

## ARTICLE

<https://doi.org/10.1038/s41467-022-29178-8>

OPEN

# All-fiber high-speed image detection enabled by deep learning

Zhoutian Liu<sup>1</sup>, Lele Wang<sup>1</sup>, Yuan Meng<sup>1</sup>, Tiantian He<sup>1</sup>, Sifeng He<sup>1</sup>, Yousi Yang<sup>1</sup>, Liuyue Wang<sup>1</sup>, Jiading Tian<sup>1</sup>, Dan Li<sup>1,2</sup>, Ping Yan<sup>1,2</sup>, Mali Gong<sup>1,2</sup>, Qiang Liu<sup>1,2</sup> & Qirong Xiao<sup>1,2</sup>✉

Ultra-high-speed imaging serves as a foundation for modern science. While in biomedicine, optical-fiber-based endoscopy is often required for in vivo applications, the combination of high speed with the fiber endoscopy, which is vital for exploring transient biomedical phenomena, still confronts some challenges. We propose all-fiber imaging at high speeds, which is achieved based on the transformation of two-dimensional spatial information into one-dimensional temporal pulsed streams by leveraging high intermodal dispersion in a multimode fiber. Neural networks are trained to reconstruct images from the temporal waveforms. It can not only detect content-aware images with high quality, but also detect images of different kinds from the training images with slightly reduced quality. The fiber probe can detect micron-scale objects with a high frame rate (15.4 Mfps) and large frame depth (10,000). This scheme combines high speeds with high mechanical flexibility and integration and may stimulate future research exploring various phenomena in vivo.

<sup>1</sup>State Key Laboratory of Precision Measurement Technology and Instruments, Department of Precision Instrument, Tsinghua University, Beijing 100084, China. <sup>2</sup>Key Laboratory of Photonic Control Technology, Ministry of Education, Tsinghua University, Beijing 100084, China.

✉email: [xiaoqirong@mail.tsinghua.edu.cn](mailto:xiaoqirong@mail.tsinghua.edu.cn)

Ultra-high-speed imaging is vital for observing microscopic and transient physical phenomena<sup>1</sup>. To date, silicon-based imaging sensors, including charge-coupled device (CCD) and complementary metal-oxide-semiconductor (CMOS) cameras, have achieved imaging speeds of up to millions of frames per second (fps)<sup>2</sup>. Some advanced systems have also been invented for even faster transient imaging, reaching trillions of fps, including sequentially timed all-optical mapping photography (STAMP)<sup>3</sup>, frequency-domain tomography<sup>4</sup>, femtosecond time-resolved optical polarimetry<sup>5</sup>, and compressed ultrafast spectral photography<sup>6</sup>. These advanced technologies have helped researchers better understand various transient phenomena, such as lattice dynamics<sup>1</sup>, hot-electron diffusion<sup>7</sup>, the evolution of laser ablation<sup>8</sup>, and the production of electronic plasmas<sup>9</sup>. However, in some other fields, especially in vivo applications<sup>10</sup>, high-speed detection requires imaging in narrow spaces, for which the emerging fiber-based imaging technology has unique advantages.

In contrast to bulk imaging systems, fiber-based imaging systems feature high mechanical flexibility, compact sizes, and resistance to ambient interference. These features have made fiber-based imaging a competitive candidate for detecting images under special circumstances, for example, in environments with high temperatures, pressures, or radiation levels. Fiber probes can also penetrate deep into narrow spaces for endoscopy, which is essential in fields such as biomedicine<sup>11</sup> and microfluidics<sup>12</sup>. Fiber endoscopy with a high frame rate is especially necessary in some special scenarios. For instance, a fiber probe can be inserted into the cerebral cortex to examine the fast signals of neural activation<sup>13</sup> or used in vivo to observe chemical dynamics in living tissues<sup>14</sup>. In physics and engineering, such probes can also be used for observing transient physical reactions in closed containers<sup>15</sup> or exploring fuel injection dynamics in internal combustion engines.

For currently prevalent fiber-based imaging systems, the basic principle involves analyzing the light fields at the output fiber facet and reconstructing two-dimensional (2D) images using transmission matrix methods and deep learning methods<sup>16–19</sup>. Due to this principle, they must detect different frame fields at a fixed position, which means that they can only use conventional single-sensor cameras (special cameras such as rotating-mirror cameras<sup>20</sup> and framing cameras with higher frame rates are inapplicable). However, the traditional cameras generally require a balance between the imaging speed and the frame depth (number of frames that can be captured in a single shot) due to a limited readout speed from the pixel arrays to memory<sup>2</sup>. To the best of our knowledge, the world's fastest single-sensor camera has a frame rate of 10 Mfps and frame depth of 256 frames<sup>21</sup>, which places an upper limit on performance of the current fiber-based systems in high-speed imaging. Moreover, the silicon-based cameras are sensitive only to wavelengths below 1.1  $\mu\text{m}$ <sup>22</sup>, which also limits the applications of these systems in longer infrared bands. In addition, free-space optical elements are commonly required in the collection of output fields from the fiber end, which reduces the level of integration and makes these systems susceptible to environmental disturbances.

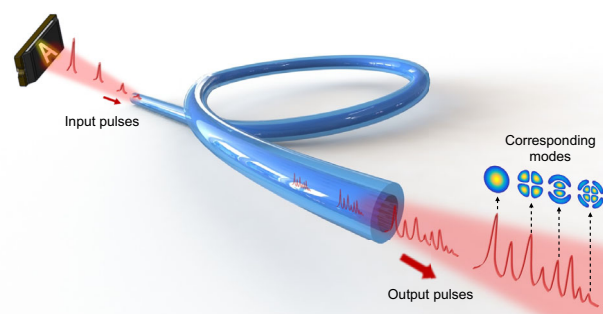
A single-pixel imaging method, termed serial time-encoded amplified microscopy<sup>23–25</sup>, has been proposed to eliminate pixelated sensors by encoding the spatial information of objects into time-domain signals, which requires only a one-pixel detector. Since each optical pulse can carry the information of one image frame, a high frame rate can be achieved by recording the temporal signals of a pulse train with a high repetition rate. Moreover, the use of one-pixel detectors, such as InGaAs photodiodes can extend the detection wavelengths to longer infrared bands. However, such systems require bulk spatial dispersers, which are not compatible with fiber endoscopy.

Here, we combine the advantages of the time-stretching method and fiber endoscopy and propose a one-pixel method to enable all-fiber high-speed detection of images. Using a single multimode fiber (MMF) as the probe, real-time image acquisition with a frame rate of over 15 Mfps and a shutter time of 45.1 ps was experimentally demonstrated, in which 10,000 frames could be recorded in a single shot. We also verified that the maximum frame rate of the system can be further enhanced to 53.5 Mfps. Leveraging the intermodal dispersion effect in an MMF, we transformed 2D spatial information into one-dimensional (1D) time-domain pulsed waveforms. A neural network model was trained to reconstruct images from the temporal waveforms recorded by an ultrafast photodiode connected to the output end of the fiber. In addition, we propose an all-fiber structure by combining a fiber-output pulse laser, a triple-cladding fiber probe, and a side-pump coupler. This scheme enables high levels of integration and system stability.

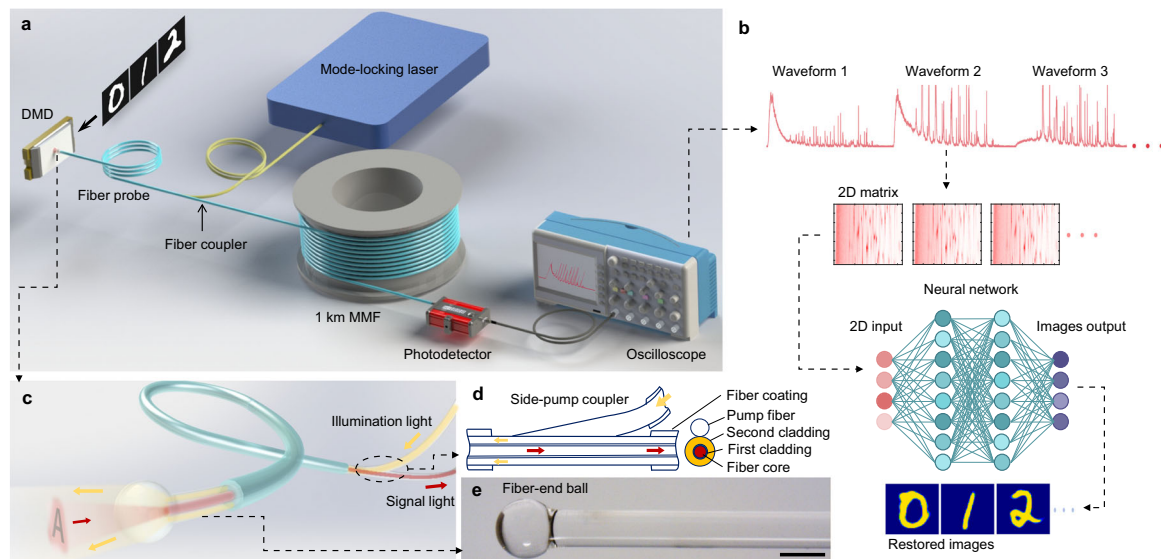
## Results

**Principles.** The light fields in an MMF can be resolved into a set of orthogonal spatial modes<sup>26</sup> that enable the transmission of spatial information. It has been verified that the information contained in images with  $4N$  resolvable features, where  $N$  is the number of fiber modes per polarization, can be carried in a single MMF<sup>27</sup>. When light scattered by an object is collected by an MMF, various fiber modes are excited to different degrees. When an ultrafast pulse laser is used as the illumination source, the energy of each pulse entering the MMF can be dispersed into different modes. Because the different modes have different group velocities, the pulses in these modes will arrive at the opposite end of the MMF with different time delays. If the intermodal dispersion of the MMF is sufficiently large, after transmission through the MMF, a pulse with a temporal duration of less than the delay difference between different modes will be split into a number of isolated subpulses in the time domain, as schematically shown in Fig. 1. If the power of the pulse is sufficiently low and its wavelength bandwidth is sufficiently narrow, both the chromatic dispersion and nonlinear effects in the MMF can be ignored, resulting in the pulse evolution being dominated by intermodal dispersion<sup>28,29</sup> (see Supplementary Note 1 for details). Therefore, the temporal distribution of the train of subpulses depends on the mode composition of the original pulse, which is determined by the spatial distribution of the object. Hence, the spatial information of objects can be encoded into the time waveforms of the output pulses.

**Experimental setup.** The structure of the system is illustrated in Fig. 2a. The illumination pulses from a mode-locked fiber laser are directly coupled into the fiber probe by a side-pump



**Fig. 1 Evolution of ultrashort pulses.** The pulses entering the MMF will be split into lots of subpulses due to the large intermodal dispersion in the MMF. Each subpulse contains the energy of a certain mode in principle.



**Fig. 2 Experimental layout.** **a** Schematic of the experimental setup. **b** Flow of the reconstruction process from waveforms to images. **c** Schematic of the fiber-end ball, the side-pump fiber coupler, and the flow of illumination light (orange arrows) and signal light (red arrows). **d** Structures of the side-pump coupler and triple-cladding fiber probe, where illumination light (orange arrows) is coupled into and transmitted through the cladding of the fiber, while signal light (red arrows) is collected and transmitted through the fiber core in the opposite direction. The cross section of this coupling region is also shown. The triple-cladding fiber consists of the fiber core (red), first cladding layer (blue), and second cladding layer (orange). **e** Micrograph of the fiber-end ball. Scale bar: 500  $\mu\text{m}$ .

coupler<sup>30</sup>. After approximately 2 m of transmission, the illumination pulses emerge from the fiber probe to illuminate the intensity patterns displayed by a digital micromirror device (DMD). The pulse laser operates at a wavelength of 1064 nm with a 3 dB bandwidth of 0.14 nm. While the full width at half maximum of the output pulses is 26.4 ps, it broadens to 45.1 ps after the pulses transmit through the fiber probe and emerge from the fiber-end-ball (see Supplementary Note 1). Then, the light reflected from the patterns reenters the fiber probe, as shown in Fig. 2c. In this way, the illumination and reception of light are integrated into a single fiber probe. The other end of the fiber probe is spliced with a 1 km MMF (50/125  $\mu\text{m}$ , numerical aperture (NA) = 0.22), in which the spatial information carried in the signal pulses is transformed into temporal waveforms. The step-index core of the MMF can provide much greater intermodal dispersion than a graded-index core<sup>31</sup>. In such a long MMF, the delay differences between different modes are sufficiently large to cause each signal pulse to split into a burst of subpulses. The temporal waveforms of the pulses at the other end of the MMF are detected by an ultrafast InGaAs photodetector (spectral response 750–1650 nm, bandwidth 30 GHz) and instantly stored in the memory of an oscilloscope (100 G samples/s.). In the training stage, different displayed images and the corresponding waveforms are used to train the neural network model. After training, the network is capable of recovering new images directly from the acquired waveforms, as shown in Fig. 2b.

The fiber probe is a triple-cladding fiber. The diameters of its core, first cladding, and second cladding are 50, 70, and 360  $\mu\text{m}$  respectively. The core is step-index with an NA of 0.2, and the second cladding has an NA of 0.46. Both the core and the second cladding layer can transmit light. The structure of the side-pump coupler, where the illumination light is coupled into the second cladding layer of the fiber probe (see Supplementary Note 2 for detailed structure), is schematically shown in Fig. 2d. Although the light reflected by the DMD enters both the core and cladding of the fiber probe, only the light in the core (signal light) can enter the MMF due to the matching NA and diameter between the fiber probe core and the MMF core. The end part of the fiber probe is

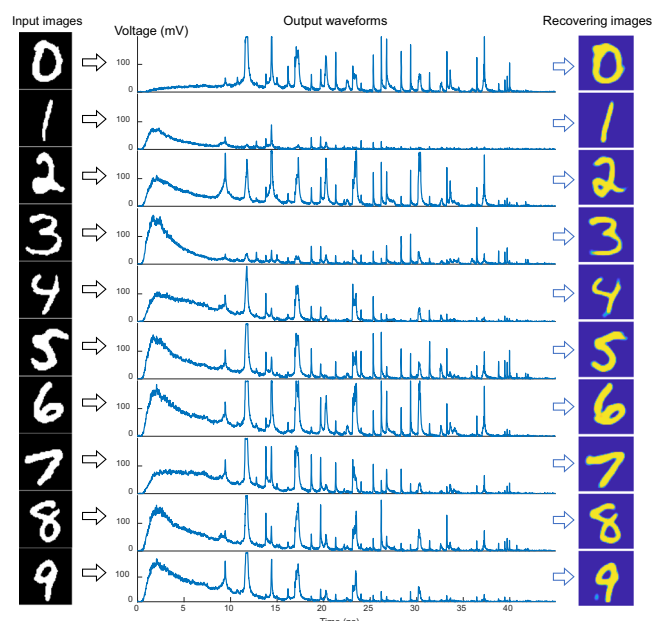
fused into a microball with a 580  $\mu\text{m}$  diameter, as shown in Fig. 2e, which serves to produce more uniform and focused illumination (in the absence of this microball, the beam emerging from the cladding of the fiber would have an annular shape). This probe can be directly moved very close to microscale objects for imaging, with no requirement of objectives that are vital for conventional cameras. To demonstrate this, the fiber-end ball probe was placed very close to the surface of the DMD such that it could only receive light returning from a very small region of the DMD. The area of this small region measured approximately  $200 \times 200 \mu\text{m}^2$ , in which images of approximately  $28 \times 28$  pixels could be displayed.

**Image recovery.** Figure 3 shows several example images from the MNIST dataset<sup>32</sup> and their corresponding temporal waveforms. We see that after transmission through the long MMF, a single input pulse splits into a burst of subpulses spanning approximately 45 ns (see Supplementary Note 3 for more waveform details). A U-Net model was trained on 19,000 waveform/image pairs to learn the corresponding mapping. Using the trained model, we could directly recover other new images from the corresponding acquired waveforms. The recovery results corresponding to these example images are shown in the right side of Fig. 3. The results for 1000 test images showed an average fidelity (calculated as the 2D correlation) of 81.8% and an average structural similarity index measure (SSIM, which correlates well with human perception) of 0.78. Compared with previous fiber endoscopy technologies, which generally operate at low frame rates<sup>16,33,34</sup>, our scheme showed comparable performance in terms of image quality.

We also tested the reconstruction performance for several different types of images, including handwritten letters from the EMNIST dataset<sup>35</sup> and patterns of clothes from the Fashion-MNIST dataset<sup>36</sup>. After similar training processes, some examples are shown in Fig. 4a, along with the average fidelities and SSIMs. The results indicate high practicability of our scheme. While one waveform corresponds to one image and there is no mutual

interference between neighboring waveforms (see Fig. 4c), the successive pulses can enable detection of images at a frame rate of 15.4 Mfps, which is consistent with the repetition rate of the pulse source. Moreover, the shutter time of the system is equal to the time duration of a single pulse irradiating the DMD. Thus, the shutter time can be as low as 45.1 ps, consistent with the pulse width.

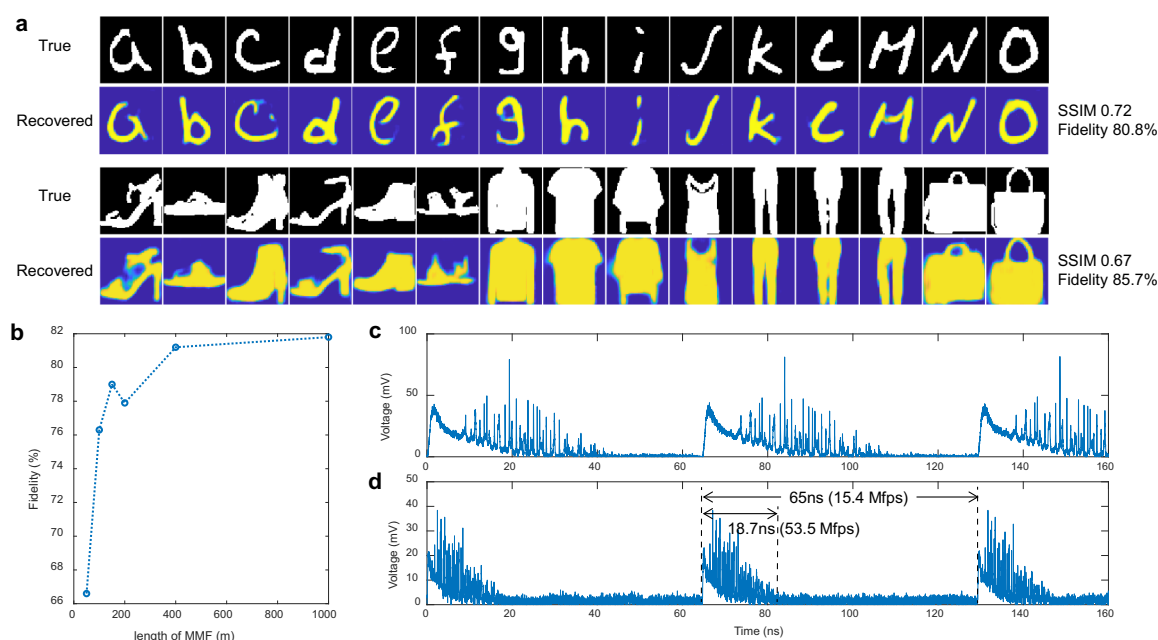
Additionally, the MMF length of 1 km can be further reduced without significantly deteriorating the system performance. The recovery quality under different MMF lengths in the system is



**Fig. 3 Reconstruction process.** The input images of digits 0–9 selected from the test database are transformed into output waveforms. Then the images are recovered from the waveforms.

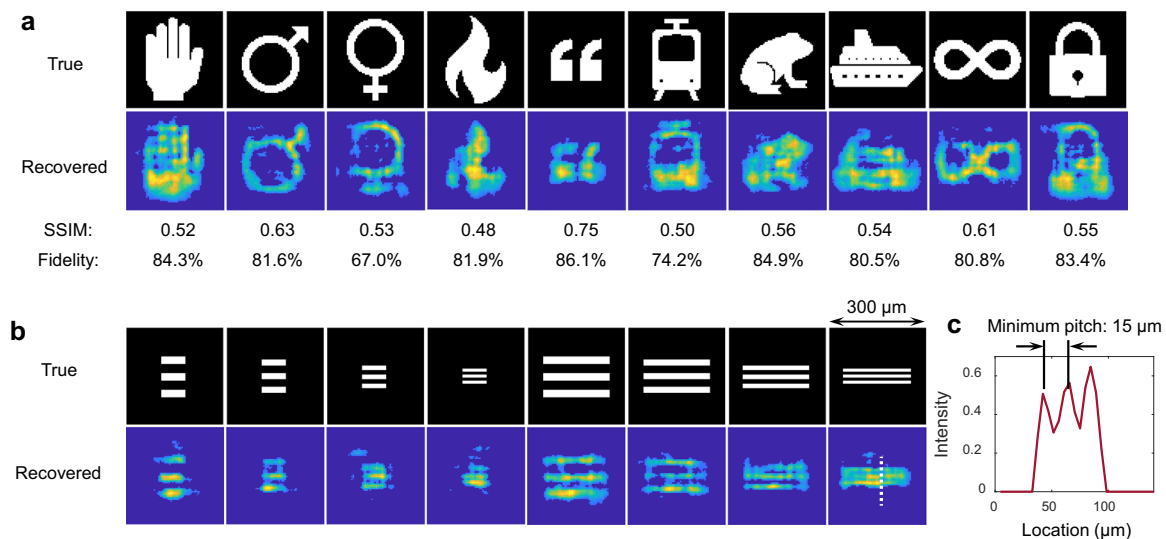
shown in Fig. 4b. The experimental processes using different MMF lengths were consistent with those for the 1 km MMF described above and the same number of digit images were used. There was almost no loss of fidelity as the MMF length was reduced from 1 km to 400 m. After the length was reduced to below 150 m, the image quality deteriorated obviously, indicating that such lengths were too short to split the pulse adequately for separating the information in different modes. In addition, when the MMF length was reduced from 1000 to 400 m, the width of the waveforms was much compressed as shown in Fig. 4c, d. For the 400 m length, a single waveform was much narrower than the period of the pulses, indicating that the temporal space was not fully utilized. Thus, the frame rate could be further increased by lowering the pulse period. Because a single waveform had a length of 18.7 ns in this case, the pulse repetition rate could be increased to 53.5 MHz without any overlap between the neighboring waveforms as shown in Fig. 4d. Thus, it is feasible to increase the frame rate of the current system to 53.5 Mfps by changing the repetition rate of the illumination pulses. Furthermore, if the system is modified to detect a larger image with more pixels, it would require an MMF with more modes. In this case, a larger modal dispersion is required that makes the waveforms becoming broader so that the pulses can be split adequately for separating the information in more modes. However, the broadening of the waveforms will cause the reduction of the frame rate. Thus, considering that the number of modes is in direct proportion to the number of resolvable pixels in the images, the frame rate will be approximately inversely proportional to the number of resolvable pixels.

In the imaging experiments discussed above, the detected images are of the same type as the images used to train the network. Here, we verified that this system could also recover images of different types. To validate this, we replaced the U-Net network with a fully connected network. We found that for the current system, although the U-Net model could realize high-quality imaging of type-aware objects, its ability to image random types of objects, i.e., generalization ability, was not as high as that of the fully connected network (see Supplementary Note 4 for



**Fig. 4 Imaging performance analysis.** **a** Example images of the letters and clothes and the recovered images. **b** Average fidelities when using different MMF lengths in the system. Collected time signals when the MMF length is **c** 1000 m and **d** 400 m respectively. The period of the laser pulses is 65 ns, corresponding to a repetition rate of 15.4 MHz. The width of the waveforms is 18.7 ns, which allows the repetition rate to be further increased to 53.5 MHz.





**Fig. 5** Imaging of random types of objects. **a** Several recovered images of different types. **b** Recovering results of images containing resolution targets similar to the USAF 1951. The width of the images displayed on the DMD is approximately 300  $\mu\text{m}$ . **c** Vertical intensity profile of the smallest recovered bars along the white dotted line in (**b**). The pitch of 15  $\mu\text{m}$  can be distinguished.

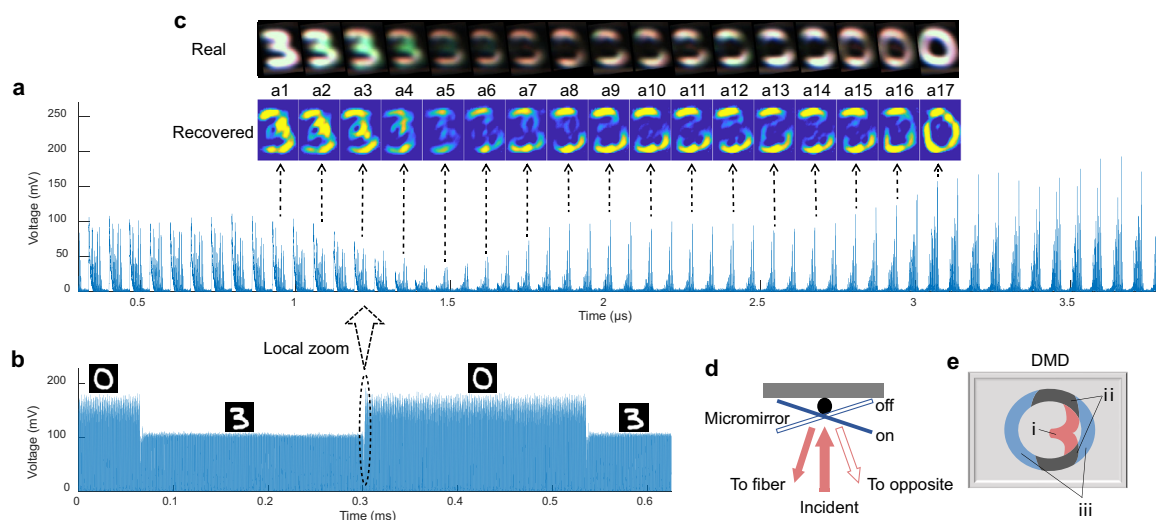
details). Moreover, given that the generalization of a trained neural network is closely connected to the complexity of the training data, we used the images from the Omniglot dataset<sup>37</sup> for training. These images contain patterns made up of nearly random lines. In addition to the original images of the Omniglot dataset, we also generated some other training images based on the original ones by shifting, rotating, or scaling the original patterns to increase the complexity of the training data. Finally, 20,000 images with different complex patterns were used for the training. Then, the trained model was used to restore images of completely different types from the training images. Exemplary results are shown in Fig. 5a and we see that the test images could be recovered with high fidelities. Furthermore, we also verified that this system can be used for detecting grayscale images (see Supplementary Note 9 for details).

Next, we used the same trained model to test the spatial resolution of the system by recovering images of resolution targets similar to the USAF 1951, which contains white bars with different pitches. Here, we adjusted the location of the fiber-end-ball relative to the DMD until it could receive light from a larger region of the DMD surface. The area of this region measured approximately  $300 \times 300 \mu\text{m}^2$  and included  $40 \times 40$  pixels (pixel size of 7.56  $\mu\text{m}$ ), which was larger than the previous  $28 \times 28$  pixels and thus could help explore the minimum resolution of the system. The recovered results are shown in Fig. 5b, indicating that the smallest bars with pitches of 15  $\mu\text{m}$  (occupying two pixels on the DMD surface) could be distinguished, which is shown more clearer in Fig. 5c.

This system can also be used for high-speed classification, which has great value in fields such as microfluidics<sup>12</sup>. We tested this ability via the classification of handwritten digits based on the acquired waveforms (see Supplementary Note 5 for details). A high accuracy of 91.5% was achieved. We note that image detection through such long fibers has been a major challenge for conventional multimode imaging systems<sup>34</sup> because the disturbance grows more severe as the fiber length increases<sup>33</sup>, making the recovery more difficult (the accuracy of digits classification is less than 70% for the speckle-based imaging through a 1 km MMF). However, in our scheme, the classification accuracy remains at such a high level under the same length, indicating high interference immunity and practicability. This superiority can probably be attributed to low crosstalk between different

modes when the pulse energy in the modes is separated after transmission over a certain distance in the MMF, thus, the energy coupling between different modes is suppressed. This feature makes our scheme suitable for long-distance detection.

**High-speed detection.** To verify the feasibility of high-speed detection, we adjusted the time scale of the oscilloscope to the maximum (625  $\mu\text{s}$ ), allowing it to store approximately 10,000 waveforms in a single record. Although the highest refresh rate of the DMD used here is limited to 4.3 kHz, preventing it from displaying an ultrahigh-speed video that matches our detection frame rate of 15.4 Mfps, the refresh processes when the DMD switches from one image to another are nearly transient and spend only 3  $\mu\text{s}$  (see the recorded waveforms in Fig. 6b). Thus, we chose to detect this refresh process using our system to reveal the detailed process over such a short time. We set the DMD to periodically display two images and simultaneously record the time signals, as shown in Fig. 6b. The detailed waveforms corresponding to one refresh process (marked with a black circle) are shown in Fig. 6a, where we can see the process of the waveforms corresponding to the image 3 gradually changing to the waveforms corresponding to the image 0 within 3  $\mu\text{s}$ . The retrieved successive frames are shown in the insets (a1–a17), from which we can understand the refresh process of the DMD. The whole refresh process can be divided into three stages. In stage 1 (insets a1–a5), the DMD initially displays the image 3, which means that the micromirrors in regions (i) and (ii) of the DMD (see Fig. 6e) are in the on state, while the others are in the off state. The states of the micromirrors are explained in Fig. 6d, where region (ii) represents the overlap between the patterns of 0 and 3. When the DMD starts to refresh to the image 0, the micromirrors in regions (i) and (iii) rotate in opposite directions<sup>38</sup>, causing the light in region (i) to fade away. In stage 2 (insets a6–a14), only the light from region (ii) can be observed because the micromirrors there maintain in the on state. In stage 3 (insets a15–a17), the light from region (iii) appears, indicating that the corresponding micromirrors have rotated into on state. Thus, the image 3 has been refreshed to the image 0. For comparison, we also used a commercial high-speed camera to record the refresh process (see the “Methods” section for details), and the real images captured are shown in Fig. 6c. We can see that the change process of the



**Fig. 6 Ultra-high-speed imaging.** **a** Waveforms collected during the transient time when the DMD refreshes from an image 3 to an image 0. Insets (a1–a17) show the images reconstructed from some of these waveforms. **b** Waveforms recorded with the DMD periodically displaying two images, from which we see that each image has an exposure time of 235  $\mu$ s, which is used by the DMD to load the data of the next image into the memory cells beneath the pixels. After all data have been loaded, the DMD switches to the other image within only 3  $\mu$ s, as marked by the dashed circle. The details of this region are shown in (a). **c** Real images of the DMD captured during the refresh process using a commercial high-speed camera. **d** Structure of one pixel of the DMD. A DMD chip has many micromirrors on its surface, which correspond to the pixels in the image to be displayed. These micromirrors can be individually rotated by approximately  $\pm 12^\circ$  to an on or off state. In the on state, the illumination light is reflected back to the fiber probe. In the off state, the light is directed in another direction. Thus, the array of micromirrors can produce intensity modulations on the light field. When the DMD refreshes its current image, all micromirrors move to their assigned states at the same moment. **e** Schematic of the patterns of digits 3 and 0 displayed on the DMD chip. When the DMD switches from 3 to 0, the micromirrors in the overlapping region of the two patterns (region ii) do not change their states, while those in regions (i) and (iii) rotate in the opposite directions.

DMD patterns is consistent with what observed using the proposed system.

**System robustness.** To analyze the robustness of the system, we investigated the influence of temperature and fiber bending on the imaging performance. For temperature effect, we changed the environmental temperature from 23 to 27  $^\circ$ C by adjusting the air conditioners. We tested the imaging performance at different temperatures in two different cases. In the first case, the neural network was trained with image/waveform pairs collected with the temperature fixed at approximately 25  $^\circ$ C. In the other case, the network was trained with the data collected at different temperatures, called joint training. The details of the experiment and test results are shown in Supplementary Note 6. While the average fidelity of the recovered images remains above 70% within a temperature variation of 0.5  $^\circ$ C in the first case, this variation range increases to 3  $^\circ$ C for the joint training case. This temperature sensibility is mainly caused by the temperature-induced index-distribution change in the MMF which will influence the temporal distribution of the subpulses. To investigate the bending effect, we fixed the fiber-end ball and bent the fiber probe into a semicircle with variable radii as shown in the inset of Supplementary Fig. 11b. Similarly, we tested the imaging performance under different bending states in two cases: training with the data collected under one bending state and joint training with the data collected under different bending states. The results (see Supplementary Fig. 11) show that in the first case, the average fidelity can remain above 70% when the bending radius changes from 28 to 22 cm. And for the joints training, the 70% fidelity is obtained in the range of 28–19 cm, showing a higher robustness. The sensibility to the bending is mainly caused by the fiber-stress-induced modal crosstalk inside the MMF. In summary, we verified that this system has certain robustness for practical applications.

## Discussion

Because the proposed scheme requires only a single photodiode rather than pixelated sensors, it can be easily applied to other wavelengths. For example, considering that the InGaAs-based photodiode used here has high sensitivity over a broad band from 1 to 1.6  $\mu$ m<sup>39</sup>, while the silica fiber has very low attenuation in this band, our scheme can be easily extended to other wavelengths within this band. This will be highly valuable for real applications because conventional Si-based CCD and CMOS cameras are sensitive only to wavelengths below 1.1  $\mu$ m<sup>22</sup>. Our method also has the potential to operate in the mid-infrared or THz bands, considering the development of photodetectors in these bands. In the mid-infrared band, the use of a fluoride-glass fiber can significantly reduce optical loss, which makes it possible to develop long waveguides with high intermodal dispersion in these bands. In addition, we can see from Fig. 4b that the MMF length may be potentially reduced to 150 m with little degradation of image quality. The required MMF length can be further reduced by increasing the NA of the MMF, which increases the intermodal dispersion. Thus, although current technology can only fabricate fluoride-glass fibers with losses on the order of 0.1 dB/m<sup>40</sup>, the total loss can be controlled to an acceptable range. In addition, there is still room to lower the loss in fluoride-glass fibers according to the theoretical predictions of Shibata et al.<sup>41</sup>. Thus, it is possible to extend the wavelength of this system to the mid-infrared region. Also, in the THz band, much work has focused on the development of waveguides with reduced loss and dispersion<sup>42</sup>, which provides a certain possibility to apply the method to this band. This will be helpful for detecting certain materials that have strong responses only at these wavelengths or for detection under special conditions in which only light in these bands can be transmitted with low loss. In summary, our scheme offers an alternative approach for observing vivid physical phenomena in a vast number of scenarios.

The performance of our demonstrated proof-of-principle system may be further improved. The wavelength of the source used here (1064 nm) is much longer than those adopted in most previous studies<sup>16,26,33,34</sup>, which resulted in a much smaller number of excited modes and, thus, much less spatial information carried in the MMF. Hence, upon using an MMF with a larger core and higher NA, more spatial information can be collected, and the resolution of the recovered images will be much higher. In addition, the shutter time can be further shortened to enable the detection of faster events by using shorter pulses. More importantly, with a fiber amplifier spliced to the end of the MMF, the pulse signals can be significantly amplified, which will greatly enhance the sensitivity of the detection system to make it suitable for detecting very weak signals. Moreover, because the illumination zone and intensity of the applied fiber probe are limited, the current system is only suitable for detecting small objects. For larger-object detection, an objective can be used in front of the fiber probe to couple more light from the object into the probe. Additionally, for brighter illumination, auxiliary illumination can be adopted as discussed in Supplementary Note 7.

Our scheme can be further modified to detect 3D objects by combining it with the existing time-of-flight technique<sup>43–45</sup>, in which ultrafast pulses are generally used to illuminate objects of interest and an ultrafast camera is used to detect the reflected light at different arrival times. Because the light reflected from different depths on the object will arrive at the camera with different time delays, the variations in 2D images captured over time can reveal the 3D information of the object. The system presented in this paper is naturally compatible with the time-of-flight method because we also adopt an ultrafast pulse laser for illumination. If the fiber probe is used to detect a 3D object, the temporal waveforms will contain both depth information and 2D spatial information. Thus, the use of specific reconstruction algorithms will make it possible to recover the 3D information encoded in these ultrafast time signals.

## Methods

**Experiments.** The laser source is a homemade Yb-doped mode-locked fiber laser with an average output power of approximately 1 W. The fiber probe is a triple-cladding fiber with diameters of 50/70/360  $\mu\text{m}$  (NUFERN FUD-4658, BD-S50/70/360-22FA-HP). The homemade fiber coupler couples light from the source into the second cladding layer of the probe. The fiber-end ball at the end of the probe was produced via fusion with a fusion splicer. Because we adopted a pulse laser, considering that the DMD can only perform grayscale modulation on continuous light, all images were binarized before loading into the DMD. The DMD (Texas Instruments DLP4500) consists of  $912 \times 1140$  micromirrors, each being 7.56  $\mu\text{m}$  in size. The photodetector (Thorlabs DXM30BF) has a 30 GHz response bandwidth and a 15 ps impulse response with a sensible spectrum of 750–1650 nm. The photodetector receives light through an OM4 (50/125  $\mu\text{m}$ ) fiber, which is connected to the other end of the MMF. The oscilloscope (Tektronix MSO73304DX) has a 33 GHz analog bandwidth and a sample rate of 100 G/s. Its maximum record length is 62.5 Mega samples.

During the process of collecting waveforms of the training images, we set the oscilloscope to automatically save waveforms at a speed of 4 waveforms per second. We have verified that training with 10,000 samples was adequate for the network to achieve the optimal performance (see Supplementary Note 8). Thus, the whole collection procedure spanned 42 min. The collected raw data was processed before fed to the neural network. Because a recorded time signal included several periods of pulses, one period would be selected and extracted as one waveform. Finally, the waveforms of all images were put together and converted into a matrix data. This was processed by a MATLAB program, requiring approximately 5 min. Training the U-Net network with 10,000 sample data required approximately 4 min. We used the online computing resource from Google Colab that provides a Tesla P100-PCIE GPU. In conclusion, the whole calibration process, including sample collecting, processing, and training, required approximately 51 min.

The high-speed camera (MotionBLITZ EoSens<sup>®</sup> mini) used in the high-speed imaging experiment has a frame rate of 40 kfps, which is too slow to record a refresh process of the DMD in real time. Thus, the images showing this transient process in Fig. 6c were actually not captured during a single refresh process. Instead, they were obtained using the following method. First, we acquired a large number of images while the DMD periodically switched between two images. Because the time of a single refresh process occupies only a very small part of the

switching period, as shown in Fig. 6b, only a small number of images were captured exactly during the refresh processes. Because these images tended to record different states of the process, they could be combined to present a continuous refresh process. The exposure time of the camera was set to the minimum to capture these transient states.

**Neural networks.** The structure of the U-Net network is shown in Supplementary Note 5. In training, the original images were all interpolated into  $64 \times 64$  matrixes as the output of the network, and the 4096-point waveforms were reshaped into  $64 \times 64$  matrixes as the input. The Fully-Connected network consists of five fully connected layers, which are one dimensional, and thus the matrixes of images should be reshaped into vectors and thus the reshaping of the waveforms is not required.

## Data availability

The image and waveform data that are necessary to evaluate the conclusions in this study are available in the Tsinghua cloud [<https://cloud.tsinghua.edu.cn/f/f7e530af7c6c44caaf74/?dl=1>].

## Code availability

The python codes used in this study are available in the Tsinghua cloud [<https://cloud.tsinghua.edu.cn/f/f7e530af7c6c44caaf74/?dl=1>].

Received: 28 May 2021; Accepted: 24 February 2022;

Published online: 17 March 2022

## References

- Feist, A., Silva, N. R. D., Liang, W., Ropers, C. & Schfer, S. Spatio-temporal probing of lattice dynamics in graphite by ultrafast TEM. In European Microscopy Congress 2016: Proceedings, 330–331 (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2016).
- El-Desouki, M., Deen, M. J., Fang, Q., Liu, L. & Tse, F. CMOS image sensors for high speed applications. *Sensors* **9**, 430–444 (2009).
- Nakagawa, K. Sequentially timed all-optical mapping photography for observation of ultrafast phenomena. In *2015 Opto-Electronics and Communications Conference (OECC)* 15650105 (IEEE, Miami, FL, 2015).
- Li, Z., Zgadzaj, R., Wang, X., Chang, Y. Y. & Downer, M. C. Single-shot tomographic movies of evolving light-velocity objects. *Nat. Commun.* **5**, 3085 (2014).
- Wang, X., Yan, L., Si, J., Matsuo, S. & Xu, H. High-frame-rate observation of single femtosecond laser pulse propagation in fused silica using an echelon and optical polarigraphy technique. *Appl. Opt.* **53**, 8395–8399 (2014).
- Wang, P., Liang, J. & Wang, L. V. Single-shot ultrafast imaging attaining 70 trillion frames per second. *Nat. Commun.* **11**, 2091 (2020).
- Block, A., Liebel, M., Yu, R., Spector, M. & Hulst, N. F. V. Tracking ultrafast hot-electron diffusion in space and time by ultrafast thermomodulation microscopy. *Sci. Adv.* **5**, eaav8965 (2019).
- Zyung, T., Kim, H., Postlewaite, J. C. & Dlott, D. D. Ultrafast imaging of 0.532  $\mu\text{m}$  laser ablation of polymers: Time evolution of surface damage and blast wave generation. *J. Appl. Phys.* **65**, 4548–4563 (1989).
- Gawelda, W. et al. Ultrafast imaging of transient electronic plasmas produced in conditions of femtosecond waveguide writing in dielectrics. *Appl. Phys. Lett.* **93**, 231115 (2008).
- Osmanski, B. F. et al. Ultrafast imaging of blood flow dynamics in the myocardium. *IEEE Trans. Med. Imaging* **31**, 1661–1668 (2012).
- Deffieux, T., Gennisson, J.-L., Tanter, M., Fink, M. & Nordez, A. Ultrafast imaging of in vivo muscle contraction using ultrasound. *Appl. Phys. Lett.* **89**, 184107 (2006).
- Li, Y. et al. Deep cytometry: deep learning with real-time inference in cell sorting and flow cytometry. *Sci. Rep.* **9**, 11088 (2019).
- Mehta, A. D., Jung, J. C., Flusberg, B. A. & Schnitzer, M. J. Fiber optic in vivo imaging in the mammalian nervous system. *Curr. Opin. Neurobiol.* **14**, 617–628 (2004).
- Petty, H. R. Spatiotemporal chemical dynamics in living cells: From information trafficking to cell physiology. *Biosystems* **83**, 217–224 (2006).
- Dufour, J., Murat, D., Dufour, X. & Foos, J. Experimental observation of nuclear reactions in palladium and uranium—possible explanation by hydrex mode. *Fusion Sci. Technol.* **40**, 91–106 (2001).
- Rahmani, B., Loterie, D., Konstantinou, G., Psaltis, D. & Moser, C. Multimode optical fiber transmission with a deep learning network. *Light.: Sci. Appl.* **7**, 69 (2018).
- Loterie, D. et al. Digital confocal microscopy through a multimode fiber. *Opt. Express* **23**, 23845–23858 (2015).



18. Choi, Y., Yoon, C., Kim, M., Yang, T. D. & Choi, W. Scanner-free and wide-field endoscopic imaging by using a single multimode optical fiber. *Phys. Rev. Lett.* **109**, 203901 (2012).
19. Caramazza, P., Moran, O., Murray-Smith, R. & Faccio, D. Transmission of natural scene images through a multimode fibre. *Nat. Commun.* **10**, 2029 (2019).
20. Wu, L. et al. Analysis and design of a CMOS ultra-high-speed burst mode imager with in-situ storage topology featuring in-pixel CDS amplification. *Sensors* **18**, 3683 (2018).
21. Shimadzu Corporation. *Hyper Vision HPV-X2*, <https://www.shimadzu.com/an/products/materials-testing/high-speed-video-camera/hyper-vision-hpv-x2/index.html> (2021).
22. El Gamal, A. & Eltoukhy, H. CMOS image sensors. *IEEE Circuits Devices Mag.* **21**, 6–20 (2005).
23. Goda, K., Tsia, K. & Jalali, B. Serial time-encoded amplified imaging for real-time observation of fast dynamic phenomena. *Nature* **458**, 1145–1149 (2009).
24. Karpf, S. et al. Spectro-temporal encoded multiphoton microscopy and fluorescence lifetime imaging at kilohertz frame-rates. *Nat. Commun.* **11**, 2026 (2020).
25. Liao, R., Hon, N. K., Buckley, B. W., Diebold, E. D. & Jalali, B. Chromo-modal dispersion for optical communication and time-stretch spectroscopy. *Opt. Lett.* **46**, 500–503 (2021).
26. Zhu, C. et al. Image reconstruction through a multimode fiber with a simple neural network architecture. *Sci. Rep.* **11**, 896 (2021).
27. Mahalati, R. N., Gu, R. Y. & Kahn, J. M. Resolution limits for imaging through multi-mode fiber. *Opt. Express* **21**, 1656–1668 (2013).
28. Lee, J. & Kim, D. Determination of the differential mode delay of a multimode fiber using Fourier-domain intermodal interference analysis. *Opt. Express* **14**, 9016–9021 (2006).
29. Cheng, J. et al. Time-domain multimode dispersion measurement in a higher-order-mode fiber. *Opt. Lett.* **37**, 347–349 (2012).
30. Xiao, Q., Yan, P., Ren, H., Chen, X. & Gong, M. A side-pump coupler with refractive index valley configuration for fiber lasers and amplifiers. *J. Lightwave Technol.* **31**, 3015–3022 (2013).
31. Keiser, G. *Optical Fiber Communications* 3rd edn (McGraw Hill, 2000).
32. Deng, L. The mnist database of handwritten digit images for machine learning research [best of the web]. *IEEE Signal Process. Mag.* **29**, 141–142 (2012).
33. Borhani, N., Kakkava, E., Moser, C. & Psaltis, D. Learning to see through multimode fibers. *Optica* **5**, 960–966 (2018).
34. Li, Y. et al. Image reconstruction using pre-trained autoencoder on multimode fiber imaging system. *IEEE Photonics Technol. Lett.* **32**, 779–782 (2020).
35. Cohen, G., Afshar, S., Tapson, J. & Van Schaik, A. EMNIST: Extending MNIST to handwritten letters. In *2017 International Joint Conference on Neural Networks (IJCNN)* 2921–2926 (IEEE, 2017).
36. Xiao, H., Rasul, K. & Vollgraf, R. Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms. Preprint at <https://arxiv.org/abs/1708.07747> (2017).
37. Lake, B. M., Salakhutdinov, R. & Tenenbaum, J. B. Human-level concept learning through probabilistic program induction. *Science* **350**, 1332–1338 (2015).
38. Conner, J. L., Overlaur, M. & Bhuva, R. L. Spatial light modulator with buried passive charge storage cell array. US patent 5,671,083 (1997).
39. Joshi, A. M., Heine, F. & Feifel, T. Rad-hard ultrafast InGaAs photodiodes for space applications. *Proc. SPIE* **6220**, 622003 (2006).
40. Cozic, S., Poulain, S. & Poulain, M. Low loss fluoride optical fibers: Fabrication and applications. In *Specialty Optical Fibers SoM2H.3* (Optica Publishing Group, Washington, DC, 2018).
41. Shibata, S. et al. Prediction of loss minima in infra-red optical fibres. *Electron. Lett.* **17**, 775–777 (2007).
42. Atakaramians, S., Afshar, S., Monro, T. M. & Abbott, D. Terahertz dielectric waveguides. *Adv. Opt. Photonics* **5**, 169–215 (2013).
43. Velten, A. et al. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nat. Commun.* **3**, 745 (2012).
44. Turpin, A., Musarra, G., Kapitany, V., Tonolini, F. & Faccio, D. Spatial images from temporal data. *Optica* **7**, 900–905 (2020).
45. Jalali, B., Jiang, Y. & Karpf, S. Time stretch lidar: A fast spectrally scanned time-of-flight 3D camera. *Proc. SPIE* **11684**, 116841B (2021).

## Acknowledgements

Q.X. acknowledges financial support from National Natural Science Foundation of China (Grants No. 62122040, 62075113, and 61875103). We thank Dr. Zeyi Li and Dr. Yvze Lu for guidance in neural-network algorithms.

## Author contributions

Z.L. and Q.X. conceived and performed the most experiments and calculations. L.W. (Lele Wang) made the schematic diagrams. T.H., Y.Y. and L.W. (Liuyue Wang). contributed to the deep-learning algorithms. Q.L., M.G. and P.Y. discussed the results and contributed to the writing of the paper. J.T., D.L., S.H. and Y.M. contributed to part of the experiments. Q.X. conceived and led the project.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-29178-8>.

**Correspondence** and requests for materials should be addressed to Qirong Xiao.

**Peer review information** *Nature Communications* thanks Christophe Moser and the other anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission 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) 2022