PAPER • OPEN ACCESS

Process tomography of structured optical gates with convolutional neural networks

To cite this article: Tareq Jaouni *et al* 2024 *Mach. Learn.: Sci. Technol.* **5** 045071

IOP Publishing Mach. Learn.: Sci. Technol. 5 (2024) 045071 https://doi.org/10.1088/2632-2153/ad9ba8

MACHINE LEARNING
Science and Technology

OPEN ACCESS

RECEIVED
23 May 2024

REVISED
28 October 2024

ACCEPTED FOR PUBLICATION
22 November 2024

PUBLISHED
23 December 2024

PAPER

Process tomography of structured optical gates with convolutional neural networks

Tareq Jaouni[1], Francesco Di Colandrea[1,4,*], Lorenzo Amato[2,3], Filippo Cardano[4] and Ebrahim Karimi[1,5]

[1] Nexus for Quantum Technologies, University of Ottawa, K1N 5N6 Ottawa, ON, Canada
[2] Condensed Matter Theory Group, Paul Scherrer Institute, CH-5232 Villigen PSI, Switzerland
[3] Laboratory for Solid State Physics, ETH Zurich, CH-8093 Zurich, Switzerland
[4] Dipartimento di Fisica "Ettore Pancini", Università di Napoli Federico II, Complesso Universitario di Monte Sant'Angelo, Via Cintia, Napoli 80126, Italy
[5] National Research Council of Canada, 100 Sussex Drive, K1A 0R6 Ottawa, ON, Canada
[*] Author to whom any correspondence should be addressed.

E-mail: francesco.dicolandrea@uottawa.ca

Keywords: quantum process tomography, quantum optics, machine learning, convolutional neural networks

Abstract

Efficient and accurate characterization of an experimental setup is a critical requirement in any physical setting. In the quantum realm, the characterization of an unknown operator is experimentally accomplished via Quantum Process Tomography (QPT). This technique combines the outcomes of different projective measurements to reconstruct the underlying process matrix, typically extracted from maximum-likelihood estimation. Here, we exploit the logical correspondence between optical polarization and two-level quantum systems to retrieve the complex action of structured metasurfaces within a QPT-inspired context. In particular, we investigate a deep-learning approach that allows for fast and accurate reconstructions of space-dependent SU(2) operators by only processing a minimal set of measurements. We train a convolutional neural network based on a scalable U-Net architecture to process entire experimental images in parallel. Synthetic processes are reconstructed with average fidelity above 90%. The performance of our routine is experimentally validated in the case of space-dependent polarization transformations acting on a classical laser beam. Our approach further expands the toolbox of data-driven approaches to QPT and shows promise in the real-time characterization of complex optical gates.

1. Introduction

Quantum Process Tomography (QPT) is the extension of system identification to the quantum realm [1]. A quantum operator can be characterized by measuring how a set of inputs evolve under its action, the same way a dynamical system is classified based on measured outputs [2]. For experimental purposes, the tomography is crucial in verifying the proper functionality of a quantum device. It found applications in various experiments, from nuclear magnetic resonances [3] to cold atoms [4], trapped ions [5, 6], and photonic setups [7–18]. As in the case of state tomography, some optimization routine is typically needed to solve the problem of non-physical reconstructions, which are a consequence of experimental noise [19].

In photonic setups, the characterization of a set of waveplates acting on light polarization is a practical challenge. The well-established analogy between polarized light and two-level quantum systems can be used to map the optical problem into the reconstruction of an SU(2) gate acting on a single qubit, allowing it to be formulated within the mathematical framework of QPT [20].

In this paper, we address the more challenging scenario of characterizing optical SU(2) gates that are space-dependent. These operations are relevant in all applications requiring a local control of light polarization, such as polarization imaging [21], multiplexing [22], as well as the generation and manipulation of vector beams [23, 24]. Assuming photons propagate along z, QPT is required at each

transverse position (x, y), hereafter referred to as *pixel*. Iterative pixel-by-pixel solutions prove to be not optimal, as the computation time grows linearly with the number of pixels. Moreover, similar approaches overlook the overall pixel distributions, preventing the possibility of extracting richer information from the entire experimental images.

In order to tackle the increasing complexity, optimization strategies based on evolutionary methods and supervised learning have been recently illustrated in [17], where an optimal set of five polarimetric measurements has also been demonstrated. Here, instead, we adopt a convolutional neural network (CNN) to deliver real-time space-resolved reconstructions of complex unitary operators. Specifically, we train a fully convolutional network based on the U-Net architecture to associate a minimal set of polarimetric images with the process parameters. Originally used for bio-imaging segmentation [25], U-Nets have been adopted for image-to-image regression tasks in optics [26], such as for super-resolution algorithms [27], aberration correction [28], and phase retrieval [29–31]. In the context of quantum tomography, convolutional architectures have been employed for efficient implementations of Quantum State Tomography [32].

First, we test our scheme on synthetic experiments. Then, we compare its performance with the genetic routine devised in [17], taking into account both the reconstruction's timing and accuracy. Finally, our architecture is validated experimentally on different combinations of liquid-crystal metasurfaces (LCMSs) [33, 34], realizing space-dependent polarization transformations. This ultimately demonstrates the robustness of our approach to real experiment noise.

2. Theory

A qubit rotation of an angle 2Θ around the axis $\boldsymbol{n} = (n_x, n_y, n_z)$, with $0 \leqslant \Theta < \pi$ and $|\boldsymbol{n}| = 1$, is described by an SU(2) operator

$$\hat{U} = e^{-i\Theta \boldsymbol{n}\cdot\boldsymbol{\sigma}} = \cos(\Theta)\,\sigma_0 - i\sin(\Theta)\,(\boldsymbol{n}\cdot\boldsymbol{\sigma}), \tag{1}$$

where σ_0 is the 2×2 identity matrix and $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ is the vector of the three Pauli matrices. The gate tomography is typically performed by processing an overcomplete set of projective measurements of the form

$$I_{ab} = \left|\langle b|\hat{U}|a\rangle\right|^2, \tag{2}$$

where $|a\rangle$ and $|b\rangle$ are extracted from the sets of eigenstates of the Pauli matrices, forming three sets of Mutually Unbiased Bases in two dimensions [1]. The process matrix is then retrieved via a maximum-likelihood approach, i.e. by minimizing a cost function expressing the distance between the experimental outcomes I_{ab}^{exp} and the corresponding theoretical predictions I_{ab}^{th} [9, 20, 35]:

$$\mathcal{L} = \sum_{ab}\left(I_{ab}^{\mathrm{exp}} - I_{ab}^{\mathrm{th}}\right)^2. \tag{3}$$

This routine is inefficient when it is independently executed on multiple gates depending on some external parameters, such as space-dependent processes.

Specifically, a space-dependent process can be modeled as a functional $\hat{U}(x, y)$ mapping the transverse plane to the SU(2) group:

$$\hat{\mathcal{U}} = \sum_{(x,y)} \hat{U}(x,y)\,|x,y\rangle\langle x,y|, \tag{4}$$

where we have decomposed the complex unitary operator $\hat{\mathcal{U}}$ in local SU(2) operators.

The polarization of photons provides a natural way of encoding qubits, which can be manipulated via optical waveplates. In the circular polarization basis, where $|L\rangle = (1,0)^T$ and $|R\rangle = (0,1)^T$ are left and right circular polarization states, respectively, a waveplate $R_{\delta,\alpha}$ having birefringence δ and optic axis oriented at an angle α with respect to the horizontal direction can be expressed in the matrix form

$$R_{\delta,\alpha} = \begin{pmatrix} \cos(\delta/2) & i\sin(\delta/2)e^{-2i\alpha} \\ i\sin(\delta/2)e^{2i\alpha} & \cos(\delta/2) \end{pmatrix}. \tag{5}$$

Here, T stands for the transpose operator. A single waveplate thus implements a qubit rotation of an angle $-\delta$ around the equatorial axis $\boldsymbol{n} = (\cos 2\alpha, \sin 2\alpha, 0)$ (cf equation (1)). Nevertheless, more general operations can be realized by cascading multiple waveplates [36, 37]. Accordingly, characterizing a set of waveplates is mathematically equivalent to performing the QPT of an SU(2) operator. In this case, the measurements of

equation (2) are realized as polarimetric measurements, involving the Stokes states $|L\rangle$ and $|R\rangle$, $|H\rangle = (|L\rangle + |R\rangle)/\sqrt{2}$ and $|V\rangle = (|L\rangle - |R\rangle)/\sqrt{2}i$ (horizontal and vertical polarizations, respectively), $|D\rangle = (|L\rangle + i|R\rangle)/\sqrt{2}$ and $|A\rangle = (|L\rangle - i|R\rangle)/\sqrt{2}$ (diagonal and antidiagonal polarizations, respectively). Space-dependent processes can be realized via waveplates exhibiting a patterned optic-axis orientation, $\alpha = \alpha(x,y)$. The tomography thus consists of determining the process parameters from a set of polarimetric measurements. In our synthetic and experimental realizations, a minimal set of five measurements is considered: $\{I_{LL}, I_{LH}, I_{LD}, I_{HH}, I_{HD}\}$ [17]. To the best of our knowledge, this is the first implementation of an optimal QPT of SU(2) gates.

As discussed in previous studies [17, 38–40], the main limitation of non-interferometric approaches is that these cannot capture any global phase. For this reason, we cannot distinguish a process $\hat{U}$, having parameters $(\Theta, \boldsymbol{n})$, from $e^{i\pi}\hat{U} = -\hat{U}$, having parameters $(\pi - \Theta, -\boldsymbol{n})$, as these generate the same experimental outcomes. We discuss how to solve this ambiguity in the next section.

3. Methodology

3.1. Parametrization of the training samples

Our network processes a set of five polarimetric images having resolution $N \times N$, where $N = 64$ pixels. For each parameter describing an SU(2) process, we generate a random continuous two-dimensional function $f(x,y)$ via a discrete Fourier decomposition:

$$\begin{aligned} f(x,y) = \sum_{i=0}^{\Omega_x}\sum_{j=0}^{\Omega_y} & c_{ij}^{(1)} \cos\frac{2\pi i x}{N}\cos\frac{2\pi j y}{N} + \\ & c_{ij}^{(2)} \cos\frac{2\pi i x}{N}\sin\frac{2\pi j y}{N} + \\ & c_{ij}^{(3)} \sin\frac{2\pi i x}{N}\cos\frac{2\pi j y}{N} + \\ & c_{ij}^{(4)} \sin\frac{2\pi i x}{N}\sin\frac{2\pi j y}{N}, \end{aligned} \tag{6}$$

where all the coefficients $c_{ij}^{(m)}$ are uniformly extracted from the range $[-1,1]$. The maximal frequencies Ω_x and Ω_y, respectively for the x and y axes, are uniformly sampled from the range $[0,5]$, thereby ensuring high-resolution pixelation: $\max(\Omega_{x,y}) \ll N$. The function is then rescaled to the ranges $[0,\pi]$ and $[-1,1]$, respectively for the rotation angle Θ and the components of the vector $\boldsymbol{n}$. Finally, the vector $\boldsymbol{n} = (n_x, n_y, n_z)$ is normalized at each transverse position.

We also include a possible rotation of the reference frame:

$$\begin{aligned} x \to x' &= \cos\xi\, x - \sin\xi\, y, \\ y \to y' &= \sin\xi\, x + \cos\xi\, y, \end{aligned} \tag{7}$$

where ξ is an angle extracted from a uniform distribution in the range $[-\xi_M, \xi_M]$, with $\xi_M = 5°$. To remove the ambiguity between processes U and $-U$, if the first pixel of the n_z component turns out to be negative, the following rule is applied:

$$\begin{aligned} \Theta(x,y) &\to \pi - \Theta(x,y), \\ n_x(x,y) &\to -n_x(x,y), \\ n_y(x,y) &\to -n_y(x,y), \\ n_z(x,y) &\to -n_z(x,y). \end{aligned} \tag{8}$$

We compute the minimal set of five polarimetric measurements associated with each process (cf section 2), which represents the input layer of the network. To simulate experimental noise, each pixel of each polarimetric measurement is perturbed with a value extracted from a Gaussian distribution with zero mean and standard deviation $\sigma = 0.02$. The output layer contains the parameter $\Theta(x,y)$ and the spherical representation of the vector $\boldsymbol{n}$:

$$\begin{aligned} \theta(x,y) &= \arccos n_z(x,y) \\ \phi(x,y) &= \mathrm{mod}\left[\mathrm{atan2}(n_y(x,y), n_x(x,y)), 2\pi\right], \end{aligned} \tag{9}$$

where $\theta(x,y)$ and $\phi(x,y)$ are the polar and azimuthal angles on the Poincaré sphere, respectively, and $\mathrm{atan2}(x,y)$ is the two-argument arctangent function, which distinguishes between diametrically opposite directions.

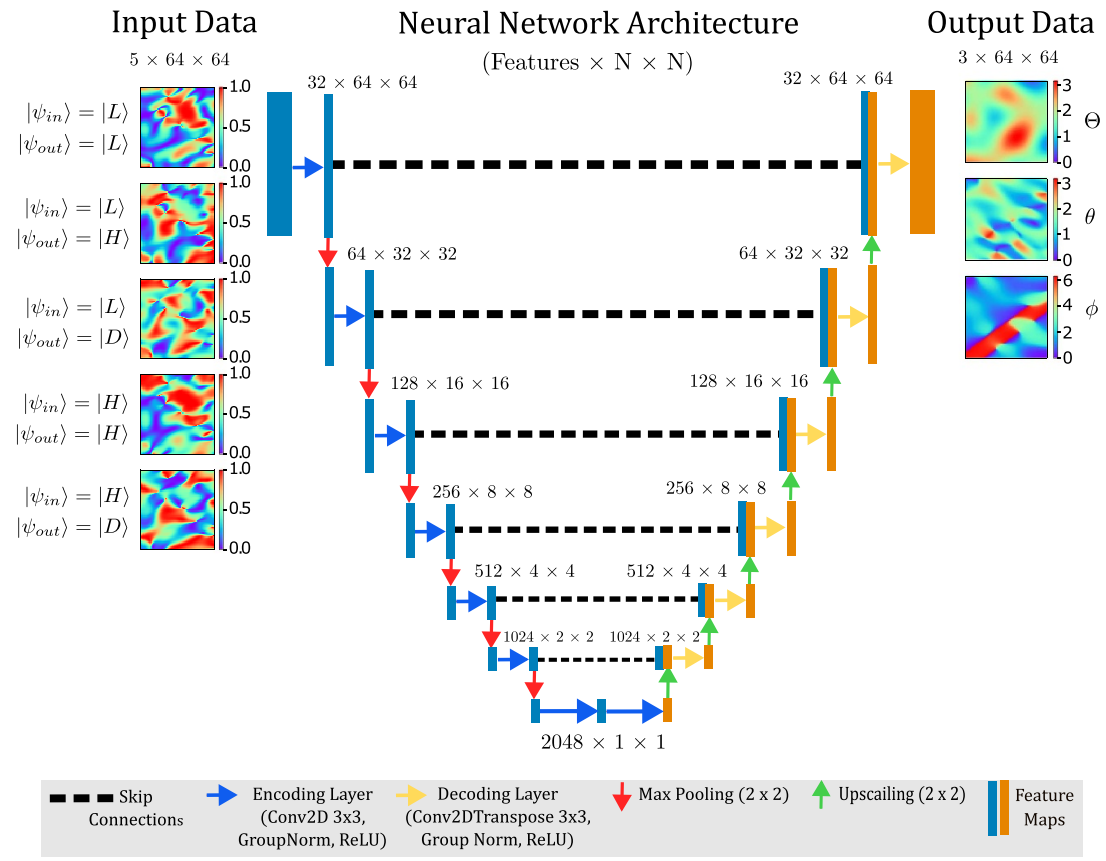

Figure 1. Neural network architecture. A fully convolutional neural network based on the auto-encoder U-Net is employed. Our input consists of five polarimetric measurements with the target resolution. The data is processed through a series of compression steps, each implemented as a 2 × 2 Max Pooling layer. After the bottommost layer, the data is decompressed with 2 × 2 upsampling layers. The output of each encoding layer is passed as input to the decoding layer within the same level of the architecture via skip connections, which help circumvent vanishing gradients during training. The feature maps are colored blue and orange to indicate, respectively, the encoding and decoding blocks of the U-Net.

3.2. Neural network architecture

To perform space-resolved process tomography, we employ a fully CNN based on the U-Net architecture. Figure 1 outlines the structure of the architecture, which can be divided into an encoder and decoder network. The polarimetric measurements are fed into the input channels and processed through a sequence of encoding layers, wherein they are compressed into feature representations in a latent space, capturing global information about the images. Crucially, such a routine is capable of recognizing the local continuity of the images. The feature representations are then fed into the decoder network, where they are eventually converted into the three $N \times N$ unitary parameters. To overcome the issue of vanishing gradients during training, skip connections send residual data from the encoder to the decoder network. We also apply Dropout regularization to prevent overfitting. The structure of the network, together with the size of the training set, has been optimized to achieve satisfactory reconstructions on unseen data (see section 4.1). In principle, our routine can be scaled to arbitrary spatial resolutions, at the obvious cost of increasing computational complexity.

3.3. Training strategy

We prepare a fixed dataset of input-output training examples to perform supervised learning. The agreement between the predicted unitary process, $\hat{\mathcal{U}}_{\text{exp}}$, and the theoretical operator, $\hat{\mathcal{U}}_{\text{th}}$, is quantified in terms of the *map* fidelity:

$$\mathcal{F} = \frac{1}{2N^2}\left|\mathrm{Tr}\left(\hat{\mathcal{U}}_{\text{th}}^{\dagger}\hat{\mathcal{U}}_{\text{exp}}\right)\right| = \frac{1}{2N^2}\left|\sum_{x=1}^{N}\sum_{y=1}^{N}\mathrm{Tr}\left(U_{\text{th}}^{\dagger}(x,y)\,U_{\text{exp}}(x,y)\right)\right|. \tag{10}$$

The loss function used for training is the infidelity $1-\mathcal{F}$. This metric constitutes the natural choice for our tomographic problem and crucially differs from the average *pixel* fidelity:

$$\bar{F}=\frac{1}{2N^2}\sum_{x=1}^{N}\sum_{y=1}^{N}\left|\mathrm{Tr}\left(U_{\mathrm{th}}^{\dagger}(x,y)\,U_{\mathrm{exp}}(x,y)\right)\right|. \tag{11}$$

The latter is insensitive to relative phases between pixels. Conversely, high values of $\mathcal{F}$ certify that the routine delivers accurate predictions, also embedding the continuity constraint, which was enforced by hand in previous implementations [17]. Further details on the training hyperparameters are provided in appendix A.

4. Results

4.1. Synthetic experiments

The performance of our network is first validated on a test set of 10^3 ideal numerical experiments. These synthetic processes are generated as outlined in section 3.1. The map fidelity (see equation (10)) is used to evaluate the network predictions. In view of the assessment of the network on actual experimental data, where there is no preliminary knowledge of the 'true' unitary process, we also compute what we call *polarimetric infidelity*. It quantifies the discrepancy between the input polarimetric measurements and the ones computed from the network predictions:

$$\Delta=\frac{1}{5N^2}\sum_{p=1}^{5}\sum_{x=1}^{N}\sum_{y=1}^{N}\left|I_{\mathrm{th}}^{(p)}(x,y)-I_{\mathrm{exp}}^{(p)}(x,y)\right|^2, \tag{12}$$

where p labels individual measurements. Figure 2(a) report the distributions of the infidelity $1-\mathcal{F}$ and Δ. The average map infidelity, $1-\bar{\mathcal{F}}<0.1$, proves that our network has been successfully trained to provide accurate reconstructions of a variety of complex processes. Correspondingly, a low value is also reported for the average polarimetric infidelity: $\bar{\Delta}<0.05$.

To assess the network performance in the presence of noise, we also compute the average infidelity when the polarimetric measurements are affected by varying levels of Gaussian noise, with zero mean and $\sigma=0.01,0.02,0.05,$ and 0.1. The results shown in figure 2(b) certify that the network performance remains stable even as noise strength increases. This suggests that our network effectively distinguishes relevant process features from the noise.

A subset of 100 processes is then extracted to compare the performance of our network with the genetic algorithm (GA) described in [17], by adapting the original version to the minimal case of five measurements. Figures 2(c) and (d) report the infidelity distribution of the two routines when processing ideal measurements and noisy data, respectively. In the second case, a Gaussian noise with $\sigma=0.02$ is considered. The average performance of our routine is closely aligned with the GA. With ideal input measurements, the average infidelity is $1-\bar{\mathcal{F}}_{\mathrm{CNN}}=0.140$ and $1-\bar{\mathcal{F}}_{\mathrm{GA}}=0.139$ for the network and the GA, respectively. With noise, we obtain $1-\bar{\mathcal{F}}_{\mathrm{CNN}}=0.140$ and $1-\bar{\mathcal{F}}_{\mathrm{GA}}=0.168$. This further certifies the robustness of the network to noise. Interestingly, a poor correlation between map and pixel fidelity is observed for the GA, as shown in the insets of figures 2(c) and (d). This is ascribed to the intrinsic pixel-by-pixel approach, which cannot capture the continuity of the physical parameters. The network is, in fact, immune to this effect.

On an AMD Ryzen 4500U @ 2.38 GHz CPU, each reconstruction from the network takes 150 ms on average, compared to ≈ 60 s required by the GA.

4.2. Complex polarization transformations

Process tomography is experimentally engineered within the setup sketched in figure 3(a). The source is the output of a Ti:Sa laser (central wavelength $\lambda=810$ nm), spatially cleaned through a single-mode fiber (not shown in the figure). Complex polarization transformations are realized via LCMSs, acting as waveplates with patterned optic-axis modulation. The birefringence of these devices is electrically controlled [41]. In particular, our experiments involve different combinations of linear polarization gratings, known as g-plates [42], and plates exhibiting a uniform optic-axis orientation. In the case of g-plates acting along x, $T_{x,\Lambda}(\delta)$, the optic-axis distribution is given by $\alpha(x,y)=\pi x/\Lambda$, with $\Lambda=2.5$ mm (and similarly along y). Uniform plates have $\alpha(x,y)=0$, and act as $W=(\sigma_0+i\sigma_x)/\sqrt{2}$. The beam waist is adjusted to $w_0\simeq 2.5$ mm to cover at least one spatial period of the optical operator. A linear polarizer, a half-wave plate, and a quarter-wave plate (LP_1-HWP_1-QWP_1) are needed to prepare an arbitrary polarization state. The polarimetric measurements are completed with a mirror sequence (QWP_2-HWP_2-LP_2) implementing the projection. The resulting intensity distributions are collected on a camera. The latter is placed immediately after the projection stage, so that the decomposition of equation (4) is verified with good approximation.

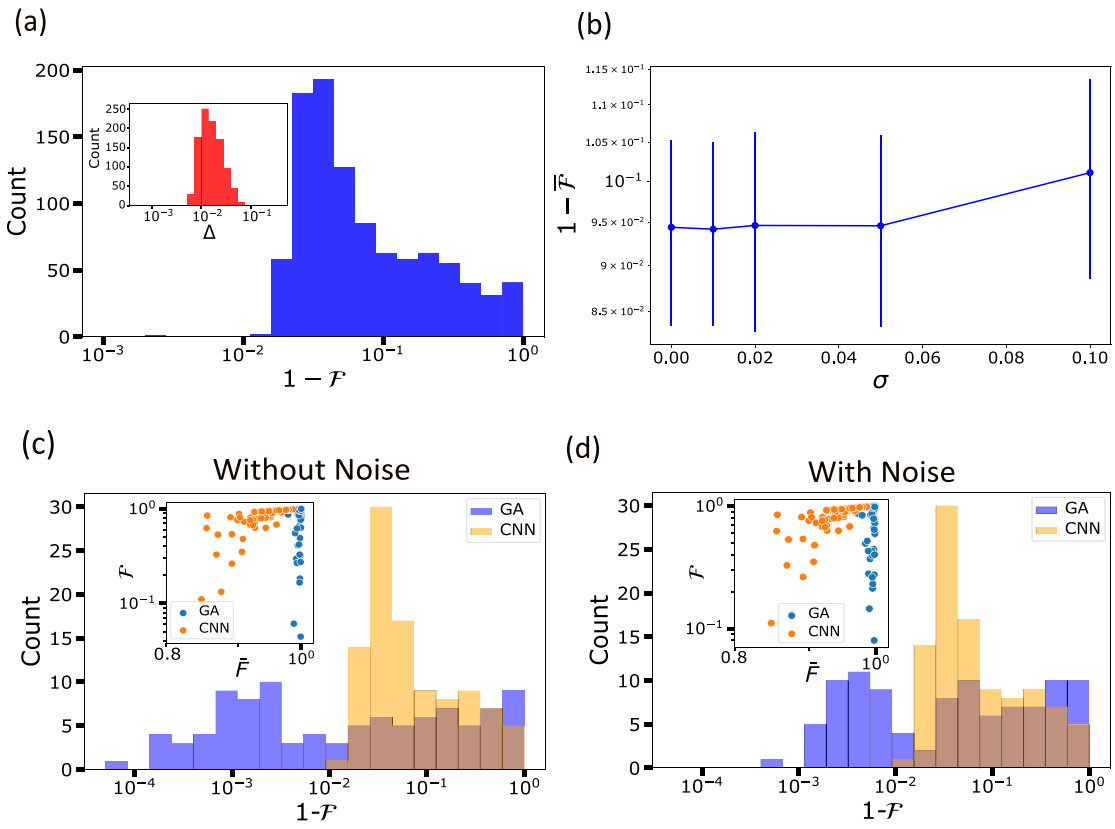

Figure 2. Synthetic experiments. (a) Distribution of the map infidelity $1-\mathcal{F}$ and the polarimetric infidelity Δ (inset) resulting from the network reconstruction of 10^3 complex synthetic processes. (b) Average map infidelity for varying levels of Gaussian noise (with standard deviation σ). The network performance does not vary significantly with increasing noise strength. Error bars are computed as the standard deviation over 10^3 processes. The performance of the network is compared with the genetic algorithm of [17] on a subset of 100 processes, in the case of (c) ideal measurements and (d) noisy input data ($\sigma = 0.02$). The insets show the correlation between pixel fidelity $\bar{F}$ and map fidelity. Limited by a pixel-by-pixel approach, the GA cannot capture global features, which leads to no correlations.

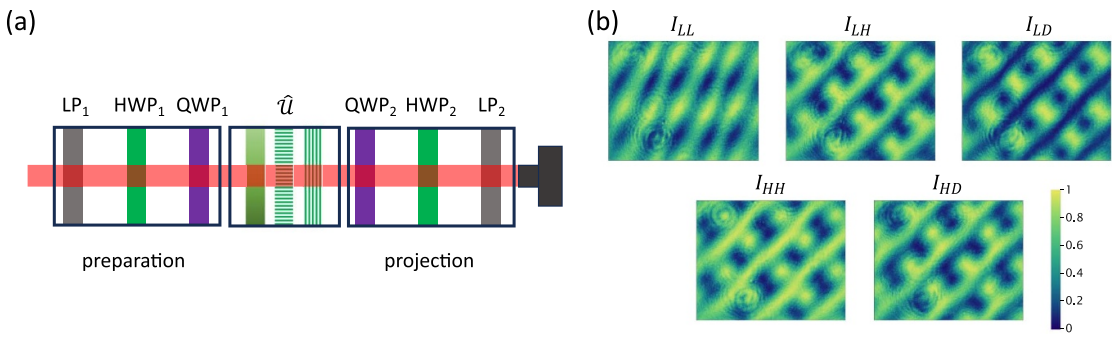

Figure 3. Experimental process tomography. (a) A space-dependent optical operator is implemented via LCMSs. Polarimetric measurements are realized by adjusting the preparation and projection stage. The resulting intensity distribution is collected on a camera. (b) The set of five polarimetric measurements collected for the process $U = T_{y,\Lambda}(\pi/2)T_{x,\Lambda}(\pi/2)W$. The normalization factor is the total light intensity.

This corresponds to neglecting any effect resulting from light propagation, which can be addressed via, e.g. Fourier QPT [43].

In a first experiment, we engineer the three-plate process $U = T_{y,\Lambda}(\pi/2)T_{x,\Lambda}(\pi/2)W$. Figure 3(b) shows the intensity distributions resulting from the complete set of five polarimetric measurements. In figure 4(a), we report the comparison between the theoretical unitary parameters and the network predictions on both synthetic and experimental data. The network successfully captures the distinctive features of all the modulations. The polarimetric infidelities are $\Delta_s = 0.0123$ (corresponding to a fidelity $\mathcal{F} = 96.6\%$) and $\Delta_e = 0.0339$ for the synthetic and experimental reconstructions, respectively. The second experiment is realized by cascading six LCMSs as follows: $U = T_{y,\Lambda/2}(1)T_{x,\Lambda/2}(1)WT_{y,\Lambda}(\pi/2)T_{x,\Lambda}(\pi/2)W$. The network

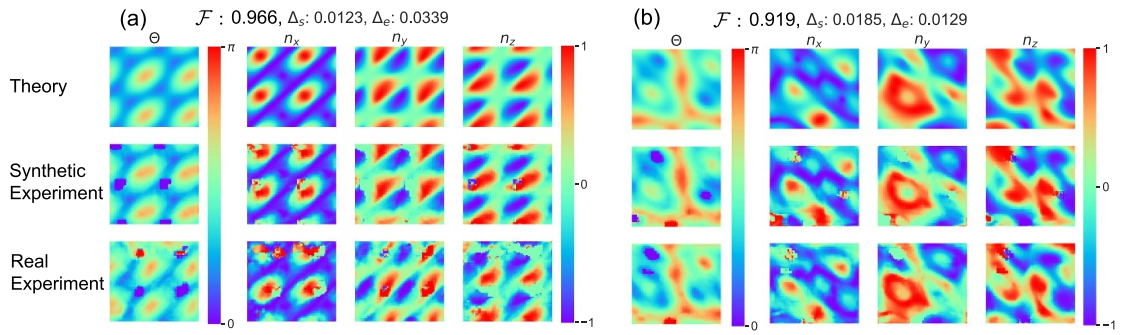

Figure 4. Complex polarization transformations. By combining different LCMSs, we realize the processes (a) $T_{y,\Lambda}(\pi/2)T_{x,\Lambda}(\pi/2)W$ and (b) $T_{y,\Lambda/2}(1)T_{x,\Lambda/2}(1)WT_{y,\Lambda}(\pi/2)T_{x,\Lambda}(\pi/2)W$ (see main text for the description of individual optical operators). The parameters of the theoretical process are plotted with the tomographic reconstruction resulting from both synthetic and real experimental data.

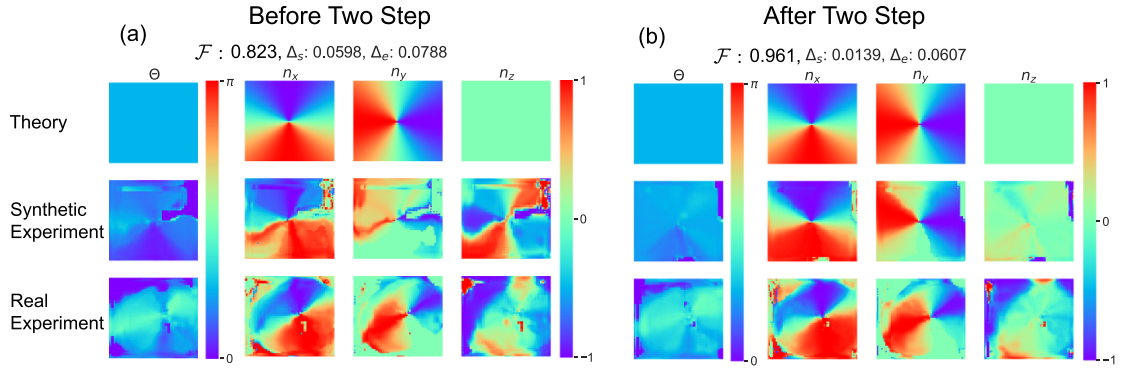

Figure 5. Single-plate process. We perform QPT on a q-plate. As expected, the first reconstruction is poor (a), but the network performance significantly improves after re-training the original architecture (b).

still provides satisfactory reconstructions, with very low polarimetric infidelities: $\Delta_s = 0.0185$ ($\mathcal{F} = 91.9\%$) and $\Delta_e = 0.0129$ (see figure 4(b)). We finally prepare a third experiment reproducing a process that does not fall into the training category examples. In particular, we perform QPT on a single q-plate [44], with $\delta = \pi$. Such a device is characterized by an azimuthal pattern featuring a singularity at the center: $\alpha(x,y) = \alpha(\varphi) = q\varphi$, where $\varphi = \text{atan2}(y,x)$ and q is the topological charge. In our case, we set $q = 1/2$. The network reconstruction is generally poor, as shown in figure 5(a). This is expected for two main reasons: (i) the unitary parameters do not feature a simple decomposition as in equation (6), and (ii) single-plate processes feature the maximal global phase ambiguity [17], as $n_z(x,y) = 0$. However, these problematic cases can be better handled with a second stage of training, whereby the network is further trained on an augmented dataset containing a small portion of such examples. After the re-training stage, the reconstruction of the q-plate process significantly improves (see figure 6(b)), with $\sim 80\%$ and $\sim 25\%$ relative improvement on the synthetic and experimental predictions, respectively. Further details on the re-training procedure are provided in appendix B, where we also investigate the general improvement of the network performance on single-plate processes. Reduced performance in the real experimental case is mainly ascribed to fabrication issues, especially the finite-size resolution which affects the pattern quality across the singularity.

In appendix C, we show the reconstruction of these processes performed by the same GA routine used before [17], providing a further comparison of the two approaches.

5. Conclusions

The CNN we have trained to perform space-resolved process tomography is capable of outputting fast and accurate reconstructions, by parallel-processing polarimetric data and extracting global features of complex polarization transformations. We have also proven that further re-training on specialized datasets improves the performance on out-of-training examples.

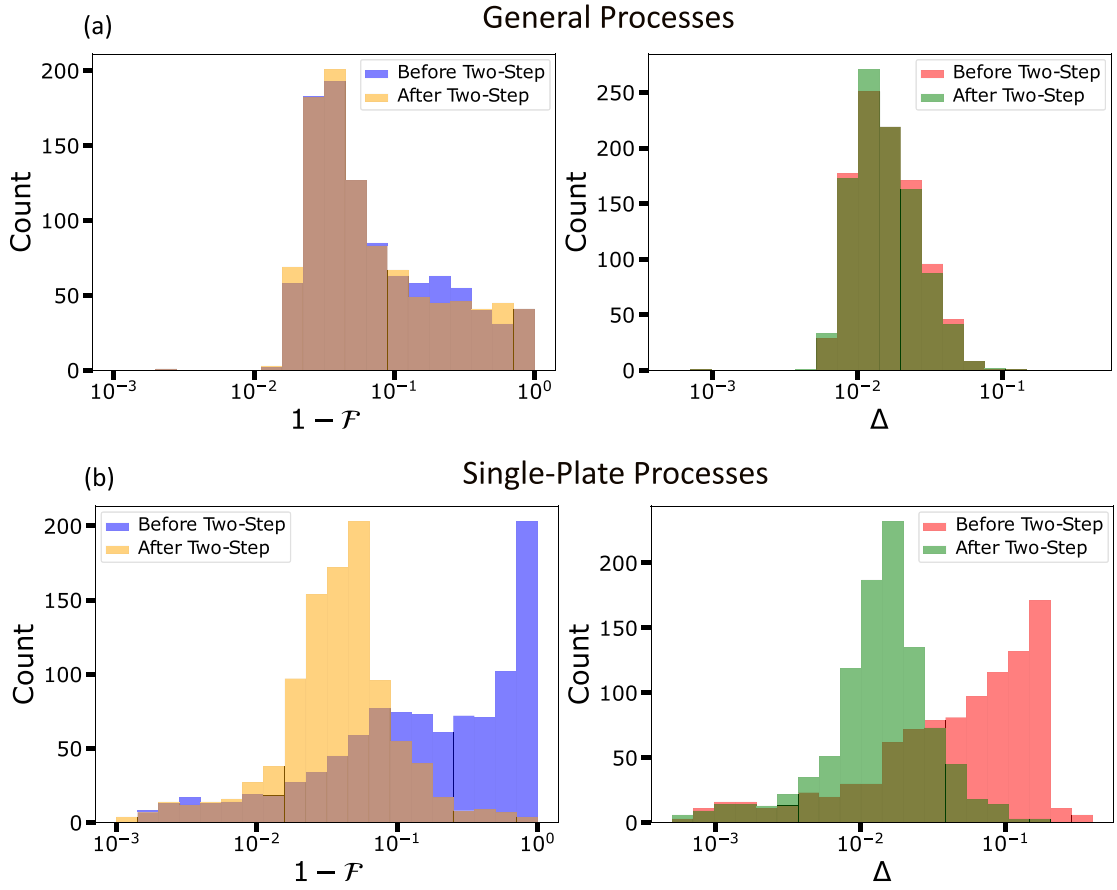

Figure 6. Two-step training. Distribution of the map and the polarimetric infidelity, before and after the two-step training, resulting from the network reconstruction of 10^3 synthetic processes, generated as (a) typical training samples, and (b) single-plate processes. After the two-step training, the network performance on single-plate processes significantly improves, without attacking the performance on general processes.

Our method has been validated on a vast set of synthetic and experimental data. The comparison with an alternative genetic approach suggests that these methods can be used as complementary approaches to address a wider range of problems.

Despite our implementation, these findings are not limited to optical polarization, and could be straightforwardly adapted to other fields by specializing the training set to the particular class of experiments. This study sets the baseline for the development of further optimization routines for tomographic problems both in the classical and quantum domains. For instance, adaptive approaches represent a promising direction, where future measurements are adjusted depending on the outcomes of previous ones [45, 46]. Future generalizations of this work could be applied to non-unitary evolutions [47] and multi-photon gates [48] in high-dimensional Hilbert spaces, where the increasing complexity poses significant computational challenges.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://github.com/TareqJ1000/QPT_NN/ [49].

Acknowledgments

This work was supported by the Canada Research Chair (CRC) Program, NRC-uOttawa Joint Centre for Extreme Quantum Photonics (JCEP) via the Quantum Sensors Challenge Program at the National Research Council of Canada, and Quantum Enhanced Sensing and Imaging (QuEnSI) Alliance Consortia Quantum grant. FDC and FC further acknowledge support from the PNRR MUR project PE0000023-NQSTI.

Conflict of interest

The authors declare no conflicts of interest.

Appendix A. Training hyperparameters

The complete set of hyperparameters is listed in table 1. We generate 50 000 training samples with the parameterization described in section 3.1. An 85:15 split between the training and the validation dataset is used. The validation stage prevents the issue of overfitting, certifying that the network is not merely adapting its parameters to the specific training set. The input consists of five 64×64 polarimetric images, while the unitary parameters form the output layer. Artificial noise is introduced on each pixel from a Gaussian distribution with zero mean and $\sigma = 0.02$.

Table 1. Hyperparameters.

Image Resolution	64×64
Size of Dataset	50000
Batch Size	8
Train-Validation Split	85:15
Initial Learning Rate	10^{-3}
Dropout Rate	0.2
Number of Epochs	200
Patience	15 Epochs
Reduction Factor	0.1
Training Error at Convergence	0.07
Validation Error at Convergence	0.11

The training process is realized within the Tensorflow library [50]. Backpropagation of the network after each epoch was performed using the Adam optimizer [51]. We adopt an adaptive training strategy, with the learning rate that is initialized at 10^{-3} and is reduced by a factor of 10^{-1} if, after 15 epochs, no appreciable change in the validation loss is observed. We define an early stopping criterion, where training is stopped if the validation loss does not change appreciably within 20 epochs. Accordingly, the training was stopped after 200 epochs, in correspondence with the convergence of the validation loss. Training is carried out on the *Narval* supercluster using two NVidia A100SXM4 GPUs, each with 40 GB of memory. The process was completed in roughly 7 h, with each epoch taking ≈ 2 min.

Appendix B. Training specialization

An architecture that has been trained on a specific class of processes may struggle to reconstruct the features of more exotic examples. In our case, this occurs for single-plate optical operators. This limitation can be overcome by re-training the original network on a specialized dataset containing a portion of problematic examples, in reminiscence of fine-tuning Language Learning Models towards desired properties [52]. This is known as Transfer Learning [53]. Specifically, this is accomplished by augmenting the original dataset with 6000 examples consisting of single- and two-plate processes. For each waveplate, the birefringence δ is uniformly extracted from the range $[0, 2\pi]$, while the optic-axis orientation follows a decomposition similar to equation (6):

$$\begin{aligned}\alpha(x,y) = \sum_{i=0}^{\Omega_x}\sum_{j=0}^{\Omega_y} & c_{ij}^{(1)} \cos\frac{2\pi i x}{N}\cos\frac{2\pi j y}{N} + \\ & c_{ij}^{(2)} \cos\frac{2\pi i x}{N}\sin\frac{2\pi j y}{N} + \\ & c_{ij}^{(3)} \sin\frac{2\pi i x}{N}\cos\frac{2\pi j y}{N} + \\ & c_{ij}^{(4)} \sin\frac{2\pi i x}{N}\sin\frac{2\pi j y}{N},\end{aligned} \tag{B1}$$

with $0 \leqslant \Omega_x \leqslant 3$, $0 \leqslant \Omega_y \leqslant 3$, and each coefficient $c_{ij}^{(m)}$ is uniformly extracted from the range $[-1, 1]$.

Importantly, the ratio of these examples with the original dataset is kept reasonably small to avoid biasing the network towards these processes. Indeed, the ability of our network to reconstruct general processes is not diminished after Transfer Learning, as shown in figure 6(a). Conversely, the network demonstrates improved performance in predicting complex single-plate processes, as can be seen in figure 6(b) by the shift of the initial infidelity distributions towards smaller values. Finally, we observe that the two-step training process proves successful due to the robustness conferred by the initial pre-training of the network (see section 4.1), which allows it to effectively absorb the features of the new dataset while mitigating the risk of overfitting or bias towards it.

Appendix C. Genetic tomography

The GA implements a pixel-by-pixel optimization. Crucially, the initial population [54] undergoing the genetic workflow is not picked at random (except for the first pixel), but sampled from neighboring pixels that have already been reconstructed. This enforces a local continuity constraint, which can greatly improve the quality of final reconstructions. Further details can be found in [17], the only difference being that the implementation discussed here only takes 5 input polarimetric measurements.

The tomographic reconstruction operated by the GA on the same experimental processes considered in the main text is shown in figures 7 and 8. In general, we observe a better quality of the experimental reconstructions, also quantified by the low polarimetric infidelity. This is mainly due to enhanced performances on the single-pixel level, at the cost of a much longer (three orders of magnitude) processing time.

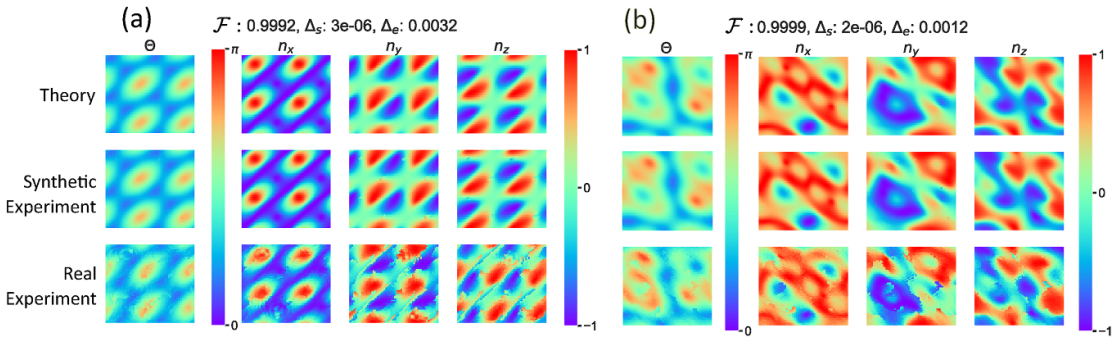

Figure 7. GA tomography of complex polarization transformations. We perform QPT for the same processes of figure 4 using the GA of [17].

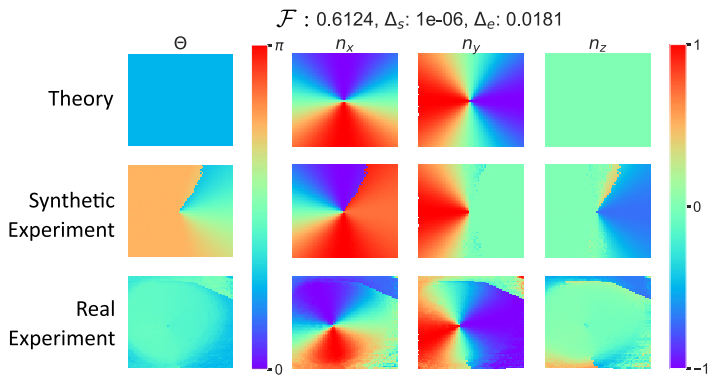

Figure 8. GA tomography of a single-plate process. QPT results for the GA reconstruction of the q-plate process. In the synthetic implementation, the global phase continuity is broken, resulting in a poor reconstruction.

An interesting case is provided by the synthetic q-plate experiment, which proves extremely challenging for the GA. The low polarimetric infidelity and high map infidelity imply that there has been a flip of the global phase across a large domain, clearly visible in figure 8. This result further proves that our architecture, while not necessarily outperforming alternative methods under all possible metrics, is better suited for identifying global features of the process.

ORCID iDs

Tareq Jaouni https://orcid.org/0009-0006-5661-2403
Francesco Di Colandrea https://orcid.org/0000-0002-4863-1448
Lorenzo Amato https://orcid.org/0000-0001-6126-6740
Filippo Cardano https://orcid.org/0000-0002-7828-3819
Ebrahim Karimi https://orcid.org/0000-0002-8168-7304

References

[1] Chuang I L and Nielsen M A 1997 *J. Mod. Opt.* **44** 2455
[2] Söderström T and Stoica P 1989 *System Identification Prentice-Hall International Series in Systems and Control Engineering* (Prentice Hall)
[3] Childs A M, Chuang I L and Leung D W 2001 *Phys. Rev.* A **64** 012314
[4] Myrskog S H, Fox J K, Mitchell M W and Steinberg A M 2005 *Phys. Rev.* A **72** 013615
[5] Roos C F, Lancaster G P T, Riebe M, Häffner H, Hänsel W, Gulde S, Becher C, Eschner J, Schmidt-Kaler F and Blatt R 2004 *Phys. Rev. Lett.* **92** 220402
[6] Riebe M, Kim K, Schindler P, Monz T, Schmidt P O, Körber T K, Hänsel W, Häffner H, Roos C F and Blatt R 2006 *Phys. Rev. Lett.* **97** 220407
[7] Mitchell M W, Ellenor C W, Schneider S and Steinberg A M 2003 *Phys. Rev. Lett.* **91** 120402
[8] Altepeter J B, Branning D, Jeffrey E, Wei T C, Kwiat P G, Thew R T, O'Brien J L, Nielsen M A and White A G 2003 *Phys. Rev. Lett.* **90** 193601
[9] O'Brien J L, Pryde G J, Gilchrist A, James D F V, Langford N K, Ralph T C and White A G 2004 *Phys. Rev. Lett.* **93** 080502
[10] Lobino M, Korystov D, Kupchak C, Figueroa E, Sanders B C and Lvovsky A I 2008 *Science* **322** 563
[11] Bongioanni I, Sansoni L, Sciarrino F, Vallone G and Mataloni P 2010 *Phys. Rev.* A **82** 042307
[12] Rahimi-Keshari S, Broome M A, Fickler R, Fedrizzi A, Ralph T C and White A G 2013 *Opt. Express* **21** 13450
[13] Zhou X-Q, Cable H, Whittaker R, Shadbolt P, O'Brien J L and Matthews J C F 2015 *Optica* **2** 510
[14] Antón C *et al* 2017 *Optica* **4** 1326
[15] Jacob K V, Mirasola A E, Adhikari S and Dowling J P 2018 *Phys. Rev.* A **98** 052327
[16] Bouchard F, Hufnagel F, Koutný D, Abbas A, Sit A, Heshami K, Fickler R and Karimi E 2019 *Quantum* **3** 138
[17] Di Colandrea F, Amato L, Schiattarella R, Dauphin A and Cardano F 2023 *Opt. Express* **31** 31698
[18] Goel S, Leedumrongwatthanakun S, Valencia N H, McCutcheon W, Tavakoli A, Conti C, Pinkse P W and Malik M 2024 *Nat. Phys.* **20** 1
[19] James D F V, Kwiat P G, Munro W J and White A G 2001 *Phys. Rev.* A **64** 052312
[20] Aiello A, Puentes G, Voigt D and Woerdman J P 2006 *Opt. Lett.* **31** 817
[21] Solomon J E 1981 *Appl. Opt.* **20** 1537
[22] Davis J A, Evans G H and Moreno I 2005 *Appl. Opt.* **44** 4049
[23] Zhan Q 2009 *Adv. Opt. Photonics* **1** 1
[24] Rosales-Guzmán C, Ndagano B and Forbes A 2018 *J. Opt.* **20** 123001
[25] Ronneberger O, Fischer P and Brox T 2015 *U-Net: Convolutional Networks for Biomedical Image Segmentation* (Springer) pp 234–41
[26] Wang K *et al* 2024 *Light Sci. Appl.* **13** 4
[27] Luo Z, Yurt A, Stahl R, Lambrechts A, Reumers V, Braeken D and Lagae L 2019 *Opt. Express* **27** 13581
[28] Zhang G, Guan T, Shen Z, Wang X, Hu T, Wang D, He Y and Xie N 2018 *Opt. Express* **26** 19388
[29] Nguyen T, Xue Y, Li Y, Tian L and Nehmetallah G 2018 *Opt. Express* **26** 26470
[30] Sinha A, Lee J, Li S and Barbastathis G 2017 *Optica* **4** 1117
[31] Proppe A H, Thekkadath G, England D, Bustard P J, Bouchard F, Lundeen J S and Sussman B J arXiv:2402.06063
[32] Schmale T, Reh M and Gärttner M 2022 *npj Quantum Inf.* **8** 115
[33] Rubano A, Cardano F, Piccirillo B and Marrucci L 2019 *J. Opt. Soc. Am.* B **36** D70
[34] Di Colandrea F, Babazadeh A, Dauphin A, Massignan P, Marrucci L and Cardano F 2023 *Optica* **10** 324
[35] Fiurášek J and Hradil Z 2001 *Phys. Rev.* A **63** 020101
[36] Simon R and Mukunda N 1990 *Phys. Lett.* A **143** 165
[37] Sit A, Giner L, Karimi E and Lundeen J S 2017 *J. Opt.* **19** 094003
[38] Fläschner N, Rem B S, Tarnowski M, Vogel D, Lühmann D-S, Sengstock K and Weitenberg C 2016 *Science* **352** 1091
[39] Tarnowski M, Ünal F N, Fläschner N, Rem B S, Eckardt A, Sengstock K and Weitenberg C 2019 *Nat. Commun.* **10** 1728
[40] Yi C-R, Yu J, Yuan H, Jiao R-H, Yang Y-M, Jiang X, Zhang J-Y, Chen S and Pan J-W 2023 *Phys. Rev. Res.* **5** L032016
[41] Piccirillo B, D'Ambrosio V, Slussarenko S, Marrucci L and Santamato E 2010 *Appl. Phys. Lett.* **97** 241104
[42] D'Errico A, Cardano F, Maffei M, Dauphin A, Barboza R, Esposito C, Piccirillo B, Lewenstein M, Massignan P and Marrucci L 2020 *Optica* **7** 108
[43] Di Colandrea F, Dehghan N, D'Errico A and Karimi E 2024 *npj Quantum Inf.* **10** 49

[44] Marrucci L, Manzo C and Paparo D 2006 *Phys. Rev. Lett.* **96** 163905
[45] Quek Y, Fort S and Ng H K 2021 *npj Quantum Inf.* **7** 105
[46] Lange H, Kebrič M, Buser M, Schollwöck U, Grusdt F and Bohrdt A 2023 *Quantum* **7** 1129
[47] Wang C, Fu Z, Mao W, Qie J, Stone A D and Yang L 2023 *Adv. Opt. Photonics* **15** 442
[48] Zhong H-S *et al* 2020 *Science* **370** 1460
[49] Jaouni T, Di Colandrea F, Amato L, Cardano F and Karimi E (available at: https://github.com/TareqJ1000/QPT_NN)
[50] Abadi M *et al* 2015 TensorFlow: large-scale machine learning on heterogeneous systems (available at: https://www.tensorflow.org/?hl=it)
[51] Kingma D P and Ba J (arXiv: 1412.6980)
[52] Taylor R, Kardas M, Cucurull G, Scialom T, Hartshorn A, Saravia E, Poulton A, Kerkez V and Stojnic R 2022 (arXiv:2211.09085)
[53] Weiss K, Khoshgoftaar T M and Wang D 2016 *J. Big Data* **3** 1
[54] By population, we refer to multiple instances of parameters $\Theta(x,y), \boldsymbol{n}(x,y)$ describing the single-pixel unitary $\hat{U}(x,y)$. The population is iteratively updated by the GA.