Quantum optical classifier with superexponential speedup

Simone Roncallo [1,2,*] Angela Rosy Morgillo [1,2,†] Chiara Macchiavello [1,2,‡] Lorenzo Maccone [1,2,§] and Seth Lloyd [3,¶]

[1] *Dipartimento di Fisica, Università degli Studi di Pavia, Via Agostino Bassi 6, I-27100, Pavia, Italy*
[2] *INFN Sezione di Pavia, Via Agostino Bassi 6, I-27100, Pavia, Italy*
[3] *Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

We present a quantum optical pattern recognition method for binary classification tasks. Without direct image reconstruction, it classifies an object in terms of the rate of two-photon coincidences at the output of a Hong-Ou-Mandel interferometer, where both the input and the classifier parameters are encoded into single-photon states. Our method exhibits the same behaviour of a classical neuron of unit depth. Once trained, it shows a constant $\mathcal{O}(1)$ complexity in the number of computational operations and photons required by a single classification. This is a superexponential advantage over a classical neuron (that is at least linear in the image resolution). We provide simulations and analytical comparisons with analogous neural network architectures.

Keywords: Quantum image classifier; Quantum optical neuron; Quantum neural networks; Hong-Ou-Mandel effect;

I. INTRODUCTION

Image classification has been significantly affected by the introduction of deep learning algorithms, providing several architectures that can learn and extract image features [1–4]. The large number of parameters involved is motivating a consistent effort in reducing the cost of these methods, e.g. by leveraging all-optical implementations that bypass hardware usage [5–11], or quantum mechanical effects that can provide a significant speedup in these computations [12–16]. Quantum optical neural networks harness the best of both worlds, i.e. deep learning capabilities from quantum optics [17–21].

In this paper, we introduce a quantum optical setup to classify objects without reconstructing their images. Our approach relies on the Hong-Ou-Mandel effect, for which the probability that two photons exit a beam splitter in different modes, depends on their distinguishability [22]. In our implementation, an input object is targeted by a single-photon source, and eventually followed by an arbitrary lens system. The single-photon state interferes with another one, which encodes a set of trainable parameters, e.g. through a spatial light modulator. Classification occurs by measuring the rate of two-photon coincidences at the Hong-Ou-Mandel output (see Fig. 1). The Hong-Ou-Mandel effect has been successfully applied to quantum kernel evaluation [23], which can compute distances between pairs of data points in the feature space. In this case, each point is sent to one branch of the interferometer, encoded in the temporal modes of a single-photon state. In our method, the interferometer has only one independent branch, which takes the spatial modes of a single-photon state reflected off the target object. The other branch remains fixed after training, and contains the layer of parameters. After the measurement, the response function of our apparatus mathematically resembles that of a classical neuron. For this reason, we refer to our setup as quantum optical neuron. By analytically comparing the resource cost of the classical and quantum neurons, we show that our method requires constant $\mathcal{O}(1)$ computational operations and injected photons, whereas the classical methods are at least linear in the image resolution: a superexponential advantage.

II. METHOD

In this section, we discuss the apparatus of Fig. 1, without explicitly modelling the probe. Two single-photon states are fed into the left and top branches of a $50:50$ beam splitter, acting as input and processing layers, re-
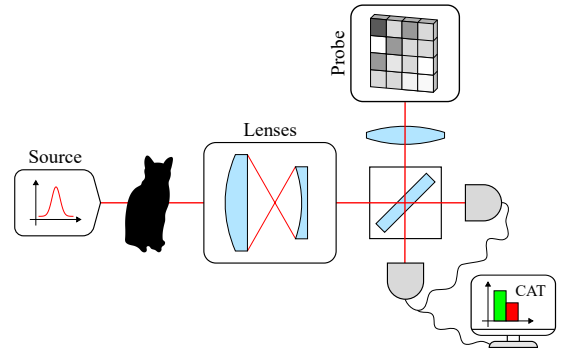

FIG. 1. Quantum optical neuron implemented by the Hong-Ou-Mandel interferometric setup. An object is targeted by a single-photon source and classified through the rate of two-photon coincidences at the interferometer output, without reconstructing its full image. In the top branch, an additional thin lens can translate the classification problem to the Fourier domain.

* Simone Roncallo: simone.roncallo01@ateneopv.it
† Angela Rosy Morgillo: angelarosy.morgillo01@ateneopv.it
‡ Chiara Macchiavello: chiara.macchiavello@unipv.it
§ Lorenzo Maccone: lorenzo.maccone@unipv.it
¶ Seth Lloyd: slloyd@mit.edu

spectively. In the left branch, the single-photon source reflects off the object, and reaches the beam splitter after a linear optical system. In the top branch, we consider a generic single-photon state, which depends on a set of trainable real parameters. We count the two-photon coincidences at the beam splitter output. We show how to interpret the Hong-Ou-Mandel response as the one produced by a single-layer neural network-like operation on the object image.

We call input and probe modes, i.e. a and b, those fed into the left and top branches of the interferometer. In the input branch, a single photon with spectrum ϕ is generated at the longitudinal origin $z = 0$, followed by an object with two-dimensional shape $\mathcal{O}$. An imaging system with transfer function $\mathcal{L}_d$, e.g a pinhole or a linear optical apparatus, is placed after the object. Here, z_o and z_i are the longitudinal positions of the object and the image plane, respectively, and $d = z_i - z_o$ their displacement.

The output of the imaging optics reads (see Appendix A)

$$|\Psi_{\mathcal{I}}\rangle = \int \mathrm{d}^2 k \, \hat{\mathcal{I}}_\omega(k|\mathcal{O}) a_\omega^\dagger(k) |0\rangle \ , \tag{1}$$

with $\hat{\mathcal{I}}_\omega(\cdot|\mathcal{O}) = [(\hat{\phi}_\omega \hat{\mathfrak{H}}_{z_o}) * \hat{\mathcal{O}}]\hat{\mathcal{L}}_d$ the total transfer function from the single-photon source to the image plane, and $a_\omega^\dagger(k)$ the creation operator of a photon in the input mode, acting on the vacuum state $|0\rangle$. The hat operator denotes the two-dimensional Fourier transform on the transverse coordinates plane, $*$ the convolution operation, $\mathfrak{H}_{z_o}$ the transfer function from the source to the object plane, $k = (k_x, k_y)$ the transverse momentum, and ω the frequency conjugated to the temporal degree of freedom of the electromagnetic potential.

In the probe branch, a generic quantum state is prepared, eventually followed by a linear optical system. At the beam splitter plane, the probe state reads

$$|\Psi_{\mathcal{U}}\rangle = \int \mathrm{d}^2 k \, \hat{\mathcal{U}}_\omega(k|\lambda) b_\omega^\dagger(k) |0\rangle \ , \tag{2}$$

with $\lambda = \{\lambda_{i_1 \ldots i_n}\}$ a collection of (trainable) parameters, $\mathcal{U}$ the spatial spectrum of the probe, and $b_\omega^\dagger(k)$ the creation operator of a photon in the probe mode.

A photodetector is placed at the output of each branch. After feeding both states into a 50:50 beam splitter, the rate of two-photon coincidences reads

$$p(1_a \cap 1_b|\lambda, \mathcal{O}) = \frac{1}{2} \left[\alpha_\lambda(\mathcal{O}) - f_\lambda(\mathcal{O}) \right] \ , \tag{3}$$

$$\text{with} \quad \begin{aligned} \alpha_\lambda(\mathcal{O}) &= ||\mathcal{I}_\omega(\cdot|\mathcal{O})||^2 ||\mathcal{U}_\omega(\cdot|\lambda)||^2 \ , \\ f_\lambda(\mathcal{O}) &= |\langle \mathcal{I}_\omega(\cdot|\mathcal{O}), \mathcal{U}_\omega(\cdot|\lambda)\rangle|^2 \ , \end{aligned} \tag{4}$$

where $||\cdot||$ and $\langle \cdot, \cdot \rangle$ denote the L^2-norm and inner product, respectively. Here, $\alpha_\lambda(\mathcal{O})$ depends on the normalization of the input and probe states, which can be $\alpha_\lambda < 1$ in the presence of optical losses. Whenever the two spectra are indistinguishable, i.e. when $\mathcal{U}$ perfectly matches

$\mathcal{I}$, coincidences are not observed. On the other hand, the more distinguishable the input and the probe states are, the smaller $\langle \mathcal{I}(\cdot|\mathcal{O}), \mathcal{U}(\cdot|\lambda)\rangle$ becomes and the rate of coincidences increases. See Appendix B for a derivation, and Appendix D, for a similar result in the Fourier domain.

At the image plane I, with transverse coordinates $r = (x, y)$, we have

$$f_\lambda(\mathcal{O}) = \left| \int_I \mathrm{d}r \, \mathcal{I}_\omega(r|\mathcal{O}) \mathcal{U}_\omega^*(r|\lambda) \right|^2 \ . \tag{5}$$

This integral measures the point-wise overlap between the input image and the probe. We interpret it as the prediction of our classification model, where $f_\lambda \in [0, 1]$ represents the probability that $\mathcal{I}$ belongs to the class of $\mathcal{U}$. In particular, $f_\lambda \to 0$ ($f_\lambda \to 1$) when the class of $\mathcal{I}$ is orthogonal to (is the same of) $\mathcal{U}$. In the next section, we show how to encode a generic class in $\mathcal{U}$, by means of the optimization of the set of parameters λ.

The output measurement introduces a non-linear operation after the beam splitter, represented by the squared absolute value in the left-hand side of Eq. (5). We increase the predictability of our model, by enhancing this non-linearity through the following post-processing operations. Consider the sigmoid (logistic) function

$$\sigma(x) := \frac{1}{1 + e^{-\beta x + \gamma}} \ , \tag{6}$$

where β, γ are hyperparameters, i.e. constants with respect to the training process. We introduce an additional trainable parameter $b \in \mathbb{R}$, called bias, which, combined with f_λ and σ, yields

$$F_{b\lambda}(\mathcal{O}) = \sigma(f_\lambda(\mathcal{O}) + b) \ , \tag{7}$$

which determines the label predicted by the Hong-Ou-Mandel apparatus. These modifications can improve the performance of the neuron. The sigmoid increases the non-linearity introduced by the squared absolute value, and so the predictability of the model. In addition, the bias is introduced on heuristic motivations: it compensates the constraint given by the normalization in Eq. (3), while enhancing the robustness of our protocol against optical losses (which may affect the above-mentioned normalizability, yielding $\alpha_\lambda < 1$).

We now discuss the training stage. Consider a training set, i.e. an ensemble of objects $\{\mathcal{O}_j\}$ with target labels $\{y_j \in \{0, 1\}\}$. We separately feed each object into the input branch of the interferometer. Predicted and target classes are compared in terms of their binary cross-entropy, which is used as loss function of a gradient descent optimizer. The optimizer updates λ through the derivative of the loss function, whose only model-dependent contribution is

$$\partial_\lambda f = 2 \, \mathrm{Re} \left[\langle \mathcal{I}_\omega, \mathcal{U}_\omega \rangle \langle \mathcal{I}_\omega, \partial_\lambda \mathcal{U}_\omega \rangle^* \right] \ . \tag{8}$$

Ideally, the training is complete after finding a set of parameters that minimizes the loss. Notice that our model
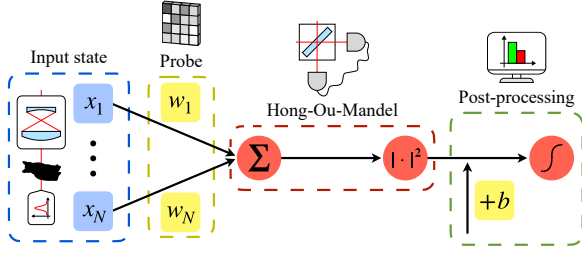
FIG. 2. Mathematical relationship between the Hong-Ou-Mandel apparatus of Fig. 1 and the classical neuron of Eq. (9). Each operation is identified with the corresponding component of the optical interferometer.

		QON	Classical
Computational (# of operations)		$\mathcal{O}(1)$	N
Optical (# of photons)	Imaging	None	$\Theta(\varsigma^{-2}\langle x\rangle N)$
	Classification	$\mathcal{O}(\varepsilon^{-2})$	$\Omega\left(\varepsilon^{-2}\langle x\rangle N\right)$

TABLE I. Computational and optical resources comparison between the quantum optical neuron (QON) and its classical counterparts, when reconstructing and classifying an image x of N pixels. Here, ς and $\langle x\rangle$ are the standard deviation and the average brightness of the image (which depend on the reflectivity of the object), while ε is the uncertainty on the classification outcome. Our method achieves a superexponential speedup over its classical counterpart: $\mathcal{O}(1)$ vs. $\mathcal{O}(N)$.

is resilient against the issue of gradient explosion [24], since it depends on physical data and functions only (see Appendix C for a discussion).

There is a formal relationship between the post-processed output of the Hong-Ou-Mandel interferometer of Eq. (7) and that of a classical neuron. Consider $f_\lambda(\mathcal{O})$ discretized and vectorized in a mesh of N cells, either in the spatial or in the Fourier domain. Then, Eq. (7) corresponds to the composition of a real-valued neuron, with N trainable weights, square absolute value activation function and no bias, and a second neuron, with a scalar unit weight, sigmoid activation function and a trainable bias. Namely

$$G_{bw}(x) = \sigma\left(|w\cdot x|^2 + b\right) , \qquad (9)$$

where $x \in \mathbb{C}^N$ is the input, while $w \in \mathbb{C}^N$ and $b \in \mathbb{R}$ are the weights and bias, respectively. We can formally identify $G_{bw}(x)$ with $F_{b\lambda}(\mathcal{O})$ under the substitution

$$(x,w) \xleftarrow{\sim} \left(\mathcal{I}_\omega(r|\mathcal{O}), \mathcal{U}_\omega(r|\lambda)\right) , \qquad (10)$$

where $\xleftarrow{\sim}$ is the discretization and vectorization to $\mathbb{C}^N$. This analogy is represented in Fig. 2.

A classical neuron requires at least N photons and N computational operations to classify an image composed of N pixels. Our setup bypasses both costs, by leveraging two essential features. On the one hand, it is completely optical, avoiding the computational need of processing the image. On the other hand, it classifies patterns through the Hong-Ou-Mandel effect, reducing the photon cost of imaging. In both ways, it provides a superexponential speedup, from $\mathcal{O}(N)$ to $\mathcal{O}(1)$. Photon losses due to absorption introduce a constant overhead in both the classical and quantum strategies, which depends on the total reflectivity of the object. We summarize this discussion in Table I. See Appendix E for a detailed derivation.

III. AMPLITUDE MODULATED PROBE

We specialize our discussion by replacing the generic probe state $\mathcal{U}$ with a toy model of an amplitude spatial light modulator (SLM), placed in the top branch of the Hong-Ou-Mandel interferometer, e.g. a liquid crystal (LC) grid with negligible losses [25]. Different approaches can be investigated, such as phase-only SLM [26], which may exhibit superior resiliency against losses.

Consider a pattern on a greyscale LC grid with N real amplitudes $\{\lambda_{\mu\nu}\}$. Each pixel, labelled by (μ,ν), is represented by an $L\times L$ square with center $r_{\mu\nu} = (\mu+1/2, \nu+1/2)L$. Upon an overall parameter-independent normalization, the probe can be approximated as a combination of top-hat functions

$$\mathcal{U}_\omega(r|\lambda) = \sum_{\mu,\nu} u(r - r_{\mu\nu})\frac{\lambda_{\mu\nu}}{||\lambda||} , \qquad (11)$$

where $||\lambda||^2 = \sum_{\mu,\nu}\lambda_{\mu\nu}^2$ and $u(r) := \theta(r + L/2) - \theta(r - L/2)$, with θ the two-dimensional Heaviside step function. Under this choice, Eq. (5) simplifies to

$$f_\lambda(\mathcal{O}) = \left|\sum_{\mu,\nu}(u \star \mathcal{I}_\omega)(r_{\mu\nu})\frac{\lambda_{\mu\nu}}{||\lambda||}\right|^2 , \qquad (12)$$

where $\star$ is the cross-correlation operation. We introduce a bias and a sigmoid activation function, so that the post-processed output reads $F_{b\lambda}(\mathcal{O}) = \sigma(f_\lambda(\mathcal{O}) + b)$. Assuming that $\mathcal{I}$ is real, Eq. (8) simplifies to

$$\partial_{\mu\nu}f \simeq 2\frac{\sqrt{f}}{||\lambda||}\left[(u \star \mathcal{I}_\omega)(r_{\mu\nu}) - \sqrt{f}\frac{\lambda_{\mu\nu}}{||\lambda||}\right] , \qquad (13)$$

with $\partial_{\mu\nu}f := \partial f/\partial\lambda_{\mu\nu}$. This expression can be evaluated in an all-optical way, by taking the amplitude measurement of $\mathcal{I}$ directly in the left branch of the interferometer, before the beam splitter. This operation can be done offline, and once per training object. In the next section, we present a simulation of these results, for different choices of the dataset.
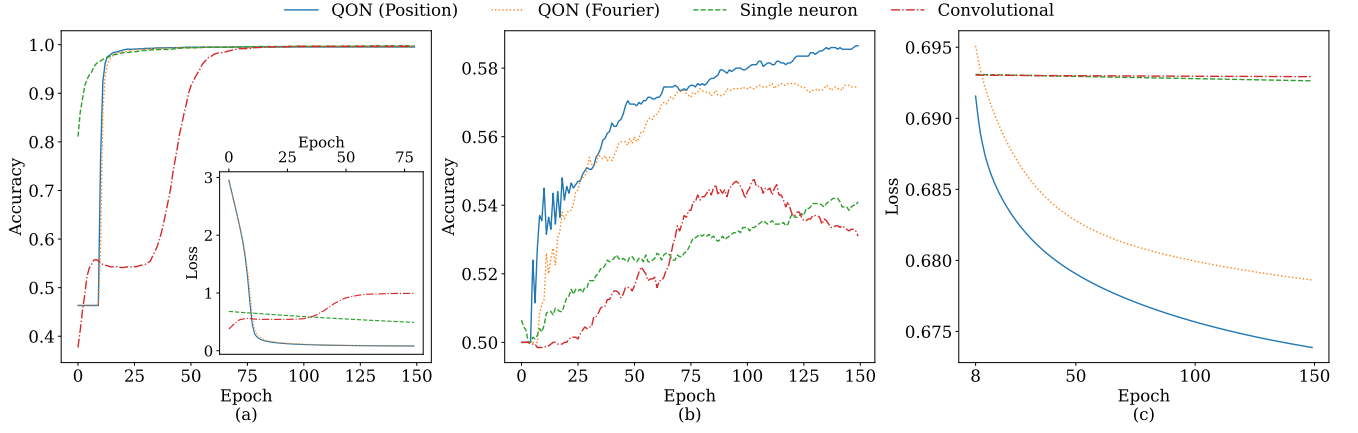
FIG. 3. Comparison between the quantum optical neuron (QON), a single classical neuron and a convolutional network, all trained with the same number of ~ 1024 parameters, optimizer and learning rates. The quantum optical neuron is modelled by an amplitude modulated probe with resolution of 32×32 pixels, both in the spatial and in the Fourier domains. The optimization is performed with learning rates $\eta_\lambda = 0.075$ and $\eta_b = 0.005$. (a) Accuracy versus the number of training epochs for the MNIST dataset. The models are trained to distinguish among images of *zeros* and *ones*, showing compatible results in terms of trainability and accuracy, whose final value is above 99%. The inset is a history plot of the binary cross-entropy, used as loss function in the gradient descent optimization. (b-c) Accuracy and binary cross-entropy plots versus the number of training epochs for the CIFAR-10 dataset. The models are trained to classify images of *cats* and *dogs*. Our method reaches an asymptotic accuracy above 58%, showing an advantage with respect to its classical counterparts.

A. Simulations

We present a simulation of the model introduced above, comparing its performance against those of classical neural network-based techniques, for different datasets. All the simulations are run in Python and TensorFlow [27], and summarized in Fig. 3.

We tested our model using two widely recognized datasets: the MNIST, which contains 28×28 images of handwritten digits from 0 to 9, and the CIFAR-10, comprised of 32×32 color images, distributed across 10 different classes. We guaranteed a fair comparison by increasing the MNIST resolution to 32×32 pixels (separately padding each image of the dataset), while converting the CIFAR-10 to greyscale. We represent each element of the dataset as $(x_j, \ y_j)$, where $y_j \in \{0, 1\}$ is the true class label, and x_j is the input vector, obtained by discretizing and vectorizing either the amplitudes $\mathcal{I}$ or their Fourier spectrum $\hat{\mathcal{I}}$, thus bypassing the simulation of the imaging optics. We adopt the binary cross-entropy as loss function, combined with the standard (non-stochastic) gradient descent optimizer. We use the accuracy, i.e. the proportion of correct predictions over the total ones, as figure of merit of our results.

Our model demonstrates significative performances in both datasets (see Fig. 3). In the MNIST, it achieves accuracy rates exceeding 99%, when discerning between *zeros* and *ones*. In the CIFAR-10, it reaches accuracy above 58%, when distinguishing between *cats* and *dogs*. This difference reflects the complexity of the two classification tasks.

We compared our model against conventional neural network designs with a similar number of parameters. Specifically, we considered a single neuron and a convolutional neural network, commonly employed in pattern recognition tasks [2, 28, 29]. Adopting the TensorFlow notation, the convolutional structure is: Conv2D (10, 3×3) $\rightarrow$ Conv2D (4, 2×2) $\rightarrow$ MaxPooling2D (2×2). Roughly, all the architectures have $\sim 10^3$ trainable parameters. All the models equally perform in the MNIST dataset, both in terms of trainability and final accuracy. When applied to the CIFAR-10 dataset, our classifier outperforms the conventional ones, showing superior efficiency under a strongly-constrained parameters count. All the findings emphasize the competitive accuracy of our method, and also its comparative advantage in pattern recognition tasks with a limited number of parameters.

IV. CONCLUSIONS

In summary, we introduced an interferometric setup of a quantum optical classifier, with the Hong-Ou-Mandel effect as cornerstone of our classification method. We demonstrated the mathematical relation between our model and a classical neuron, constrained to unit depth, showing their similarity in terms of structure and response function. Our design is completely optical and single-photon based: it provides a superexponential speedup with respect to its classical counterpart, in terms of number of photons and computational resources. After modelling the classifier in terms of a spatial light modulator, we numerically compared our performances against

those of standard neural network architectures, showing compatible to superior capabilities in terms of accuracy and training convergence, under the same number of parameters and depending on the pattern complexity.

DATA AVAILABILITY

The underlying code that generated the data for this study is openly available in GitHub [30].

APPENDIX A: SINGLE-PHOTON ENCODING

In this section, we consider the single-photon state obtained at the output of the left branch of the Hong-Ou-Mandel apparatus, providing a detailed discussion of Eq. (1). We adopt units in which $c = 1$.

Consider a generic single-photon state, generated by a monochromatic source with longitudinal position z

$$|\Psi\rangle = \int \mathrm{d}^3 k \ \hat{\Phi}(\mathbf{k}) a^\dagger(\mathbf{k}) |0\rangle \ , \tag{A1}$$

with momentum spectrum Φ and $\mathbf{k} = (k_x, k_y, k_z)$. We neglect the polarization of the photon and consider the single-frequency-mode assumption [31], i.e. we assume that the wavefront propagates along definite-sign z-directions only. Then, $k = (k_x, k_y)$ represents the only independent degrees of freedom of the single-photon state, which reads

$$|\Psi\rangle = \int \mathrm{d}^2 k \ \hat{\phi}_\omega(k) a_\omega^\dagger(k) |0\rangle \ , \tag{A2}$$

$$= \int_S \mathrm{d}^2 r \ \phi_\omega(r) a_\omega^\dagger(r) |0\rangle \ , \tag{A3}$$

where $\hat{\phi}_\omega(k) = \hat{\Phi}\left(k_x, k_y, \sqrt{\omega^2 - k_x^2 - k_y^2}\right)$ and $r = (r_x, r_y)$ labels the transverse coordinates on the source plane S.

For simplicity, we assume that the source is placed at the longitudinal origin $z = 0$. Consider an object with two dimensional shape $\mathcal{O}$, placed at longitudinal position z_o. After free-space propagation occurs, the single-photon spectrum undergoes spatial amplitude modulation [32], that is $\Psi_\mathcal{O}(r) = \mathcal{O}(r)\Psi(r)_\rightarrow$, with $\Psi(r)_\rightarrow$ the spatial input wavefront on the object plane O. Namely

$$|\Psi_\mathcal{O}\rangle = \int_O \mathrm{d}^2 r \ [\phi_\omega * \mathfrak{H}_{z_o}](r)\mathcal{O}(r) a_\omega^\dagger(r) |0\rangle \ , \tag{A4}$$

where $\mathfrak{H}_{z_o}$ denotes the free-space transfer function between the S and O planes. Using twice the convolution theorem, it follows that

$$|\Psi_\mathcal{O}\rangle = \int \mathrm{d}^2 k \ [(\hat{\phi}_\omega \mathfrak{H}_{z_o}) * \hat{\mathcal{O}}](k) a_\omega^\dagger(k) |0\rangle \ . \tag{A5}$$

Consider a linear optical system with transfer function $\mathcal{L}$, with image plane at longitudinal position z_i. By applying again the convolution theorem to $\mathcal{I}_\omega(\cdot|\mathcal{O}) = ((\phi_\omega * \mathfrak{H}_{z_o})\mathcal{O}) * \mathcal{L}_{z_o - z}$, we obtain

$$|\Psi_\mathcal{I}\rangle = \int \mathrm{d}^2 k \ \hat{\mathcal{I}}_\omega(k|\mathcal{O}) a_\omega^\dagger(k) |0\rangle \ , \tag{A6}$$

with $\hat{\mathcal{I}}_\omega(k|\mathcal{O}) = [(\hat{\phi}_\omega \mathfrak{H}_{z_o}) * \hat{\mathcal{O}}]\hat{\mathcal{L}}_d$. Notice that $\mathcal{I}_\omega(r|\mathcal{O})$ describes the image formed on a screen placed at distance d from the object.

APPENDIX B: HONG-OU-MANDEL COINCIDENCES

In this section, we compute the rate of coincidences at the output of the Hong-Ou-Mandel interferometer of Fig. 1, with left and top branch states given by Eqs. (1) and (2), respectively. We write the input-probe bipartite state as

$$|\Psi_{\mathcal{I}}\rangle \otimes |\Psi_{\mathcal{U}}\rangle = \int d^2 k_1 \, d^2 k_2 \; \hat{\Psi}(k_1, k_2) a^\dagger(k_1) b^\dagger(k_2) |0\rangle \;, \tag{B1}$$

with $\hat{\Psi}(k_1, k_2) = \hat{\mathcal{I}}(k_1|\mathcal{O})\hat{\mathcal{U}}(k_2|\lambda)$, where we dropped the ω subscript for simplicity. The $50{:}50$ beam splitter acts as the unitary operation [33]

$$\begin{cases} a^\dagger \to \frac{1}{\sqrt{2}}\left(a^\dagger + b^\dagger\right) \\ b^\dagger \to \frac{1}{\sqrt{2}}\left(a^\dagger - b^\dagger\right) \end{cases}, \tag{B2}$$

yielding

$$|\Psi_{\mathcal{I}}\rangle \otimes |\Psi_{\mathcal{U}}\rangle \to |\Phi\rangle = \frac{1}{2}\int d^2 k_1 \, d^2 k_2 \; \hat{\Psi}(k_1, k_2) \left[a^\dagger(k_1) + b^\dagger(k_1)\right]\left[a^\dagger(k_2) - b^\dagger(k_2)\right]|0\rangle \;. \tag{B3}$$

Detection of mode $m \in \{a, b\}$ is described by the projector $\Pi_m = \int d^2 k \; m^\dagger(k)|0\rangle\langle 0| m(k)$. The rate of coincidences, i.e. the probability that one and only one photon is detected in each mode, reads

$$p(1_a \cap 1_b) = \text{Tr}[|\Phi\rangle\langle\Phi| \, \Pi_a \otimes \Pi_b] \;, \tag{B4}$$

$$\text{with } \Pi_a \otimes \Pi_b = \int d^2 k_3 \, d^2 k_4 \; a^\dagger(k_3) b^\dagger(k_4) |0\rangle\langle 0| a(k_3) b(k_4) \;. \tag{B5}$$

By substitution of Eq. (B3), we get

$$p(1_a \cap 1_b) = \frac{1}{4}\int \prod_{i=1}^{6} d^2 k_i \; \hat{\Psi}(k_1, k_2)\hat{\Psi}^*(k_5, k_6) W_1(k_1, k_2, k_3, k_4) W_2(k_3, k_4, k_5, k_6) \;, \tag{B6}$$

where

$$\begin{aligned} W_1(k_1, k_2, k_3, k_4) &= \langle 0| a(k_3) b(k_4) \left[a^\dagger(k_1) a^\dagger(k_2) - a^\dagger(k_1) b^\dagger(k_2) + b^\dagger(k_1) a^\dagger(k_2) - b^\dagger(k_1) b^\dagger(k_2)\right]|0\rangle \\ &= \delta(k_2 - k_3)\delta(k_1 - k_4) - \delta(k_1 - k_3)\delta(k_2 - k_4) \;, \end{aligned} \tag{B7}$$

$$\begin{aligned} W_2(k_3, k_4, k_5, k_6) &= \langle 0| \left[a(k_6) a(k_5) - b(k_6) a(k_5) + a(k_6) b(k_5) - b(k_6) b(k_5)\right] a^\dagger(k_3) b^\dagger(k_4) |0\rangle \\ &= \delta(k_3 - k_6)\delta(k_4 - k_5) - \delta(k_3 - k_5)\delta(k_4 - k_6) \;. \end{aligned} \tag{B8}$$

By integrating out the Dirac deltas in Eq. (B6), we obtain

$$p(1_a \cap 1_b) = \frac{1}{2}\int d^2 k_1 \, d^2 k_2 \, d^2 k_5 \, d^2 k_6 \; \hat{\Psi}(k_1, k_2)\hat{\Psi}^*(k_5, k_6) \left[\delta(k_1 - k_5)\delta(k_2 - k_6) - \delta(k_1 - k_6)\delta(k_2 - k_5)\right] \;. \tag{B9}$$

Finally, the rate of coincidences reads

$$p(1_a \cap 1_b|\lambda, \mathcal{O}) = \frac{1}{2}\int d^2 k_1 \; |\hat{\mathcal{I}}(k_1|\mathcal{O})|^2 \int d^2 k_2 \; |\hat{\mathcal{U}}(k_2|\lambda)|^2 - \frac{1}{2}\left|\int d^2 k \; \hat{\mathcal{I}}(k|\mathcal{O})\hat{\mathcal{U}}^*(k|\lambda)\right|^2 \;. \tag{B10}$$

More compactly,

$$p(1_a \cap 1_b|\lambda, \mathcal{O}) = \frac{1}{2}\left[||\mathcal{I}_\omega(\cdot|\mathcal{O})||^2 ||\mathcal{U}_\omega(\cdot|\lambda)||^2 - |\langle\mathcal{I}_\omega(\cdot|\mathcal{O}), \mathcal{U}_\omega(\cdot|\lambda)\rangle|^2\right] \;, \tag{B11}$$

with $||\cdot||$ and $\langle\cdot,\cdot\rangle$ denoting the L^2-norm and inner product, which is precisely the results of Eq. (3).

APPENDIX C: TRAINING

In this section, we discuss how to train the Hong-Ou-Mandel interferometer as a binary classifier. We separately feed each element of the training set (an ensemble of objects with known labels) into the input branch of Fig. 1, comparing the predicted classes with the target ones. We optimize the probe parameters λ by means of the gradient descent algorithm, and using the binary cross-entropy as loss function.

Consider a training set made of M objects $\{\mathcal{O}_j\}$, each associated to a binary target label $y_j \in \{0, 1\}$, with $0 \leq j \leq M - 1$. We denote $f_\lambda^{(j)} = f_\lambda(\mathcal{O}_j)$ our model prediction. After feeding $\mathcal{O}_j$ into the input branch of the interferometer

$$f_\lambda^{(j)} = C - 2p(1_a \cap 1_b | \lambda, \mathcal{O}_j) \ , \tag{C1}$$

$$F_{b\lambda}^{(j)} = \sigma(f_\lambda^{(j)} + b) \ , \tag{C2}$$

where $p \in [0, 1/2]$. For simplicity, we assumed that the losses are independent on both the input and the probe, that is $C := \alpha_\lambda(\mathcal{O}_j) \ \forall \lambda, j$.

Given a sample object, the binary cross-entropy between the target label and the predicted one reads

$$H\left(y_j, F_{b\lambda}^{(j)}\right) = -y_j \log\left(F_{b\lambda}^{(j)}\right) - (1 - y_j) \log\left(1 - F_{b\lambda}^{(j)}\right) \ . \tag{C3}$$

We optimize the probe parameters by means of the gradient descent algorithm, where the binary cross-entropy, averaged on the training set, is used as loss function. Namely

$$\lambda \to \lambda - \frac{\eta_\lambda}{M} \sum_{j=0}^{M-1} \partial_\lambda H\left(y_j, F_{b\lambda}^{(j)}\right) \ , \tag{C4}$$

$$b \to b - \frac{\eta_b}{M} \sum_{j=0}^{M-1} \partial_b H\left(y_j, F_{b\lambda}^{(j)}\right) \ , \tag{C5}$$

with η_λ, η_b the learning rates of the probe and bias parameters, respectively. The derivatives with respect to the parameters and the bias yield

$$\partial_\lambda H = (\partial_F H)(\partial_\xi \sigma)\partial_\lambda f \ , \tag{C6}$$

$$\partial_b H = (\partial_F H)\partial_\xi \sigma \ , \tag{C7}$$

with $\xi_{b\lambda} = f_\lambda + b$. Then,

$$\partial_F H = \frac{F - y}{F(1 - F)} \ , \tag{C8}$$

$$\partial_\xi \sigma = \beta F(1 - F) \ , \tag{C9}$$

with β the hyperparameter of Eq. (6). For any complex function of real variable $h : \mathbb{R} \to \mathbb{C}$, it follows that $\partial_\lambda |h(\lambda)| = \mathrm{Re}\left[h(\lambda)(\partial_\lambda h(\lambda))^*\right] / |h(\lambda)|$. Hence,

$$\partial_\lambda f = 2\,\mathrm{Re}\left[\langle \mathcal{I}_\omega, \mathcal{U}_\omega \rangle \langle \mathcal{I}_\omega, \partial_\lambda \mathcal{U}_\omega \rangle^*\right] \ . \tag{C10}$$

Neglecting the phase of $\langle \mathcal{I}, \mathcal{U} \rangle$,

$$\partial_\lambda f \simeq 2\sqrt{f}\,\mathrm{Re}\left[\langle \mathcal{I}_\omega, \partial_\lambda \mathcal{U}_\omega \rangle\right] \ . \tag{C11}$$

This assumption, which we verified in our simulations under a self-consistency test, simplifies the computation of the first factor of Eq. (C10), which is directly determined at the output of the Hong-Ou-Mandel interferometer.

APPENDIX D: CLASSIFICATION IN THE FOURIER DOMAIN

In this section, we discuss the effect of adding a single lens in the probe branch of the Hong-Ou-Mandel interferometer, as shown in Fig. 1. We summarize the main calculations, which closely follow that of Section II.

A thin lens is placed at one focal length ℓ from both the probe image plane and the beam splitter. In the near-field limit, the lens performs a Fourier transform of the probe state [31], yielding $|\Psi_{\mathcal{U}}\rangle \to |\Psi_{\mathcal{U}'}\rangle$, where

$$\mathcal{U}'_\omega(r|\lambda) = -i\frac{\omega}{\ell}e^{2i\omega f}\hat{\mathcal{U}}_\omega\left(\frac{\omega}{\ell}r|\lambda\right) . \tag{D1}$$

After the beam splitter, the rate of coincidences is

$$p(1_a \cap 1_b|\lambda, \mathcal{O}) = \frac{1}{2}\left[\alpha_\omega(\mathcal{O}) - \widetilde{f}_\lambda(\mathcal{O})\right] , \tag{D2}$$

$$\widetilde{f}_\lambda(\mathcal{O}) = \left|\langle\mathcal{I}_\omega(\cdot|\mathcal{O}), \hat{\mathcal{U}}_\omega(\cdot|\lambda)\rangle\right|^2 , \tag{D3}$$

yielding

$$\widetilde{F}_{b\lambda}(\mathcal{O}) = \sigma(\widetilde{f}_\lambda(\mathcal{O}) + b) , \tag{D4}$$

$$\widetilde{f}_\lambda(\mathcal{O}) = \left|\int_I dr\, dr'\, \mathcal{I}_\omega(r|\mathcal{O})\mathcal{U}_\omega^*(r'|\lambda)e^{ir\cdot r'}\right|^2 , \tag{D5}$$

with σ and b the sigmoid activation function and bias, already introduced in Eq. (7). In contrast to Eq. (5), $\widetilde{f}_\lambda(\mathcal{O})$ is not a point-wise evaluation: it combines the image spatial modes with the momentum spectrum of the probe state. Using the duality of the Fourier transform, it follows that

$$\widetilde{f}_\lambda(\mathcal{O}) = \left|\langle\hat{\mathcal{I}}_\omega(\cdot|\mathcal{O}), \mathcal{U}_\omega(\cdot|\lambda)\rangle\right|^2 , \tag{D6}$$

which corresponds to the output of the same scheme of Fig. 1, but with the thin lens placed in the left branch, before the beam splitter. Equivalently, this takes the Fourier transform of the image, instead of that of the probe. In the next section, we leverage this symmetry to simplify both the training process and the numerical simulations.

The training of the model follows the same procedure of Appendix C. By placing the lens on the top branch of the interferometer, while using the duality of the Fourier transform, we get

$$\partial_\lambda\widetilde{f} \simeq 2\sqrt{\widetilde{f}}\,\mathrm{Re}\left[\langle\hat{\mathcal{I}}_\omega, \partial_\lambda\mathcal{U}_\omega\rangle\right] . \tag{D7}$$

Under the same conditions of Section III and Appendix C, the last two equations become

$$\widetilde{f}_\lambda(\mathcal{O}) = \left|\sum_{\mu,\nu}(u \star \hat{\mathcal{I}}_\omega^*)(r_{\mu\nu})\frac{\lambda_{\mu\nu}}{||\lambda||}\right|^2 , \tag{D8}$$

$$\partial_{\mu\nu}\widetilde{f} \simeq 2\frac{\sqrt{\widetilde{f}}}{||\lambda||}\,\mathrm{Re}\left[(u \star \hat{\mathcal{I}}_\omega^*)(r_{\mu\nu}) - \sqrt{\widetilde{f}}\frac{\lambda_{\mu\nu}}{||\lambda||}\right] , \tag{D9}$$

where in the last step we neglected the phase of $\langle\hat{\mathcal{I}},\mathcal{U}\rangle$. Similarly to Eq. (13), Eq. (D9) can be evaluated in an all-optical way through the characterization of the real part of $\hat{\mathcal{I}}$, namely, by performing an amplitude and phase measurement at the output of a thin lens, placed in the left branch, before the beam splitter. In Fig. 3, we compare the predictability of the neuron in the spatial and Fourier domains.

APPENDIX E: OPTICAL AND COMPUTATIONAL ADVANTAGE

In this section, we discuss the optical and computational advantage as the number of photons and operations required by a single image classification. Assuming that all the parameters have been previously trained with optimal accuracy, we show that our protocol requires a constant number of resources, i.e. $\mathcal{O}(1)$ complexity, independently of the input image resolution: it provides a superexponential speedup over its classical counterpart.

We first discuss the computational advantage when substituting a classical neuron with a quantum optical one. From now on, we denote Ω, Θ and $\mathcal{O}$, respectively the lower, tight and upper bounds on the number of resources needed by a certain (optical or computational) operation. Consider a digital image x of N pixels, fed into a neuron

$$G_{bw}(x) = \sigma(w \cdot x + b) , \tag{E1}$$

where $x, w \in \mathbb{R}^N$, $b \in \mathbb{R}$ and σ is the sigmoid activation of Eq. (6), with hyperparameters $\beta = 1$ and $\gamma = 0$. Eq. (E1) costs N operations to compute $w \cdot x$. The Hong-Ou-Mandel interferometer performs the same operation in an all-optical way, leaving the computational cost of the activation function and bias only, which is $\mathcal{O}(1)$.

We now discuss the optical advantage when using coincidences to classify single-photon states instead of a classical neuron on fully reconstructed images. After targeting an object with light, a digital image x is an ensemble of grey levels obtained by counting the number of photons collected by different pixels on a sensor grid, e.g. a charge-coupled device [34]. Let n_p be the average number of photons in the input state, and μ_i the average number of photons collected by the i-th pixel of the grid, with $i \in \{0, \ldots, N\}$. Assuming perfect quantum efficiency and sufficiently low exposure times to neglect the saturation of the sensor, the grey values at each pixel read

$$x_i = \frac{\mu_i L}{\mu_w} \ , \tag{E2}$$

with L the number of grey levels, i.e. the depth of the image, and $\mu_w = \max_i \mu_i$ the maximum number of photons collected in a single pixel. Indeed, $x_i \in \{0, 1, \ldots, L-1\}$ with 0 and L labelling the *black* and *white* colors, respectively. Each pixel has variance $\varsigma_i^2 = \Delta\mu_i^2 L^2 / \mu_w^2$, with $\Delta\mu_i^2$ the variance on the number of collected photons. For coherent light, the photo-detection process undergoes the standard quantum limit (SQL) [35, 36], with Poissonian fluctuations that satisfy $\Delta\mu_i^2 \simeq \mu_i$. The average uncertainty reads

$$\varsigma^2 := \frac{1}{N} \sum_{i=0}^{N} \varsigma_i^2 \overset{\text{SQL}}{\simeq} \langle x \rangle^2 N n_p^{-1} \ , \tag{E3}$$

with $\langle x \rangle = N^{-1} \sum_i x_i \in [0, L]$ the average brightness of the image, which we assume to be independent of its resolution. Hence, the number of photons n_p required by a full image reconstruction with *average* variance ς^2 is $\Theta(\varsigma^{-2}\langle x \rangle N)$. This is the cost of image reconstruction only. We now take into account the information propagation through the neuron of Eq. (E1).

Proposition 1. *Consider a neuron with sigmoid activation function. Suppose that there exists a sequence of parameters $\{(w_N, b_N) \in \mathbb{R}^{N+1}\}_{N \gg 1}$ that optimally solve the N-pixel image classification task, with b_N and the ℓ^1-norm $||w_N||_1$ asymptotically bounded for $N \to \infty$. Then, the number of photons n_p required to classify an image x with uncertainty ε, is $\Omega\left(\varepsilon^{-2}\langle x \rangle N\right)$.*

Proof. Consider the output of the neuron $G_{bw}(x) = \sigma(w_N \cdot x + b_N)$, and its derivative $\partial G(x) = G_{bw}(x)(1 - G_{bw}(x))$. By neglecting the spatial neighbourhood correlations, which may introduce at most a constant overhead in our estimation, we propagate the uncertainty of x as

$$\varepsilon^2 = \langle x \rangle (\partial G)^2(x) \sum_{i=0}^{N-1} (w_N)_i^2 x_i N \tilde{n}_p^{-1} \ , \tag{E4}$$

where $\tilde{n}_p = n_r n_p$, with n_r is the number of independent image acquisition and classification. Since black pixels do not contribute to this summation, we get

$$\sum_{i=0}^{N-1} (w_N)_i^2 x_i \geq \sum_{i \notin \mathcal{B}} (w_N)_i^2 = ||w_N||^2 - \sum_{i \in \mathcal{B}} (w_N)_i^2 \ , \tag{E5}$$

with $\mathcal{B} = \{i \in \mathbb{N} \mid x_i = 0 \text{ for } 0 \leq i \leq N-1\}$ the set of black pixels labels. However, $||w_N||^2 \gg \sum_{i \in \mathcal{B}} (w_N)_i^2$. Otherwise, $||w_N||^2 \simeq \sum_{i \in \mathcal{B}} (w_N)_i^2$ would imply either that the image is mostly black, independently of its resolution, or that $(w_N)_i \simeq 0$ for all non-black pixels, which are both conditions that prevent the learnability of the neuron. By substitution into Eq. (E4) we get

$$\tilde{n}_p \geq \varepsilon^{-2} \langle x \rangle (\partial G)^2(x) ||w_N||^2 N \ . \tag{E6}$$

Since w_N is a sequence of non-trivial solutions of the classification problem, the ℓ^2-norm $||w_N||^2$ cannot go to zero for $N \to \infty$. Finally, we show that $(\partial G(x))^2$ does not converge to 0 for $N \to \infty$. Consider

$$(\partial G)^2(x) = \frac{e^{-2(w_N \cdot x + b_N)}}{[1 + e^{-(w_N \cdot x + b_N)}]^4} \ . \tag{E7}$$

If b_N is asymptotically limited, $(\partial G)^2$ converges to zero if and only if $w_N \cdot x \to \pm\infty$. By splitting this scalar product into positive and negative contributions $w_N \cdot x = \sum_{(w_N)_i > 0}(w_N)_i x_i - \sum_{(w_N)_i < 0}|(w_N)_i|x_i$, it follows that

$$w_N \cdot x \leq \sum_{(w_N)_i > 0}(w_N)_i x_i \leq L||w_N||_1 \ , \tag{E8}$$

$$w_N \cdot x \geq - \sum_{(w_N)_i < 0}|(w_N)_i|x_i \geq -L||w_N||_1 \ , \tag{E9}$$

namely that $|w_N \cdot x| \leq L||w_N||_1$. Since the ℓ^1-norm is limited, $(\partial G)^2$ admits strictly positive lower bound for $N \to \infty$. Finally, this imply that $\tilde{n}_p = \Omega\left(\varepsilon^{-2}\langle x\rangle N\right)$. □

In the previous discussion, two conditions lead to the above lower bound. On the one hand, that $||w_N||^2 \not\to 0$ for $N \to \infty$, which is essential to guarantee that the neuron is trainable at any resolution. On the other hand, that $||w_N||_1$ is bounded for $N \to \infty$, which is compatible with LASSO and Tikhonov's regularization techniques [37, 38].

We show that our protocol exponentially reduces this cost, requiring only the estimation of the rate of coincidences of the Hong-Ou-Mandel interferometer of Fig. 1. Let $\tilde{n}_p = 2n_p$ be the number input photons, and $\tilde{p} \in [0, 1/2]$ the empirical rate of coincidences. Under the normal approximation, with the 95% confidence level [39], the estimation uncertainty reads

$$\varepsilon = 2\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}_p}} \ . \tag{E10}$$

Since $4\tilde{p}(1-\tilde{p}) \leq 1$, the total number of photons is $\mathcal{O}(\varepsilon^{-2})$, which is constant with respect to the resolution of the image.

In conclusion, the quantum optical neuron provides a superexponential advantage over its classical counterpart, both in the number of operations and photons saved to classify a single image. We summarize these results in Table I.

[1] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE **86**, 2278 (1998).

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet classification with deep convolutional neural networks, Commun. ACM **60**, 84–90 (2017).

[3] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in *IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '16, p. 770.

[4] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, An image is worth 16x16 words: Transformers for image recognition at scale, in *International Conference on Learning Representations*, ICLR '21, arXiv:2010.11929 [cs.CV].

[5] B. J. Shastri, A. N. Tait, T. Ferreira de Lima, W. H. P. Pernice, H. Bhaskaran, C. D. Wright, and P. R. Prucnal, Photonics for artificial intelligence and neuromorphic computing, Nat. Photon. **15**, 102–114 (2021).

[6] X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, All-optical machine learning using diffractive deep neural networks, Science **361**, 1004–1008 (2018).

[7] Y. Zuo, B. Li, Y. Zhao, Y. Jiang, Y.-C. Chen, P. Chen, G.-B. Jo, J. Liu, and S. Du, All-optical neural network with nonlinear activation functions, Optica **6**, 1132 (2019).

[8] S. Colburn, Y. Chu, E. Shilzerman, and A. Majum-

dar, Optical frontend for a convolutional neural network, Appl. Opt. **58**, 3179 (2019).

[9] S. Li, B. Ni, X. Feng, K. Cui, F. Liu, W. Zhang, and Y. Huang, All-optical image identification with programmable matrix transformation, Opt. Express **29**, 26474 (2021).

[10] Y. Luo, Y. Zhao, J. Li, E. Çetintaş, Y. Rivenson, M. Jarrahi, and A. Ozcan, Computational imaging without a computer: seeing through random diffusers at the speed of light, eLight **2**, 4 (2022).

[11] P. L. McMahon, The physics of optical computing, Nat. Rev. Phys. **5**, 717–734 (2023).

[12] F. Benatti, S. Mancini, and S. Mangini, Continuous variable quantum perceptron, Int. J. Quantum Inf. **17**, 1941009 (2019).

[13] F. Tacchino, C. Macchiavello, D. Gerace, and D. Bajoni, An artificial neuron implemented on an actual quantum processor, Npj Quantum Inf. **5**, 26 (2019).

[14] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, Nat. Rev. Phys. **3**, 625–644 (2021).

[15] A. Senokosov, A. Sedykh, A. Sagingalieva, B. Kyriacou, and A. Melnikov, Quantum machine learning for image classification, Mach. Learn.: Sci. Technol. **5**, 015040 (2024).

[16] M. Cerezo, M. Larocca, D. García-Martín, N. L. Diaz, P. Braccia, E. Fontana, M. S. Rudolph, P. Bermejo,

A. Ijaz, S. Thanasilp, E. R. Anschuetz, and Z. Holmes, Does provable absence of barren plateaus imply classical simulability? Or, why we need to rethink variational quantum computing (2024), arXiv:2312.09121 [quant-ph].

[17] G. R. Steinbrecher, J. P. Olson, D. Englund, and J. Carolan, Quantum optical neural networks, Npj Quantum Inf. **5**, 60 (2019).

[18] N. Killoran, T. R. Bromley, J. M. Arrazola, M. Schuld, N. Quesada, and S. Lloyd, Continuous-variable quantum neural networks, Phys. Rev. Res. **1**, 033063 (2019).

[19] X. Sui, Q. Wu, J. Liu, Q. Chen, and G. Gu, A review of optical neural networks, IEEE Access **8**, 70773 (2020).

[20] D. Stanev, N. Spagnolo, and F. Sciarrino, Deterministic optimal quantum cloning via a quantum-optical neural network, Phys. Rev. Res. **5**, 013139 (2023).

[21] C. Wood, S. Shrapnel, and G. J. Milburn, A Kerr kernel quantum learning machine (2024), arXiv:2404.01787 [quant-ph].

[22] C. K. Hong, Z. Y. Ou, and L. Mandel, Measurement of subpicosecond time intervals between two photons by interference, Phys. Rev. Lett. **59**, 2044 (1987).

[23] C. Bowie, S. Shrapnel, and M. J. Kewming, Quantum kernel evaluation via Hong–Ou–Mandel interference, Quantum Sci. Technol. **9**, 015001 (2023).

[24] X. Glorot and Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, PMLR, Vol. 9 (2010) pp. 249–256.

[25] J. Neff, R. Athale, and S. Lee, Two-dimensional spatial light modulators: a tutorial, Proc. IEEE **78**, 826 (1990).

[26] Z. Zhang, Z. You, and D. Chu, Fundamentals of phase-only liquid crystal on silicon (LCOS) devices, Light Sci. Appl. **3**, e213 (2014).

[27] TensorFlow Developers, TensorFlow (2023).

[28] R. Collobert and S. Bengio, Links between perceptrons, MLPs and SVMs, in *Proceedings of the Twenty-First International Conference on Machine Learning*, ICML '04, p. 23.

[29] F. Rosenblatt, The perceptron: A probabilistic model for information storage and organization in the brain, Psychol. Rev. **65**, 386 (1958).

[30] A. R. Morgillo and S. Roncallo, `https://github.com/simoneroncallo/quantum-optical-neuron`.

[31] M. Rezai and J. A. Salehi, Fundamentals of quantum Fourier optics, IEEE Trans. Quantum Eng. **4**, 1 (2022).

[32] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics* (Wiley, 1991).

[33] A. M. Brańczyk, Hong-Ou-Mandel interference (2017), arXiv:1711.00080 [quant-ph].

[34] W. S. Boyle and G. E. Smith, Charge coupled semiconductor devices, Bell Syst. Tech. J. **49**, 587 (1970).

[35] M. I. Kolobov, The spatial behavior of nonclassical light, Rev. Mod. Phys. **71**, 1539 (1999).

[36] V. Delaubert, N. Treps, C. Fabre, H. A. Bachor, and P. Réfrégier, Quantum limits in image processing, EPL **81**, 44001 (2008).

[37] F. Santosa and W. W. Symes, Linear inversion of band-limited reflection seismograms, SIAM J. Sci. Stat. Comput. **7**, 1307–1330 (1986).

[38] A. Tikhonov and V. Glasko, Use of the regularization method in non-linear problems, USSR Comput. Math. Math. Phys. **5**, 93 (1965).

[39] A. Rotondi, P. Pedroni, and A. Pievatolo, *Probability, Statistics and Simulation: With Application Programs Written in R* (Springer, 2022).