binary spatial aperture function imaged onto a dispersive optical element through relay optics, encoding both the spatial and spectral features contained within the input FOV into an intensity pattern collected by a monochrome focal-plane array[26]. Since this initial proof-of-concept demonstration, various improvements have been reported on coded aperture-based snapshot spectral imaging systems based on, e.g., the use of color-coded apertures[27], compressive sensing techniques[28–31] and others[32]. On the other hand, these systems still require the use of optical relay systems and dispersive optical elements such as prisms, and diffractive elements, resulting in relatively bulky form factors; furthermore, their frame rate is often limited by the computationally intense iterative recovery algorithms that are used to digitally retrieve the multispectral image cube from the raw data. Recent studies have also reported using diffractive lens designs, addressing the form factor limitations of multispectral imaging systems[33–35]. These approaches provide restricted spatial and spectral encoding capabilities due to their limited degrees of freedom without coded apertures, causing relatively poor spectral resolution. Recent work also demonstrated the use of feedforward deep neural networks to achieve better image reconstruction quality, addressing some of the limitations imposed by the iterative reconstruction algorithms typically employed in multispectral imaging and sensing[36–38]. On the other hand, deep learning-enabled computational multispectral imagers require access to powerful graphics processing units (GPUs)[39] for rapid inference of each spectral image cube and rely on training data acquisition or a calibration process to characterize their point-spread functions[33].

With the development of high-resolution image sensor-arrays, it has become more practical to compromise spatial resolution to collect richer spectral information. The most ubiquitous form of a relatively primitive spectral imaging device designed around this trade-off is a color camera based on the Bayer filters (R, G, B channels, representing the red, green and blue spectral bands, respectively). The traditional RGB color image sensor is based on a periodically repeating array of 2 × 2 pixels, with each subpixel containing an absorptive spectral filter (also known as the Bayer filters) that transmits the red, green, or blue wavelengths while partially blocking the others. Despite its frequent use in various imaging applications, there has been a tremendous effort to develop better alternatives to these absorptive filters that suffer from a relatively high-cross-talk, low power efficiency, and poor color representation[40]. Towards this end, numerous engineered optical material structures have been explored, including plasmonic antennas[41], dielectric metasurfaces[42–46] and 3D porous materials[47–49]. While the intrinsic losses associated with metallic nanostructures limit their optical efficiency, multispectral imager designs based on dielectric metasurfaces and 3D porous compound optical elements have been reported to achieve higher power efficiencies with lower color cross-talk[40]. However, these structured material-based approaches, including various metamaterial designs, were all limited to 4 or fewer spectral channels, and did not demonstrate a large array of spectral filters for multispectral imaging. Independent from these spectral filtering techniques based on optimized meta-designs, increasing the number of unique spectral channels in conventional multispectral filters was also demonstrated, which, in general, poses various design and implementation challenges for scale-up[23,50].

Here, we introduce the design of a snapshot multispectral imager that is based on a diffractive optical network (also known as D^2NN, diffractive deep neural network[51–60]) and demonstrate its performance with 4 (2 × 2), 9 (3 × 3) and 16 (4 × 4) unique spectral bands that are periodically repeating at the output image FOV to form a *virtual* multispectral filter array. This diffractive network-based multispectral imager (Fig. 1) is trained to project the spatial information of an object onto a grid of virtual pixels, with each one carrying the information of a pre-determined spectral band, performing snapshot multispectral imaging via engineered diffraction of light through passive transmissive layers that axially span ~72 λ_m, where λ_m is the mean wavelength of the entire spectral band of interest. This unique multispectral imager design based on diffractive optical networks achieves two tasks simultaneously: (1) its acts as a broadband spatially-coherent relay optics achieving the optical imaging task between the input and the output FOVs over a wide spectral range; and (2) it spatially separates the input spectral channels into distinct pixels at the same output image plane, serving as a *virtual spectral filter array* that preserves the spatial information of the scene/object, instantaneously yielding an image cube without image reconstruction algorithms, except the standard demosaicing of the virtual filter array pixels. Stated differently, we demonstrate diffractive optical networks that virtually convert a monochrome focal-plane array or an image sensor into a snapshot multispectral imaging device without the need for conventional spectral filters.

We present different numerical diffractive network designs that achieve multispectral coherent imaging with 4, 9 and 16 unique spectral bands within the visible spectrum based on passive diffractive layers that are laterally engineered at a feature size of ~225 nm, spanning ~43 μm in the axial direction from the first layer to the last, forming a compact and scalable design. Our numerical analyses on the spectral signal contrast provided by these diffractive multispectral imagers reveal that

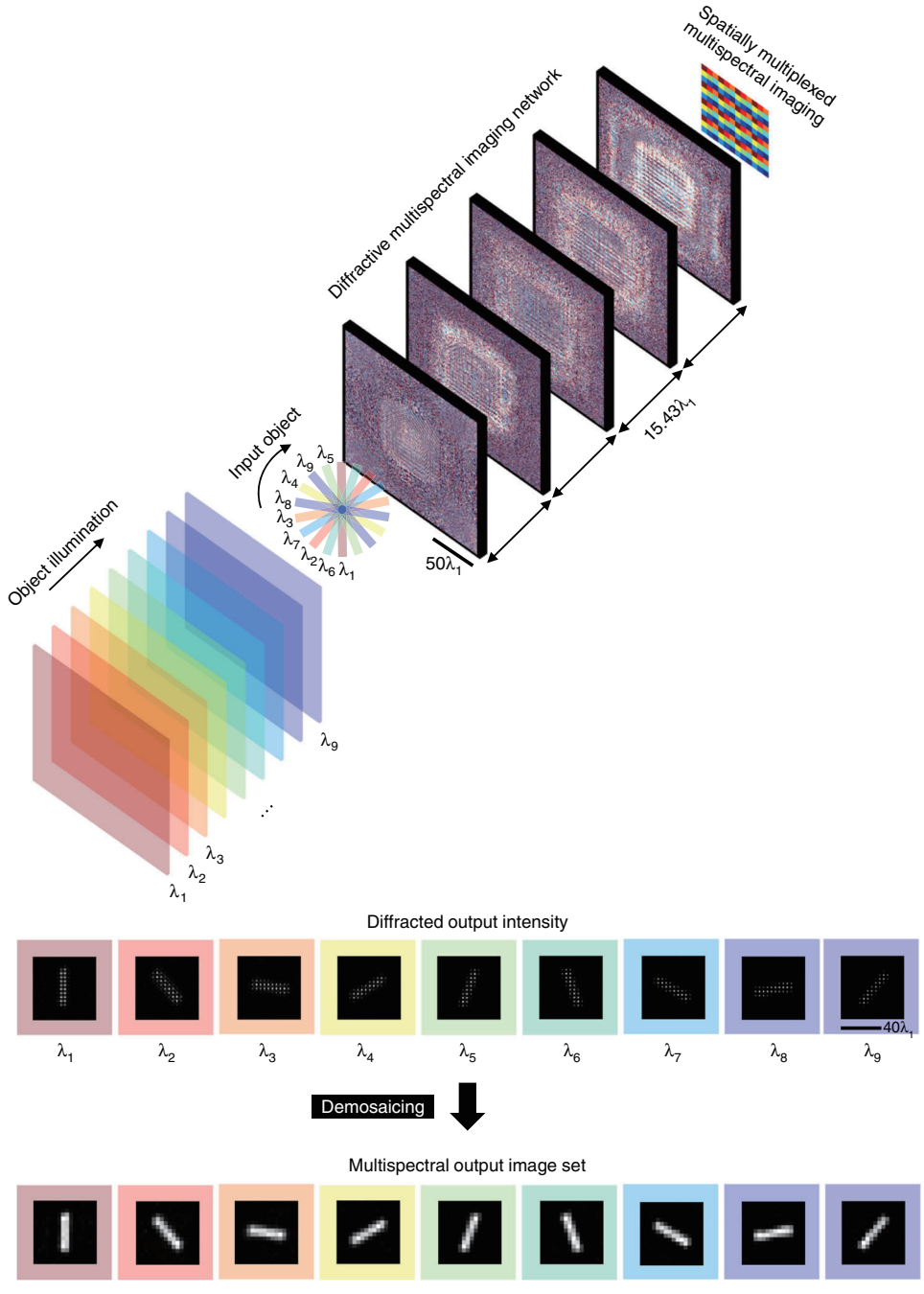

Fig. 1 The schematic of a diffractive multispectral imager. The depicted diffractive optical network simultaneously performs coherent optical imaging and spectral routing/filtering to achieve multispectral imaging by creating a periodic virtual filter array at the output. In this example, $3 \times 3 = 9$ spectral bands per virtual filter array are illustrated; Figs. 4–5 report $4 \times 4 = 16$ spectral bands per virtual filter array. In alternative implementations, the diffractive multispectral network can also be placed behind the image plane of a camera (before the image sensor), transferring the multispectral image of an object onto the plane of a monochrome image sensor

for a given array of virtual filter pixels (covering, e.g., 4, 9 and 16 spectral bands), the mean optical power of each one of the targeted spectral bands is approximately an order of magnitude larger compared to the average optical power of the other wavelengths.

Furthermore, we experimentally demonstrate the success of our diffractive multispectral imaging framework using a 3D-printed diffractive network operating at terahertz wavelengths. Targeting peak frequencies at 0.375, 0.400, 0.425 and 0.450 THz, the fabricated diffractive

network with 3 structured transmissive layers can successfully route each spectral component onto a corresponding array of virtual pixels at the output image plane, forming a multispectral coherent imager with 4 spectral channels. Although we focused on spatially-coherent multispectral imaging in this work, phase-only diffractive layers can also be optimized using deep learning to create spatially incoherent snapshot multispectral imagers, following the same design principles outlined here. With its compact form factor and snapshot operation without any image cube reconstruction algorithms, we believe the presented diffractive multispectral imaging framework can be transformative in various imaging and sensing applications. Since the presented diffractive multispectral imagers utilize isotropic dielectric materials, their virtual spectral filter arrays are not sensitive to the input polarization state of the illumination light, which provides an additional advantage. Finally, due to its scalability, it can drive the development of multispectral imagers at any part of the electromagnetic spectrum, which would be especially important for bands where high-density and large-format spectral filter arrays are not widely available or too costly.

Results

Figure 1 depicts the optical layout and the forward model of a 5-layer diffractive multispectral imager that can spatially separate N_B distinct spectral bands into a virtual spectral filter array on a monochrome image sensor located at the output image plane; in this illustration of Fig. 1, $N_B = 9$ is shown as an example, although it can be further increased, as will be reported below. The input FOV in Fig. 1 exemplifies a hypothetical object where the amplitude channel of the object's light transmission is composed of intersecting lines, and each line strictly transmits only one wavelength. Our multispectral imaging diffractive network aims to spatially separate the optical signal carried by each wavelength component on the output sensor plane so that a simple demosaicing operation would reveal the wavelength-dependent images of the input object. Such a forward optical transformation can be defined using a linear spatial mapping ($y = x$) between the input intensity describing the amplitude transmission properties of the input object at a given wavelength and the corresponding monochromatic pixels of the output sensor assigned to that targeted spectral band. This indicates that for a diffractive network-based spatially-coherent multispectral imager, there is a phase degree of freedom at the output image plane, making it easier to learn the desired multispectral imaging task through, e.g., deep learning. For a diffractive multispectral imager, as shown in Fig. 1, N_i and N_o indicate the number of effective pixels at the input and output FOVs, respectively, which are dictated by the extent of the input and output FOVs along with the desired spatial resolution (within the diffraction limit)[55,56]. The number of spectral channels (N_B) as part of the targeted multispectral imaging design depends on the cross-talk among different spectral bands of the virtual filter array created at the diffractive network output, which will be quantified in our analysis reported below. Although not demonstrated here, in alternative implementations, the diffractive multispectral network can also be placed right behind the image plane of a camera, transferring the multispectral image of an object onto the plane of the monochrome focal-plane array, converting an existing monochrome imaging system into a multispectral imager.

To train (and design) our diffractive multispectral imager, we created input objects, where the transmission field amplitude of a given object at each spectral band was represented by an image randomly selected from the 101.6 K training images of the EMNIST dataset (see the "Methods" section). The phase profiles of the five diffractive layers (containing ~0.76 million trainable diffractive features in total) were optimized through the error-backpropagation and stochastic gradient descent using a loss function based on the spatial mean-squared error (MSE) that includes all the desired spectral channels; see the "Methods" section. This deep learning-based optimization used 100 epochs, where the ground-truth multispectral output images were generated using the EMNIST dataset randomly assigned to different spectral bands of interest. Figure 2a illustrates the resulting material thickness profiles of a $K = 5$-layer diffractive multispectral imager trained to operate within the visible spectrum, evenly covering the wavelength range from $\lambda_9 = 450$ nm to $\lambda_1 = 700$ nm based on the optical layout shown in Fig. 1, i.e., $\lambda_9 < \lambda_8 < \ldots < \lambda_1$. For simplicity and without loss of generality, we assume the input light spectrum to lie between 450 nm and 700 nm; modern CMOS image sensors cover a slightly wider bandwidth than considered here. The forward optical training model of this diffractive network assumes a monochrome image sensor at the output plane with a pixel size of 0.9 μm × 0.9 μm (~$1.28\lambda_1 \times 1.28\lambda_1$), which is typical for today's CMOS image sensor technology widely deployed in, e.g., smartphone cameras[61]. This diffractive design spatially extends ~43 μm in the axial direction (from the first diffractive layer to the last layer), and is optimized to route $N_B = 9$ distinct spectral lines (i.e., 700 nm, 668.75 nm, 637.5 nm, 606.25 nm, 575 nm, 543.75 nm, 512.5 nm, 481.25 nm, and 450 nm) onto a 3 × 3 monochrome sensor pixel array, that is repeating in space for snapshot multispectral imaging without any digital image reconstruction algorithm. Without loss of generality, we assumed unit magnification between the object/input FOV and the monochrome image sensor plane (output FOV); hence, the size of the smallest feature size of the

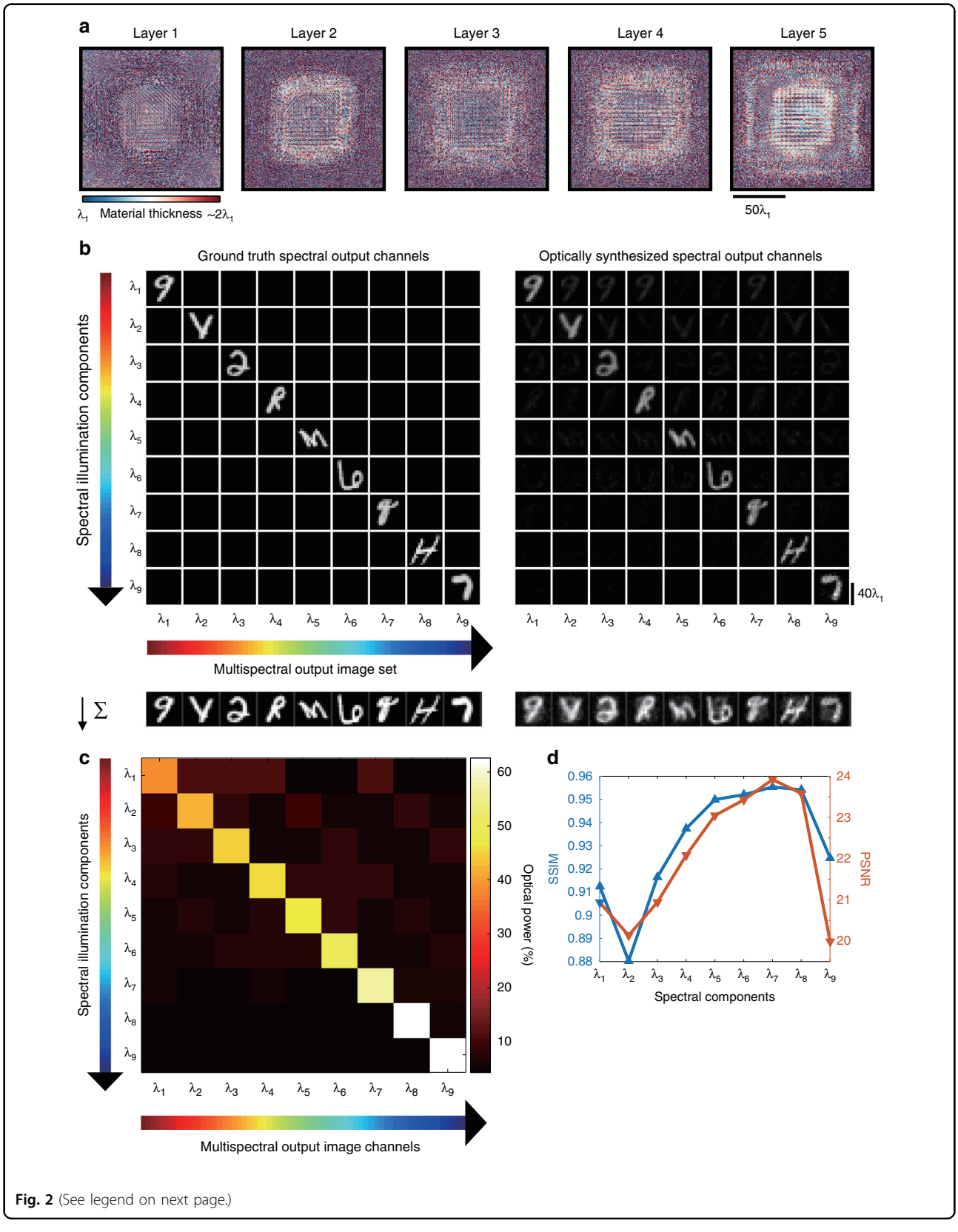

Fig. 2 (See legend on next page.)

(see figure on previous page)
Fig. 2 Performance of a diffractive multispectral imager with $N_B = 9$ spectral bands. a The material thickness distribution of the diffractive layers trained using deep learning to spatially separate 9 distinct spectral bands, creating a periodic virtual filter array. **b** Cross-talk image matrix showing the output images at different illumination wavelengths. Off-diagonal images indicate that the level of spectral cross-talk is minimal. By summing up all the images in each column, the impact of the spectral cross-talk from the other 8 spectral channels on each target wavelength is visualized at the bottom of the image matrix, as a separate row. **c** Output optical power distribution as a function of the illumination wavelength. Each row in this matrix adds up to 100%, and the off-diagonal optical power percentages indicate the level of spectral cross-talk between different bands. **d** SSIM and PSNR values of the resulting images at the output of the diffractive network; these image quality metrics were calculated between the diagonal images shown in (**b**) (the ground-truth images on the left diagonal vs. the diffractive network output images on the right diagonal)

input images was set to be 3×0.9 μm, i.e., equal to the width of a virtual spectral filter array (3×3).

Following the deep learning-based training and design phase (see the "Methods" section for further details), a multicolor image test set with a total of 2080 distinct objects (never seen during the training) was used to quantify the multispectral imaging performance of the trained diffractive network design. For each object in our blind test set, the field amplitude of the object transmission function at each spectral band was modeled based on an image randomly selected from the test dataset. An example of the imaging results corresponding to a multispectral test object never used during the training is shown in Fig. 2b. Based on the checkerboard-like output intensity patterns synthesized by the diffractive multispectral imager in response to the 2080 different test objects, the spectral image contrast of the diffractive network output can be quantified as shown in Fig. 2c; each row of the matrix in Fig. 2c corresponds to a different illumination wavelength and all the rows sum up to 100% (optical power). Hence, the rows of this matrix represent the percentage of the output optical power that resides within the designated group of virtual pixels for a given wavelength channel, calculated as an average of all the 2080 blind test objects. The columns of the matrix in Fig. 2c, on the other hand, illustrate the signal contrast and the spectral leakage over a given array of virtual spectral filters assigned to a spectral band. Our analyses show that for a given set of virtual spectral pixels assigned to a particular spectral band (a column of the matrix in Fig. 2c), the power of the desired signal band is on average (8.57 ± 1.59)-fold larger compared to the mean power of the other spectral bands (leakage) collected by the same array of virtual spectral filter pixels.

Based on the data shown in Fig. 2c, we see that the performance of the diffractive multispectral imager is inversely proportional to the wavelength. In other words, the diffractive optical network designed using deep learning can route smaller wavelengths onto their corresponding virtual spectral filter locations better than larger wavelengths. A similar conclusion can also be observed in the spectral responsibility curves of the 3×3 virtual filter array, periodically assigned to $N_B = 9$ (see Fig. 3); the responsivity curves of these virtual filter arrays get narrower as the wavelength gets smaller, with the narrowest filter response achieved for $\lambda_9 = 450$ nm. These observations can be explained based on the degrees of freedom available at each wavelength: due to the diffraction limit of light, the *effective* number of trainable diffractive features seen/controlled by larger wavelengths is smaller than the total number of trainable features within the entire diffractive network, $N = 5 \times 392 \times 392$. For example, a given diffractive layer depicted in Fig. 2a contains $N_L = 392 \times 392$ diffractive features, each with a size 225 nm × 225 nm, i.e., $\lambda_9/2 \times \lambda_9/2$, which also corresponds to $\lambda_1/3.11 \times \lambda_1/3.11$. Considering that our diffractive network operates based on traveling/propagating waves, the longer wavelengths experience reduced degrees of freedom due to the diffraction limit of light, which restricts the independent (useful) feature size on a diffractive layer to half of the wavelength in each spectral band.

Next, we further quantified the multispectral imaging quality provided by the diffractive network design shown in Fig. 2a using two additional performance metrics: Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR); see the "Methods" section. Figure 2d illustrates the average SSIM and PSNR values achieved by the diffractive network as a function of the desired spectral bands. These image quality metrics were calculated between the diagonal images shown in Fig. 2b (the ground-truth images on the left diagonal vs. the diffractive network output images on the right diagonal). Although there are some variations in the multispectral imaging quality of the diffractive network depending on the spectral band of the input light, the SSIM (PSNR) values have a very high lower bound (worst case performance) of 0.88 (19.8 dB). In addition, the mean SSIM and PSNR values are found as 0.93 and 22.06 dB, respectively. By summing up all the images in each column of Fig. 2b, we can create an image that visualizes the impact of the spectral cross-talk from the other $N_B - 1 = 8$ spectral channels on each target wavelength, which is shown at the bottom of the image matrix in Fig. 2b, as a separate row. Due to this spectral power cross-talk among channels (quantified in Fig. 2c), the average values of the SSIM and PSNR of the output multispectral image cube (computed across all the bands) drop to 0.65 and 16.24 dB,

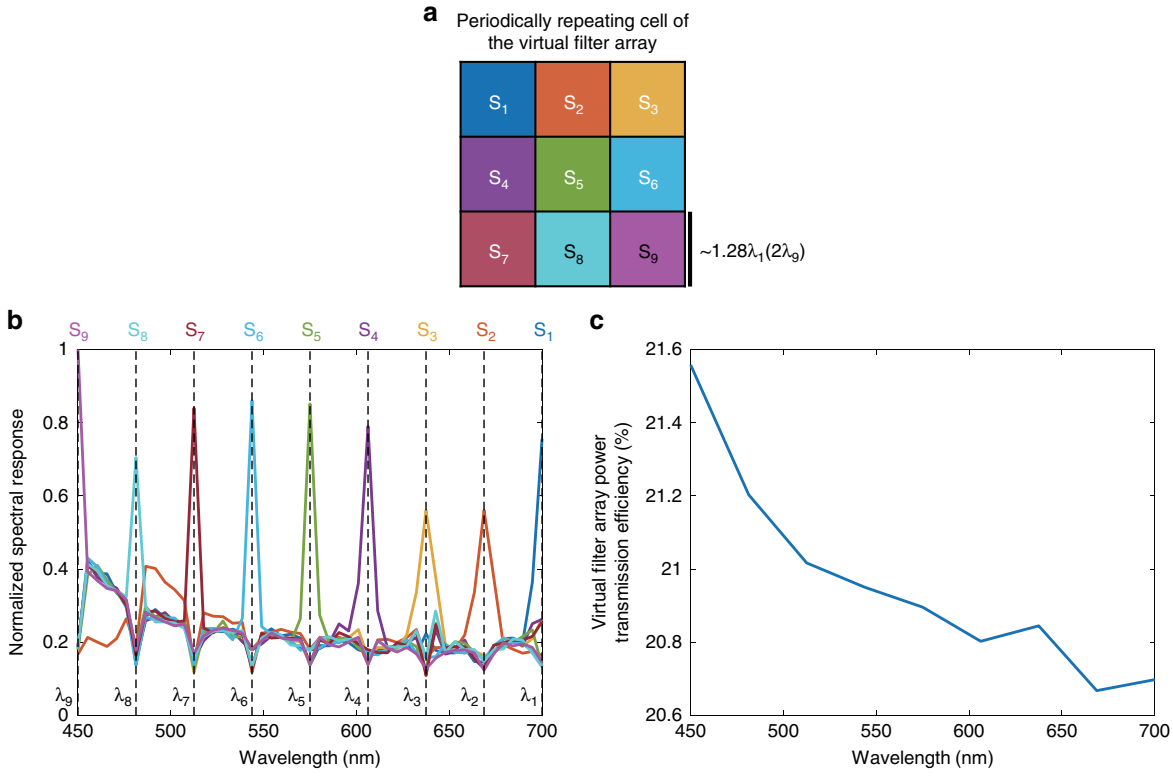

Fig. 3 Spectral responsivity and power efficiency achieved by the diffractive multispectral imager shown in Fig. 2 ($N_B = 9$). a The virtual spectral filter array periodically repeats at the output field of view of the diffractive network. The colored labels, S_i, $i = 1, 2, 3, \ldots, 9$, denote the virtual pixels assigned to the wavelength λ_i. **b** Average normalized spectral response of the virtual filter array. **c** The wavelength-dependent transmission power efficiency of the virtual filter array created by the diffractive optical network

respectively. Also see Supplementary Fig. 1 for the cross-talk matrix and multispectral imaging performance of a diffractive multispectral imager designed for $N_B = 4$ spectral bands in the visible spectrum. Due to the reduced number of target spectral bands compared to the $N_B = 9$ case, the spectral power cross-talk is reduced for the $N_B = 4$ diffractive imager as quantified in Supplementary Fig. S1b; as a result, the diffractive network can synthesize multispectral image cubes with improved mean SSIM (0.82) and mean PSNR (19.29 dB) calculated across all the $N_B = 4$ bands.

To demonstrate diffractive multispectral imaging with an increased number of spectral channels, Fig. 4a demonstrates the material thickness profiles of the diffractive layers constituting a new multispectral imager that was trained for $N_B = 16$, evenly distributed between $\lambda_{16} = 450$ nm to $\lambda_1 = 700$ nm, mapped onto a 4×4 monochrome pixel array repeating in space for snapshot multispectral imaging. Compared to the diffractive multispectral imager depicted in Fig. 2, this new diffractive design targets a lower spatial resolution due to the trade-off between N_B and the spatial resolution of the snapshot multispectral imager. Similar to Fig. 2b, the output images on the diagonals of the multispectral image cube shown in Fig. 4b closely match the ground-truth multispectral images at the input, highlighting the success of the diffractive imaging design. The off-diagonal images that are dark (see Fig. 4b) further illustrate the success of the spectral routing performed by the diffractive multispectral imager, minimizing the cross-talk among channels. Figure 4c also illustrates the average spectral signal contrast synthesized by the diffractive network at its output for $N_B = 16$ spectral bands. Compared to the signal contrast map of the previous diffractive network design ($N_B = 9$ shown in Fig. 2c), the values in Fig. 4c point to a slight decrease in the average spectral contrast at the output of this new diffractive multispectral imager with $N_B = 16$. However, the output image quality of the diffractive multispectral imager with $N_B = 16$ is still outstanding: the output SSIM (PSNR) values have a very good lower bound of 0.88 (19.62 dB), and the mean SSIM and PSNR values are 0.92 and 22.0 dB, respectively (see Fig. 4d). Same as in Fig. 2, these image quality metrics were calculated between the diagonal images shown in Fig. 4b (left vs. right). By summing up all the images in each column of Fig. 4b, we can create an image that visualizes the impact

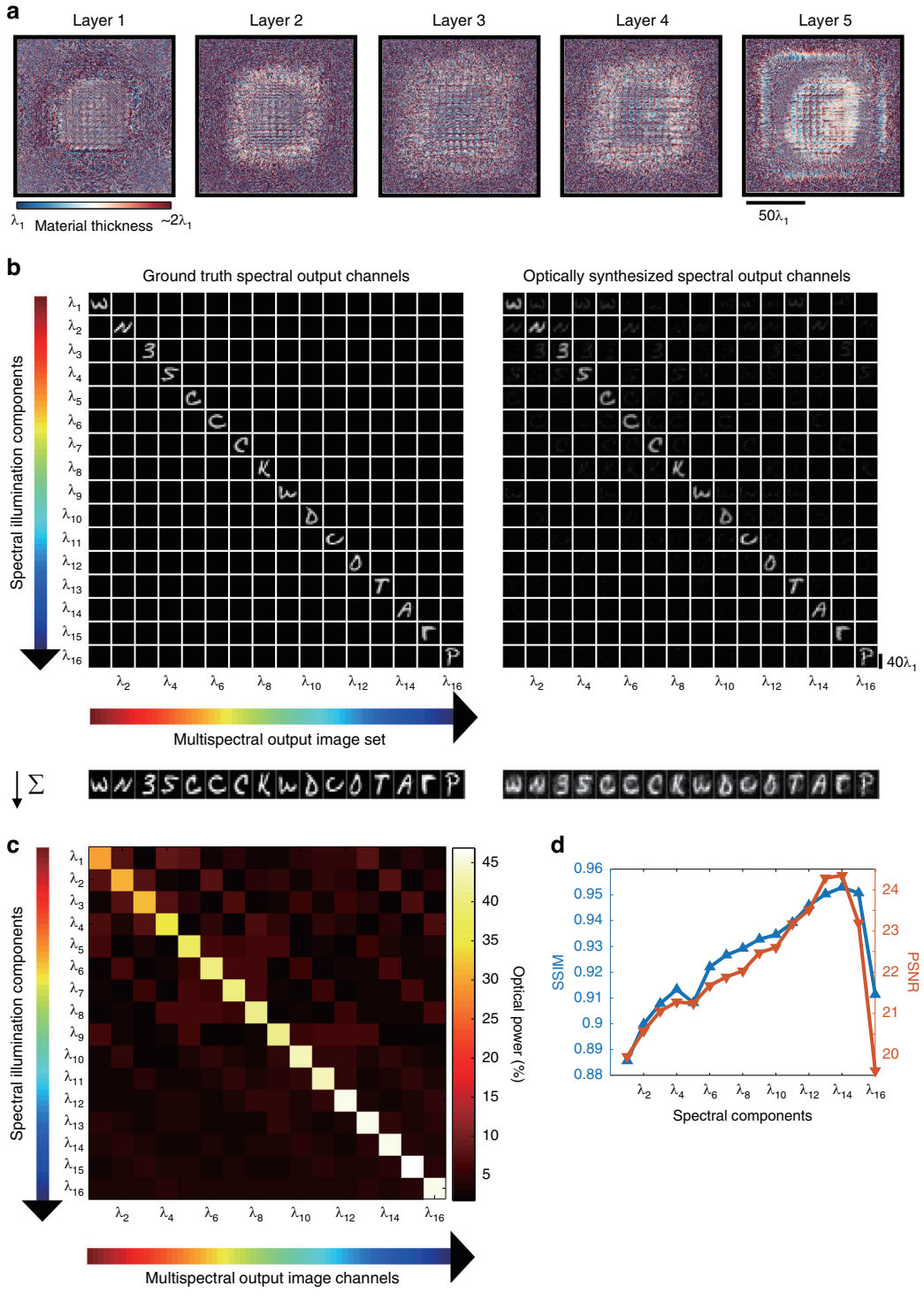

Fig. 4 Performance of a diffractive multispectral imager with $N_B = 16$ spectral bands. a The material thickness distribution of the diffractive layers trained using deep learning to spatially separate 16 distinct spectral bands, creating a periodic virtual filter array. **b** Cross-talk image matrix showing the output images at different illumination wavelengths. Off-diagonal images indicate that the level of spectral cross-talk is minimal. By summing up all the images in each column, the impact of the spectral cross-talk from the other 15 spectral channels on each target wavelength is visualized at the bottom of the image matrix, as a separate row. **c** Output optical power distribution as a function of the illumination wavelength. Each row in this matrix adds up to 100%, and the off-diagonal optical power percentages indicate the level of spectral cross-talk between different bands. **d** SSIM and PSNR values of the resulting images at the output of the diffractive network; these image quality metrics were calculated between the diagonal images shown in Fig. 4b (the ground-truth images on the left diagonal vs. the diffractive network output images on the right diagonal)

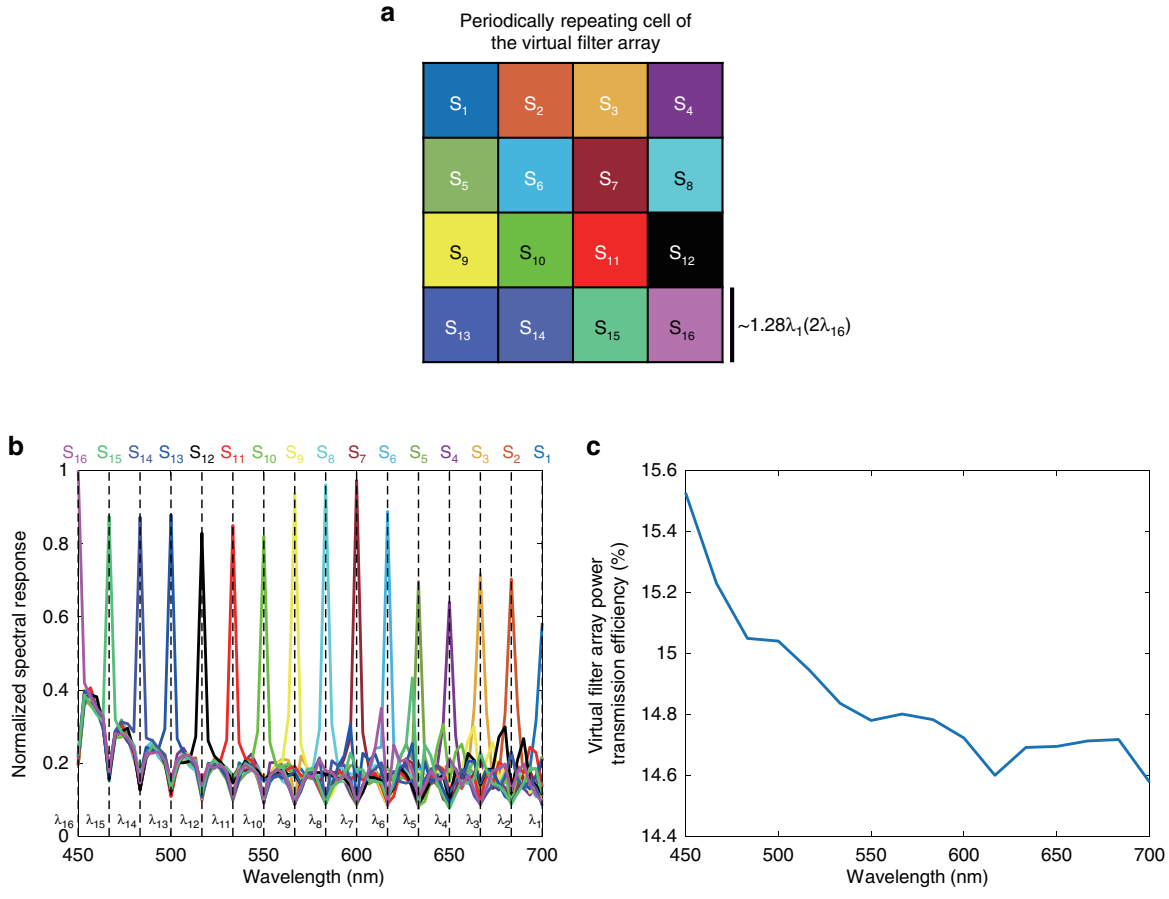

Fig. 5 Spectral responsivity and power efficiency achieved by the diffractive multispectral imager shown in Fig. 4 ($N_B = 16$). a The virtual spectral filter array periodically repeats at the output field of view of the diffractive network. The colored labels, S_i, $i = 1, 2, 3, \ldots, 16$, denote the virtual pixels assigned to the wavelength λ_i. **b** Average normalized spectral response of the virtual filter array. **c** The wavelength-dependent transmission power efficiency of the virtual filter array created by the diffractive optical network

of the power cross-talk from the other $N_B - 1 = 15$ spectral bands on each target wavelength, which is shown at the bottom of the image matrix in Fig. 4b, as a separate row. As a manifestation of the spectral power cross-talk quantified in Fig. 4c, the average values of SSIM and PSNR of the output multispectral image cube drop to 0.60 and 15.33 dB, respectively, calculated across all the $N_B = 16$ target spectral channels. Furthermore, this diffractive multispectral imager with $N_B = 16$ can route the input spectral bands onto designated output pixels with an average power contrast that is 11.06× larger with respect to the mean power carried by the remaining $N_B - 1 = 15$ spectral channels. Figure 5b also reports the spectral responsivity curves of the 4 × 4 virtual filter array at the output image FOV of the diffractive network.

Next, to experimentally demonstrate the presented diffractive multispectral imaging framework, we designed a diffractive network that can process terahertz wavelengths. This terahertz-based diffractive multispectral imager uses $K = 3$ layers (see Fig. 6) to form a virtual filter array at its output plane with periodically repeating 2×2 spectral pixels targeting 0.375, 0.4, 0.425 and 0.45 THz (i.e., $N_B = 4$). For the input object 'U' shown in Fig. 6b, the demosaiced output images predicted by the numerical forward model of our diffractive terahertz multispectral imager are depicted in Fig. 7a. In the 4-by-4 image matrix shown in Fig. 7a, the diagonal images represent the correct match between the spectral content of the illumination and the corresponding demosaiced pixels within each 2×2 cell of the virtual filter array; in other words, they represent the channels of the multispectral image cube, while the off-diagonal images show the cross-talk between different spectral bands. To quantify the performance of our diffractive multispectral imager, we compared each spectral channel of the multispectral image cube predicted by the numerical forward model of our diffractive terahertz multispectral imager with respect to the ground-truth image of the input object 'U', which achieved PSNR values of 15.12, 14.93, 15.03 and 13.30 dB for the spectral bands at 0.375, 0.4, 0.425 and

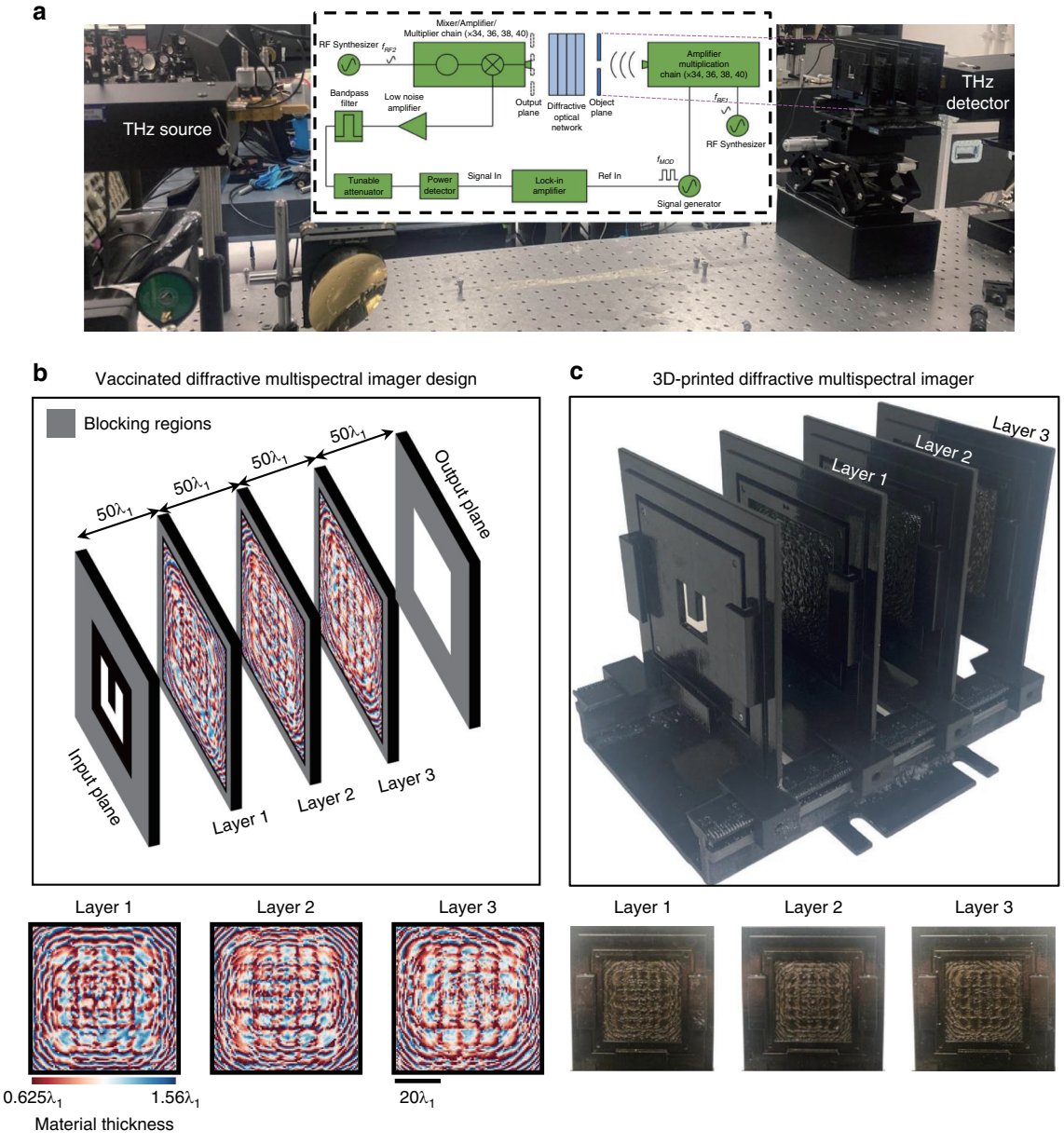

Fig. 6 Experimental setup of the multispectral terahertz imager ($N_B = 4$)**. a** Schematic of the experimental setup using terahertz illumination and signal detection. **b** Optical design layout of the fabricated diffractive multispectral imager and the material thickness profiles of the 3 diffractive surfaces. **c** Fabricated diffractive network and the input object, the letter 'U'

0.45 THz, respectively. To compare our numerical results with their experimental counterparts, Fig. 7b illustrates the experimentally measured multispectral imaging results obtained through the 3D-printed multispectral diffractive imager shown in Fig. 6c, which provided a decent agreement between our numerical and experimental multispectral images. Quantitative evaluation of the experimental multispectral imaging results reveals PSNR values of 13.02, 13.71, 13.02 and 12.64 dB PSNR at 0.375, 0.4, 0.425 and 0.45 THz, respectively. Compared to our numerical results, these PSNR values point to ~1−2 dB loss of image quality which can be largely attributed to the limited lateral resolution and potential misalignments of the 3D-printed diffractive multispectral imager shown in Fig. 6c.

Beyond the multispectral image quality, we also quantified the spectral cross-talk performance of the experimentally tested diffractive multispectral imager. Figure 7c, d illustrate the spectral cross-talk matrices generated by the numerical forward model of the diffractive

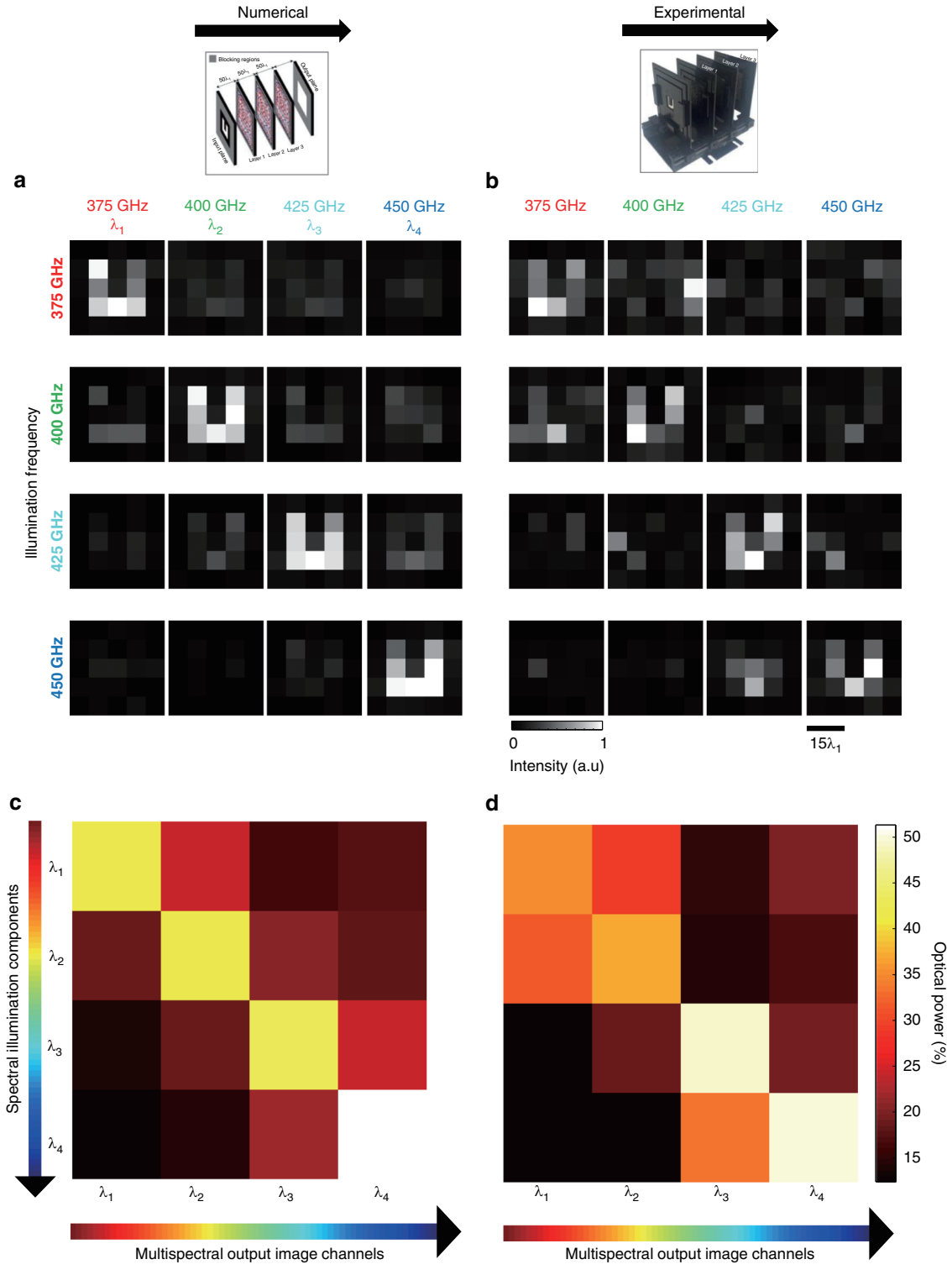

Fig. 7 Experimental results for the diffractive multispectral imager shown in Fig. 6. Multispectral imaging of letter 'U' at 4 distinct wavelengths in terahertz part of the spectrum based on the 3-layer diffractive optical network shown in Fig. 6. a Multispectral image cube synthesized by the numerical forward model of the diffractive optical network, after the demosaicing step. **b** Same as (**a**), except that the images are extracted from the experimentally measured output optical intensity profiles. **c** Cross-talk matrix predicted by the numerical forward model of the diffractive multispectral imager in response to the input object 'U'. **d** Same as (**c**), except that the entries in the cross-talk matrix represent the experimentally measured percentages of the optical power for each spectral band

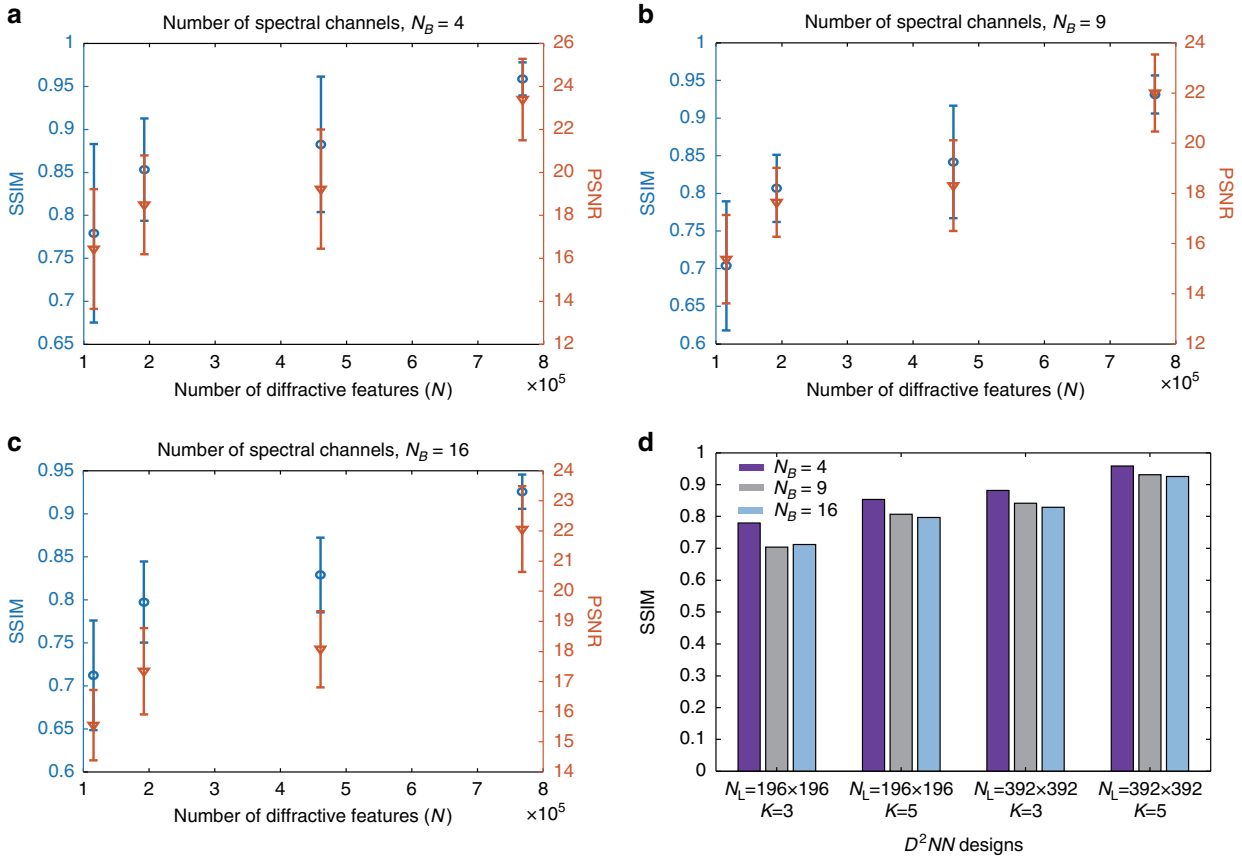

Fig. 8 The impact of the number of diffractive features (N) and the targeted spectral bands (N_B) on the quality of the multispectral imaging using a diffractive optical network. Mean SSIM and PNSR values as a function of N are reported for **a** $N_B = 4$, **b** $N_B = 9$, **c** $N_B = 16$ spectral channels, forming a spatially repeating virtual spectral filter array. **d** The impact of N_B on the SSIM of the output multispectral images for different diffractive network architectures. K refers to the number of successive diffractive layers jointly trained for multispectral imaging, and N_L refers to the number of trainable diffractive features per diffractive layer

multispectral imager shown in Fig. 6b and its experimentally measured counterpart using the 3D-printed diffractive design shown in Fig. 6c, respectively. For a given virtual filter array designated to a particular spectral band, the ratio between the mean power of the target spectral band and the mean power of all the other 3 undesired spectral bands was found to be 2.42 (numerical) and 2.21 (experimental) based on the matrices shown in Fig. 7c, d, respectively, providing a decent agreement between the numerical and experimental (3D-fabricated) models of our diffractive multispectral imager.

In general, a key design parameter in diffractive optical networks is the number of diffractive features, N, that are engineered using deep learning since it directly determines the number of independent degrees of freedom in the system[55,56,62]. Figure 8a–c compare the multispectral imaging quality achieved by four different diffractive network architectures as a function of N for $N_B = 4$, 9 and 16, respectively. For example, the diffractive multispectral imager designs for $N_B = 9$ and $N_B = 16$ shown in Figs. 2 and 4, respectively, contain in total $N = 392 \times 392 \times 5$ trainable diffractive features equally distributed over $K = 5$ diffractive layers, i.e., the number of diffractive features per layer is, $N_L = 392 \times 392$. While these two diffractive multispectral imagers can achieve average SSIM (PSNR) values of 0.93 (22.06 dB) and 0.92 (22.00 dB) at their output images, respectively, the diffractive multispectral imager architectures with fewer N cannot match their performance. For instance, in the case of a diffractive multispectral imager design based on $N_B = 9$, $N_L = 196 \times 196$ and $K = 3$ (see Fig. 8b), the average output SSIM and PSNR values drop to 0.7 and 15.38 dB, respectively. Figure 8d further illustrates the impact of N_B on the multispectral imaging performance of diffractive networks for four different combinations of N_L and K. One can observe in Fig. 8d that for a fixed N_L and K combination, the multispectral imaging performance is inversely proportional to N_B, which is expected due to the increased level of spectral multiplexing. As a comparison, the average SSIM (PSNR) values attained by the diffractive

multispectral imager with the smallest $N = 196 \times 196 \times 3$ increase from 0.7 (15.38 dB) to 0.78 (16.44 dB) when $N_B = 9$ is reduced to $N_B = 4$ spectral bands; this once again points to the relationship between N and N_B, indicating that a larger N_B would require additional diffractive degrees of freedom (a larger N) in order to perform the desired multispectral imaging task over a larger set of spectral bands.

Another critical figure of merit regarding the design of diffractive multispectral imagers is the power transmission efficiency of the optically synthesized virtual filter array. Figures 3c and 5c illustrate the power transmission efficiencies of the virtual filter arrays generated by the diffractive multispectral imager networks with $N_B = 9$ and $N_B = 16$ distinct bands within the visible spectrum. For example, based on the data depicted in Fig. 3c, the highest and lowest transmission efficiencies for $N_B = 9$, are found as 21.56% and 20.70% at 450 nm and 700 nm, respectively. On average, this diffractive multispectral imager can provide 20.96% virtual filter transmission efficiency for $N_B = 9$ spectral bands targeted by the 3×3 repeating cell of the virtual filter array. However, the deep learning-based training of this diffractive multispectral imager shown in Fig. 2 focused solely on the quality of the multispectral optical imaging, i.e., the output diffraction efficiency-related training loss term ($\mathcal{L}_e$) was dropped in Eq. 8 (see the "Methods" section). While this training strategy drives the evolution of the diffractive surfaces to maximize the multispectral imaging performance, the associated virtual filter array transmission efficiency reflects only a lower performance bound that can be achieved by a diffractive multispectral imager with the same optical architecture. To find a better balance between the multispectral imaging quality and the power efficiency of the virtual filter array, the loss function that guides the diffractive multispectral imager design during its deep learning-based training can include an additional term, $\mathcal{L}_e$, penalizing poor diffraction efficiencies (see the "Methods" section). The multiplicative constant, γ, in Eq. 8 determines the weight of the diffraction efficiency penalty, $\mathcal{L}_e$, controlling the trade-off between the multispectral imaging quality and the power efficiency of the associated virtual spectral filter array. To quantify the impact of $\mathcal{L}_e$ and γ on the performance of diffractive multispectral imagers, we trained new diffractive models that share an identical optical architecture with the diffractive multispectral imager shown in Fig. 2 ($N_B = 9$ within the visible spectrum), where each design used a different value of γ. The results of this analysis are shown in Fig. 9, which indicate that it is possible to create a 5-layer diffractive multispectral imager with $N_B = 9$, achieving an average virtual filter transmission efficiency as high as 79.32%. Furthermore, the compromise in multispectral image quality in favor of this significantly

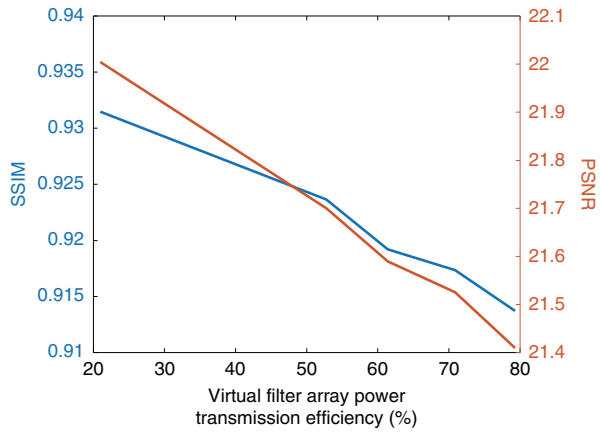

Fig. 9 The trade-off between the multispectral imaging quality and the virtual filter array power transmission efficiency. The impact of the additional power efficiency-related penalty term, $\mathcal{L}_e$ and its weight γ, on the multispectral image cube synthesized by the diffractive multispectral imagers that were trained to form 3×3 virtual spectral filter arrays assigned to $N_B = 9$ unique spectral bands within the visible spectrum. The optical architectures of these diffractive multispectral imagers are identical to the layout depicted in Fig. 1. We report an average virtual filter transmission efficiency of >79% across $N_B = 9$ spectral bands with a minimal penalty on the output image quality

increased power transmission efficiency turned out to be only minimal: while the average SSIM (PSNR) values achieved by the lower efficiency diffractive networks shown in Fig. 2 were 0.93 (22.06 dB), the more efficient diffractive multispectral imager design with 79.32% average virtual filter array transmission efficiency achieves an SSIM of ~0.91 and a PSNR of 21.42 dB (see Fig. 9).

In addition to diffraction efficiency, other practical concerns that might significantly impact the performance of the diffractive multispectral imagers include opto-mechanical misalignments and surface back-reflections. The former might be partially mitigated by using high-accuracy 3D fabrication tools such as two-photon polymerization; the latter, on the other hand, could potentially be addressed with anti-reflective coatings frequently used in the fabrication of high-quality lenses. We should also note that some of the earlier studies on multi-layer diffractive networks showed that surface reflections, in general, did not lead to a significant discrepancy between the outputs predicted by the numerical forward models/designs and their experimental counterparts[51,57,63–65]. Furthermore, some of these error sources, e.g., layer-to-layer misalignments, can directly be incorporated into the optical training forward model as random variables to drive and shape the deep learning-based evolution of the diffractive surfaces towards robust solutions that exhibit relatively flat performance curves within the possible error ranges[53]. In fact, we used this approach to 'vaccinate' the fabricated diffractive multispectral imager

shown in Fig. 6 against (1) lateral misalignments in both x and y directions, (2) axial misalignments along the optical axis and (3) in-plane diffractive layer rotations covering 4 different geometrical degrees of freedom. An important aspect of these vaccinated diffractive optical networks is that they can maintain their performance within the error ranges modeled during their training. For instance, a diffractive optical image classification network can provide a flat blind testing accuracy within the trained range of misalignments; similarly, the fabricated diffractive multispectral imager shown in Fig. 6 achieves relatively flat SSIM and PSNR curves for the output images within the error range that it was trained for. Although, this diffractive network vaccination scheme can, in principle, be extended to cover all 6 degrees of freedom, the inclusion of the two remaining rotational variations (out of the plane of each layer) brings a computational burden on the forward training model of the diffractive networks since they require the light diffraction between successive layers be accurate for tilted planes[66,67]. Beyond these sources of error discussed above, our experimental results might have also been affected by the optoelectronic detection noise and the deviation of the illumination wavefront with respect to a uniform plane wave assumed during the training.

Discussion

In the forward optical model of the presented diffractive multispectral imagers, the wave propagation in between the diffractive layers was modeled using the Rayleigh-Sommerfeld diffraction integral (see the "Methods" section), which takes into account all the propagating modes within the spatial band supported by free space, including the waves at oblique angles with respect to the optical axis; stated differently, the forward model of the presented diffractive multispectral imagers is based on a numerical aperture of 1 in air. This rich design space provided by diffractive network-based imagers optimized using deep learning opens up new avenues, such as the engineering of spatially-varying point-spread functions between an input and an output field-of-view[56,65]. We should also emphasize that the presented diffractive multispectral imager design framework using deep learning-based optimization of phase-only diffractive layers can also be extended to spatially incoherent illumination. One way to realize such a design using deep learning is to decompose each spatially incoherent wavefront at a given band into field amplitudes with random 2D input phase patterns, and the output image can be synthesized by averaging the intensities resulting from various independent random phase patterns for the same input field amplitude. The downside of such an incoherent multispectral imager design is that it would take much longer to converge using deep learning since each forward operation during the training phase would need many independent runs with random input phase patterns for each batch of the multispectral training input images. At the cost of a longer one-time training effort, phase-only diffractive layers can also be optimized using deep learning to create a spatially incoherent snapshot multispectral imager, following the same design principles outlined in this work. Therefore, the extension of the presented diffractive multispectral imager design to process spatially incoherent light is an exciting future research direction that can eventually enable the integration of these diffractive networks with existing ambient light-based camera systems for multispectral imaging and information processing.

Another interesting aspect of the presented diffractive multispectral imaging designs is that although the desired spatial distribution of different spectral bands over the output image sensor is periodic, this periodicity does not apply to the diffractive surface profiles shown in Figs. 2, 4, 6 and Supplementary Fig. S1. Despite the relatively small layer-to-layer distances used in our designs, the deep learning-based training converges to nonperiodic surface designs, one diffractive layer following another. Due to the data-driven nature of our training, the evolution of the diffractive surfaces is mainly affected by the spatial profiles of the wavelength-dependent transmission of the input objects. Stated differently, the topology of the diffractive layer designs depends on the dataset used for modeling the wavelength-dependent optical transmission of the input objects.

Finally, our diffractive designs are based on isotropic materials that do not exhibit any polarization-dependent modulation such as birefringence; therefore, a given modulation unit over a diffractive layer treats all the polarization states carried by a wavelength component equally, imposing the same phase delay regardless of the input polarization state. Hence, the multispectral imaging capability and the virtual spectral filter responses of the presented diffractive optical networks are independent of the input polarization state of the illumination light, which provides an important advantage.

In summary, we demonstrated snapshot diffractive multispectral imagers that can create a virtual spectral filter array over the pixels of a monochrome focal-plane-array or image sensor without the need for a conventional filter array, while simultaneously establishing an imaging condition between the input and output fields-of-view. Owing to their extremely compact form factor, power-efficient optical forward operation (reaching >79% filter transmission efficiency) and high-quality spectral filtering capabilities, the presented diffractive multispectral imagers can be useful for numerous imaging and sensing applications, covering different parts of the spectrum

where high-density and wide-area multispectral filter arrays are not readily available.

Materials and methods

Training forward model of diffractive multispectral imagers

Optical forward model

D^2NN framework uses deep learning to devise the transmission/reflection coefficients of diffractive features located over a series of optical modulation surfaces. The modulation coefficient over each diffractive feature/neuron is controlled through one or more physical design variables. The presented diffractive multispectral imagers in this work were designed to be fabricated based on a single dielectric material and the material thickness, h, was selected as the physical parameter for controlling the complex-valued modulation coefficient associated with each diffractive feature. For a given diffractive layer, the transmittance coefficient of a diffractive feature located on the lth layer at a coordinate of (x_q, y_q, z_l) is defined as,

$$t(x_q, y_q, z_l) = \exp\left(\frac{-2\pi\kappa h(x_q, y_q, z_l)}{\lambda}\right)\exp\left(\frac{-j2\pi(n-n_m)h(x_q, y_q, z_l)}{\lambda}\right) \tag{1}$$

where n and κ denote the real and imaginary parts of the refractive index of the fabrication dielectric material, respectively, and $n_m = 1$ corresponds to the refractive index of the propagation medium (air) between the layers. In the case of the diffractive multispectral imagers designed to operate at the visible wavelengths, the material of the diffractive layers was selected as Schott glass of type 'BK7' due to its wide availability and low absorption[68]. Since its absorption coefficient for the visible spectrum is on the order of 10^{-3} cm^{-1}, the imaginary part of the refractive index was ignored, i.e., it was assumed to be absorption-free; considering the fact that our diffractive designs extend <45 μm in the axial direction, this is a valid assumption. For the experimentally tested diffractive multispectral imaging network shown in Figs. 6–7, on the other hand, the real and imaginary parts of the diffractive materials were measured experimentally using a THz spectroscopy system, i.e., $n = 1.6524, 1.6518, 1.6512, 1.6502$, and $\kappa = 0.05, 0.06, 0.06, 0.06$, at 0.375, 0.400, 0.425 and 0.450 THz, respectively.

Each diffractive layer was modeled as a multiplicative thin modulation surface in the optical forward model. The light propagation between successive diffractive layers was implemented based on the Rayleigh-Sommerfeld scalar diffraction theory; since the smallest diffractive features considered here have a size of ~λ/2 this is a valid assumption for all-optical processing of diffraction-limited traveling/propagating fields, without any evanescent waves. According to this diffraction formulation, the free-space diffraction is interpreted as a linear, shift-invariant operator with an impulse response of,

$$w(x, y, z) = \frac{z}{r^2}\left(\frac{1}{2\pi r} + \frac{n}{j\lambda}\right)\exp\left(\frac{j2\pi n r}{\lambda}\right) \tag{2}$$

where $r = \sqrt{x^2 + y^2 + z^2}$. Based on Eq. 2, qth diffractive feature on the lth layer, at (x_q, y_q, z_l), can be described as the source of a secondary wave, generating the field in the form of,

$$w_q^l(x, y, z) = \frac{z - z_l}{\left(r_q^l\right)^2}\left(\frac{1}{2\pi r_q^l} + \frac{n}{j\lambda}\right)\exp\left(\frac{j2\pi n r_q^l}{\lambda}\right) \tag{3}$$

where $r_q^l = \sqrt{(x - x_q)^2 + (y - y_q)^2 + (z - z_l)^2}$. These secondary waves created by the diffractive features on the diffractive layer l propagate to the next layer, i.e., the $(l+1)^{th}$ layer and are spatially superimposed. Accordingly, the light field incident on the pth diffractive feature at (x_p, y_p, z_{l+1}) can be written as $\sum_q A_q^l w_q^l(x_p, y_p, z_{l+1})$, where A_q^l is the complex amplitude of the wave field right after the qth diffractive feature of the lth layer. This field is modulated through the field transmittance of the diffractive unit at (x_p, y_p, z_{l+1}), i.e., $t(x_p, y_p, z_{l+1})$, where a new secondary wave is generated, described by:

$$u_p^{l+1}(x, y, z) = w_p^{l+1}(x, y, z)t(x_p, y_p, z_{l+1})\sum_q A_q^l w_q^l(x_p, y_p, z_{l+1}) \tag{4}$$

The outlined successive modulation and the secondary wave generation processes continue until the waves propagating through the diffractive network reach the output image plane. Although the forward optical model described by Eqs. 1–4 is given over a continuous 3D coordinate system, during our deep learning-based training of the presented diffractive optical networks, all the wave fields and the modulation surfaces were represented based on their discrete counterparts. For the diffractive multispectral imager designs operating at the visible bands, the spatial sampling rate was set to be $0.5\lambda_{N_B} = 225\,\text{nm}$ for both $N_B = 4\,and\,9$, which was also equal to the size of a diffractive feature. For the experimentally tested diffractive multispectral imaging system, on the other hand, the sampling rate was selected as $0.375\lambda_{N_B}$ and the size of each diffractive feature was taken as $0.75\lambda_{N_B}$ with $N_B = 4$.

Design of diffractive multispectral imagers operating at visible bands

For a given dispersive object defined by the spectral intensity image cube, i.e., the target/ground truth, $I_{in}(x, y, \lambda)$, located at the input plane, $z = z_i$, the

underlying complex-valued field was assumed to be $U_{in}(x,y,\lambda)=\sqrt{I_{in}(x,y,\lambda)}$. In our forward model, we assumed that the input light is spatially-coherent with a constant phase front across the diffractive network input aperture (spanning a width of ~72 λ_m) at each wavelength; accordingly, the relative phase delays between different spectral components are not important, i.e., can be arbitrary, without impacting the output multispectral image intensities. Without loss of generality, diffractive multispectral imagers, depending on the application of interest, can be trained with any dispersive object model, including different input phase functions.

The size of the input/output FOVs of the diffractive multispectral imagers operating in the visible band was set to be $61.71\lambda_1\times 61.71\lambda_1$, defining a unit magnification optical imaging between the object and sensor planes. The unit magnification is not a necessary assumption for our diffractive multispectral imaging framework, and all the presented designs/methods can be extended to work under a magnification or demagnification factor, for example, by placing the diffractive layers between the image plane of a camera and a monochrome focal-plane array or image sensor. The size of each pixel at the monochrome image sensor array was assumed to be $\sim 1.28\lambda_1\times 1.28\lambda_1$, corresponding to $N_S=48$ pixels in each direction (x and y). These 48×48 pixels were grouped into 2×2, 3×3 and 4×4 blocks during the training of the diffractive multispectral imagers targeting $N_B=4$, $N_B=9$ and $N_B=16$ spectral bands, respectively. Based on these pixel grouping schemes, the EMNIST images representing the intensity patterns of the input objects were interpolated to a size 24×24, 16×16 and 12×12 pixels for the diffractive designs with $N_B=4$, $N_B=9$ and $N_B=16$ spectral bands, respectively. Note that the original size of the images in the EMNIST dataset is 28×28; hence, the ground-truth images as well as the output spectral channels shown in Figs. 2 and 4 have a slightly lower resolution than the original EMNIST data.

Each of the diffractive layers shown in Figs. 2 and 4 contains $N_L=392\times 392$ diffractive features, where the physical size of each diffractive layer was set as $126\lambda_1\times 126\lambda_1$. Since the diffractive feature size was kept identical in all the models reported in Fig. 8, the modulation surfaces constituting the diffractive multispectral imagers designed based on $N_L=196\times 196$ features per layer, occupy a smaller area of $63\lambda_1\times 63\lambda_1$. The layer-to-layer (axial) distances in all these diffractive multispectral imagers were taken as $15.43\lambda_1$.

The input intensity patterns (ground truth) describing the wavelength-dependent modulation function of the input objects, sampled at a rate $0.5\lambda_{N_B}=0.32\lambda_1$, were represented as 3D discrete vectors of size $192\times 192\times N_B$ denoted by $I_{in}[m,n,w]$ with $m=1,2,3,\ldots,192$, $n=1,2,3,\ldots,192$ and $w=1,2,3,\ldots,N_B$. The resolution of an image representing the intensity pattern of a given input object in a spectral band depends on the number of spectral bands in the system. We used a two-step interpolation to match the feature size of the input images to the size of a virtual spectral filter array. Assuming that we have N_B many images from the training dataset to represent an input object at different spectral bands, i.e., $I[x,y,w]$, each image of a given spectral band was first interpolated to a size of $N_S/\sqrt{N_B}\times N_S/\sqrt{N_B}$. These N_B low-resolution images, $I_{GT,LR}[k,r,w]$, represent the spectral channels of the ground-truth multispectral image cube extracted through the demosaicing step at the output image plane. To generate input fields matching the spatial sampling of our forward optical model, i.e., $I_{in}[m,n,w]$, in the second step, each low-resolution image was upsampled to a size of 192×192 with each pixel corresponding to an amplitude transmittance coefficient over a physical area of $0.32\lambda_1\times 0.32\lambda_1=225\times 225$ nm^2.

Based on these definitions, we used a spatial structural loss function defined as:

$$\mathcal{L}=\frac{1}{N_S}\frac{1}{N_S}\frac{1}{N_B}\sum_{q=1}^{N_S}\sum_{p=1}^{N_S}\sum_{w=1}^{N_B}\left|I_{GT}[q,p,w]-\sigma I_S[q,p,w]\right|^2 \tag{5}$$

where, I_{GT} refers to the 3D ground-truth image cube with a size of $N_S\times N_S\times N_B$, where for each spectral channel w, there are zeros introduced into proper locations representing the virtual pixels assigned to N_B-1other spectral channels for each virtual filter array period. The variable I_S in Eq. 5 denotes the optically synthesized 3D image cube at the output plane of a diffractive network that is being trained. To compute I_S based on the output optical intensity created by a diffractive optical network, $I_{out}[m,n,w]$, we applied a pixel binning based on the average pooling operator with strides on both dimensions equal to 4 (900 nm / 225 nm = 4, which refers to the ratio of the image detector pixel size to the simulation pixel size of the forward model). The multiplicative parameter, σ, in Eq. 5 is a normalization constant that accounts for the variations in the output optical power and it is updated for every batch of the training image samples based on,

$$\sigma=\frac{\sum_{q=1}^{N_S}\sum_{p=1}^{N_S}\sum_{w=1}^{N_B}I_{GT}[q,p,w]I_S[q,p,w]}{\sum_{q=1}^{N_S}\sum_{p=1}^{N_S}\sum_{w=1}^{N_B}I_S[q,p,w]^2} \tag{6}$$

To increase the output power efficiency, an additional loss term, $\mathcal{L}_e$, was utilized to balance the structural loss term defined in Eq. (5). For the power-efficient designs

depicted in Fig. 9, $\mathcal{L}_e$ was defined as $\mathcal{L}_e = e^{-\eta}$, with

$$\eta = \frac{\sum_m \sum_n \sum_w I_{out}[m,n,w]}{\sum_m \sum_n \sum_w I_{in}[m,n,w]} \times 100 \quad (7)$$

Therefore, the overall training loss function, $\mathcal{L}'$, was defined as a linear combination of $\mathcal{L}_e$ and $\mathcal{L}$, i.e.,

$$\mathcal{L}' = \mathcal{L} + \gamma \mathcal{L}_e \quad (8)$$

with the multiplicative constant γ controlling the balance between the multispectral imaging performance and the output power efficiency of the associated diffractive network model.

For a given spectral channel, w', the virtual filter array transmission efficiency, $T_{w'}$, presented in Figs. 3, 5 and 9 was calculated based on,

$$T_{w'} = \frac{\sum_k \sum_r I_{S,LR}[k,r,w']}{\sum_k \sum_r I_{GT,LR}[k,r,w']} \times 100 \quad (9)$$

where $I_{S,LR}[k,r,w']$ refers to an image of size $N_S/\sqrt{N_B} \times N_S/\sqrt{N_B}$ created by the demosaicing of $I_S[m,n,w']$. The image, $I_{GT,LR}[k,r,w']$, on the other hand, represents the $N_S/\sqrt{N_B} \times N_S/\sqrt{N_B}$ optical intensity at the spectral channel w', based on the demosaiced version of the ground-truth image, $I_{GT}[m,n,w']$.

During the training of a diffractive multispectral imager, the evolution of the phase profiles of the diffractive layers is guided through the gradients of the loss function with respect to the learnable physical parameters of the system, i.e., the material thickness values of each diffractive layer. To limit the range of the material thickness values provided by the stochastic gradient descent-based iterative updates, the thickness over each diffractive feature of a given diffractive layer was defined as a function of an associated auxiliary variable h_a,

$$h(h_a) = \frac{\sin(h_a) + 1}{2}(h_m - h_b) + h_b \quad (10)$$

where h_m and h_b denote the maximum modulation thickness and the base material thickness, respectively. For the presented diffractive multispectral imagers operating at the visible part of the electromagnetic spectrum, h_m was set to be 1.4 μm, while h_b was taken as 0.7 μm.

Design of the experimentally tested diffractive multispectral imager operating at terahertz bands

As shown in Fig. 6, the size of the input and output FOVs of the experimentally tested diffractive multispectral imager with $N_B = 4$ were set to be $37.5\lambda_1 \times 37.5\lambda_1$, where $\lambda_1 \sim 0.8$ mm is the wavelength at 0.375 THz. It was assumed that the THz output image plane has 100 (10 × 10) pixels of size $3.75\lambda_1 \times 3.75\lambda_1$. Since $N_B = 4$, these 10 × 10 pixels were divided into groups of 2 × 2 virtual spectral filters repeating in space. The fabricated diffractive multispectral imager was trained using randomly generated intensity patterns, representing the amplitude transmission of the input objects. The 3D-printed blind test object is the letter 'U' designed based on a 5 × 5 binary image with each pixel corresponding to an area of $7.5\lambda_1 \times 7.5\lambda_1$.

The size of each diffractive feature on the 3D-printed diffractive layers shown in Fig. 6 equals ~0.5 mm × 0.5 mm. Each of the 3 fabricated diffractive surfaces processes the incoming waves based on 100 × 100 optimized diffractive features, extending over $62.5\lambda_1 \times 62.5\lambda_1$. In the optical forward model of this diffractive network, all the axial distances between (1) the input FOV and the first diffractive surface, (2) two successive diffractive surfaces and (3) the last diffractive layer and the output FOV were set to be 40 mm, i.e., $\sim 50\lambda_1$. The variables h_m and h_b in Eq. 10 were taken to be $1.56\lambda_1$ and $0.625\lambda_1$, respectively.

The fabricated diffractive multispectral imager shown in Fig. 6 was trained based on $\mathcal{L}'$ depicted in Eq. 8 with $\gamma = 0.15$. Based on this γ value, the $K = 3$ layer diffractive optical network shown in Fig. 6 provides 5.68%, 5.32%, 5.2% and 5.01% virtual filter array transmission efficiency (T) for the spectral components at 0.375 THz, 0.4 THz, 0.425 THz and 0.45 THz, respectively.

The forward model of a 3D-printed diffractive network is prone to physical errors, e.g., layer-to-layer misalignments. To mitigate the impact of these experimental error sources, such misalignments were modeled as random variables and incorporated into the forward training model so that the deep learning-based evolution of the diffractive surfaces is enforced to converge to solutions that show resilience against implementation errors[53]. Accordingly, the diffractive network design shown in Fig. 6 was vaccinated against random 3D layer-to-layer misalignments in the form of lateral and axial translations as well as in-plane rotations. For this, we introduced 4 uniformly distributed random variables, D_x^l, D_y^l, D_z^l and D_θ^l, representing the random errors in the 3D location and orientation of a diffractive layer, l, i.e.,

$$\begin{aligned} D_x^l &\sim U(-\Delta_x, \Delta_x) \\ D_y^l &\sim U(-\Delta_y, \Delta_y) \\ D_z^l &\sim U(-\Delta_z, \Delta_z) \\ D_\theta^l &\sim U(-\Delta_\theta, \Delta_\theta) \end{aligned} \quad (11)$$

where Δ_x, Δ_y, Δ_z and Δ_θ denote the error range anticipated based on the fabrication margins of our experimental system. For the 3D-printed diffractive optical network shown in Fig. 6, the range of the random errors for the lateral misplacement of the diffractive surfaces was taken

as $\Delta_x = \Delta_y = 0.625\lambda_1$. The variable, Δ_z, which controls the maximum axial displacement of each layer, was set to be $2.5\lambda_1$. The range of errors in the orientation of each layer around the optical axis was assumed to be within $(-2^\circ, 2^\circ)$, i.e., $\Delta_\theta = 2^\circ$. During the training stage, D_x^l, D_y^l, D_z^l and D_θ^l were updated for each layer, l, independently for every batch of input objects, introducing a new set of random misalignment errors into the forward optical model at each error-backpropagation step.

The numerically computed and experimentally measured power cross-talk matrices shown in Fig. 7c, d, were computed based on the images of the letter 'U' at 4 different illumination wavelengths: ~0.8 mm, ~0.75 mm, ~0.7 mm and ~0.66 mm.

Details of the experimental setup

The schematic diagram of the experimental setup is given in Fig. 6. In this system, the THz wave incident on the object was generated through a horn antenna compatible with the source WR2.2 modular amplifier/multiplier chain (AMC) from Virginia Diode Inc. (VDI). Electrically modulated with a 1 kHz square wave, the AMC received an RF input signal that is a 16 dBm sinusoidal waveform at 11.111 GHz (f_{RF1}). This RF signal is multiplied 34, 36, 38 and 40 times to generate a continuous-wave (CW) radiation at ~0.375 THz, ~0.4 THz, ~0.425 THz and ~0.45 THz, corresponding to ~0.8 mm, ~0.75 mm, ~0.7 mm and ~0.66 mm in wavelength, respectively. The exit aperture of the horn antenna was placed ~60 cm away from the object plane of the 3D-printed diffractive optical network so that the beam profile of the THz illumination closely approximates a uniform plane wave. The diffracted THz light at the output plane was collected using a single-pixel Mixer/AMC from Virginia Diode Inc. (VDI). A 10 dBm sinusoidal signal at 11.083 GHz was sent to the detector as a local oscillator for mixing so that the down-converted signal is at 1 GHz. The $37.5\lambda_1 \times 37.5\lambda_1$ output FOV was scanned by placing the single-pixel detector on an XY stage that was built by combining two linear motorized stages (Thorlabs NRT100). The scanning step size was set to be 1 mm~$1.25\lambda_1$. The down-converted signal of a single-pixel detector at each scan location was sent to low-noise amplifiers (Mini-Circuits ZRL-1150-LN+) to amplify the signal by 80 dBm and a 1 GHz (+/−10 MHz) bandpass filter (KL Electronics 3C40-1000/T10-O/O) to clean the noise coming from unwanted frequency bands. Following the amplification, the signal was passed through a tunable attenuator (HP 8495B) and a low-noise power detector (Mini-Circuits ZX47-60), and then the output voltage was read by a lock-in amplifier (Stanford Research SR830). The modulation signal was used as the reference signal for the lock-in amplifier and accordingly, we conducted a calibration by tuning the attenuation and recording the lock-in amplifier readings. The lock-in amplifier readings at each scan location were converted to a linear scale according to the calibration.

The diffractive multispectral imager was fabricated using a 3D printer (Objet30 Pro, Stratasys Ltd). The optical architecture of the 3D-printed diffractive optical network consisted of an input object and 3 diffractive layers (see Fig. 6). While the active modulation area of our 3D-printed diffractive layers was 5 cm × 5 cm ($62.5\lambda_1 \times 62.5\lambda_1$), they were printed as light-modulating insets surrounded by a uniform slab of the printing material with a thickness of 2.5 mm.

Training details and image quality metrics

The image quality metrics SSIM and PSNR were computed based on the comparison between the low-resolution ground-truth image cube, $I_{GT,LR}[k, r, w]$, and the output image cube formed through the demosaicing of the optical intensity patterns collected by the image sensor, $I_{S,LR}[k, r, w]$. Both PSNR and SSIM metrics were computed separately for each spectral channel. The PSNR achieved by a diffractive multispectral imager for the spatial information in a spectral band, w', was computed based on,

$$PSNR_{w'} = 20\log_{10}\left(\frac{1}{\sqrt{\sum_k \sum_r \left|I_{GT,LR}[k, r, w'] - I_{S,LR}[k, r, w']\right|^2}}\right) \tag{12}$$

To compute the SSIM metric, we used the built-in tf.image.ssim() function in TensorFlow based on its default parameters. Each data point in SSIM and PSNR values shown in Figs. 2 and 4 represents the average value calculated using 2080 blind test objects created in a way that the amplitude channel of the spatial transmission function at each spectral band was modeled based on an image randomly selected from the 18.8 K test images of the EMNIST dataset.

The deep learning-based training of the diffractive networks was implemented using Python (v3.6.5) and TensorFlow (v1.15.0, Google Inc.). The backpropagation updates were calculated using the Adam optimizer[69], and its parameters were taken as the default values in TensorFlow and kept identical in each model. The learning rates of the diffractive optical networks were set to be 0.001. The training batch size was taken as 8 during the deep learning-based training of all the presented diffractive multispectral imagers. The training of a 5-layer diffractive multispectral imager network with 392 × 392 diffractive features per layer (for 100 epochs) takes approximately 2 weeks using a computer with a GeForce GTX 1080 Ti Graphical Processing Unit (GPU, Nvidia Inc.) and Intel® Core ™ i7-8700 Central Processing Unit

(CPU, Intel Inc.) with 64 GB of RAM, running Windows 10 operating system (Microsoft). Although the training time for the deep learning-based design of a diffractive multispectral imager is relatively long, it should be noted that this is a one-time effort. Once the diffractive multispectral imager is fabricated following the training stage, its physical forward optical operation consumes no power except the illumination beam.

Acknowledgements
This research is supported by the U.S. Department of Energy (DOE), Office of Basic Energy Sciences, Division of Materials Sciences and Engineering under Award # DE-SC0023088.

Author details
[1]Electrical and Computer Engineering Department, University of California, Los Angeles, CA 90095, USA. [2]Bioengineering Department, University of California, Los Angeles, CA 90095, USA. [3]California NanoSystems Institute, University of California, Los Angeles, CA 90095, USA

Author contributions
A.O. conceived and initiated the research. D.M. and A.T. conducted experiments. D.M designed the diffractive multispectral imagers. D.M. processed the resulting data. All the authors contributed to the preparation of the manuscript. A.O. and M.J. supervised the research.

Conflict of interest
The authors declare no competing interests.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41377-023-01135-0.

Received: 30 November 2022 Revised: 14 March 2023 Accepted: 17 March 2023
Published online: 06 April 2023

References

1. Kargel, J. S. et al. Multispectral imaging contributions to global land ice measurements from space. *Remote Sens. Environ.* **99**, 187–219 (2005).
2. Bell, J. F. et al. Pancam multispectral imaging results from the opportunity rover at Meridiani Planum. *Science* **306**, 1703–1709 (2004).
3. Dinguirard, M. & Slater, P. N. Calibration of space-multispectral imaging sensors: a review. *Remote Sens. Environ.* **68**, 194–205 (1999).
4. Berry, S. et al. Analysis of multispectral imaging with the AstroPath platform informs efficacy of PD-1 blockade. *Science* **372**, eaba2609 (2021).
5. Boelt, B. et al. Multispectral imaging – a new tool in seed quality assessment? *Seed Sci. Res.* **28**, 222–228 (2018).
6. De Oca, A. M. et al. Low-cost multispectral imaging system for crop monitoring. In *Proc 2018 International Conference on Unmanned Aircraft Systems (ICUAS)*, 443–451. (IEEE, Dallas, 2018).
7. Rouse, A. R. & Gmitro, A. F. Multispectral imaging with a confocal microendoscope. *Opt. Lett.* **25**, 1708–1710 (2000).
8. Levenson, R. M. & Mansfield, J. R. Multispectral imaging in biology and medicine: Slices of life. *Cytometry* **69A**, 748–758 (2006).
9. McGrath, K. E., Bushnell, T. P. & Palis, J. Multispectral imaging of hematopoietic cells: where flow meets morphology. *J. Immunol. Methods* **336**, 91–97 (2008).
10. Halicek, M. et al. In-vivo and ex-vivo tissue analysis through hyperspectral imaging techniques: revealing the invisible features of cancer. *Cancers* **11**, 756 (2019).
11. Levenson, R. M., Fornari, A. & Loda, M. Multispectral imaging and pathology: seeing and doing more. *Expert Opin. Med. Diagnostics* **2**, 1067–1081 (2008).
12. Feng, C. H. et al. Hyperspectral imaging and multispectral imaging as the novel techniques for detecting defects in raw and processed meat products: current state-of-the-art research advances. *Food Control* **84**, 165–176 (2018).
13. Qin, J. et al. Hyperspectral and multispectral imaging for evaluating food safety and quality. *J. Food Eng.* **118**, 157–171 (2013).
14. Elias, M. & Cotte, P. Multispectral camera and radiative transfer equation used to depict Leonardo's sfumato in Mona Lisa. *Appl. Opt.* **47**, 2146–2154 (2008).
15. Pelagotti, A. et al. Multispectral imaging of paintings. *IEEE Signal Process. Mag.* **25**, 27–36 (2008).
16. Cosentino, A. Identification of pigments by multispectral imaging; a flowchart method. *Herit. Sci.* **2**, 8 (2014).
17. Easton, R. L., Knox, K. T. & Christens-Barry, W. A. Multispectral imaging of the Archimedes palimpsest. In *Proc 32nd Applied Imagery Pattern Recognition Workshop, 2003*, 111–116 (IEEE, Washington, 2003).
18. Ortega, S. et al. Use of hyperspectral/multispectral imaging in gastroenterology. shedding some–different–light into the dark. *J. Clin. Med.* **8**, 36 (2019).
19. Eisenbeiß, W., Marotz, J. & Schrade, J. P. Reflection-optical multispectral imaging method for objective determination of burn depth. *Burns* **25**, 697–704 (1999).
20. Shimoni, M., Haelterman, R. & Perneel, C. Hypersectral imaging for military and security applications: combining myriad processing and sensing techniques. *IEEE Geosci. Remote Sens. Mag.* **7**, 101–117 (2019).
21. Zhang, C. et al. A novel 3D multispectral vision system based on filter wheel cameras. In *Proc 2016 IEEE International Conference on Imaging Systems and Techniques (IST)*, 267–272 (IEEE, Chania, 2016).
22. Thompson, L. L. *Remote Sensing Using Solid-state Array Technology* (NTRS, 1979).
23. Chen, Z. Y., Wang, X. & Liang, R. G. RGB-NIR multispectral camera. *Opt. Express* **22**, 4985–4994 (2014).
24. Fletcher-Holmes, D. W. & Harvey, A. R. Real-time imaging with a hyperspectral fovea. *J. Opt. A Pure Appl. Opt.* **7**, S298–S302 (2005).
25. Weitzel, L. et al. 3D: the next generation near-infrared imaging spectrometer. *Astron. Astrophys. Suppl. Ser.* **119**, 531–546 (1996).
26. Wagadarikar, A. et al. Single disperser design for coded aperture snapshot spectral imaging. *Appl. Opt.* **47**, B44–B51 (2008).
27. Arguello, H. & Arce, G. R. Colored coded aperture design by concentration of measure in compressive spectral imaging. *IEEE Trans. Image Process.* **23**, 1896–1908 (2014).
28. Correa, C. V., Arguello, H. & Arce, G. R. Snapshot colored compressive spectral imager. *J. Opt. Soc. Am. A* **32**, 1754–1763 (2015).
29. Lin, X. et al. Dual-coded compressive hyperspectral imaging. *Opt. Lett.* **39**, 2044–2047 (2014).
30. Lin, X. et al. Spatial-spectral encoded compressive hyperspectral imaging. *ACM Trans. Graph.* **33**, 233 (2014).
31. August, Y. et al. Compressive hyperspectral imaging by random separable projections in both the spatial and the spectral domains. *Appl. Opt.* **52**, D46–D54 (2013).
32. Wu, Y. H. et al. Development of a digital-micromirror-device-based multishot snapshot spectral imaging system. *Opt. Lett.* **36**, 2692–2694 (2011).
33. Wang, P. & Menon, R. Ultra-high-sensitivity color imaging via a transparent diffractive-filter array and computational optics. *Optica* **2**, 933–939 (2015).
34. Heide, F. et al. Encoded diffractive optics for full-spectrum computational imaging. *Sci. Rep.* **6**, 33543 (2016).
35. Jeon, D. S. et al. Compact snapshot hyperspectral imaging with diffracted rotation. *ACM Trans. Graph.* **38**, 117 (2019).
36. Arguello, H. et al. Shift-variant color-coded diffractive spectral imaging system. *Optica* **8**, 1424–1434 (2021).
37. Dun, X. et al. Learned rotationally symmetric diffractive achromat for full-spectrum computational imaging. *Optica* **7**, 913–922 (2020).
38. Ballard, Z. et al. Machine learning and computation-enabled intelligent sensor design. *Nat. Mach. Intell.* **3**, 556–565 (2021).
39. Wetzstein, G. et al. Inference in artificial intelligence with deep optics and photonics. *Nature* **588**, 39–47 (2020).
40. Chen, M. J. et al. Full-color nanorouter for high-resolution imaging. *Nanoscale* **13**, 13024–13029 (2021).
41. Shegai, T. et al. A bimetallic nanoantenna for directional colour routing. *Nat. Commun.* **2**, 481 (2011).

42. Chen, B. H. et al. GaN Metalens for pixel-level full-color routing at visible light. *Nano Lett.* **17**, 6345–6352 (2017).
43. Zou, X. J. et al. Pixel-level Bayer-type colour router based on metasurfaces. *Nat. Commun.* **13**, 3288 (2022).
44. Li, J. H. et al. Single-layer Bayer metasurface via inverse design. *ACS Photonics* **9**, 2607–2613 (2022).
45. Miyata, M., Nakajima, M. & Hashimoto, T. High-sensitivity color imaging using pixel-scale color splitters based on dielectric metasurfaces. *ACS Photonics* **6**, 1442–1450 (2019).
46. Nishiwaki, S. et al. Efficient colour splitters for high-pixel-density image sensors. *Nat. Photonics* **7**, 240–246 (2013).
47. Sell, D. et al. Periodic dielectric metasurfaces with high-efficiency, multiwavelength functionalities. *Adv. Opt. Mater.* **5**, 1700645 (2017).
48. Camayd-Muñoz, P. et al. Multifunctional volumetric meta-optics for color and polarization image sensors. *Optica* **7**, 280–283 (2020).
49. Zhao, N., Catrysse, P. B. & Fan, S. H. Perfect RGB-IR color routers for sub-wavelength size CMOS image sensor pixels. *Adv. Photonics Res.* **2**, 2000048 (2021).
50. Brauers, J. & Aach, T. A color filter array based multispectral camera. In *12. Workshop Farbbildverarbeitung*, 55–64 (Ilmenau, 2006).
51. Lin, X. et al. All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).
52. Mengu, D. et al. Analysis of diffractive optical neural networks and their integration with electronic neural networks. *IEEE J. Sel. Top. Quantum Electron.* **26**, 3700114 (2020).
53. Mengu, D. et al. Misalignment resilient diffractive optical networks. *Nanophotonics* **9**, 4207–4219 (2020).
54. Bai, B. J. et al. To image, or not to image: class-specific diffractive cameras with all-optical erasure of undesired objects. *eLight* **2**, 14 (2022).
55. Kulce, O. et al. All-optical information-processing capacity of diffractive surfaces. *Light Sci. Appl.* **10**, 25 (2021).
56. Kulce, O. et al. All-optical synthesis of an arbitrary linear transformation using diffractive surfaces. *Light Sci. Appl.* **10**, 196 (2021).
57. Luo, Y. et al. Computational imaging without a computer: seeing through random diffusers at the speed of light. *eLight* **2**, 4 (2022).
58. Mengu, D. & Ozcan, A. All-optical phase recovery: diffractive computing for quantitative phase imaging. *Adv. Opt. Mater.* **10**, 2200281 (2022).
59. Goi, E. et al. Nanoprinted high-neuron-density optical linear perceptrons performing near-infrared inference on a CMOS chip. *Light Sci. Appl.* **10**, 40 (2021).
60. Li, J. X. et al. Massively parallel universal linear transformations using a wavelength-multiplexed diffractive optical network. *Adv. Photonics* **5**, 016003 (2023).
61. Hasegawa, T. et al. A new 0.8 μm CMOS image sensor with low RTS noise and high full well capacity. *IISW Dig. Tech. Pap.* **1**, 24–27 (2019).
62. Li, J. X. et al. Polarization multiplexed diffractive computing: all-optical implementation of a group of linear transformations through a polarization-encoded diffractive network. *Light Sci. Appl.* **11**, 153 (2022).
63. Li, J. X. et al. Spectrally encoded single-pixel machine vision using diffractive networks. *Sci. Adv.* **7**, eabd7690 (2021).
64. Veli, M. et al. Terahertz pulse shaping using diffractive surfaces. *Nat. Commun.* **12**, 37 (2021).
65. Mengu, D. et al. Diffractive interconnects: all-optical permutation operation using diffractive networks. *Nanophotonics*. https://doi.org/10.1515/nanoph-2022-0358 (2022).
66. Matsushima, K., Schimmel, H. & Wyrowski, F. Fast calculation method for optical diffraction on tilted planes by use of the angular spectrum of plane waves. *J. Opt. Soc. Am. A* **20**, 1755–1762 (2003).
67. Delen, N. & Hooker, B. Free-space beam propagation between arbitrarily oriented planes based on full diffraction theory: a fast Fourier transform approach. *J. Opt. Soc. Am. A* **15**, 857–867 (1998).
68. N-BK7 | SCHOTT advanced optics. http://www.schott.com/shop/advanced-optics/en/Optical-Glass/N-BK7/c/glass-N-BK7 (2022).
69. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. In *Proc 3rd International Conference on Learning Representations* (ICLR, San Diego, CA, USA, 2015).