PHOTONICS Research

Entanglement quantification via weak measurements assisted by deep learning

Mu Yang,[1,†] Ya Xiao,[2,†] Ze-Yan Hao,[1,3] Yu-Wei Liao,[1,3] Jia-He Cao,[1,3] Kai Sun,[1,3,4] En-Hui Wang,[1,5] Zheng-Hao Liu,[1,3] Yutaka Shikano,[6,7,8,9,10,11] Jin-Shi Xu,[1,3,4,12] Chuan-Feng Li,[1,3,4,13] and Guang-Can Guo[1,3,4]

[1]*CAS Key Laboratory of Quantum Information, University of Science and Technology of China, Hefei 230026, China*
[2]*College of Physics and Optoelectronic Engineering, Ocean University of China, Qingdao 266100, China*
[3]*CAS Center for Excellence in Quantum Information and Quantum Physics, University of Science and Technology of China, Hefei 230026, China*
[4]*Hefei National Laboratory, University of Science and Technology of China, Hefei 230088, China*
[5]*Electric Power Research Institute, State Grid Anhui Electric Power Co., Ltd., Hefei 230601, China*
[6]*Institute of Systems and Information Engineering, University of Tsukuba, Ibaraki 305-8573, Japan*
[7]*Center for Artificial Intelligence Research, University of Tsukuba, Ibaraki 305-8577, Japan*
[8]*Graduate School of Science and Technology, Gunma University, Gunma 371-8510, Japan*
[9]*Institute for Quantum Studies, Chapman University, Orange, California 92866, USA*
[10]*JST PRESTO, Saitama 332-0012, Japan*
[11]*e-mail: yshikano@cs.tsukuba.ac.jp*
[12]*e-mail: jsxu@ustc.edu.cn*
[13]*e-mail: cfli@ustc.edu.cn*
[†]*These authors contributed equally to this work.*

Entanglement has been recognized as being crucial when implementing various quantum information tasks. Nevertheless, quantifying entanglement for an unknown quantum state requires nonphysical operations or post-processing measurement data. For example, evaluation methods via quantum state tomography require vast amounts of measurement data and likely estimation. Although a direct entanglement determination has been reported for the unknown pure state, it is still tricky for the mixed state. In this work, assisted by weak measurement and deep learning technology, we directly detect the entanglement (namely, the concurrence) of a class of two-photon polarization-entangled mixed states both theoretically and experimentally according to the local photon spatial distributions after weak measurement. In this way, the number of projective bases is much smaller than that required in quantum state tomography. © 2024 Chinese Laser Press

https://doi.org/10.1364/PRJ.498498

1. INTRODUCTION

Entanglement is an essential resource for nonclassical information tasks in quantum communication [1,2] and quantum computation [3,4], and has also been applied to nonlinear molecular spectroscopy [5]. However, quantifying entanglement requires nonphysical operations or post-processing of the measurement data [6].

One of the most common quantifications of entanglement is concurrence [7,8], and its determination generally requires quantum state tomography [9,10], which relies on projective measurements to form a set of projective bases on the Hilbert space of the system, providing all the information about the state. The number of projective bases increases exponentially with the system scaleup. When the scale of the system is very large, it is impossible to perform projective measurements. To circumvent this impossibility, several approaches have been proposed to quantify the concurrence, including collective

measurements [11] and the interference between two copies of the quantum states [12]. However, such approaches contrapose two-qubit pure states and cannot be directly applied to the case of the complicated mixed state.

Recently, weak measurement technology has been used to directly detect quantum states, including the transverse wavefunction of the photon states [13], the density of a mixed two-qubit state [14], and even entangled states [15]. The weak measurement can extract a small amount of information from a single outcome. Although each outcome of the weak measurement is uncertain, the average value can build a certain value, named weak value, to reveal the state information.

In this study, we experimentally use the wake values to demonstrate the feasibility of entanglement quantification associated with a deep learning two-photon polarization-entangled system. Specifically, we compress the weak value information of two-qubit photonic states into local conditional states, which

are illustrated via the spatial distributions of photons. An end-to-end relationship between the concurrence and the distribution of photons is determined by a well-trained convolutional neural network (CNN) [16]. We find that only two projective bases are needed to predict the concurrence, greatly reducing the operation time. Besides, the generalization of this method can quantify the entanglement of high-dimensional multiparticle pure states, and the advantage of this method is exponentially increased.

2. THEORETICAL FRAMEWORK AND CNN PERFORMANCE

Here, we consider a two-qubit (A and B) system, which is in a class of quantum states ρ_{AB} as [17]

$$\rho_{AB}(p,\theta) = p|\psi_\theta\rangle\langle\psi_\theta| + (1-p)I_A/2 \otimes \rho_B^\theta, \quad (1)$$

where $|\psi_\theta\rangle = \cos\theta|00\rangle + \sin\theta|11\rangle$ and $\rho_B^\theta = \mathrm{Tr}_A[|\psi_\theta\rangle\langle\psi_\theta|]$. $\theta \in [0, 45°]$ and $p \in [0,1]$ are two state parameters. These states are of the form X-type, and the parameter p controls the components of pure states and mixed states. Figure 1 (insert) shows the main processes involved in our approach, in which we perform a local projective measurement Γ_A on A, where $\Gamma_A = |\frac{|0\rangle+|1\rangle}{\sqrt{2}}\rangle\langle\frac{|0\rangle+|1\rangle}{\sqrt{2}}|$. The information about entanglement is compressed into the local conditional state of B, written as

$$\rho_B^{\Gamma_A} = \frac{\mathrm{Tr}_A[\rho_{AB}\Gamma_A \otimes I_B]}{\mathrm{Tr}[\rho_{AB}\Gamma_A \otimes I_B]}. \quad (2)$$

To extract the information about entanglement, one can execute a weak measurement of the Pauli observable $\hat{\sigma}_x = |0\rangle\langle1| + |1\rangle\langle0|$ on B [18,19]. Via post-selecting on the state $|k\rangle$ ($k \in \{0,1\}$), the obtained weak value [20,21] can be written as

$$_{\langle k|}\langle\hat{\sigma}_x\rangle_{\Gamma_A}^w = \frac{\mathrm{Tr}[\hat{\sigma}_x\rho_B^{\Gamma_A}|k\rangle\langle k|]}{\mathrm{Tr}[\rho_B^{\Gamma_A}|k\rangle\langle k|]}, \quad (3)$$

which is closely related to the density matrix elements of $\rho_B^{\Gamma_A}$. This value reflects the distance between $\rho_B^{\Gamma_A}$ and the maximally mixed state $\frac{1}{2}I_B$, which is visually illustrated in the Bloch sphere in Fig. 1 (see Appendixes A and B for more details). The concurrence of the class states can be expressed as

$$C(\rho_{AB}) = \left(3\sqrt{|_{\langle0|}\langle\hat{\sigma}_x\rangle_{\Gamma_A}^w||_{\langle1|}\langle\hat{\sigma}_x\rangle_{\Gamma_A}^w|} - 1\right)$$
$$\times \frac{\sqrt{|_{\langle0|}\langle\hat{\sigma}_x\rangle_{\Gamma_A}^w||_{\langle1|}\langle\hat{\sigma}_x\rangle_{\Gamma_A}^w|}}{|_{\langle0|}\langle\hat{\sigma}_x\rangle_{\Gamma_A}^w| + |_{\langle1|}\langle\hat{\sigma}_x\rangle_{\Gamma_A}^w|}. \quad (4)$$

In particular, when ρ_{AB} is the pure state, $C(\rho_{AB})$ is reduced to $\sqrt{1 - |_{\langle0|}\langle\hat{\sigma}_x\rangle_{\Gamma_A}^w|^2}$, and Γ_A is reduced to identity measurement [7].
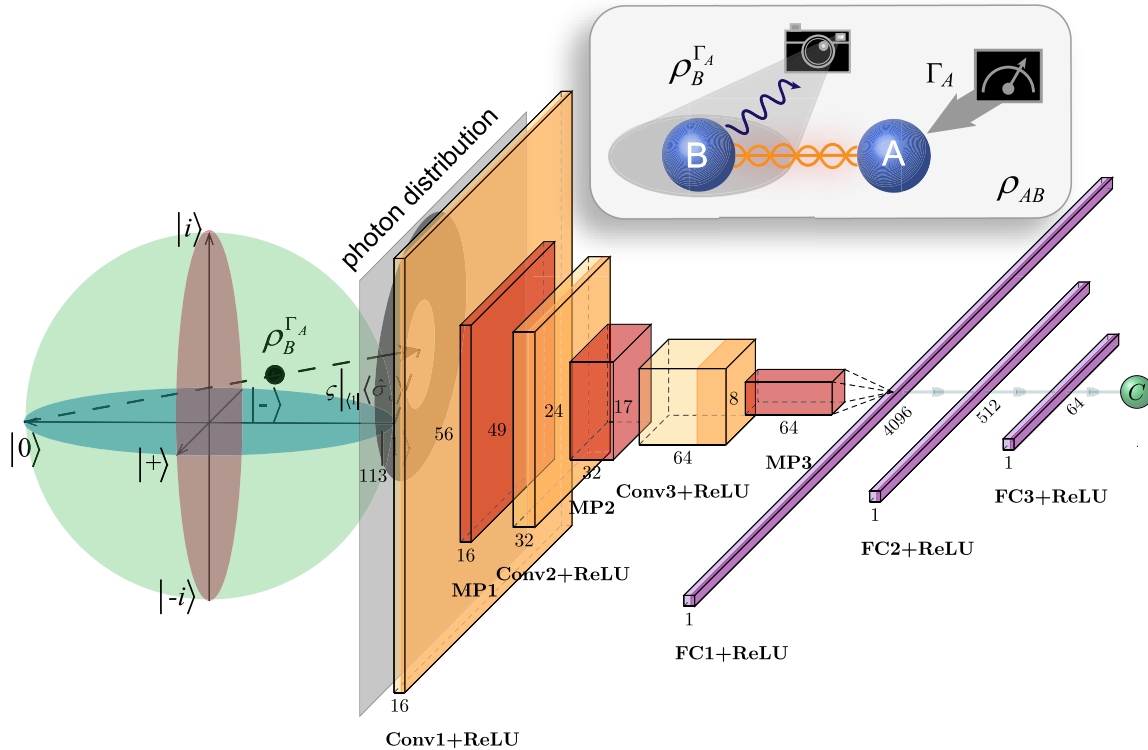

Fig. 1. Theoretical framework and performance of the convolutional neural network (CNN). The weak values $_{\langle1|}\langle\hat{\sigma}_x\rangle_{\Gamma_A}^w$ and $_{\langle0|}\langle\hat{\sigma}_x\rangle_{\Gamma_A}^w$ of B on the state of $\rho_B^{\Gamma_A}$ (as shown in the Bloch sphere) are encoded in the central position of a Laguerre–Gaussian mode. The projected particle distribution is sent to the CNN to extract the concurrence. The boxes indicate the dimension (width of the box) and number (height of the box) of the feature maps. Convolution layers (Conv) with one stride are shown in a buff color. Activation functions (ReLU) of the convolution layers and full connection layers (FC) are denoted by the orange and purple colors, respectively. The red boxes indicate max-pooling layers (MPs) with two strides, and a small green ball represents the output concurrence. The arrows represent the flow of data. Insert: ρ_{AB} is a two-qubit entangled state. The spatial distribution of B's conditional state $\rho_B^{\Gamma_A}$ is obtained, while a local projective measurement Γ_A is performed on A.

To conveniently access the weak values, here we introduce orbital angular momentum (OAM) as a pointer, where the real and imaginary parts of $_{\langle k|}\langle\hat{\sigma}_x\rangle^w_{\Gamma_A}$ are encoded as the coordinate x- and y-positions, respectively, of the central singular point [18,19] of the Laguerre–Gaussian (LG) mode on the local spatial distribution (see Appendix C for more details). This singular center can be extracted by straightforward optimal estimation or the least-square fit method. However, this weak shifting of the singular center is too susceptible to be effectively identified because of the experimental noise. To solve this problem, we attempt to directly extract the concurrence from the local spatial distribution by using a deep learning (DL) method [22], which has been used to solve many-body [23], large-scale quantum tomography [24], quantum state [25], and nonclassical correlation [26] problems. Specifically, we use a CNN to establish end-to-end mapping from the photon distribution to the concurrence, which avoids taking account of the form of the fitting functions or the relations between the weak values and concurrence. To train the CNN, we prepare a large number of two-qubit states ρ_{AB}. The corresponding local spatial distributions are used as features, and the concurrences determined from the tomographic data are used as the label. After being trained with a substantial number of distribution-concurrence pairs, the CNN establishes a statistical functional mapping between the local spatial distributions and the concurrence. The concurrence of the training set is labeled using the traditional method [7,8] by reconstructing the density matrix ρ_{AB}. The CNN also considers the imperfection of the local spatial distribution, so this method has high accuracy and strong robustness.

As shown in Fig. 1, the CNN consists of two paths: the convolution path and the full connection (FC) path. There are three blocks in the convolution path. The input images are mapped into the feature space by convolution layers and downsampled by max-pooling layers to magnify the weak variance in each block. To fit the concurrence, the output results of the convolution path are flattened into a one-dimensional vector and sent to the FC path, which contains four layers with different neurons. The activation function of each layer is selected to be ReLU [$y = \max(0, x)$] so that the network can fit the nonlinear function. We use the Adam optimizer [27], which can effectively prevent local optimization. The mean squared error (MSE), which is used as the loss function, is defined as $\frac{1}{n}\sum_1^n (C^i_{\text{act}} - C^i_{\text{pre}})^2$, where n is the total number of samples, C^i_{act} represents the actual concurrence labeled by the traditional method, and C^i_{pre} means the predicted concurrence. Some neurons were deleted to prevent overfitting (see Appendix D for more details).

3. EXPERIMENTAL DEMONSTRATION

The experimental setup is shown in Fig. 2. A 20 mm long periodic KTiOPO$_4$ (PPKTP) crystal located in the Sagnac interferometer [28] is pumped by a 404 nm continuous-wave diode laser to create a pair of polarization-entangled photons via a type-II spontaneous parametric down-conversion process. These two photons are then sent to Alice and Bob. To generate the entangled state $\cos\theta|HH\rangle + \sin\theta|VV\rangle$ (where H and V represent horizontal and vertical polarization, respectively), a half-wave plate HWP$_1$ is used to control the parameter θ by rotating the polarization of the pump laser.

On Alice's side, the photon beam passes through an unbalanced interferometer (UI), first separated into two paths by a beam splitter (BS). Two sufficiently long calcite crystals (CCs) with HWP$_4$ set to 22.5° between them are placed on one of the paths to destroy the coherence between the different
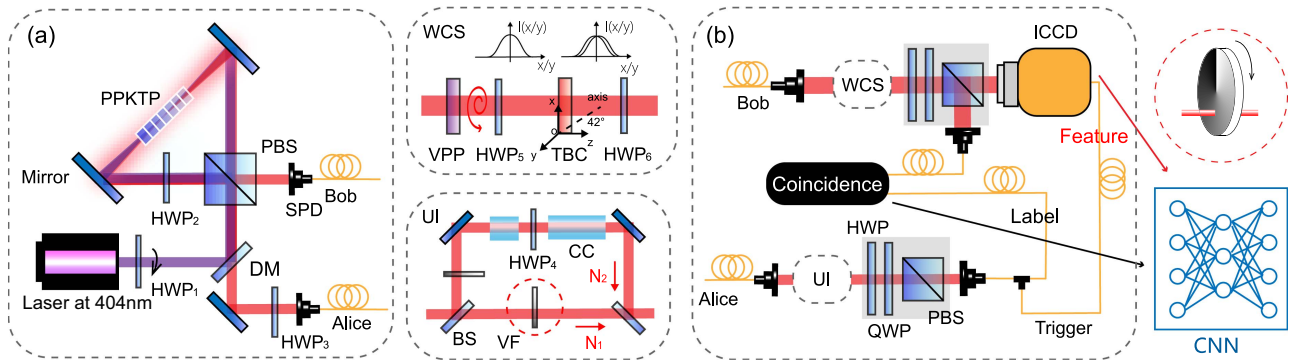

Fig. 2. Experimental setup. (a) A pair of polarization-entangled photons are generated by pumping a type-II PPKTP crystal in a Sagnac interferometer with a 404 nm ultraviolet laser in the preparation stage. A half-wave plate HWP$_1$ is set in front of the pump laser to rotate the polarizations. The polarizations of the pump light and down-converted photons are exchanged by the dual-wavelength HWP$_2$, which is set to 45° in the Sagnac interferometer. HWP$_3$ is used to change the form of the entangled state and is set to 45°. Bob's and Alice's photons pass through a weak measurement system (WM, shown in the gray dotted-line square) and an unbalanced interferometer (UI), respectively. The UI separates the photon into two paths by a beam splitter (BS). There are two sufficiently long calcite crystals (CCs) with the second length being twice larger than that of the first. In-between these crystals, HWP$_4$ is set to 22.5° in one of the paths. This setup destroys the coherence in the different polarization components. Two variable filters (VFs, in the red dotted line circles) control the relative photon counts between the two arms. In the WM, the photon passes through a vortex phase plate (VPP, $l = 2$) and is shaped into the Laguerre–Gaussian mode. HWP$_5$, HWP$_6$, and a thin birefringent crystal (TBC) with its axis set to 42° in the x-o-z plane are used to weakly couple the polarizations and momentum of the photon. (b) Quarter-wave plates (QWPs), HWPs, and polarization beam splitters (PBSs) on both sides of Alice and Bob (shown in the gray squares) are used to perform the projective measurements. On Bob's side, the photons are detected by a single-photon detector (SPD) in the reflected path or by an intensified charge coupled device (ICCD) camera in the transmitted path. The signals detected by the SPD on Alice's side are sent for coincidence or to trigger the ICCD camera. To train the convolutional neural network (CNN), the concurrence determined from the tomographic data is used as the label and the images recorded by the ICCD camera are used as the features, as indicated by the black and red arrows, respectively.

polarization components. In contrast, the state in the other path remains unchanged. Two variable filters (VFs) are used to change the relative photon counts in these two paths (N_1 and N_2), where the parameter p can be written as $p = N_1/(N_1 + N_2)$. The two-qubit states $\rho_{AB}(p, \theta)$ can be prepared when these two paths are combined.

On Bob's side, the photon passes through a weak coupling system (WCS). A vortex phase plate (VPP) is placed to generate an LG wavefunction $\phi_l(x, y)$ with an OAM of $l = 2$. Thereafter, a thin calcite crystal (TBC) with a thickness of 0.7 mm is employed to introduce a weak interaction between the polarization and the momentum of the photon; its optical axis is set in the x–o–z plane, and it is oriented at 42° with respect to the z axis [29]. Combining a pair of HWPs (HWP$_5$ and HWP$_6$) with optical axes set to 22.5°, the interaction Hamiltonian is $\hat{H} = \zeta \hat{\sigma}_x \otimes P_x$, where $P_x = -i\hbar \frac{\partial}{\partial x}$ represents the momentum operator, ζ is the coupling strength, and $\hat{\sigma}_x = |H\rangle\langle V| + |V\rangle\langle H|$.

The polarization states on both sides are measured by a group of quarter-wave plates (QWPs), an HWP, and a polarization beam splitter (PBS). The coincidence from tomographic measurement is used as the label of CNN while the photon distributions are used as the input features of CNN. On Alice's side, the photon is detected by a single-photon detector (SPD), and the signal is sent for coincidence or to trigger the intensified charge-coupled device (ICCD) camera on Bob's side. In our experiment, we use a coincidence window of 3 ns and an exposure time of 1 s for all measurements. The projective direction on Alice's side only needed to be selected as $\Gamma_A = |D\rangle\langle D|$ ($|D\rangle = \frac{|H\rangle + |V\rangle}{\sqrt{2}}$) in this experiment.

On Bob's side, the photon is detected by an SPD and an ICCD camera in the reflected path and the transmitted path, respectively. The signal from SPD is sent for coincidence to reconstruct the experimental density matrix ρ_{exp} (the reflected path of the PBS). The fidelities $\left(\text{Tr}\left[\sqrt{\sqrt{\rho_{AB}(p,\theta)}\rho_{\text{exp}}\sqrt{\rho_{AB}(p,\theta)}}\right]^2\right)$ between the theoretical physical states $\rho_{AB}(p, \theta)$ and the experimental states ρ_{exp} are more than 92%. The concurrences C_{act} are calculated from ρ_{exp} [7,8]. The photon spatial distributions are detected by the ICCD camera. A small down-rightward deflection ($\phi_l(x, y) \rightarrow \phi_l^{H(V)}(x - \text{Re}[_{\langle H(V)|}\langle \hat{\sigma}_x \rangle^w], y - \text{Im}[_{\langle H(V)|}\langle \hat{\sigma}_x \rangle^w])$; see Appendix C for more details) is introduced [18,19] after post-selecting the photon state onto the basis $|H\rangle\langle H|$ ($|V\rangle\langle V|$). $\langle H(V)|\langle \hat{\sigma}_x \rangle^w$ can be obtained according to Eq. (3) with $\rho_B^{\Gamma_A} = \text{Tr}_A[\rho_{AB}|D\rangle\langle D| \otimes I_B]/\text{Tr}[\rho_{AB}|D\rangle\langle D| \otimes I_B]$ representing the corresponding local conditional state. The final intensities $I_H = |\phi_l^H|^2$ and $I_V = |\phi_l^V|^2$ of the photon spatial distributions are detected by the ICCD camera (Fig. 3). Obviously, only a single projective base is needed to detect the weak value $\langle H(V)|\langle \hat{\sigma}_x \rangle^w$ while quantum state tomography (QST) needs 16 projective bases. The total number of photon pairs that contribute to each base is approximately 600,000 (600,000 repetitions) with integral time 1 s. If the number of repetitions is too small, the estimation error will be increased, and more data will be required to train the CNN.

In our experiment, we prepared 415 states, with 349 of the feature-label pairs being randomly chosen to be in the training
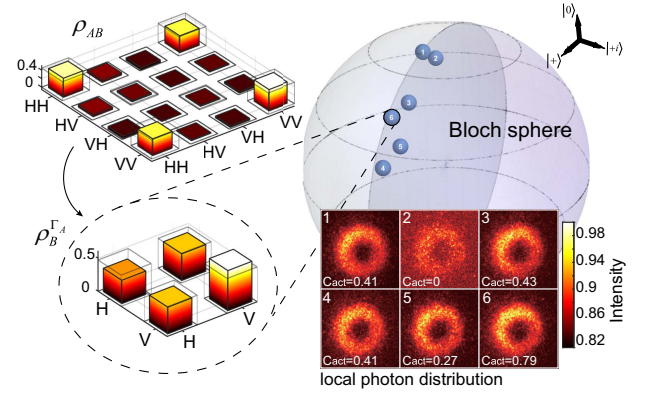

Fig. 3. Conditional states and photon spatial distributions. The numbered dots in the Bloch sphere represent the conditional projective states of Bob. Local photon distribution (I_H) recorded by the ICCD camera with the corresponding concurrence (C_{act}) is shown in front of the Bloch sphere. The density matrix $\rho_B^{\Gamma_A}$ of the number 6 dot and the corresponding input state ρ_{AB} with $p = 0.9$ and $\theta = 0.81$ are shown; here, the solid and transparent bars represent the experimental and theoretical results, respectively.

data and the remaining 66 pairs serving as test data. Some photon distributions with different concurrences in the training set and an example of matrices $\rho_{AB}(p = 0.9, \theta = 0.81)$ and $\rho_B^{\Gamma_A}$ are shown in Fig. 3.

We first investigated the convergence of CNN. The performance with training time (epoch) is shown in Fig. 4(a). The weight of the CNN is optimized with the batch size 10 for 200 epochs. MSE (brown line) falls below 0.07 and PCC (green line) reaches 0.92 at the end of the training process, which demonstrates small prediction errors and strong relevance. The CNN learns more and more information concerning the entanglement with increasing epochs and is well trained when the epoch number is 200.

To demonstrate the predicted concurrence (C_{pre}) accuracy of the trained CNN, we compared it to the concurrence (actual concurrence C_{QST}) calculated by QST. The predicted concurrences (brown dots) for the test data are shown in Fig. 4(b). The point distribution is close to the optimal curve (blue line), which means that the predicted concurrences approximate the actual concurrences. We can also find the errors (uncertainty) between the concurrence C_{pre} predicted by CNN and the actual concurrence C_{QST} is less than 0.2 ($|C_{\text{pre}} - C_{\text{QST}}| < 0.2$). With further optimization of the CNN parameters and an increase in the training data, the error can be further reduced.

The predicted concurrence distributions in the parameter space of p–θ are shown in Fig. 4(c). The contour lines represent the theoretical boundaries of $\rho_{AB}(p, \theta)$, and the predicted concurrence agrees with the theoretical trend. The (p, θ) values of the dots are calculated by minimizing the value of $\text{Tr}[\rho_{AB}(p, \theta) - \rho_{\text{exp}}]$. We further investigated the source of the error shown in Fig. 4(d). With an increase in the training time, the average MSE of the separable states in the test set approaches 10^{-4} (red bars), while the average MSE of the entangled states in the test set approaches 10^{-2} (blue bars); these values are reached when epoch is greater than 160.
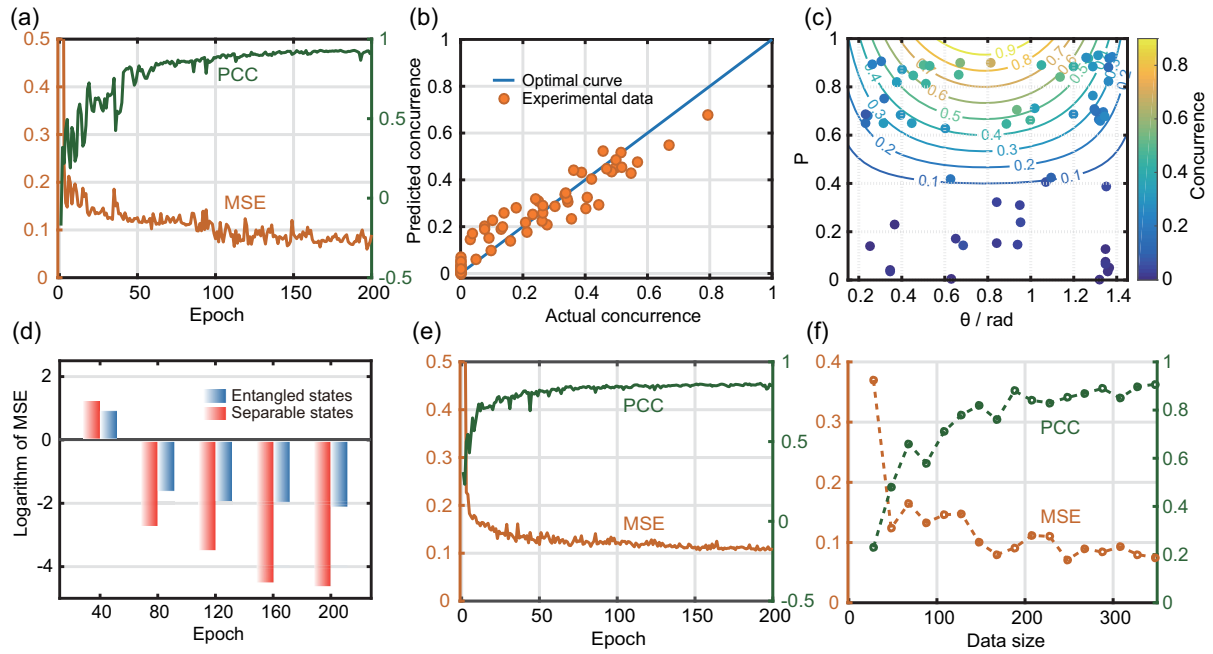
Fig. 4. Experimental results. (a) CNN performance versus epoch. The brown curve represents the MSE value, and the green line represents the PCC between the actual concurrence C_{act}^i and the predicted concurrence C_{pre}^i. (b) Distribution of predicted concurrences. The blue line represents the optimal curve. (c) Concurrence distributions in the p – θ space. The dots and contour lines represent the experimental and theoretical results, respectively. (d) Logarithm of MSE values of the entangled and separated states versus epoch. (e) CNN fivefold cross-validation performance. The result is averaged five folds. (f) CNN performance versus data size.

These results imply that the CNN is more accurate when determining separated states.

To estimate how the model is expected to perform in general cases with limited samples, a fivefold cross-validation was implemented with our CNN. The data set was split into five parts, and each part was chosen in turn to be the test set, while the rest were used as the training set sent to the CNN. The mean MSE and PCC values of all five results converged to 0.09 and 0.87, respectively, when the epoch number is 200, as shown in Fig. 4(e).

Figure 4(f) demonstrates that the result is dependent on the amount of data. With an increasing data size, we should be able to predict the entanglement with greater accuracy.

4. SCALABILITY

It is worth noting that our entanglement quantification method remains efficient even when scaling to pure states with larger numbers of qubits or higher dimensions. To illustrate this, we present the results of three-qubit pure states and two-qutrit pure states below. The concurrence for multipartite pure states $\rho_P = |\psi\rangle\langle\psi|$ in arbitrary dimensions can be expressed as [30]

$$C(|\psi\rangle) = \min_{\gamma_i \in \gamma} \sqrt{2(1 - \mathrm{Tr}[\rho_{\gamma_i}^2])}, \quad (5)$$

where $\gamma = \{\gamma_i\}$ represents the set of all possible bipartitions $\{A_i \bar{A}_i\}$ of $\{1, 2, ..., n\}$, and $\rho_{\gamma_i} = \mathrm{Tr}_{\bar{A}_i}\rho_P$ is the reduced density matrix across these bipartitions.

For three-qubit states ρ_{ABC}, there are three possible bipartitions: $\gamma = \{\{A|B, C\}, \{B|A, C\}, \{C|A, B\}\}$. The concurrence of an arbitrary three-qubit pure state $|\psi_{ABC}\rangle$ can be

rewritten as $C(|\psi_{ABC}\rangle) = \min\{2\sqrt{\det(\rho_A)}, 2\sqrt{\det(\rho_B)},$ $2\sqrt{\det(\rho_C)}\} = \min\{\sqrt{1 - |\langle 0|\langle\hat{\sigma}_x\rangle_{\Gamma_{B,C}}^w|^2}, \sqrt{1 - |\langle 0|\langle\hat{\sigma}_x\rangle_{\Gamma_{A,C}}^w|^2},$ $\sqrt{1 - |\langle 0|\langle\hat{\sigma}_x\rangle_{\Gamma_{A,B}}^w|^2}\}$, where $\Gamma_{A,B}$, $\Gamma_{B,C}$, and $\Gamma_{C,A}$ represent the measurement on the corresponding subsystem (A, B), (B, C), and (C, A), respectively. Clearly, using the weak measurement method, only three projective bases are required to quantify the entanglement of any three-qubit pure state. However, 64 projective bases are required for tomography.

In the computational basis $\{|0\rangle, |1\rangle, |2\rangle\}$, an arbitrary two-qutrit pure state can be expressed as

$$|\psi_{AB}\rangle = \sum_{i,j=0}^{2} a_{ij}|ij\rangle_{AB}, \quad \sum_{i,j=0}^{2} |a_{ij}|^2 = 1. \quad (6)$$

By combining Eq. (5) and Eq. (6), we can obtain $C(|\psi_{AB}\rangle)^2 = 4\det(\rho_{B01}) + 4\det(\rho_{B02}) + 4\det(\rho_{B12})$, where $\rho_{Bmn} = \rho_B(mm)|m\rangle\langle m| + \rho_B(mn)|m\rangle\langle n| + \rho_B(nm)|n\rangle\langle m| + \rho_B(nn)|n\rangle\langle n|$ represents the reduced density matrix of Bob in the qubit subsystem. $\rho_B(mn)$ is the matrix element of $\rho_B = \mathrm{Tr}_A[|\psi_{AB}\rangle\langle\psi_{AB}|]$ in the m-th row and n-th column, where $m, n \in \{0, 1, 2\}$ and $m \neq n$. Furthermore, it can be easily demonstrated that $C(|\psi_{AB}\rangle) = \sqrt{1 - |\langle 0|\langle\hat{\sigma}_{01}\rangle_{\Gamma_B}^w|^2} + \sqrt{1 - |\langle 0|\langle\hat{\sigma}_{02}\rangle_{\Gamma_B}^w|^2} + \sqrt{1 - |\langle 1|\langle\hat{\sigma}_{12}\rangle_{\Gamma_B}^w|^2}$, where $\hat{\sigma}_{mn} = |m\rangle\langle n| + |n\rangle\langle m|$, and Γ_B represents the identity measurement. Clearly, for any two-qutrit pure state, only three projective bases are needed to quantify the entanglement. However, when employing tomography, 81 projective bases are required.

As for universal N-partite D-dimensional pure state, there are $\sum_{j=1}^{\lceil N/2\rceil} C_N^j$ possible bipartitions and $\sum_{k=1}^{\lceil D/2\rceil} C_D^k$ possible qubit subsystems; thus, no more than $\sum_{j=1}^{\lceil N/2\rceil} \sum_{k=1}^{\lceil D/2\rceil} C_N^j C_D^k$ projective bases are needed to quantify the entanglement. $\lceil N/2\,(D/2)\rceil$ represents round $N/2\,(D/2)$ down to the nearest integer, and $C_N^j\,(C_D^k)$ represents the number of all combinations of taking $j(k)$ distinct elements from $N\,(D)$ distinct elements. In contrast, tomography requires D^{2N} projective bases. Clearly, the benefits of our method become increasingly apparent as the scale and the number of dimensions of the system increase.

5. CONCLUSION AND DISCUSSION

In this study, we established the relations between the concurrence and weak values, where the weak values were encoded into the transverse Laguerre–Gaussian (LG) photon distributions. By detecting the shift of singular center, we can directly get the concurrence. This weak measurement method will greatly reduce the number of projective bases and save the measurement resources. For a class of mixed two-qubit states, two projective bases are sufficient to quantify the entanglement, while 16 bases are required for tomography. Assisted by the DL method, we can achieve an "end-to-end" mapping between the photon distribution and concurrence without considering the form of the mapping functions. Moreover, this method, which takes the whole photon distribution as the input, can avoid estimation errors because it extracts the singular center weak shift.

We also show that the advantages of our method become more pronounced when scaling to larger numbers of qubits or higher dimensions, which demonstrates that our entanglement quantization method exhibits increasingly evident advantages over quantum state tomography as the system scales up. This highlights the potential of our method to efficiently characterize entanglement in large-scale quantum systems.

APPENDIX A: RELATIONSHIP BETWEEN CONDITIONAL STATE $\rho_B^{\Gamma_A}$ AND THE VALUES $_{\langle 0(1)|}\langle\hat{\sigma}_x\rangle^w$

For a two-qubit system ρ_{AB}, the local conditional state $\rho_B^{\Gamma_A}$ can be generally represented as

$$\rho_B^{\Gamma_A} = \sum_{i,j=0,1} \rho_{ij}|i\rangle\langle j|. \tag{A1}$$

The weak values of σ_x with the post-selected state being $|0\rangle$ and $|1\rangle$ are, respectively, denoted as

$$_{\langle 0|}\langle\hat{\sigma}_x\rangle^w = \frac{\mathrm{Tr}[\hat{\sigma}_x \rho_B^{\Gamma_A}|0\rangle\langle 0|]}{\mathrm{Tr}[\rho_B^{\Gamma_A}|0\rangle\langle 0|]} = \frac{\rho_{10}}{\rho_{00}},$$

$$_{\langle 1|}\langle\hat{\sigma}_x\rangle^w = \frac{\mathrm{Tr}[\hat{\sigma}_x \rho_B^{\Gamma_A}|1\rangle\langle 1|]}{\mathrm{Tr}[\rho_B^{\Gamma_A}|1\rangle\langle 1|]} = \frac{\rho_{01}}{\rho_{11}}. \tag{A2}$$

For any physical state, $\mathrm{Tr}[\rho_B^{\Gamma_A}] = 1$ and $\rho_B^{\Gamma_A} = \rho_B^{\Gamma_A\dagger}$. The elements of the density matrix can be expressed by

$$\rho_{00} = \frac{|_{\langle 1|}\langle\hat{\sigma}_x\rangle^w|}{|_{\langle 0|}\langle\hat{\sigma}_x\rangle^w| + |_{\langle 1|}\langle\hat{\sigma}_x\rangle^w|},$$

$$\rho_{01} = \frac{|_{\langle 0|}\langle\hat{\sigma}_x\rangle^w|_{\langle 1|}\langle\hat{\sigma}_x\rangle^w|}{_{\langle 0|}\langle\hat{\sigma}_x\rangle^w| + |_{\langle 1|}\langle\hat{\sigma}_x\rangle^w|},$$

$$\rho_{10} = \frac{|_{\langle 1|}\langle\hat{\sigma}_x\rangle^w|_{\langle 0|}\langle\hat{\sigma}_x\rangle^w|}{_{\langle 0|}\langle\hat{\sigma}_x\rangle^w| + |_{\langle 1|}\langle\hat{\sigma}_x\rangle^w|},$$

$$\rho_{11} = \frac{|_{\langle 0|}\langle\hat{\sigma}_x\rangle^w|}{|_{\langle 0|}\langle\hat{\sigma}_x\rangle^w| + |_{\langle 1|}\langle\hat{\sigma}_x\rangle^w|}, \tag{A3}$$

which are determined by these two weak values. In our work, CNN needs only original information. The photon distribution on Bob's side with the post-selected state $|0\rangle$ and $|1\rangle$ is used as the input feature.

APPENDIX B: REPRESENTATION OF THE VALUES $|_{\langle 1|}\langle\hat{\sigma}_x\rangle^w|$ IN THE BLOCH SPHERE

It is well known that the conditional state of $\rho_B^{\Gamma_A}$ can be expressed as

$$\rho_B^{\Gamma_A}(r) = \frac{1}{2}(I + \boldsymbol{r}\cdot\boldsymbol{\sigma}), \tag{B1}$$

where $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ and $\boldsymbol{r} = r(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$. $\boldsymbol{r}$ indicates the position of the conditional state of $\rho_B^{\Gamma_A}$ in spherical coordinates. As shown in Fig. 5, the connection between the pure state $|0\rangle$ and $\rho_B^{\Gamma_A}$ intersects with the x-o-y plane. Obviously, the distance between these intersected points and the center of the sphere (corresponding to maximally mixed state $\frac{1}{2}I$) is

$$d = \frac{r\sin\theta}{1 - r\cos\theta}. \tag{B2}$$

It is worth noting that d is related to the distance between the conditional state $\rho_B^{\Gamma_A}$ and maximally mixed state $\frac{1}{2}I$.

According to Eq. (A2), the values $\langle 1|\langle\hat{\sigma}_x\rangle^w$ can be expressed as

$$\langle 1|\langle\hat{\sigma}_x\rangle^w = \frac{\rho_{10}}{\rho_{11}} = \frac{re^{i\phi}\sin\theta}{1 - r\cos\theta}. \tag{B3}$$

As a result, the value $|_{\langle 1|}\langle\hat{\sigma}_x\rangle^w| = d$ reflects the distance between the conditional state $\rho_B^{\Gamma_A}$ and maximally mixed state
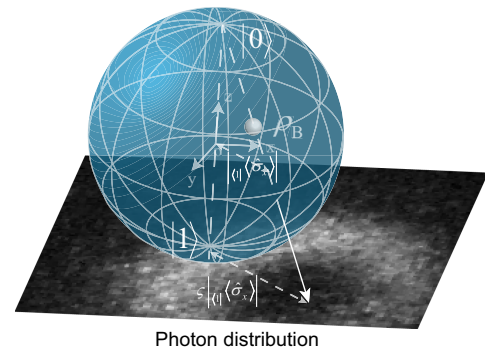

Photon distribution

Fig. 5. Value $_{\langle 1|}\langle\hat{\sigma}_x\rangle^w$ of photon B on the state of $\rho_B^{\Gamma_A}$ is shown in the Bloch sphere, which is encoded in the photon B's spatial distribution carrying orbital angular momentum.

$\frac{1}{2}I$. In addition, the representation of the weak value $|_{\langle 0|}\langle\hat{\sigma}_x\rangle^w|$ is similar to $|_{\langle 1|}\langle\hat{\sigma}_x\rangle^w|$.

APPENDIX C: WEAK VALUE EXTRACTION FROM THE CONDITIONAL SPATIAL DISTRIBUTIONS

For a two-qubit system ρ_{AB}, the local conditional state $\rho_B^{\Gamma_A}$ can be generally represented as

$$\rho_B^{\Gamma_A} = \sum_{i,j=0,1} \rho_{ij}|i\rangle\langle j|. \quad \text{(C1)}$$

The weak value of σ_x with the post-selected state $|k\rangle$ ($k \in \{0,1\}$) is denoted as

$$_{\langle k|}\langle\hat{\sigma}_x\rangle^w = \frac{\text{Tr}[\hat{\sigma}_x\rho_B^{\Gamma_A}|k\rangle\langle k|]}{\text{Tr}[\rho_B^{\Gamma_A}|k\rangle\langle k|]} = \frac{\sum_{i,j=0,1}\rho_{ij}\langle k|\hat{\sigma}_x|i\rangle\langle j|k\rangle}{\sum_{i,j=0,1}\rho_{ij}\langle k|i\rangle\langle j|k\rangle}. \quad \text{(C2)}$$

Note that this definition can be taken as the formal formalism of the density-operator-based weak value [31]. In our experiment, $\rho_B^{\Gamma_A}$ and the transverse field $\phi_l(x,y)$ are chosen to be the system and the pointer, respectively. The whole density matrix is

$$\rho_{\text{all}} = \sum_{i,j=0,1}\rho_{ij}|i\rangle\langle j| \otimes |\phi_l(x,y)\rangle\langle\phi_l(x,y)|. \quad \text{(C3)}$$

After weak coupling, described by the interaction Hamiltonian $\hat{H} = \zeta\hat{\sigma}_x \otimes P_x$, the whole system is obtained as

$$\rho_{\text{all}}^f = U\rho_{\text{all}}U^\dagger, \quad \text{(C4)}$$

where $U = e^{-i\hat{H}\Delta t/\hbar} = e^{-i\zeta\hat{\sigma}_x\otimes P_x}$ (assuming $\Delta t/\hbar = 1$). By post-selecting the system on the state $|k\rangle\langle k|$, we get the pointer state

$$rcl\rho_P^f = \text{Tr}[\rho_{\text{all}}^f|k\rangle\langle k| \otimes I_p]$$
$$= \sum_{i,j=0,1}\rho_{ij}\langle k|e^{-i\zeta\hat{\sigma}_x\otimes P_x}|i\rangle$$
$$\times |\phi_l(x,y)\rangle\langle\phi_l(x,y)| \times \langle j|e^{i\zeta\hat{\sigma}_x\otimes P_x}|k\rangle$$
$$= \sum_{i,j=0,1}\rho_{ij}\langle k|i\rangle \frac{\sum_{i,j=0,1}\rho_{ij}\langle k|e^{-i\zeta\hat{\sigma}_x\otimes P_x}|i\rangle\langle j|k\rangle}{\sum_{i,j=0,1}\rho_{ij}\langle k|i\rangle\langle j|k\rangle}$$
$$\times |\phi_l(x,y)\rangle\langle\phi_l(x,y)|$$
$$\times \frac{\sum_{i,j=0,1}\rho_{ij}\langle j|e^{-i\zeta\hat{\sigma}_x\otimes P_x}|k\rangle\langle k|i\rangle}{\sum_{i,j=0,1}\rho_{ij}\langle k|i\rangle\langle j|k\rangle}\langle j|k\rangle. \quad \text{(C5)}$$

Taking Eq. (C2) into the pointer state, the pointer state becomes

$$rcl\rho_P^f = \sum_{i,j=0,1}\rho_{ij}\langle k|i\rangle\langle j|k\rangle e^{-i\zeta_{\langle k|}\langle\hat{\sigma}_x\rangle^w\otimes P_x}$$
$$\times |\phi_l(x,y)\rangle\langle\phi_l(x,y)|e^{i\zeta_{\langle k|}\langle\hat{\sigma}_x\rangle^w\otimes P_x}$$
$$= \sum_{i,j=0,1}\rho_{ij}\langle k|i\rangle\langle j|k\rangle$$
$$\times |\phi_l(x-\zeta_{\langle k|}\langle\hat{\sigma}_x\rangle^w,y)\rangle\langle\phi_l(x-\zeta_{\langle k|}\langle\hat{\sigma}_x\rangle^w,y)|. \quad \text{(C6)}$$

For the LG beam $|\phi_l(x,y)\rangle \propto (x+iy)e^{-x^2+y^2}$, the pointer state satisfies

$$\phi_l(x-\zeta_{\langle k|}\langle\hat{\sigma}_x\rangle^w,y)$$
$$= \phi_l(x-\zeta\times\text{Re}[_{\langle k|}\langle\hat{\sigma}_x\rangle^w],y-\zeta\,\text{Im}[_{\langle k|}\langle\hat{\sigma}_x\rangle^w]). \quad \text{(C7)}$$

Therefore, the weak value information is encoded into the pointer state and detected by ICCD as conditional spatial distributions [32,33]. The differentiation due to the approximation was discussed in Ref. [34].

APPENDIX D: MECHANISM OF THE CNN

In the convolutional (Conv) path, the matrix of input image I is transferred into a feature space through three convolutional-max pooling layers. To illustrate the convolution operation, assume that the filter of the Conv layer is

$$\text{conv} = \begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \\ \theta_{31} & \theta_{32} & \theta_{33} \end{pmatrix}. \quad \text{(D1)}$$

The output feature matrix $O = I * \text{conv}$ is denoted as

$$O_{i,j} = \sum_{k,l=1}^3 I_{i+k,j+l}\theta_{k,l}, \quad \text{(D2)}$$

where i, j represent the element indices of the matrix. To reduce the size of feature space, a max pooling (MP) layer is set followed by the Conv layer. Assuming the filter of the MP layer is 2×2, the output matrix M can be written as

$$M_{i,j} = \max(O_{i,j}, O_{i+1,j}, O_{i,j+1}, O_{i+1,j+1}). \quad \text{(D3)}$$

By implementing the same operations three times, the Conv path would output the final feature map, which is flattened into a one-dimensional vector $\vec{x}$ and sent to full connection (FC) layers. The FC consists of four layers, which is constructed through the relation

$$\vec{x}_1 = \sigma_{\text{RL}}(W_1\vec{x} + \vec{b}_1),\ \vec{x}_2 = \sigma_{\text{RL}}(W_2\vec{x}_1 + \vec{b}_2),$$
$$\vec{x}_3 = \sigma_{\text{RL}}(W_3\vec{x}_2 + \vec{b}_3),\ C = W_4\vec{x}_3 + b_4. \quad \text{(D4)}$$

$W_{i=1,2,3,4}$ and $b_{i=1,2,3,4}$ represent the weight matrix and bias, respectively, which would be updated in the training process. The nonlinear function ReLU is defined by $\sigma_{\text{RL}}(z_1,z_2,z_3,\ldots,z_n) = (\max(z_1,0),\max(z_2,0),\max(z_3,0),\ldots,\max(z_n,0))$. The output C is compared with the actual concurrence. The errors are sent back to the networks to optimize the weight matrix $W_{i=1,2,3,4}$ and bias $b_{i=1,2,3,4}$.

Disclosures. The authors declare no conflicts of interest.

Data Availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

REFERENCES

1. C. H. Bennett, G. Brassard, C. Crepeau, *et al.*, "Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels," Phys. Rev. Lett. **70**, 1895 (1993).
2. A. K. Ekert, "Quantum cryptography based on Bell's theorem," Phys. Rev. Lett. **67**, 661 (1991).
3. D. P. DiVincenzo, "Quantum computation," Science **270**, 255 (1995).
4. R. Horodecki, P. Horodecki, M. Horodecki, *et al.*, "Quantum entanglement," Rev. Mod. Phys. **81**, 865 (2009).
5. K. E. Dorfman, F. Schlawin, and S. Mukamel, "Nonlinear optical signals and spectroscopy with quantum light," Rev. Mod. Phys. **88**, 045008 (2016).
6. D. F. V. James, P. G. Kwiat, W. J. Munro, *et al.*, "Measurement of qubits," Phys. Rev. A **64**, 052312 (2001).
7. W. K. Wootters, "Entanglement of formation of an arbitrary state of two qubits," Phys. Rev. Lett. **80**, 2245 (1998).
8. R. Hildebrand, "Concurrence revisited," J. Math. Phys. **48**, 102108 (2007).
9. D. T. Smithey, M. Beck, M. G. Raymer, *et al.*, "Measurement of the Wigner distribution and the density matrix of a light mode using optical homodyne tomography: application to squeezed states and the vacuum," Phys. Rev. Lett. **70**, 1244–1247 (1993).
10. G. Breitenbach, S. Schiller, and J. Mlynek, "Measurement of the quantum states of squeezed light," Nature **387**, 471–475 (1997).
11. P. Horodecki, "Measuring quantum entanglement without prior state reconstruction," Phys. Rev. Lett. **90**, 167901 (2003).
12. S. P. Walborn, P. S. Ribeiro, L. Davidovich, *et al.*, "Experimental determination of entanglement with a single measurement," Nature **440**, 1022–1024 (2006).
13. M. Yang, Y. Xiao, Y. W. Liao, *et al.*, "Zonal reconstruction of photonic wavefunction via momentum weak measurement," Laser Photonics Rev. **14**, 1900251 (2020).
14. J. Z. Salvail, M. Agnew, A. S. Johnson, *et al.*, "Full characterization of polarization states of light via direct measurement," Nat. Photonics **7**, 316–321 (2013).
15. W. W. Pan, X. Y. Xu, Y. Kedem, *et al.*, "Direct measurement of a nonlocal entangled quantum state," Phys. Rev. Lett. **123**, 150402 (2019).
16. K. O'Shea and R. Nash, "An introduction to convolutional neural networks," arXiv, arXiv:1511.08458 (2015).
17. Y. Xiao, X. J. Ye, K. Sun, *et al.*, "Demonstration of multisetting one-way Einstein-Podolsky-Rosen steering in two-qubit systems," Phys. Rev. Lett. **118**, 140404 (2017).
18. M. Tukiainen, H. Kobayashi, and Y. Shikano, "Quantification of concurrence via weak measurement," Phys. Rev. A **95**, 052301 (2017).
19. H. Kobayashi, K. Nonaka, and Y. Shikano, "Stereographical visualization of a polarization state using weak measurements with an optical-vortex beam," Phys. Rev. A **89**, 053816 (2014).
20. Y. Aharonov, D. Z. Albert, and L. Vaidman, "How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100," Phys. Rev. Lett. **60**, 1351–1354 (1988).
21. J. Dressel, M. Malik, F. M. Miatto, *et al.*, "Colloquium: understanding quantum weak values: basics and applications," Rev. Mod. Phys. **86**, 307–316 (2014).
22. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature **521**, 436–444 (2015).
23. G. Carleo and M. Troyer, "Solving the quantum many-body problem with artificial neural networks," Science **355**, 602–606 (2017).
24. G. Torlai, G. Mazzola, J. Carrasquilla, *et al.*, "Neural-network quantum state tomography," Nat. Phys. **14**, 447–450 (2018).
25. J. Gao, L. F. Qiao, Z. Q. Jiao, *et al.*, "Experimental machine learning of quantum states," Phys. Rev. Lett. **120**, 240501 (2018).
26. M. Yang, C. L. Ren, Y. C. Ma, *et al.*, "Experimental simultaneous learning of multiple nonclassical correlations," Phys. Rev. Lett. **123**, 190401 (2019).
27. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," arXiv, arXiv:1412.6980 (2014).
28. A. Fedrizzi, T. Herbst, A. Poppe, *et al.*, "A wavelength-tunable fiber-coupled source of narrowband entangled photons," Opt. Express **15**, 15377–15386 (2007).
29. S. Kocsis, B. Braverman, S. Ravets, *et al.*, "Observing the average trajectories of single photons in a two-slit interferometer," Science **332**, 1170–1173 (2011).
30. Z. H. Ma, Z. H. Chen, J. L. Chen, *et al.*, "Measure of genuine multipartite entanglement with computable lower bounds," Phys. Rev. A **83**, 062325 (2011).
31. Y. Shikano and A. Hosoya, "Weak values with decoherence," J. Phys. A: Math. Theor. **43**, 025304 (2010).
32. G. Puentes, N. Hermosa, and J. P. Torres, "Weak measurements with orbital-angular-momentum pointer states," Phys. Rev. Lett. **109**, 040401 (2012).
33. H. Kobayashi, G. Puentes, and Y. Shikano, "Extracting joint weak values from two-dimensional spatial displacements," Phys. Rev. A **86**, 053805 (2012).
34. Y. Turek, H. Kobayashi, T. Akutsu, *et al.*, "Post-selected von Neumann measurement with Hermite–Gaussian and Laguerre–Gaussian pointer states," New J. Phys. **17**, 083029 (2015).