

Identification of Light Sources using Machine Learning

Chenglong You,^{1,*} Mario A. Quiroz-Juárez,² Aidan Lambert,¹ Narayan Bhusal,¹ Chao Dong,¹ Armando Perez-Leija,³ Amir Javadi,¹ Roberto de J. León-Montiel,² and Omar S. Magaña-Loaiza¹

¹*Quantum Photonics Laboratory, Department of Physics and Astronomy,
Louisiana State University, Baton Rouge, LA 70803, USA*

²*Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México,
Apartado Postal 70-543, 04510 Cd. Mx., México*

³*Max-Born-Institut, Max-Born-Straße 2A, 12489 Berlin, Germany*

The identification of light sources represents a task of utmost importance for the development of multiple photonic technologies. Over the last decades, the identification of light sources as diverse as sunlight, laser radiation and molecule fluorescence has relied on the collection of photon statistics or the implementation of quantum state tomography. In general, this task requires an extensive number of measurements to unveil the characteristic statistical fluctuations and correlation properties of light, particularly in the low-photon flux regime. In this article, we exploit the self-learning features of artificial neural networks and naive Bayes classifier to dramatically reduce the number of measurements required to discriminate thermal light from coherent light at the single-photon level. We demonstrate robust light identification with tens of measurements at mean photon numbers below one. Our work demonstrates an improvement in terms of the number of measurements of several orders of magnitude with respect to conventional schemes for characterization of light sources. Our work has important implications for multiple photonic technologies such as LIDAR and microscopy.

INTRODUCTION

The underlying statistical fluctuations of the electromagnetic field have been widely utilized to identify diverse sources of light [1, 2]. In this regard, the Mandel parameter constitutes an important metric to characterize the excitation mode of the electromagnetic field and consequently to classify light sources [3]. Similarly, the degree of optical coherence has also been extensively utilized to identify light sources [3–6]. Despite the fundamental importance of these quantities, they require large amounts of data which impose practical limitations [6–10]. This problem has been partially alleviated by incorporating statistical methods, such as bootstrapping, to predict unlikely events that are hard to measure experimentally [6, 8–10]. Unfortunately, the constraints of these methods severely impact the realistic implementation of photonic technologies for metrology, imaging, remote sensing and microscopy [10–15].

The potential of machine learning has motivated novel families of technologies that exploit self-learning and self-evolving features of artificial neural networks to solve a large variety of problems in different branches of science [16, 17]. Conversely, quantum mechanical systems have provided new mechanisms to achieve quantum speedup in machine learning [17, 18]. In the context of quantum optics, there has been an enormous interest in utilizing machine learning to optimize quantum resources in optical systems [19–21]. As a tool to characterize quantum systems, machine learning has been successfully employed to reduce the number of measurements required to perform quantum state discrimination, quantum separability and quantum state tomography [22–24].

In this article, we demonstrate the potential of ma-

chine learning to perform discrimination of light sources at extremely low-light levels. This is achieved by training single artificial neurons with the statistical fluctuations that characterize coherent and thermal states of light. The self-learning features of artificial neurons enable the dramatic reduction in the number of measurements and the number of photons required to perform identification of light sources. For the first time, our experimental results demonstrate the possibility of using less than ten measurements to identify light sources with mean photon numbers below one. In addition, we demonstrate similar experimental results using the naive Bayes classifier, which are outperformed by our single neuron approach. Finally, we present a discussion on how a single artificial neuron based on an ADaptive LINear Element (ADALINE) model can dramatically reduce, by several orders of magnitude, the number of measurements required to discriminate signal photons from ambient photons. This possibility has strong implications for realistic implementation of LIDAR, remote sensing and microscopy.

EXPERIMENTAL SETUP AND MODEL

As shown in Fig. 1 (a), we utilize a continuous-wave (CW) laser beam that is divided by a 50:50 beam splitter. The transmitted beam is focused onto a rotating ground glass which is used to generate pseudo-thermal light with super-Poissonian statistics. The beam emerging from the ground glass is collimated using a lens and attenuated by neutral-density (ND) filters to mean photon numbers below one. The attenuated beam is then coupled into a single-mode fiber (SMF). The fiber directs photons to a superconducting nanowire single-photon detec-

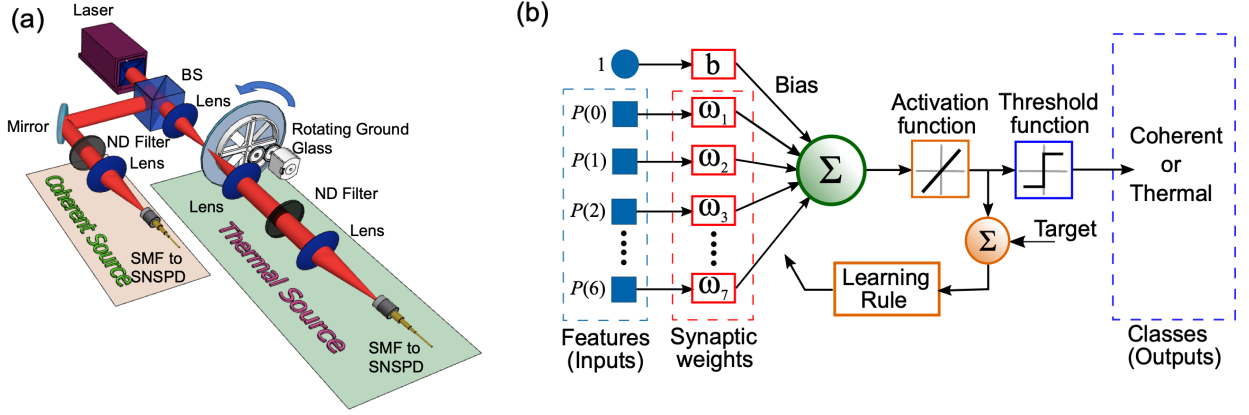


FIG. 1. (a) Schematic representation of the experimental setup. A laser beam is divided by a beam splitter (BS); the two replicas of the beam are used to generate light with Poissonian (coherent) and super-Poissonian (thermal) statistics. The thermal beam of light is generated by a rotating ground glass. Neutral density (ND) filters are utilized to attenuate light to the single-photon level. Coherent and thermal light beams are measured by superconducting nanowire single-photon detectors (SNSPDs). (b) Flow diagram of the ADALINE neuron used for demonstration of light source identification. Additional details are discussed in the body of the article.

tor (SNSPD). The beam reflected by the beam splitter is used as a source of coherent light. This beam, characterized by Poissonian statistics, is also attenuated, coupled into a SMF and detected by a SNSPD. The brightness of the coherent beam is matched to that of the pseudo-thermal beam of light.

In order to perform photon counting from our SNSPDs, we use the surjective photon counting method described in ref. [25]. The TTL pulses produced by our SNSPDs were detected and recorded by an oscilloscope. The data was divided in time bins of 1 μ s, which corresponds to the coherence time of our CW laser. Voltage peaks above ~ 0.5 V were considered as one photon event. The number of photons (voltage peaks) in each time bin was counted to retrieve photon statistics. These events were then used for training and testing our ADALINE neuron and naive Bayes classifier.

The probability of finding n photons in coherent light is given by $P_{\text{coh}}(n) = e^{-\bar{n}}(\bar{n}^n/n!)$, where $\bar{n}$ denotes the mean photon number of the beam. Furthermore, the photon statistics of thermal light is given by $P_{\text{th}}(n) = \bar{n}^n/(\bar{n}+1)^{n+1}$. It is worth noting that the photon statistics of thermal light is characterized by random intensity fluctuations with a variance greater than the mean number of photons in the mode. For coherent light, the maximum photon-number probability sits around $\bar{n}$. For thermal light, the maximum is always at vacuum. However, when the mean photon number is low, the photon number distribution for both kinds of light becomes similar. Consequently, it becomes extremely difficult to discriminate one source from the other. The conventional approach to discriminate light sources makes use of histograms generated through the collection of millions of measurements [7, 9, 26, 27]. Unfortunately, this method

is not only time consuming, but also imposes practical limitations.

In order to dramatically reduce the number of measurements required to identify light sources, we make use of an ADALINE neuron. ADALINE is a single neural network model based on a linear processing element, proposed by Bernard Widrow [28], for binary classification. In general, the neural networks undergo two stages: training and test. In the training stage, ADALINE is capable of learning the correct outputs (named as output labels or classes) from a set of inputs, so-called features, by using a supervised learning algorithm. In the test stage, this neuron produces the outputs of a set of inputs that were not in the training data, taking as reference the acquired experience in the training stage. Although we tested architectures far more complex than a single neuron for the identification of light sources, we found that a simple ADALINE offers a perfect balance between accuracy and simplicity (for more details, see the Supplementary Information). The structure of the ADALINE model is shown in Fig. 1(b). The neuron input features are denoted by $P(n)$, which corresponds to the probability of detecting n photons, in a single measurement event, for a given light source, namely coherent or thermal. Furthermore, the parameters ω_i are the synaptic weights and b is a bias term. In the training period, these parameters are optimized through the learning rule by using the error between the target output and neuron's output as reference. For the binary classification (coherent or thermal), the neuron's output is fed into the identity activation function, and subsequently to the threshold function.

To train the ADALINE, we make use of the so-called delta learning rule [29], in combination with a database of experimentally measured photon-number dis-

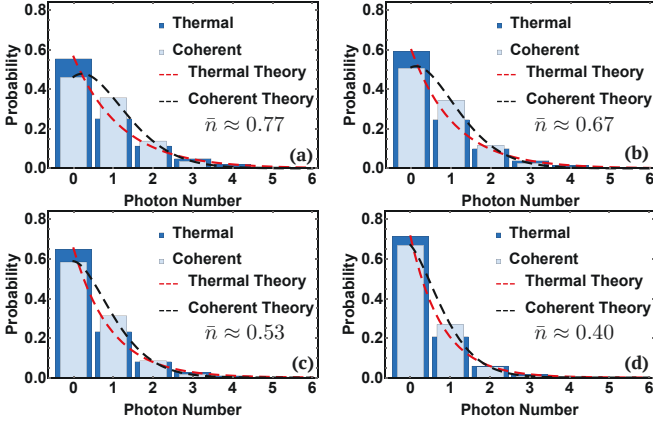


FIG. 2. A set of histograms displaying theoretical and experimental photon number probability distributions for coherent and thermal light beams with different mean photon numbers. Our experimental results are in excellent agreement with theory. The photon number distributions illustrate the difficulty in discriminating light sources at low-light levels even when large sets of data are available.

tributions, considering different mean photon numbers: $\bar{n} = 0.44, 0.53, 0.67, 0.77$. The database for each mean photon number was divided into subsets comprising 10, 20, ..., 150, 160 data points. The ADALINE neurons are thus prepared by using one thousand of those subsets, where 70% are devoted to training and 30% to testing. In all cases, the training was stopped after 50 epochs.

We have established the baseline performance for our ADALINE neuron by using naive Bayes classifier. This is a simple classifier based on Bayes' theorem [31]. Throughout this article, we assume that each measurement is independent. Moreover, we represent the measurement of the photon number sequence as a vector $\mathbf{x} = (x_1, \dots, x_k)$. Then, the probability of this sequence generated from coherent or thermal light is given by $p(C_j | x_1, \dots, x_k)$, where C_j could denote either coherent or thermal light. Using Bayes' theorem, the conditional probability can be decomposed as $p(C_j | \mathbf{x}) = p(C_j)p(\mathbf{x} | C_j)/p(\mathbf{x})$. By using the chain rule for conditional probability, we have $p(C_k | x_1, \dots, x_k) = p(C_j) \prod_{i=1}^k p(x_i | C_j)$. Since our light source is either coherent or thermal, we assume $p(C_j) = 0.5$. Thus, it is easy to construct a naive Bayes classifier, where one picks the hypothesis with the highest conditional probability $p(C_j | \mathbf{x})$. We used theoretically generated photon-number probability distributions as the prior probability $p(x_i | C_j)$, and used the experimental data as the test data.

RESULTS

In Fig. 2, we compare the histograms for the theoretical and experimental photon number distributions for different mean photon numbers $\bar{n} = 0.40, 0.53, 0.67$

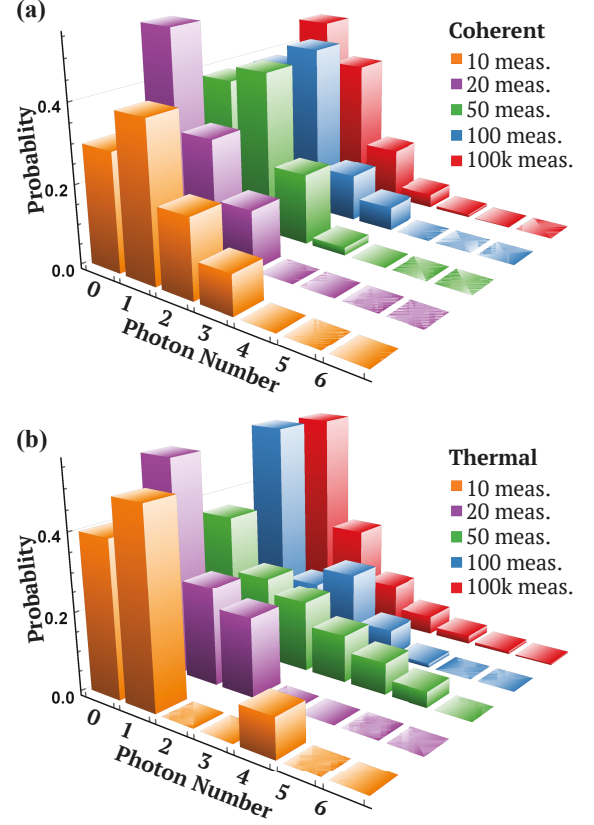


FIG. 3. Probability distributions of coherent and thermal light, for varying dataset sizes (10, 20, 50, 100, 10k). Data used here is randomly selected from of the measurement presented in Fig. 2 (a).

and 0.77. The bar plots are generated by experimental data with one million measurements for each source; the curves in each of the panels represent the expected theoretical photon number distributions for the corresponding mean photon numbers. Fig. 2 shows excellent agreement between theory and experiment which demonstrates the accuracy of our surjective photon counting method. Furthermore, from Fig. 2 (a)-(d), we can also observe the effect of the mean photon number on the photon number probability distributions. As shown in Fig. 2 (a), it is evident that millions of measurements enable one to discriminate light sources. On the other hand, Fig. 2 (d) shows a situation in which the source mean-photon number is low. In this case, the discrimination of light sources becomes cumbersome, even with millions of measurements. In order to illustrate the difficulty of using limited sets of data to discriminate light sources at low mean photon numbers, we restrict the size of our dataset to 10, 20, 50, 100 and 100000. As shown in Fig. 3, the photon number distributions obtained with limited number of measurements do not resemble those in the histograms shown in Fig. 2 (a), for both coherent

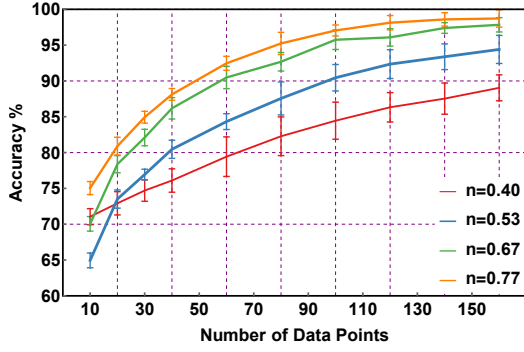


FIG. 4. Overall accuracy of light discrimination versus the number of data points used in naive Bayes classifier. The curves represent the accuracy of light discrimination for $\bar{n} = 0.40$ (red line), $\bar{n} = 0.53$ (blue line), $\bar{n} = 0.67$ (green line) and $\bar{n} = 0.77$ (orange line). The error bars are generated by dividing the data into ten subsets.

and thermal light beams.

In Fig. 4, we show the overall accuracy for light discrimination using naive Bayes classifier. The accuracy increases with the number of data points. For example, when $\bar{n} = 0.40$, the accuracy of discrimination increases from approximately 72% to 88% as we increase the number of data points from 10 to 160. It is worth noting that even with small increase in number of measurements, the naive Bayes classifier starts to capture the characteristic feature of different light sources, given by distinct sequences of photon number events. This is obvious since larger sets of data contain more information pertaining to the probability distribution. Furthermore, mean photon number of the light field significantly changes the discrimination accuracy profile. As the mean photon number increases, the overall accuracy converges faster towards 100% as expected. This is due to the fact that the photon number probability distributions become more distinct at higher mean photon number.

The overall accuracy of light-source discrimination with respect to the number of data points is shown in Fig. 5. Using only 10 data points, ADALINE leads to an average accuracy between 61%-65% for $\bar{n} = 0.40$; whereas for 160 data points, the accuracy is greater than 90%. The comparison of Fig. 4 and Fig. 5 reveals that ADALINE and naive Bayes classifier exhibit similar accuracy levels. However, ADALINE requires far less computational resources than naive Bayes classifier. As one might expect, in both cases, the accuracy increases with the number of data points and mean photon numbers. Interestingly, the convergence rate for naive Bayes is slightly higher than that of ADALINE classifier. For low mean photon numbers, such as $\bar{n} = 0.40$, the improvement in accuracy scales linearly for naive Bayes classifier, as opposed to almost logistic growth that has our ADALINE. This implies that at low mean photon numbers ADALINEs outperform naive Bayes classifier in the sense that the

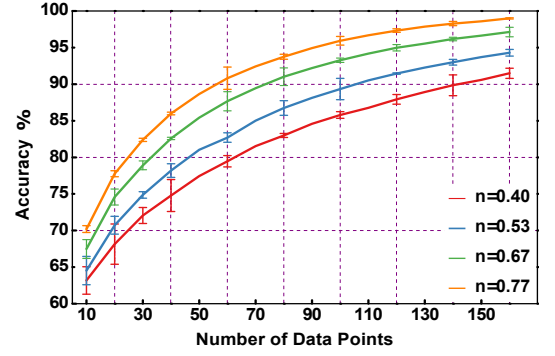


FIG. 5. Overall accuracy of light discrimination versus the number of data points used in ADALINE. The curves represent the accuracy of light discrimination for $\bar{n} = 0.40$ (red line), $\bar{n} = 0.53$ (blue line), $\bar{n} = 0.67$ (green line) and $\bar{n} = 0.77$ (orange line). The error bars represent the standard deviation of the training stages.

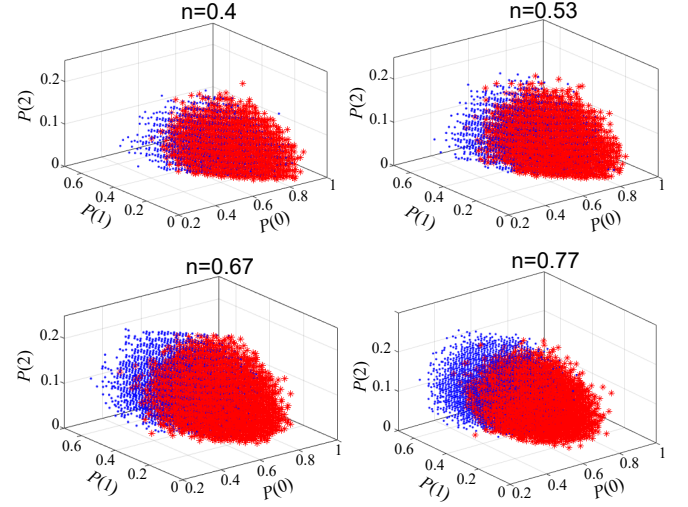


FIG. 6. Projection of the feature space on the plane $(P(0), P(1), P(2))$ for different mean photon numbers: (a) $\bar{n} = 0.4$, (b) $\bar{n} = 0.53$, (c) $\bar{n} = 0.67$, and (d) $\bar{n} = 0.77$. The blue points correspond to photon statistics of coherent light, whereas the red stars describe photon statistics of thermal light. In all cases the number of data points is fixed at $M = 60$.

former require much less computational resources than the latter.

To understand why a single ADALINE neuron is enough for light discrimination, we first realize that ADALINE is a linear classifier; therefore the decision surface is expressed by a seven-dimensional hyper-plane, defined by the seven $P(n)$ (with $n = 0, 1, \dots, 6$) features. Interestingly, one can find that the datasets at the space of probability-distribution values are linearly separable. This can be seen from Fig. 6, where we plot the projection of the feature space on a three-dimensional sub-space defined by $(P(0), P(1), P(2))$ considering different mean photon numbers $\bar{n} = 0.4, 0.53, 0.67$ and 0.77 . In all cases, the number of data points is fixed at $M = 60$. Within

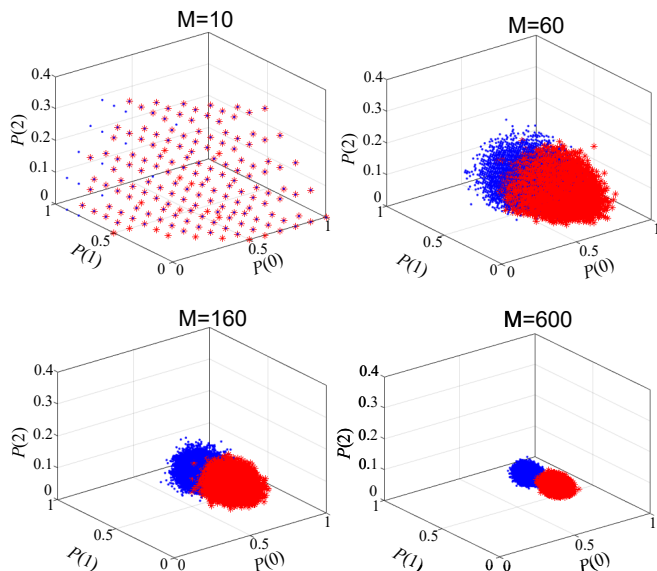


FIG. 7. Projection of the feature space on the plane ($P(0)$, $P(1)$, $P(2)$) for different number of data points: (a) $M = 10$, (b) $M = 60$, (c) $M = 160$, and (d) $M = 600$. The blue points correspond to photon statistics of coherent light, whereas the red stars describe photon statistics of thermal light. In all cases, the mean photon number is set to $\bar{n} = 0.77$

this subspace, the datasets corresponding to the photon statistics of thermal (red stars) and coherent (blue points) lights separate each other as $\bar{n}$ increases. This effect is more evident when the number of data points is increased, and the mean photon number remains fixed at $\bar{n} = 0.77$ [see Fig. 7]. Evidently, the fact that both, thermal and coherent light form two well linearly separated classes makes ADALINE the optimum classifier for light identification.

CONCLUSION

For more than twenty years, there has been an enormous interest in reducing the number of photons and measurements required to perform imaging, remote sensing and metrology at extremely low-light levels [11, 32]. In this regard, photonic technologies operating at low-photon levels utilize weak photon signals that make them vulnerable against detection of environmental photons emitted from natural sources of light. Indeed, this limitation has made unfeasible the realistic implementation of this family of technologies [6, 10, 12]. So far, this vulnerability has been tackled through conventional approaches that rely on the measurement of coherence functions, the implementation of thresholding and quantum state tomography [6, 10, 12, 33]. Unfortunately, these approaches to characterize photon-fluctuations rely on the acquisition of large number of measurements that impose constraints on the identification of light sources. Here, for

the first time, we have demonstrated a smart protocol for discrimination of light sources at mean photon numbers below one. Our work demonstrates a dramatic improvement of several orders of magnitude in both the number of photons and measurements required to identify light sources [6, 10, 12, 33]. Furthermore, our results indicate that a single artificial neuron outperforms naive Bayes classifier at low-light levels. Interestingly, this neuron has simple analytical and computational properties that enable low-complexity and low-cost implementations of our technique. We are certain that our work has important implications for multiple photonic technologies, such as LIDAR and microscopy of biological materials.

ACKNOWLEDGEMENTS

We thank the Department of Physics & Astronomy at Louisiana State University for providing Startup funding to perform this experimental work. CY would like to acknowledge support from National Science Foundation. NB would like to thank Army Research Office (ARO) for the funding. R.J.L.M. and M.A.Q.J. thankfully acknowledge financial support by CONACYT under the project CB-2016-01/284372 and by DGAPA-UNAM under the project UNAM-PAPIIT IA100718. We all thank J. P. Dowling, X. Wang, L. Cohen and H. S. Eisenberg for helpful discussions.

* cyou2@lsu.edu

- [1] R. J. Glauber, “The Quantum Theory of Optical Coherence,” *Phys. Review* **130**, 2598 (1963).
- [2] L. Mandel and E. Wolf “Optical Coherence and Quantum optics,” Cambridge: Cambridge University Press (1995).
- [3] L. Mandel, “Sub-Poissonian photon statistics in resonance fluorescence,” *Opt. Lett.* **4**, 205–207 (1979).
- [4] L. Mandel and E. Wolf, “Coherence properties of optical fields,” *Rev. Mod. Phys.* **37**, 231 (1965).
- [5] J. Liu, and Y. Shih, “Nth-order coherence of thermal light,” *Phys. Rev. A* **79**, 023819 (2009).
- [6] J. Hlousek, M. Dudka, I. Straka, and M. Jazek, “Accurate Detection of Arbitrary Photon Statistics,” *Phys. Rev. Lett.* **123**, 153604 (2019).
- [7] L. Dovrat, M. Bakstein, D. Istrati, A. Shaham, and H. S. Eisenberg, “Measurements of the dependence of the photon-number distribution on the number of modes in parametric down-conversion,” *Opt. Express* **20**, 2266–2276 (2012).
- [8] L. Dovrat, M. Bakstein, D. Istrati, E. Megidish, A. Halevy, L. Cohen and H. S. Eisenberg, “Direct observation of the degree of correlations using photon-number-resolving detectors,” *Phys. Rev. A* **87**, 053813 (2013).
- [9] G. Zambra, A. Andreoni, M. Bondani, M. Gramegna, M. Genovese, G. Brida, A. Rossi, and M. G. A. Paris, “Experimental Reconstruction of Photon Statistics without Photon Counting,” *Phys. Rev. Lett.* **95**, 063602 (2005).

- [10] L. A. Howard, G. G. Gillett, M. E. Pearce, R. A. Abraham, T. J. Weinhold, P. Kok, and A. G. White, "Optical Imaging of Remote Bodies using Quantum Detectors," *Phys. Rev. Lett.* **123**, 143604 (2019).
- [11] J. P. Dowling, and K. P. Seshadreesan, "Quantum Optical Technologies for Metrology, Sensing, and Imaging," *J. Light. Technol.* **33**, 2359 (2015).
- [12] Y. Sher, L. Cohen, D. Istrati, and H. S. Eisenberg, "Low intensity lidar using compressed sensing and a photon number resolving detector," *Emerging Digital Micromirror Device Based Systems and Applications X* **10546**, 105460J (2018).
- [13] Q. Wang, L. Hao, Y. Zhang, C. Yang, X. Yang, L. Xu, and Y. Zhao, "Optimal detection strategy for super-resolving quantum lidar," *J. Appl. Phys.* **119**, 023109 (2016).
- [14] J. P. Dowling, "Quantum optical metrology – the low-down on High-N00N states," *Contemp. Phys* **49**, 125-143 (2008).
- [15] O. S. Magaña-Loaiza, R. J. Leon-Montiel, A. Perez-Leija, A. B. U'Ren, C. You, K. Busch, A. E. Lita, S. W. Nam, R. P. Mirin, T. Gerrits, "Multiphoton Quantum-State Engineering using Conditional Measurements," *npj Quantum Information* **5**, 80 (2019).
- [16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* **521**, 436-444 (2015).
- [17] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum Machine Learning," *Nature* **549**, 195-202 (2017).
- [18] V. Dunjko, J. M. Taylor, H. J. Briegel, "Quantum-Enhanced Machine Learning," *Phys. Rev. Lett.* **117**, 130501 (2016).
- [19] A. Hentschel, and B. C. Sanders, "Machine Learning for Precise Quantum Measurement," *Phys. Rev. Lett.* **104**, 063603 (2010).
- [20] A. Lumino, E. Polino, A. S. Rab, G. Milani, N. Spagnolo, N. Wiebe, and F. Sciarrino, "Experimental Phase Estimation Enhanced by Machine Learning," *Phys. Rev. Appl.* **10**, 044033 (2018).
- [21] A. A. Melnikov, H. P. Nautrup, M. Krenn, V. Dunjko, M. Tiersch, A. Zeilinger and H. J. Briegel, "Active learning machine learns to create new quantum experiments," *PNAS* **6**, 1221-1226 (2018).
- [22] S. Lohani, R. T. Glasser, "Turbulence correction with artificial neural networks," *Opt. Lett.* **43**, 2511-2614 (2018).
- [23] J. Gao, L.-F. Qiao, Z.-Q. Jiao, Y.-C. Ma, C.-Q. Hu, R.-J. Ren, A.-L. Yang, H. Tang, M.-H. Yung, and X.-M. Jin, "Experimental Machine Learning of Quantum States," *Phys. Rev. Lett.* **120**, 240501 (2018).
- [24] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, "Neural-network quantum state tomography," *Nat. Phys.* **14**, 447-450 (2018).
- [25] S. M. H. Rafsanjani, M. Mirhosseini, O. S. Magaña-Loaiza, B. T. Gard, R. Birrittella, B. E. Koltenbah, C. G. Parazzoli, B. A. Capron, C. C. Gerry, J. P. Dowling, and R. W. Boyd, "Quantum-enhanced interferometry with weak thermal light," *Optica* **4**, 487-491 (2017).
- [26] I. A. Burenkov, A. K. Sharma, T. Gerrits, G. Harder, T. J. Bartley, C. Silberhorn, E. A. Goldschmidt, and S. V. Polyakov, "Full statistical mode reconstruction of a light field via a photon-number-resolved measurement," *Phys. Rev. A* **95**, 053806 (2017).
- [27] N. Montaut, O. S. Magaña-Loaiza, T. J. Bartley, V. B. Verma, S. W. Nam, R. P. Mirin, C. Silberhorn, and T. Gerrits, "Compressive characterization of telecom photon pairs in the spatial and spectral degrees of freedom," *Optica* **11**, 1418 (2018).
- [28] B. Windrow, M. E. Hoff, "Adaptive switching circuits," Technical Report No. 1553-1, Stanford University, Stanford-California, Stanford Electronics Laboratories (1960).
- [29] S. I. Gallant, "Neural network learning and expert systems," MIT press (1993).
- [30] A. Cauchy, "Méthode générale pour la résolution des systèmes d'équations simultanées," *Comp. Rend. Sci. Paris* **25**, 1847, pp. 536-538 (1847).
- [31] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin, "Bayesian Data Analysis," Chapman and Hall/CRC (2013).
- [32] O. S. Magaña-Loaiza, and R. W. Boyd, "Quantum Imaging and Information," *Reports on Progress in Physics*, in press (2019).
- [33] L. Cohen, E. S. Matekole, Y. Sher, D. Istrati, H. S. Eisenberg, and J. P. Dowling, "Thresholded Quantum LIDAR - Exploiting Photon-Number-Resolving Detection", *Phys. Rev. Lett.*, in press (2019).

Supplementary material: Identification of Light Sources using Machine Learning

In this supplemental material, we evaluate two additional machine-learning algorithms, namely a one-dimensional convolutional neural network (1D CNN) and a multilayer neural network (MNN). Despite both algorithms are effective to identify light sources, they are analytically and computationally more sophisticated than the simple ADALINE model, but their recognition rates do not present substantial differences. Figures 8(a) and 8(b) show the structure of the 1D-CNN and MNN, respectively.

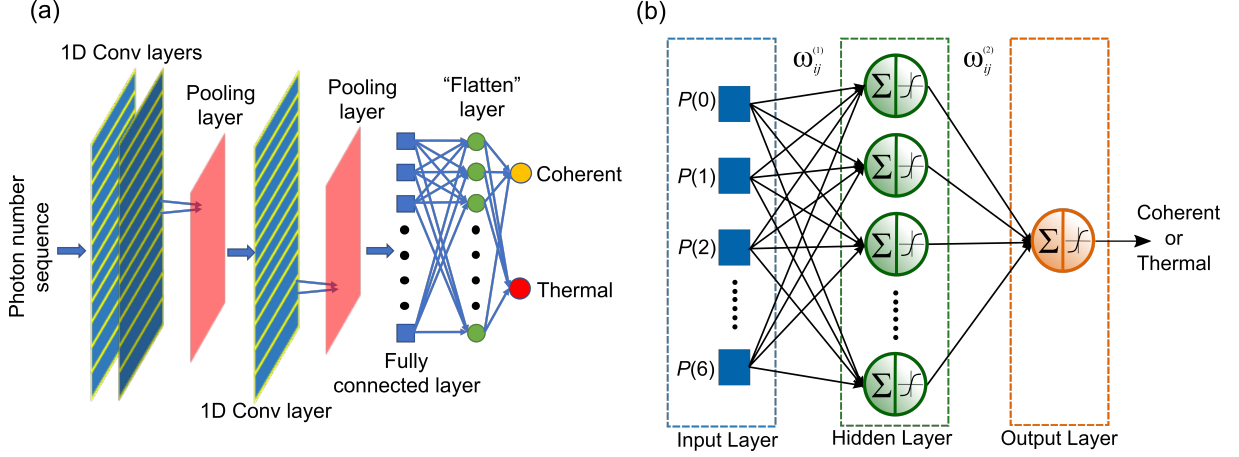


FIG. 8. Schematic representations: (a) one-dimensional convolutional neural network and (b) multilayer neural network used for demonstration of light source identification.

A convolutional neural network is a deep learning algorithm that extracts automatically relevant features of the input [1]. Here, our one-dimensional convolutional neural network is composed by two 1D-convolutional layers that extract the low and high-level features of the input. Outcomes from these two layers are subsequently fed into a convolutional layer sandwiched between two max-pooling layers. The pooling layers downsample the input representation, and therefore its dimensionality, leading to a computational simplification by removing redundant and unnecessary information. The activation function, implemented in all layers, is the rectified linear unit function (ReLU). Finally, a fully connected and a flattening layer precedes the output layer consisting of two softmax functions, whose outputs are the probability distributions over labels.

On the other hand, the multilayer neural network belongs to a classical machine learning algorithm, where the feature vector should be manually determined [2]. In our case, this vector is given by the probabilities of the photon number distribution, $P(n)$. As depicted in Fig. 8(b), the model corresponds to a two-layer feed-forward network: the hidden layer contains ten sigmoid neurons and the output layer consists of a softmax function. To determine a suitable neuron number in the hidden layer of the MNN, we trained different MNNs by changing the neuron number in the hidden layer and followed the accuracy values for each net. Figs. 9(a) and 9(b) show the overall accuracy for light discrimination versus the number of neurons in the hidden layer for different mean photon numbers, $\bar{n} = 0.4$ and 0.77 , respectively. Note that in both cases, the accuracy becomes lower as the number of neurons increases. This is because many neurons lead to over-parameterization, causing poor generalization of the test-stage data. Additionally, as the number of neurons increases, the training becomes computationally more intensive. All the MNNs were trained by using the scaled conjugate gradient backpropagation method where the cross-entropy was employed as the cost function. Since the output of sigmoid neurons is ranged in the interval $[0,1]$, the cross-entropy function is ideal for the classification task. The network training was stopped after 200 epochs.

1D-CNNs and MNNs were trained using the same training set described in the main manuscript. Despite that deep neural network should be trained with a larger amount of data, we use 70% of the dataset for the training and the rest for testing both networks. Note that the same procedure was used for the ADALINE model. Figures 10(a) and 10(b) show the overall light-discrimination accuracy for increasingly larger number of data points for (a) 1D-CNNs and (b) MNNs. In both cases, the accuracy increases with the number of data points, because larger sets of data contain more information about the probability distribution. Interestingly, the accuracy of 1D-CNNs for $\bar{n} = 0.67$ and

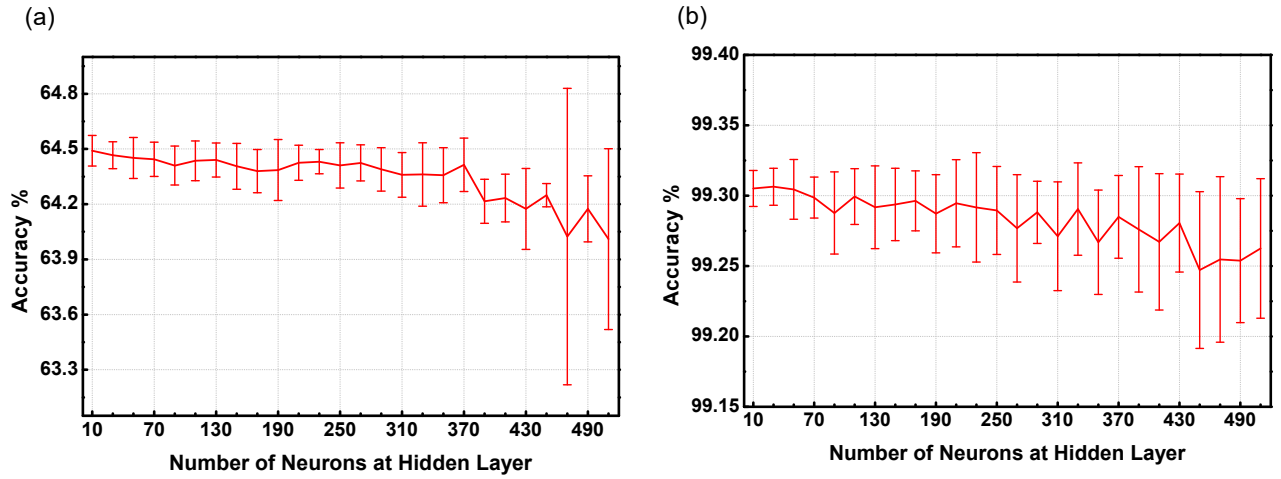


FIG. 9. Overall accuracy of light discrimination versus the number of neurons in the hidden layer of the MNN by considering two different mean photon numbers: (a) $\bar{n} = 0.4$ and (b) $\bar{n} = 0.77$. The error bars represent the standard deviation of the training stages.

$\bar{n} = 0.77$ are almost the same; this indicates that in the low mean photon-number regime, the peak performance for 1D-CNN saturates much faster than the MNN classifier. As one might expect, this fast accuracy convergence carries the cost of a much complex computation as compared to the one needed for the MNN classifier.

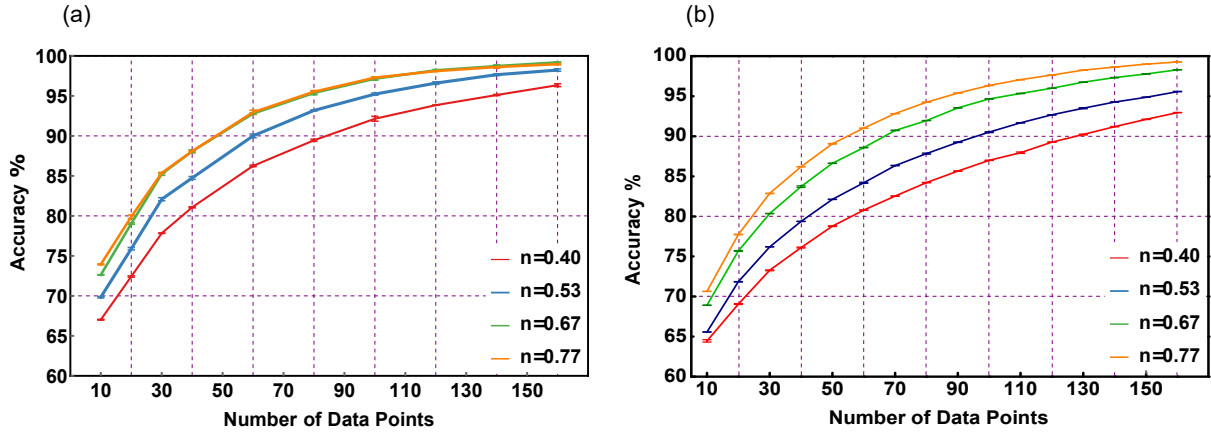


FIG. 10. Overall accuracy of light discrimination versus the number of data points used in (a) 1D-CNN and (b) MNN. The curves represent the accuracy of light discrimination for $\bar{n} = 0.40$ (red line), $\bar{n} = 0.53$ (blue line), $\bar{n} = 0.67$ (green line) and $\bar{n} = 0.77$ (orange line). The error bars represent the standard deviation of the training epochs for 1D-CNN and training stages for MNN.

REFERENCES

-
- * cyou2@lsu.edu
- [1] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.
 - [2] Bishop, Christopher M. Pattern recognition and machine learning. Springer, 2006.
-