

<https://doi.org/10.1038/s44310-024-00002-z>

Coincidence imaging for Jones matrix with a deep-learning approach

Check for updates

Jiawei Xi¹, Tsz Kit Yung¹, Hong Liang¹, Tan Li², Wing Yim Tam¹ & Jensen Li¹ ✉

Coincidence measurement has become an emerging technique for optical imaging. Based on measuring the second-order coherence g_2 , sample features such as reflection/transmission amplitude and phase delay can be extracted with developed algorithms pixel-by-pixel. However, an accurate measurement of g_2 requires a substantial number of collected photons which becomes difficult under low-light conditions. Here, we propose a deep-learning approach for Jones matrix imaging using photon arrival data directly. A variational autoencoder (β -VAE) is trained using numerical data in an unsupervised manner to obtain a minimal data representation, which can be transformed into an image with little effort. We demonstrate as few as 88 photons collected per pixel on average to extract a Jones matrix image, with accuracy surpassing previous semi-analytic algorithms derived from g_2 . Our approach not only automates formulating imaging algorithms but can also assess the sufficiency of information from a designed experimental procedure, which can be useful in equipment or algorithm designs for a wide range of imaging applications.

In quantum optics, the utilization of coincidence measurement of photon pairs has emerged as a valuable tool for diverse usages, such as distinguishing between classical and quantum light sources, identifying entangled photon pairs, quantum state characterizations, and two-photon imaging^{1–7}. In the applications of two-photon imaging with quantum interference, this technique is primarily based on the measurement of the second-order coherence, $g_2(\tau)$, which describes the distribution of coincidence events between two photons with an optical delay τ . By extracting the information stored in $g_2(\tau)$, researchers have successfully harnessed this technique for imaging applications. It has shown effectiveness in imaging the reflection/transmission amplitude and phase delay profiles of samples and can also be leveraged for hologram generation and super-resolution imaging^{8–13}. On the other hand, two-photon imaging can also employ the entanglement between photon pairs. With the adoption of heralded photons and coincidence measurements, there have been recent demonstrations of the implementation of multiplexed holography and remote control on the imaging arm^{14,15}. Such a heralding technique can improve the signal-to-noise ratio, making coincidence measurements potentially useful for imaging under low-light conditions^{16,17}.

Metasurfaces, comprised of a single layer of nanostructures, possess the ability to manipulate different degrees of freedom (DOFs) of light, including phase^{18–21}, polarization^{22–26}, wavelength^{27–31}, and orbital angular momentum^{32–36} in fine resolution. With the development of metasurfaces, they have emerged as highly effective platforms in various imaging

applications, including super-resolution imaging, computational holography, multicolor imaging, and hyperspectral imaging^{28,37–40}. In the quantum realm, metasurfaces enable the generation of high-dimensional entangled states, the control of two-photon interference, quantum state tomography, and quantum emission^{41–44}. Concerning quantum imaging applications involving metasurfaces, the use of coincidence measurements has facilitated demonstrations such as image edge detection, achieved through a high-efficiency dielectric metasurface, and the imaging of metasurface polarization responses using the heralding technique with entangled photons^{45,46}. Moreover, regarding the Jones matrix imaging which plays a pivotal role in the realm of studying light polarization in various applications, there have been demonstrations to simplify the traditional techniques for Jones matrix characterizations requiring a significant number of measurements^{47–49}, by employing Fourier ptychography^{50,51} and the method of Fourier space sharing^{52,53}. Notably, recent works have showcased metasurface-driven polarization and Jones matrix profile imaging for an unknown object using Hong-Ou-Mandel (HOM)-type interference^{54,55}. By employing reference metasurface panels with known polarization responses together with an object in pairwise coincidence measurements, the unknown polarization responses of the object can be extracted at the pixel-by-pixel level from the HOM visibilities obtained from $g_2(\tau)$. The tailor-made quantum metasurfaces hold significant potential for providing additional resources for parallelizing different optical experiments^{45–56}.

¹Department of Physics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China. ²Centre for Advances in Reliability and Safety (CAIRS), Hong Kong, China. ✉e-mail: jensenli@ust.hk

Meanwhile, machine-learning (ML) techniques have proven to be valuable in quantum optical applications, which encompass quantum device optimization, the automation of quantum experiments, and quantum state tomography^{57–59}. When compared with conventional approaches, ML techniques have exhibited higher efficiency and accuracy in classifying quantum and classical light sources, as well as in extracting spatial modes of light under low-light conditions^{60–62}. Furthermore, a deep-learning (DL)-assisted approach has been developed for antibunching super-resolution imaging by predicting $g_2(0)$, providing substantial enhancements in speed compared to conventional coincidence measurements⁶³. However, when the number of collected photons is limited, it is more difficult to identify $g_2(0)$ accurately. In fact, for many quantum imaging applications, $g_2(0)$ can be further utilized to extract sample features such as reflection/transmission amplitude and phase delay profiles with additional algorithms to be developed^{8–11}. In these cases, further developing the DL techniques has the potential to extract the sample features directly from the measured photon arrival data in a more versatile and automatic manner, even in the low-light regime.

In this work, we have developed a DL-assisted approach for coincidence imaging under low-light conditions for Jones matrix extraction. By combining a variational autoencoder (β -VAE)⁶⁴ with a regression network, our proposed method automatically obtains images of object features directly from the measured photon arrival data, eliminating the need for manually derived algorithms. The β -VAE can determine whether the data collected from our designed experimental procedure contains sufficient information for feature extraction and generate a minimal representation of data, which can be transformed into an image with little effort. Specifically, we demonstrate its application to image Jones matrix profiles of metasurfaces. The imaging outcomes highlight that our DL-assisted approach surpasses the previous semi-analytic algorithm⁵⁵ derived from $g_2(0)$ in terms of both efficiency and accuracy. The versatility and flexibility of this method allow it to be used in other imaging applications in addition to quantum imaging, such as medical imaging and ultrasound imaging^{65,66}.

Results

We commence by outlining the experimental workflow and introducing our previous semi-analytic algorithm for Jones matrix imaging, while also acknowledging its limitations⁵⁵. As illustrated in Fig. 1a, the sample is illuminated with a photon pair, with an optical delay τ and orthogonal circular polarizations. Subsequently, the transmitted photons are projected to the analyzed polarization state via a quarter-wave plate (QWP) and a polarizer,

before being detected by a single-photon avalanche diode (SPAD) camera. As depicted in Fig. 1b, the sample consists of a metasurface featuring both reference and object regions. Within the reference region, there are 4 panels schematically exhibiting varied predetermined transmission polarization responses, each of which can be represented using a Jones matrix:

$$\begin{pmatrix} t_{LL}^{(i)} & t_{LR}^{(i)} \\ t_{RL}^{(i)} & t_{RR}^{(i)} \end{pmatrix} = t_{\text{ref}} \begin{pmatrix} 1 & e^{-2i\theta_i} \\ e^{2i\theta_i} & 1 \end{pmatrix}, \quad (1)$$

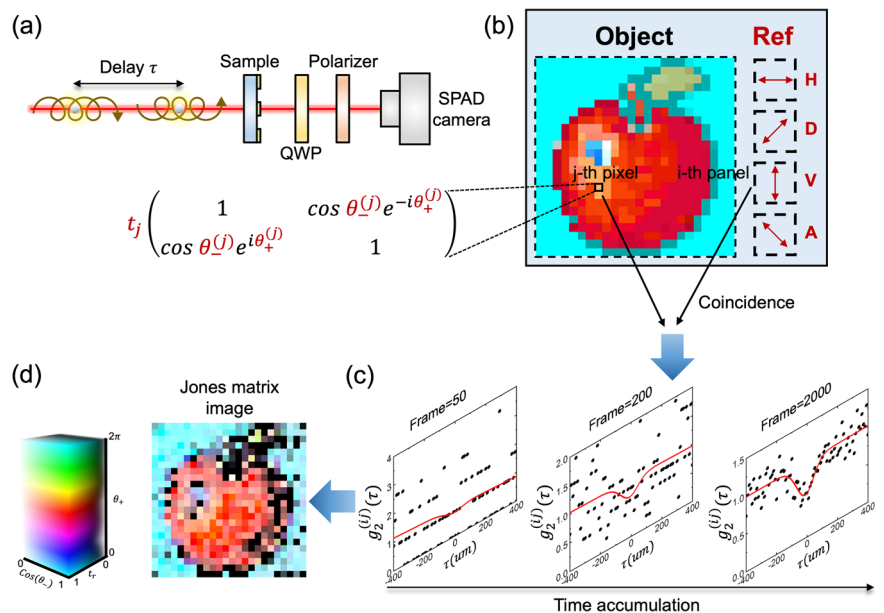
where L and R denote left-handed circular polarization (LCP) and right-handed circular polarization (RCP), respectively. Each reference panel features various passing axis angles θ_i equal to $0, \pi/4, \pi/2$ and $3\pi/4$ (designated as H, D, V, and A, respectively). These orientations are illustrated by the red double-arrows with differing directions in Fig. 1b. In the sample employed for this study, each reference panel that allows a specific polarization transmission covers 6 pixels (arranged in a 3 by 2 configuration) arriving on the single-photon camera. Conversely, the object region comprises 576 pixels (arranged in a 24 by 24 grid) on the camera, with each pixel corresponding to an unknown Jones matrix. This matrix is most generally described within the scope of this study as

$$\begin{pmatrix} t_{LL}^{(j)} & t_{LR}^{(j)} \\ t_{RL}^{(j)} & t_{RR}^{(j)} \end{pmatrix} = t_j \begin{pmatrix} 1 & \cos \theta_-^{(j)} e^{-i\theta_+^{(j)}} \\ \cos \theta_+^{(j)} e^{i\theta_-^{(j)}} & 1 \end{pmatrix} \quad (2)$$

where t_j represents the overall transmission amplitude, while $\theta_+^{(j)}$ (ranging from 0 to 2π) and $\theta_-^{(j)}$ (ranging from 0 to $\pi/2$) describe the off-diagonal elements. $\{\theta_+^{(j)}, \theta_-^{(j)}, t_j\}$ are the three parameters to be extracted at each object pixel.

During the experiment, the SPAD camera captures binary photon arrival images, indicating whether a photon arrives at each pixel during each time frame. A coincidence event arises when two distinct pixels both detect photons within the same time frame. As exemplified in Fig. 1b, the count of coincidence events between the j -th object pixel and each subpixel of the i -th reference panel can be deduced from the measured data. These counts are then averaged over the number of subpixels in the i -th reference panel to be used. Additionally, the optical delay (τ) is scanned to obtain the second-order coherence $g_2^{(ij)}(\tau)$, which emerges from the coincidence count between the j -th object pixel and the i -th reference panel, showcasing the signal of

Fig. 1 | Schematic of the experimental and fitting methods for Jones matrix imaging. **a** Schematic of the experimental setup. A photon pair with an optical delay (τ) and orthogonal circular polarizations is generated and directed onto the sample. The single-photon avalanche diode (SPAD) camera records the photon arrival data corresponding to the analyzed polarization. **b** Schematic of the sample. The sample consists of a metasurface incorporating reference and object regions, characterized by known and unknown transmission responses, respectively. The coincidence count between any combination of a reference panel and an object pixel can be obtained from the acquired data. **c** The normalized coincidence count (depicted as solid black dots) plotted against the delay (τ), alongside the corresponding fitted $g_2^{(ij)}$ curves (shown as solid red curves) utilizing varying total numbers of time frames of data. **d** Extracted Jones matrix image of the object region. The 3 DOFs in the Jones matrix are represented using an HSB color scheme.



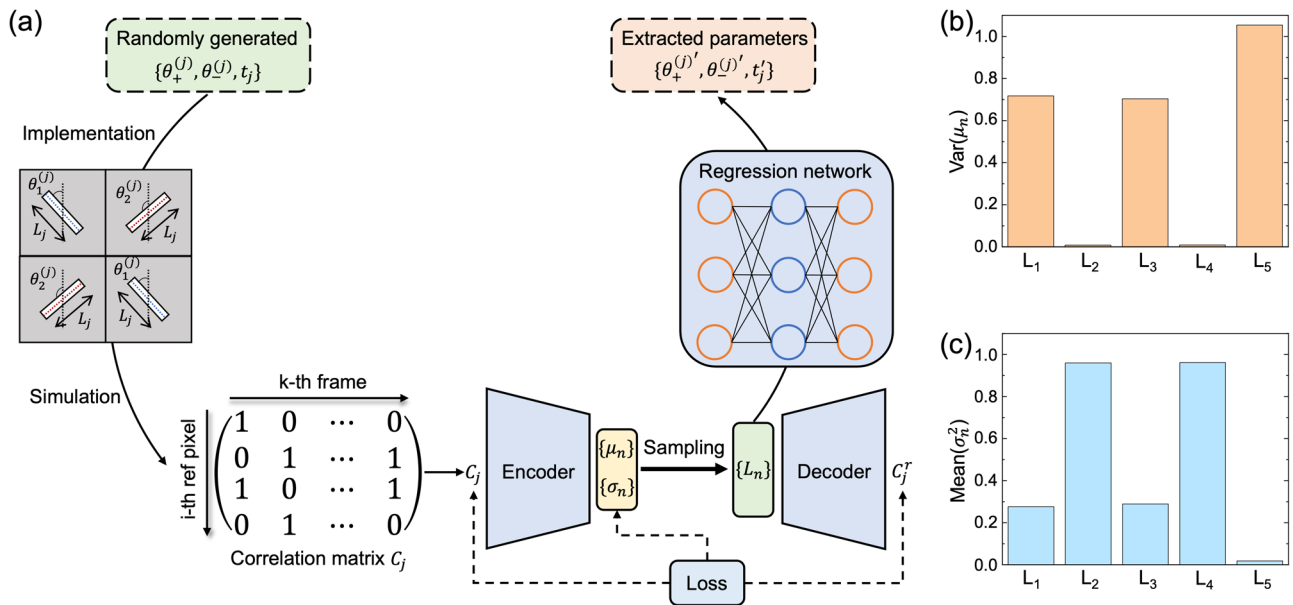


Fig. 2 | Workflow of the DL approach for Jones matrix imaging. **a** Random generation of diverse configurations of $\{\theta_+^{(j)}, \theta_-^{(j)}, t_j\}$ corresponding to various object pixels. Each object pixel is realized through an ‘interleaved’ structure featuring two pairs of nanoslots. For the j -th configuration, a correlation matrix C_j , for the different reference panels (row index) and different time frames (column index), is obtained through simulation and input into the β -VAE. The β -VAE’s objective is to uncover the DOFs within the input data, encapsulating them within meaningful latent

variables. Subsequently, these variables are converted into the DOFs of the Jones matrix via a regression network trained in the subsequent stage. **b, c** Statistics depicting distribution parameters of the extracted latent variables. The meaningful latent variables $\{L_1, L_3, L_5\}$ are characterized by the variance of μ_n proximate to 1 and the mean of σ_n^2 close to 0. Conversely, $\{L_2, L_4\}$ do not carry pertinent information about the input data.

two-photon interference. In Fig. 1c, the coincidence count (depicted as a solid black dot) is plotted against the delay (τ) for three different total numbers of time frames (50, 200, 2000), and the corresponding fitted $g_2^{(ij)}(\tau)$ curves (shown as solid red curves) are presented in the same graph. The plotted results $g_2^{(ij)}(\tau)$ are normalized by the ballistic case without two-photon interference at large τ . With ample measurement time, the fitted $g_2^{(ij)}(\tau)$ curve closely matches the experimental data. By sweeping all reference panels (each with a different θ_i as per Eq. 1) for the j -th object pixel through post-selection of the results, various fitted $g_2^{(ij)}(\tau)$ curves for different i and j can be obtained. Notably, the $g_2^{(ij)}(\tau)$ curve will exhibit a peak or dip when the delay (τ) is 0, and we define $V_{ij}(\theta_i) \triangleq 1 - g_2^{(ij)}(0)$ to characterize the visibility of the two-photon interference. Before introducing the DL approach, the approximated analytic model in ref. 55 for the interference visibility is given as

$$V_{ij}(\theta_i) \cong \beta + \gamma \cos(2\theta_i - \theta_+^{(j)}), \quad (3)$$

where

$$\beta \cong \frac{1}{4} P_{\text{ref}}^{(b)} \left(1 + \frac{2 \cos^2 \theta_-^{(j)}}{1 + \cos^2 \theta_-^{(j)}} \right) |t_j / t_{\text{ref}}|^2 \quad (4)$$

$$\gamma \cong \left(1 - \frac{1}{4} P_{\text{ref}}^{(b)} \left(2 + \left(1 + \cos^2 \theta_-^{(j)} \right) \left| \frac{t_j}{t_{\text{ref}}} \right|^2 \right) \right) \times \frac{-\cos \theta_-^{(j)}}{1 + \cos^2 \theta_-^{(j)}}. \quad (5)$$

The 3 DOFs $\{\theta_+^{(j)}, \theta_-^{(j)}, t_j\}$ of the Jones matrix of the j -th object pixel can be extracted by fitting $V_{ij}(\theta_i)$ as function depicted by Eq. 3. The constant ballistic single-photon count rate for the reference panels, denoted as $P_{\text{ref}}^{(b)}$ in Eq. 4 and Eq. 5, can be directly measured from the experiment at a large delay. By repeating the fitting approach for all object pixels and encoding the extracted 3 DOFs in an HSB color scheme, as illustrated in Fig. 1d, a Jones matrix profile of the object region can be obtained. Specifically, the hue (H), saturation (S), and brightness (B) of the HSB color scale are mapped to the 3

DOFs of the Jones matrix using $H = \pi + \theta_+ \pmod{2\pi}$, $S = \cos \theta_-$ and $B = t_r$, where t_r is defined as the relative transmission amplitude $|t_j / t_{\text{ref}}|$ between the object pixel and reference pixel, ranging from 0 to 1. More details about the experimental setup are discussed in Supplementary Information.

However, due to the probabilistic nature of photon arrival events, this fitting approach requires a substantial number of time frames in the experiment to accumulate a sufficient number of photons for obtaining well-fitted $g_2^{(ij)}(\tau)$ curves, as illustrated in Fig. 1c. Moreover, this approach demands considerable effort in developing an effective analytical model for extracting the Jones matrix (Eqs. 3 to 5), which poses challenges in adapting it to various other applications. In contrast, the present study employs DL to address these limitations. Another objective is to investigate whether the DL-assisted approach can outperform the fitting method, particularly by using a reduced number of time frames of experimental data. In other words, the aim is to achieve imaging with the smallest possible number of photons.

We now introduce a DL methodology for Jones matrix imaging by combining feature extraction and regression approaches. The training data is generated as follows. As shown in Fig. 2a, various arrangements of the 3 DOFs, $\{\theta_+^{(j)}, \theta_-^{(j)}, t_j\}$ defined by Eq. 2, are initially generated randomly from the uniform distributions $U(0, 2\pi)$, $U(0, \frac{\pi}{2})$ and $U(0.1, 0.5)$, respectively. Each configuration corresponds to a possible sample for imaging, comprising an object pixel (indexed as j) characterized by the unknown 3 DOFs and four reference pixels (H, D, V, and A) possessing predetermined Jones matrix elements outlined in Eq. 1. It is important to note that each object pixel is constructed with an ‘interleaved’ structure featuring plasmonic nanoslots, where two pairs of nanoslots possess the same length but different rotation angles, thereby controlling the 3 DOFs of the Jones matrix.

Subsequently, we simulate the probabilistic photon arrival events, including two-photon interference ($\tau = 0$), for both the object and reference pixels (total of 5 pixels) within each configuration. In particular, considering that LCP is analyzed in the experiment, the coherent amplitudes of photon pairs before detection by the camera for the reference (indexed as i) and

object (indexed as j) pixels in each time frame can be described as

$$\begin{aligned} \alpha_i &= t_{LL}^{(i)}\alpha_L + t_{LR}^{(i)}\alpha_R e^{i\phi} \\ \alpha_j &= t_{LL}^{(j)}\alpha_L + t_{LR}^{(j)}\alpha_R e^{i\phi} \end{aligned} \quad (6)$$

where $|\alpha_L| = |\alpha_R| = 1/\sqrt{2}$ (set as the coherent amplitudes of the incident photons without losing generality), and the random variable ϕ , sampled from a uniform distribution $U(0, 2\pi)$, represents the phase randomization of the incident photon pairs in the experiment. We have assumed that the photon pairs are generated through a weakly coherent light source. The number of photons (n) detected by the camera for the 5 pixels (4 reference pixels and 1 object pixel) in each time frame is sampled from a Poisson distribution $Pois(\mu_{det})$, where μ_{det} signifies the expectation value of the number of detected photons, accounting for factors like incident power, sample transmission, and detector efficiency. Each of these n detected photons is then randomly assigned to one of the 5 pixels based on the conditional probabilities $|\alpha_i|^2 / (\sum_i |\alpha_i|^2 + |\alpha_j|^2)$ for the four reference pixels and $|\alpha_j|^2 / (\sum_i |\alpha_i|^2 + |\alpha_j|^2)$ for the object pixel. Further details regarding the simulation model are discussed in the Supplementary Information.

With the simulated photon arrival data in place, we can form a corresponding correlation matrix C_j for each configuration. This matrix records the coincidence events between the object pixel and each reference pixel in each independent time frame, as illustrated in the lower panel of Fig. 2a. Rows and columns of C_j correspond to reference pixels and time frame indices, respectively. A value of '1' for each matrix element indicates a coincidence event, while '0' indicates otherwise. To streamline the use of the data with minimal post-processing in later stages, we directly input the C_j (after being flattened into a vector) into the β -VAE network, consisting of an encoder and a decoder as depicted in Fig. 2a. This probabilistic network model aims to identify the most concise representations of the input data. Specifically, during the unsupervised training process, the encoder encodes the input data into independent Gaussian distributions (with mean $\{\mu_n\}$ and variance $\{\sigma_n^2\}$) from which the compressed latent variables $\{L_n\}$ are sampled. The randomly sampled latent variables $\{L_n\}$ are then employed by the decoder to produce and parameterize the distribution from which the input data can be generated with maximal likelihood⁶⁷. Notably, a sample of latent variable $L_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$ is generated by $L_n = \mu_n + \sigma_n \epsilon_n$ to ensure the differentiability in the training process, where ϵ_n is a random number sampled from a unit Gaussian distribution $\mathcal{N}(0, 1^2)$. The loss function of the β -VAE in this work is defined as

$$\left\| C_j - C_j^r \right\|_2^2 + \beta \sum_n D_{KL}[\mathcal{N}(\mu_n, \sigma_n^2) \| \mathcal{N}(0, 1^2)] \quad (7)$$

The first term aims to minimize the mean-squared error (MSE) between the flattened input and output correlation matrices (flattened into a vector), while the second term aims to minimize the Kullback-Leibler divergence, aligning the probabilistic distributions of $\{L_n\}$ with independent unit Gaussian distributions $\mathcal{N}(0, 1^2)$. The adjustable hyperparameter β strikes a balance between reconstruction quality and the disentanglement of latent variables⁶⁴. We note that the binary cross entropy (BCE) can also be used instead of MSE to formulate the loss function in Eq. 7, with a similar performance presented in Fig. S3. More details about the β -VAE are involved in Supplementary Information. It is important to note that the β -VAE has the capability to determine the number of DOFs inherent in the input data, a revelation evident through the meaningful latent variables. In the current case, it is expected that the number of meaningful latent variables will be exactly 3. If it is larger than 3, it indicates that the network has not been adequately trained. Conversely, if it is less than 3, it suggests that the experimental procedure is not sufficiently designed to gather adequate

information, pointing to an upgrade of the experimental procedure. With a properly trained β -VAE network in place, the subsequent step involves training an additional regression network. This network is designed to map the extracted meaningful latent variables to the corresponding 3 DOFs $\{\theta_+^{(j)}, \theta_-^{(j)}, t_j\}$ of the Jones matrix. The input of such a regression network is already the minimal representation of the data, making such a network very efficient. Overall, armed with the successively trained β -VAE and regression networks, the 3 DOFs of each object pixel can be extracted from their corresponding correlation matrix.

We generate a total of 50,000 datasets, each comprising distinct configurations of $\{\theta_+^{(j)}, \theta_-^{(j)}, t_j\}$ and their corresponding correlation matrices $\{C_j\}$ via numerical simulations. Each dataset includes a total of 2,000 time frames to ensure generality. Considering the probabilistic nature of photon arrival and each time frame is independent of the others, we manually permute the sequence of time frames in $\{C_j\}$ for each dataset. This permutation does not affect the probability of coincidence events and enforces time translational symmetry in the network. This step allows us to expand the datasets to 200,000 instances, with 81% allocated for training, 9% for validation, and the remainder for testing. The β -VAE is designed with an input size of 8000 (4 reference pixels by 2000 time frames), and both the encoder and decoder consist of 3 hidden linear layers (fully connected layers), capable of extracting up to 5 latent variables. We deliberately set more than 3 latent variables in the latent space to verify the β -VAE's ability to extract the number of DOFs later. In the unsupervised training process, the loss function in Eq. 7 is employed to optimize the β -VAE. Here, to balance the training time and the capability of DOFs extraction, β is set to 1, with the ADAM optimizer and an initial learning rate of 0.0003. We note that the other values of β around 1 (ranging from 0.3 to 1.5 in our case) can also achieve a similar performance. Very high values of β may result in no meaningful latent variables and very low values of β may fail to enforce independence constraints of the latent variables⁶⁸. By applying the testing data from simulations to the trained encoder, we acquire the statistics of distribution parameters $\{\mu_n\}$ and $\{\sigma_n^2\}$ of the extracted latent variables. As demonstrated in Fig. 2b, c, the trained β -VAE correctly identifies the number of DOFs in the data. This is evident through the three latent variables $\{L_1, L_3, L_5\}$ with high variance (close to 1) in μ_n and low mean in σ_n^2 (close to 0)⁶⁸. The three meaningful latent variables are expected to represent the 3 DOFs in the Jones matrix, as illustrated in Fig. S4. Conversely, the remaining two latent variables exhibit no meaningful information, as indicated by their distribution parameter statistics. Subsequently, a regression network, constructed with fully connected layers in a 3-50-50-50-50-4 configuration, is trained using the same training data. This network transforms $\{L_1, L_3, L_5\}$ into the 3 DOFs of the Jones matrix. Given that $\theta_+^{(j)}$ is a cyclic variable ranging from 0 to 2π , we employ $\{\sin \theta_+^{(j)}, \cos \theta_+^{(j)}\}$ as output from the regression network. The loss function used to optimize the regression network is the MSE. For additional details about the architectures of both the β -VAE and regression networks, refer to Fig. S2. As shown in Fig. S5, the testing results of the regression network reveal that the Pearson correlation coefficients between the target and predicted variables average at 0.93. This outcome suggests that the extracted latent variables $\{L_1, L_3, L_5\}$ experience little information loss and can be successfully reformulated into the 3 DOFs of the Jones matrix. Notably, owing to the time translation symmetry, we can apply the above trained β -VAE to C_j matrices with any numbers of time frames (≤ 2000) through padding for convenience, instead of training a series of β -VAE networks for each different number of time frames. For instance, by replicating a C_j containing a total of 100-time frames by 20 times, we can obtain a padded C_j^r with 2000 time frames, matching the input size of the trained β -VAE. Moving forward, in subsequent stages, we utilize experimental data measured from the SPAD camera to test the trained networks and extract Jones matrix images from the samples.

It is worth noting that instead of directly training a supervised neural network to map the correlation matrix into the DOFs of the Jones matrix as a black-box process, our proposed DL approach offers distinct advantages,

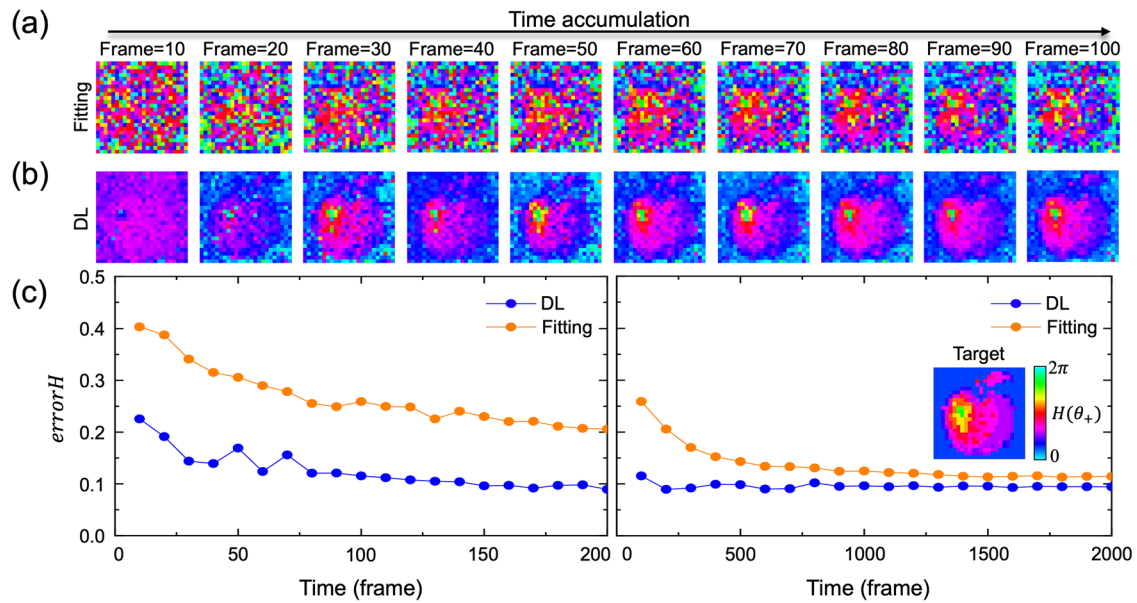


Fig. 3 | Extracted Jones matrix images and performance from the fitting algorithm and DL on one DOF case. **a, b** Extracted images of $\{\theta_+^{(j)}\}$ using different numbers of time frames of experimental data from the fitting algorithm on **a**, and DL on **b**. **c** The error of extracted images from the two different methods. The error from

DL is lower than that from the fitting method when the number of time frames is small. Also, the DL approach needs fewer number of time frames to achieve convergence. The target profile of $\{\theta_+^{(j)}\}$ is coded into hue color and shown in the insert of the right panel.

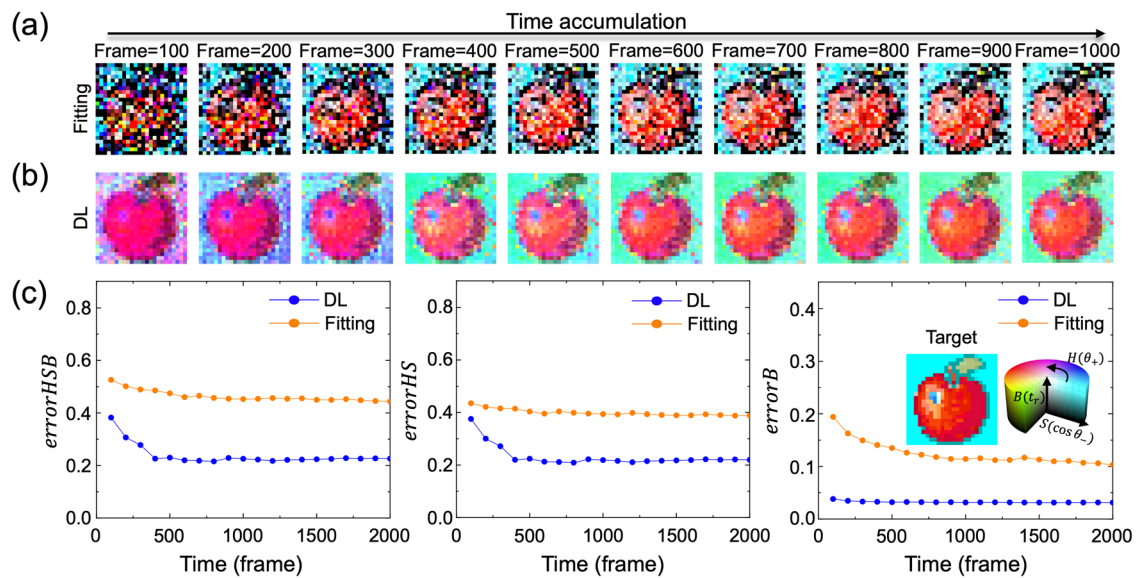


Fig. 4 | Extracted Jones matrix images and performance from the two approaches on 3 DOFs case. **a, b** Extracted Jones matrix images using different numbers of time frames of experimental data from the fitting algorithm on **a**, and DL on **b**. **c** The

performance between the target and extracted images from the two methods. The 3 DOFs in the Jones matrix are coded with an HSB color scheme.

particularly when dealing with intricate and high-dimensional datasets. In such scenarios, it can be challenging to determine what components are crucial for accurate forward prediction without any prior physical insights. In our methodology, the extensive dataset is navigated through an unsupervised network (β -VAE) that automates the quest for minimal representations of the data. This automation seeks to identify the exact number of DOFs while a relatively simple supervised network is employed in the final phase to revert these minimal representations back to the correct scale of the imaging variables. The abilities to assess the information sufficiency and to extract DOFs achieved by the β -VAE collectively bestow a potent tool for automatic information extraction and model development. This approach

empowers a more automated and efficient exploration of data, aiding in the extraction of essential physical insights.

To compare our proposed DL approach with the previous fitting algorithm (approximated analytical model), we initially showcase a single DOF case of Jones matrix imaging in Fig. 3. In this scenario, we introduce a variable profile for the argument of the off-diagonal elements $\theta_+^{(j)}$ in Eq. 2, while maintaining $\theta_-^{(j)} = 0$ and $|t_j| = |t_{ref}|$. The target phase profile $\{\theta_+^{(j)}\}$, color-coded as hue (H), is intentionally designed to resemble an apple shape, as illustrated in the inset of Fig. 3c. To ensure a fair comparison, the same experimental data is employed for both methods. By applying the

approximated analytical model from Eqs. 3 and 5 and subsequently fitting, we present the extracted images using varying total numbers of time frames (ranging from 10 to 100) for the fitting algorithm, as depicted in Fig. 3a. In addressing the presence of dead pixels within the SPAD camera, these pixels are substituted with the average value of their neighboring counterparts.

Nonetheless, the extracted images of $\{\theta_+^{(j)}\}$ from the fitting algorithm, utilizing limited total numbers of time frames, exhibit significant noise, rendering it challenging to discern the object. Conversely, the images of $\{\theta_+^{(j)}\}$ extracted from the DL approach using the same quantity of time frames are showcased in Fig. 3b. Notably, these images progressively improve in clarity and exhibit reduced noise as the time accumulation increases.

In Fig. 3c, we draw a comparison between the performance of the two distinct methods employed for imaging. As θ_+ is a cyclic variable and is color-coded as hue (H), we introduce the term “errorH” defined as

$$\text{errorH} = \frac{1}{N\pi} \sum_{x,y} \min\left(\left|H_{x,y} - H'_{x,y}\right|, 2\pi - \left|H_{x,y} - H'_{x,y}\right|\right), \quad (8)$$

which can range from 0 to 1. This metric quantifies the discrepancy between the target ($H_{x,y}$) and extracted ($H'_{x,y}$) images, where x and y denote pixel indices in two dimensions, and N is the total number of object pixels. Examining the left panel of Fig. 3c, it becomes evident that, with the constraint of a limited total number of time frames, the image error originating from the DL approach is notably lower than that produced by the fitting algorithm with around 50% improvement. This observation aligns consistently with the quality of the extracted images presented in Fig. 3a, b. Additionally, the right panel of Fig. 3c illustrates the comparative performances of the two methods based on a larger total number of time frames. When assessing the image error convergence over time frames, the DL approach achieves convergence around 200-time frames, yielding an error of approximately 0.1. The extracted images utilizing larger quantities of time frames of data are showcased in Fig. S6. Remarkably, the DL approach can successfully extract the apple-shaped object with just 30 total time frames (at 1.5 times the converged error), equivalent to merely 14 photons collected per pixel during the experiment.

Taking into account its generality, we also present the application of both approaches to Jones matrix imaging encompassing all the 3 DOFs. In this instance, all 3 DOFs $\{\theta_+^{(j)}, \theta_-^{(j)}, t_j\}$ within the object region undergo variation, and the intended Jones matrix image is encoded within an HSB color bar, showcased in the inset of Fig. 4c. Figure 4a, b are the derived Jones matrix images using varying total numbers of experimental time frames (ranging from 100 to 1000) for the fitting and DL approaches, respectively. As the total number of time frames escalates, the images obtained via the fitting algorithm continue to exhibit some level of noise. Conversely, the DL approach is able to deduce a clear image resembling the target with approximately 400 total time frames of data with converged image quality. We note that compared to the cyan background ($\theta_+^{(j)} = 0$) in the target image, the background of the extracted images from the DL approach has a slight bias ($\theta_+^{(j)}$ is 1.92π on average for these pixels) resulting from the cyclic property of θ_+ (see Figs. S5 and S7). The extracted images utilizing larger numbers of time frames of data are showcased in Fig. S6. By calculating the distance on a pixel-by-pixel basis within the HSB color space (in a cylindrical coordinate system) between the target ($H_{x,y}, S_{x,y}, B_{x,y}$) and extracted ($H'_{x,y}, S'_{x,y}, B'_{x,y}$) images, we introduce the following definition of error distance:

$$\text{errorHSB} = \frac{1}{\sqrt{5}N} \sum_{x,y} \sqrt{\left(S_{x,y} - S'_{x,y}\right)^2 + 4S_{x,y}S'_{x,y}\sin^2\frac{H_{x,y} - H'_{x,y}}{2} + \left(B_{x,y} - B'_{x,y}\right)^2}, \quad (9)$$

which ranges from 0 to 1. It serves as an evaluative measure of the quality of the extracted Jones matrix images from both methods. As illustrated in the left panel of Fig. 4c, it is evident that the error stemming from the DL approach is markedly smaller compared to the error originating from the fitting algorithm. Notably, the error from the DL approach converges to be lower than 1.5 times the converged error at around 200-time frames, corresponding to 88 photons collected per pixel during the experiment. Moreover, we can “decompose” the errorHSB into two components (averaged error distance projected into a subspace up to an arbitrary constant), each defined as follows:

$$\text{errorHS} = \frac{1}{\sqrt{5}N} \sum_{x,y} \sqrt{\left(S_{x,y} - S'_{x,y}\right)^2 + 4S_{x,y}S'_{x,y}\sin^2\frac{H_{x,y} - H'_{x,y}}{2}} \quad (10)$$

$$\text{errorB} = \frac{1}{\sqrt{5}N} \sum_{x,y} |B_{x,y} - B'_{x,y}|. \quad (11)$$

These components are illustrated in the middle and right panels of Fig. 4c, respectively. The errorHS, which pertains to θ_+ (H) and $\cos\theta_-$ (S) within the 2D polar coordinate system, constitutes the major portion of the total error for both methods. However, the errorHS resulting from the DL approach is notably smaller. Particularly significant is the observation that the errorB, associated with t_r (B), is nearly negligible in the DL approach, distinctly smaller than in the fitting algorithm. Upon analyzing the extracted profiles of the 3 DOFs separately, as demonstrated in Fig. S7, it becomes evident that the noise present in the extracted images shown in Fig. 4a predominantly originates from the t_r channel. Notably, although we utilized the semi-analytic algorithm with different approximations to extract the Jones matrix image, the proposed DL method still exhibits better performance, as illustrated in Fig. S8. The superior performance exhibited by the DL approach in contrast to the fitting method indicates that the latter’s approximation possesses room for improvement for the semi-analytic model. From this perspective, the integration of DL in algorithm formulation sets an “upper bound” on the potential attainable via an analytic algorithm, offering guidance for imaging equipment designs and algorithm developments. Moreover, we note that by using simulation data for training, there can be potential mismatches between the realistic experimental situation and the idealized simulations^{69,70}: dead pixels can appear occasionally and some of the pixels need longer setup time in our case. We envision that it is also possible to directly use experimental data for training purposes to further improve the image qualities⁷¹.

Discussion

Our DL approach can assess the sufficiency of information present within data obtained from a given experimental procedure. Such capability to differentiate the sufficiency of information in data is beneficial for us to distinguish between well-posed and ill-posed experimental procedures. Here, we have applied such an approach to find the minimal representation of photon arrival data to formulate an algorithm for Jones matrix imaging in the low-light regime. As a result, less than a hundred photons per pixel are needed to obtain an accurate image, which is a significant improvement over the previous algorithm. In low-light imaging where the imaging process becomes probabilistic or noisy in nature, a DL approach is particularly useful to obtain the optimal algorithm to get a clear image when the number of photons is limited. In other words, in benchmarking existing algorithms, the DL approach can thus tell us whether the existing algorithms are already optimal. As a whole, such an approach will be generally helpful for developing equipment and the associated algorithms for a wide range of imaging applications, including low-light imaging^{72,73}, nondestructive testing⁷⁴, ultrasound imaging⁶⁶, and depth detection⁷⁵.

Methods

The sample fabrication flow utilized in this work is as follows. Before the fabrication, the glass substrate undergoes a 24-hour immersion in concentrated sulfuric acid. It is then cleaned sequentially using distilled water and acetone. Next, a 50nm-thick silver film is deposited onto the substrate using an E-beam Evaporator (AST 600 EI Evaporator) with a deposition rate of 1 Å/s. To complete the process, the designed nanoslot pattern is written onto the silver film using a focus ion beam (FEI Helios G4 UX, 30 kV, 41 pA). At this stage, the metasurface is prepared and ready for experimentation. More details about the experimental setup are discussed in Supplementary Information.

Data availability

The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

Received: 9 October 2023; Accepted: 9 January 2024;

Published online: 06 March 2024

References

- Kok, P. & Lovett, B. W. *Introduction to Optical Quantum Information Processing*. (Cambridge University Press, 2010).
- Loudon, R. *The Quantum Theory of Light*. (OUP Oxford, 2000).
- Eisaman, M. D., Fan, J., Migdall, A. & Polyakov, S. V. Invited review article: Single-photon sources and detectors. *Rev. Sci. Instrum.* **82**, 071101 (2011).
- D'Angelo, M., Kim, Y. H., Kulik, S. P. & Shih, Y. Identifying entanglement using quantum ghost interference and imaging. *Phys. Rev. Lett.* **92**, 233601 (2004).
- Li, L. et al. Efficient photon collection from a nitrogen vacancy center in a circular bullseye grating. *Nano Lett.* **15**, 1493–1497 (2015).
- Fischer, K. A., Müller, K., Lagoudakis, K. G. & Vučković, J. Dynamical modeling of pulsed two-photon interference. *N. J. Phys.* **18**, 113053 (2016).
- Gilberte Basset, M. et al. Perspectives for applications of quantum imaging. *Laser Photon. Rev.* **13**, 1900097 (2019).
- Lyons, A. et al. Attosecond-resolution hong-ou-mandel interferometry. *Sci. Adv.* **4**, eaap9416 (2018).
- Ndagano, B. et al. Quantum microscopy based on Hong–Ou–Mandel interference. *Nat. Photon.* **16**, 384–389 (2022).
- Murray, R. & Lyons, A. Two-photon interference LiDAR imaging. *Opt. Express* **30**, 27164–27170 (2022).
- Ibarra-Borja, Z., Sevilla-Gutiérrez, C., Ramírez-Alarcón, R., Cruz-Ramírez, H. & U'Ren, A. B. Experimental demonstration of full-field quantum optical coherence tomography. *Photon. Res.* **8**, 51–56 (2020).
- Chrapkiewicz, R., Jachura, M., Banaszek, K. & Wasilewski, W. Hologram of a single photon. *Nat. Photon.* **10**, 576–579 (2016).
- Tenne, R. et al. Super-resolution enhancement by quantum image scanning microscopy. *Nat. Photon.* **13**, 116–122 (2019).
- Kong, L. J., Sun, Y., Zhang, F., Zhang, J. & Zhang, X. High-dimensional entanglement-enabled holography. *Phys. Rev. Lett.* **130**, 053602 (2023).
- Defienne, H., Ndagano, B., Lyons, A. & Faccio, D. Polarization entanglement-enabled quantum holography. *Nat. Phys.* **17**, 591–597 (2021).
- Gregory, T., Moreau, P. A., Toninelli, E. & Padgett, M. J. Imaging through noise with quantum illumination. *Sci. Adv.* **6**, eaay2652 (2020).
- Hor-Meyll, M. et al. Deterministic quantum computation with one photonic qubit. *Phys. Rev. A* **92**, 012337 (2015).
- Kwon, H., Arbabi, E., Kamali, S. M., Faraji-Dana, M. & Faraon, A. Single-shot quantitative phase gradient microscopy using a system of multifunctional metasurfaces. *Nat. Photon.* **14**, 109–114 (2020).
- Engay, E., Huo, D., Malureanu, R., Bunea, A. I. & Lavrinenko, A. Polarization-dependent all-dielectric metasurface for single-shot quantitative phase imaging. *Nano Lett.* **21**, 3820–3826 (2021).
- Balaur, E. et al. Plasmon-induced enhancement of ptychographic phase microscopy via sub-surface nanoaperture arrays. *Nat. Photon.* **15**, 222–229 (2021).
- Wesemann, L., Rickett, J., Davis, T. J. & Roberts, A. Real-time phase imaging with an asymmetric transfer function metasurface. *ACS Photon.* **9**, 1803–1807 (2022).
- Arbabi, E., Kamali, S. M., Arbabi, A. & Faraon, A. Full-Stokes imaging polarimetry using dielectric metasurfaces. *Acs Photon.* **5**, 3132–3140 (2018).
- Yang, Z. et al. Generalized Hartmann-Shack array of dielectric metalens sub-arrays for polarimetric beam profiling. *Nat. Commun.* **9**, 4607 (2018).
- Rubin, N. A. et al. Matrix Fourier optics enables a compact full-Stokes polarization camera. *Science* **365**, eaax1839 (2019).
- Rubin, N. A. et al. Imaging polarimetry through metasurface polarization gratings. *Opt. Express* **30**, 9389–9412 (2022).
- Li, L. et al. Monolithic full-Stokes near-infrared polarimetry with chiral plasmonic metasurface integrated graphene–silicon photodetector. *ACS Nano* **14**, 16634–16642 (2020).
- Faraji-Dana, M. et al. Compact folded metasurface spectrometer. *Nat. Commun.* **9**, 4196 (2018).
- Faraji-Dana, M. et al. Hyperspectral imager with folded metasurface optics. *ACS Photon.* **6**, 2161–2167 (2019).
- Yesilkoy, F. et al. Ultrasensitive hyperspectral imaging and biodetection enabled by dielectric metasurfaces. *Nat. Photon.* **13**, 390–396 (2019).
- McClung, A., Samudrala, S., Torfeh, M., Mansouree, M. & Arbabi, A. Snapshot spectral imaging with parallel metasystems. *Sci. Adv.* **6**, eabc7646 (2020).
- Zhao, W. et al. Full-color hologram using spatial multiplexing of dielectric metasurface. *Optics Lett.* **41**, 147–150 (2016).
- Ren, H. et al. Metasurface orbital angular momentum holography. *Nat. Commun.* **10**, 2986 (2019).
- Li, Y. et al. Orbital angular momentum multiplexing and demultiplexing by a single metasurface. *Adv. Opt. Mater.* **5**, 1600502 (2017).
- Karimi, E. et al. Generating optical orbital angular momentum at visible wavelengths using a plasmonic metasurface. *Light Sci. Appl.* **3**, e167–e167 (2014).
- Zhou, H. et al. Polarization-encrypted orbital angular momentum multiplexed metasurface holography. *ACS Nano* **14**, 5553–5559 (2020).
- Ren, H. et al. Complex-amplitude metasurface-based orbital angular momentum holography in momentum space. *Nat. Nanotechnol.* **15**, 948–955 (2020).
- Lee, D., Gwak, J., Badloe, T., Palomba, S. & Rho, J. Metasurfaces-based imaging and applications: from miniaturized optical components to functional imaging platforms. *Nanoscale Adv.* **2**, 605–625 (2020).
- Wei, F. et al. Wide field super-resolution surface imaging through plasmonic structured illumination microscopy. *Nano Lett.* **14**, 4634–4639 (2014).
- Zheng, P. et al. Metasurface-based key for computational imaging encryption. *Sci. Adv.* **7**, eabg0363 (2021).
- Colburn, S., Zhan, A. & Majumdar, A. Metasurface optics for full-color computational imaging. *Sci. Adv.* **4**, eaar2114 (2018).
- Li, L. et al. Metalens-array-based high-dimensional and multiphoton quantum source. *Science* **368**, 1487–1490 (2020).
- Li, Q. et al. A non-unitary metasurface enables continuous control of quantum photon–photon interactions from bosonic to fermionic. *Nat. Photon.* **15**, 267–271 (2021).
- Wang, K. et al. Quantum metasurface for multiphoton interference and state reconstruction. *Science* **361**, 1104–1108 (2018).

44. Li, C. et al. Arbitrarily structured quantum emission with a multifunctional metalens. *eLight* **3**, 19 (2023).
45. Zhou, J. et al. Metasurface enabled quantum edge detection. *Sci. Adv.* **6**, eabc4385 (2020).
46. Altuzarra, C. et al. Imaging of polarization-sensitive metasurfaces with quantum entanglement. *Phys. Rev. A* **99**, 020101 (2019).
47. del Hoyo, J., Sanchez-Brea, L. M. & Soria-Garcia, A. Calibration method to determine the complete Jones matrix of SLMs. *Opt. Lasers Eng.* **151**, 106914 (2022).
48. Park, J., Yu, H., Park, J. H. & Park, Y. LCD panel characterization by measuring full Jones matrix of individual pixels using polarization-sensitive digital holographic microscopy. *Opt. Express* **22**, 24304–24311 (2014).
49. Tiwari, V., Gautam, S. K., Naik, D. N., Singh, R. K. & Bisht, N. S. Characterization of a spatial light modulator using polarization-sensitive digital holography. *Appl. Opt.* **59**, 2024–2030 (2020).
50. Dai, X. et al. Quantitative Jones matrix imaging using vectorial Fourier ptychography. *Biomed. Opt. Express* **13**, 1457–1470 (2022).
51. Park, K. et al. Jones matrix microscopy for living eukaryotic cells. *ACS Photon.* **8**, 3042–3050 (2021).
52. Sreelal, M. M., Vinu, R. V. & Singh, R. K. Jones matrix microscopy from a single-shot intensity measurement. *Opt. Lett.* **42**, 5194–5197 (2017).
53. Chen, G. X. et al. Compact common-path polarisation holographic microscope for measuring spatially-resolved Jones matrix parameters of dynamic samples in microfluidics. *Opt. Commun.* **503**, 127460 (2022).
54. Yung, T. K. et al. Polarization coincidence images from metasurfaces with HOM-type interference. *Iscience* **25**, 104155 (2022).
55. Yung, T. K., Liang, H., Xi, J., Tam, W. Y. & Li, J. Jones-matrix imaging based on two-photon interference. *Nanophotonics* **12**, 579–588 (2022).
56. Gao, Y. J. et al. Multichannel distribution and transformation of entangled photons with dielectric metasurfaces. *Phys. Rev. Lett.* **129**, 023601 (2022).
57. Kudyshev, Z. A., Shalaev, V. M. & Boltasseva, A. Machine learning for integrated quantum photonics. *Acs Photon.* **8**, 34–46 (2020).
58. Chen, X. et al. Machine learning for optical scanning probe nanoscopy. *Adv. Mater.* <https://arxiv.org/abs/2204.098202109171> (2022).
59. Torlai, G. et al. Neural-network quantum state tomography. *Nat. Phys.* **14**, 447–450 (2018).
60. You, C. et al. Identification of light sources using machine learning. *Appl. Phys. Rev.* **7**, 021404 (2020).
61. Kudyshev, Z. A. et al. Rapid classification of quantum sources enabled by machine learning. *Adv. Quant. Technol.* **3**, 2000067 (2020).
62. Bhusal, N. et al. Spatial mode correction of single photons using machine learning. *Adv. Quant. Technol.* **4**, 2000103 (2021).
63. Kudyshev, Z. A. et al. Machine learning assisted quantum super-resolution microscopy. *Nat. Commun.* **14**, 4828 (2023).
64. Burgess, C. P. et al. Understanding disentangling in β -VAE. <https://arxiv.org/abs/1804.03599> (2018).
65. Chen, M., Shi, X., Zhang, Y., Wu, D. & Guizani, M. Deep feature learning for medical image analysis with convolutional autoencoder neural network. *IEEE Transactions on. Big Data* **7**, 750–758 (2017).
66. Van Sloun, R. J., Cohen, R. & Eldar, Y. C. Deep learning in ultrasound imaging. *Proc. IEEE* **108**, 11–29 (2019).
67. Kingma, D. P., & Welling, M. Auto-encoding variational bayes. <https://arxiv.org/pdf/1312.6114.pdf> (2013).
68. Lu, P. Y., Kim, S. & Soljačić, M. Extracting interpretable physical parameters from spatiotemporal systems using unsupervised learning. *Phys. Rev. X* **10**, 031056 (2020).
69. Liu, R., Sun, Y., Zhu, J., Tian, L. & Kamilov, U. S. Recovery of continuous 3d refractive index maps from discrete intensity-only measurements using neural fields. *Nat. Mach. Intell.* **4**, 781–791 (2022).
70. Guo, Z. et al. Physics-assisted generative adversarial network for X-ray tomography. *Opt. Express* **30**, 23238–23259 (2022).
71. Li, L. et al. Intelligent metasurface imager and recognizer. *Light Sci. Appl.* **8**, 97 (2019).
72. Lore, K. G., Akintayo, A. & Sarkar, S. LLNet: a deep autoencoder approach to natural low-light image enhancement. *Pattern Recognit.* **61**, 650–662 (2017).
73. Ren, W. et al. Low-light image enhancement via a deep hybrid network. *IEEE Trans. Image Process.* **28**, 4364–4375 (2019).
74. Bertolotti, J. et al. Non-invasive imaging through opaque scattering layers. *Nature* **491**, 232–234 (2012).
75. Godard, C., MacAodha, O., & Brostow, G. J. Unsupervised monocular depth estimation with left-right consistency. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 270–279. (IEEE, 2017).

Acknowledgements

The work is supported by the Hong Kong RGC (AoE/P-502/20-2, 16306521, R6015-18) and by Croucher Foundation (CAS20SC01, CF23SC01). T.L. acknowledges support from the Center for Advances in Reliability and Safety (CAIRS) admitted under AIR@InnoHK Research Cluster.

Author contributions

J.L. initiated the project and conceived the idea. J.X. and T.K.Y. designed and fabricated the metasurfaces. T.K.Y. conducted the experiment. J.X. developed the deep-learning method and analyzed the experimental data. J.X. prepared the figures and wrote the manuscript. All authors revised the manuscript. J.L. supervised 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/s44310-024-00002-z>.

Correspondence and requests for materials should be addressed to Jensen Li.

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

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

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

© The Author(s) 2024