Research Article Vol. 32, No. 19 / 9 Sep 2024 / Optics Express 33294
Optics EXPRESS

Object classification through heterogeneous fog with a fast data-driven algorithm using a low-cost single-photon avalanche diode array

ZHENYA ZANG AND DAVID DAY UEI LI*

Department of Biomedical Engineering, University of Strathclyde, Glasgow, G4 0NW, United Kingdom
*david.li@strath.ac.uk

Abstract: This study presents a framework for classifying a wooden mannequin's poses using a single-photon avalanche diode (SPAD) array in dynamic and heterogeneous fog conditions. The target and fog generator are situated within an enclosed fog chamber. Training datasets are continuously collected by configuring the temporal and spatial resolutions on the sensor's firmware, utilizing a low-cost SPAD array sensor priced below $5, consisting of an embedded SPAD array and diffused VCSEL laser. An extreme learning machine (ELM) is trained for rapid pose classification, as a benchmark against CNN. We quantitatively justify the selection of nodes in the hidden layer to balance the computing speed and accuracy. Results demonstrate that ELM can accurately classify mannequin poses when obscured by dynamic heavy fog to 35 cm away from the sensor, enabling real-time applications in consumer electronics. The proposed ELM achieves 90.65% and 89.58% accuracy in training and testing, respectively. Additionally, we demonstrate the robustness of both ELM and CNN as the fog density increases. Our study also discusses the sensor's current optical limitations and lays the groundwork for future advancements in sensor technology.

1. Introduction

Identifying objects obstructed by fog and smoke is crucial for safe driving, rescue and security in changing weather. Active sensing systems based on single-photon detectors are gaining rapid attention due to their outstanding single-photon sensitivity despite being used in scattering environments. In modern time-resolved sensing systems, the time-of-flight (ToF) of detected photons is recorded by a time-correlated single-photon counting (TCSPC) module to generate histograms. Single-photon detectors with raster scanning have been employed in challenging applications, such as photon-efficient imaging [1–3] non-line-of-sight imaging [4–6], long-range imaging [7,8] and time-resolved biomedical imaging [9,10]. Also, wide-field single-photon avalanche diode (SPAD) arrays discard scanning mechanisms and are efficient detectors for fluorescence lifetime reconstruction [11,12] and sensing through scattering for diffuse correlation spectroscopy [13]. In these applications, deep learning is emerging in processing histograms or photon-counting data to retrieve critical parameters, such as 3D geometries, fluorescence lifetime, and blood flow index.

Dehazing obscuration using SPADs is challenging, as dense aerosols severely affect the temporal properties and amplitudes of emitted laser and returned photons [14,15]. Exiting systems of active time-resolved sensing through obstacles rely on high-precision SPAD arrays, powerful lasers, scanners, and complex optical setups [16–19] exhibiting high cost and inadequate portability. Low-cost SPADs sharing the same essential technologies as high-precision SPADs have been rolled out to the market in recent years, exhibiting coarse spatial and temporal resolution. The compact module is promising for mobile devices in extreme environments, such as drones,

#527244 https://doi.org/10.1364/OE.527244
Journal © 2024 Received 15 Apr 2024; revised 27 Jun 2024; accepted 23 Jul 2024; published 3 Sep 2024

Optics EXPRESS

for rescue tasks in fog and smoke. Also, estimating invisible poses is crucial for human-computer interaction when facial and body images are unavailable.

Although deep learning achieves superior accuracy and speed performances compared with conventional fitting and statistical methods for SPADs-based computational imaging and sensing, the generalization is the issue, meaning the accuracy is susceptible to calibration on different setups and instruments. The deep neural networks are likely to be re-trained. Extreme learning machine (ELM) [20] is competent at processing 1D signals and saves much time during the re-training process while maintaining accuracy. Training and inference of ELM and its variance [21] have been proved hardware-friendly and implemented on system-on-chips [22].

Therefore, we integrated compact ELM and a customer SPAD array (VL53L8CH from STMicroelectronics, < \$5) to explore the sensing capabilities further. In this study, we aim to classify the poses of a mannequin through dense, dynamic fog.

To our knowledge, this is the first study to utilize a low-cost SPAD array for object classification in dynamic fog conditions. The contribution of this work is threefold:

1. We employed a low-cost commercial SPAD array to classify object poses accurately within a dynamic and heterogeneous fog chamber. Our approach includes developing both fast-training and fast-inference ELM models, as well as a high-accuracy CNN, catering to various scenarios with different latency and accuracy requirements. ELM and CNN achieve inference speeds of 0.017 ms and 0.54 ms per classification, respectively, for a mannequin 35 cm away from the sensor.

2. We evaluated how the number of neurons in the hidden layer affects latency and accuracy, providing a parameterized and flexible solution to meet varying requirements for speed and accuracy.

3. The performances of ELM and CNN classification were assessed. We presented accuracy metrics for each pose and investigated the influence of the fog density on the accuracy of ELM and CNN.

2. Related work

2.1. Obscuration dehazing using high-precision SPAD arrays

SPAD arrays have been utilized for sensing through fog, employing data-driven and statistical algorithms. Statistical models of pixel-wise dense-fog-corrupted histograms have been introduced [23], wherein depth information can be reconstructed by subtracting the background distribution from the noisy histogram for a mannequin positioned 57 cm from the sensor. A high-precision commercial SPAD array with 32×32 pixels was employed to acquire histograms. The same SPAD array was utilized to capture frames of intensity data [24]. With known characterized parameters, a Monte Carlo forward model was proposed to simulate photon propagation through the diffuser and generate synthetic data for training a convolutional neural network for classification. A SPAD array with the same spatial resolution was used for sensing through a 10-meter atmospheric obscurant, covering a total distance of 35 meters, to reconstruct depth images using a statistical algorithm [16]. Joyce *et al.* [17] developed a 3D Flash SPAD array with 32×32 resolution and reconstructed depth images of military targets through camouflage nets or fog (23 meters long in a tunnel). A linear SPAD with 16×1 pixels and a scanner were employed for underwater depth reconstruction [25] through a 1.65-meter-long water tank. Sang and Tsai proposed histogram-based [18] and point cloud-based [19] defogging algorithms using a SPAD array. Tobin *et al.* used a single-pixel SPAD and monostatic scanning transceiver to reconstruct intensity and depth images of objects occluded by different obscurants [26].

Optics EXPRESS

2.2. Sensing using low-cost SPAD arrays

High-resolution depth images were successfully retrieved using transient histograms and deep learning from a low-cost SPAD array [27]. Transient histograms were also used to recover the planar geometry and albedo of planar surfaces [28]. A portable scanning module was coupled with a low-cost SPAD array to generate high-resolution depth images and track the position of non-line-of-sight objects [29]. The histograms were also used to classify materials with different diffusion properties. Histograms from a low-cost SPAD array were fused into a deep neural network with high-resolution RGB and depth images from a conventional ToF to reconstruct distinct 3D human skeletons [30]. Another low-cost SPAD was integrated into a system-on-chip on a drone for accurate moving object sensing [31].

3. Data acquisition and pre-processing

The sensing platform, depicted in Fig. 1(a), features a SPAD array mounted on an off-the-shelf evaluation board. The sensor is positioned 1 cm away from a fog chamber measuring $90 \times 45 \times 50$ cm^3; the setup includes a hole in the chamber allowing laser light to enter and photons to return from the chamber. The SPAD sensor was configured as histogram output mode. The sensor's spatial resolution was configured as 8×8, capturing data from all pixels. Histograms from pixels other than the (0, 0) index were subsampled and aggregated to create a bucket pixel. Due to memory constraints, the number of time bins for the aggregated histogram can be configured up to 128, although we set it to 50 to cover the length of the chamber. Each bin represents a time interval of ~250 ps (equivalent to 37 mm). The integration time was configured to 5 ms. A P-NUCLEO-53L5A1 evaluation board equipped with a 32-bit microcontroller was used to execute the compiled code and configure the sensor. The compiled code was transferred to the board via an I^2C interface. Subsequently, the aggregated histogram was transmitted from the sensor to the computer via UART, where it was received by a Python script and saved as training and test datasets. The data transfer speed via UART was approximately 40 histograms per second. The SPAD array's critical parameters in this experiment are summarized in Table 1. Upon initiating data acquisition, the ultrasonic fog generator and fan are enabled to generate dynamic fog. The fog density increases as data acquisition is enabled.

Table 1. Specifications of the sensor and configured parameters in this work

Parameter's name	Value
Non-configurable field	
Laser wavelength	940 nm
Laser FWHM	1.6 nm
FoV	$45° \times 45°$
Configurable field	
# time bins	50
# merged pixels	63
# merged time bins	0
Time-bin width	~250 ps
Index of the bucket pixel	(0, 0)
Ranging frequency	15 Hz
Integration time	5 ms
Spatial resolution	8×8
# sensed targets/pixel	1

Optics EXPRESS

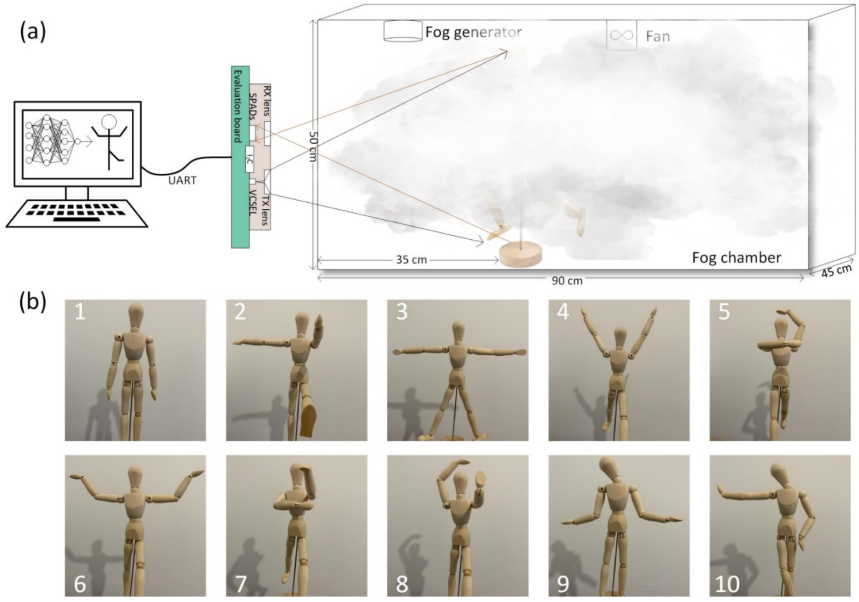

Fig. 1. (a) Platform overview of active sensing through fog: The diffused laser from the embedded VCSEL illuminates the fog chamber through a small window that preserves a sufficient angle for both emitted laser and returned photons. Decoded histograms are transferred back to a PC, which runs an ELM model for pose classification. (b). Ten classes of mannequin pose, where the number in each image is the index of each class.

The fog density increases as the number of aggregated histograms increases. We acquired 110,000 aggregated histograms (10,000 histograms for each class) to train our neural networks. The corresponding actual classes are enumerated in Fig. 1(b). As we aim to evaluate the accuracy performance of our algorithm for the sensor under different fog densities, we split 11,000 histograms, ranging from sparse to dense fog, for testing, while the remaining 99,000 histograms were used for training. All the histograms were normalized between [0, 1]. For CNN, 110,000 histograms were divided into 80,000 for training, 20,000 for validation, and 10,000 for testing. Figure 2(a) and (b) show aggregated histograms and corresponding intensity images. Intensity images were up sampled to 25×25 using OpenCV for better visualization. In Fig. 2(a), the left image depicts histograms in the fog-free condition, where the peak indicates the end of the chamber. The right image shows the histogram from the object-free, dense fog condition. Figure 2(c) shows an example (Class-2) describing how the histogram and intensity images change as the frame number increases. The peak of the histogram shifts to smaller time-bin indices as the frame (fog density) increases, indicating that more photons are reflected from the fog. Figure 2(d) depicts the corresponding real-time intensity images. When the frame number is 9,500, fog reflected most photons and target information was rarely distinguishable from the histogram and intensity image. Furthermore, compared with the no-object condition with dense fog (the right image in Fig. 2(a)), more photons are reflected due to the object's presence despite dense fog. Visualization 1 visualizes the intensity video of the mannequin (Class-2) in fog as the fog density increases.

Optics EXPRESS

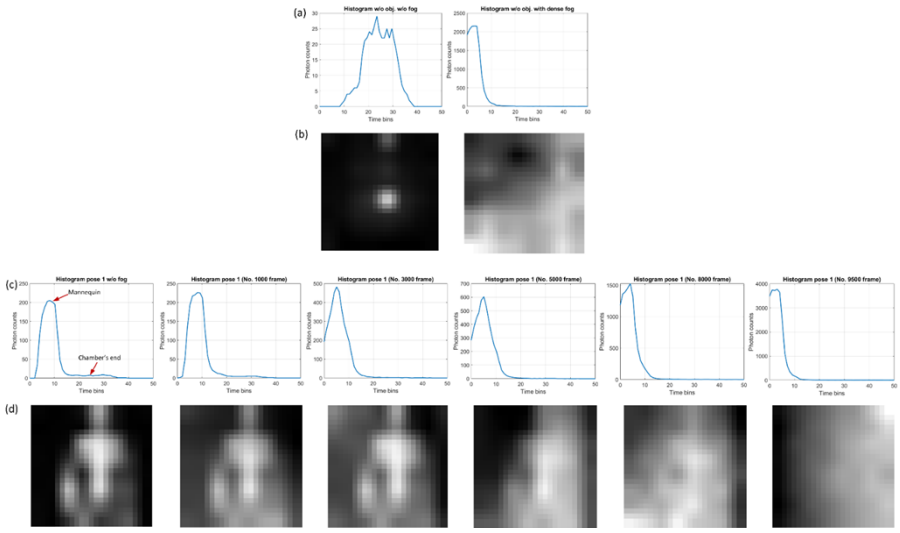

Fig. 2. (a) Aggregated histograms and (b) corresponding normalized up-sampled intensity images of the 100th frame and the 9000th frame. (a) No object scenarios without fog (left) and with (right) dense fog. (c) and (d), histograms and intensity images of the 1000^{th}, 3000^{th}, 5000^{th}, 7000^{th}, 9000^{th}, and 10000^{th} frame. Red arrows indicate the mannequin and the chamber's end.

4. Neural networks for classification

4.1. ELM training details

ELM is a back-propagation-free single-layer feedforward network that uses matrices' pseudo inversion for training. Compared with neural networks using partial differentiation for training, ELM does not suffer from over-fitting issues, and trains much faster. ELM excels in processing one-dimensional time-series-like data, like EEG [32] and ECG [33]. Our previous work shows ELM is an accurate and robust architecture for fluorescence lifetime reconstruction from histograms using a time-resolved sensor [34]. Inspired by our earlier work, the number of input nodes equals the number of time bins in the histogram, which is 50 in this ELM. We set the number of nodes in the hidden layer (L) to 2,500 to balance the trade-off between latency and accuracy for training and inference. *Sigmoid ()* is the activation function as it yields better training accuracy than other activation functions (*relu()*, *sin()*, *hardlim()*). As shown in Fig. 3(a), there are 11 nodes at the output layer, generating predicted results. A *max()* function was applied to the output vector to generate the maximum value, indicating the predicted class.

In training datasets, we define H pairs of vectors that are $\boldsymbol{x}_i = [x_{i1},\ x_{i2},\ \ldots\ ,\ x_{im}]^T \in \mathbb{R}^m$ and $\boldsymbol{y}_i = [y_{i1},\ y_{i2},\ \ldots\ ,\ y_{in}] \in \mathbb{R}^n$, incicating the ith input vector (histograms) and the ith target vector (ground truth classes), respectively. The output matrix of the hidden layer can be defined as:

$$A = \begin{bmatrix} \varphi(w_1 \cdot x_1 + b_1) & \cdots & \varphi(w_L \cdot x_1 + b_L) \\ \vdots & \ddots & \vdots \\ \varphi(w_1 \cdot x_H + b_1) & \cdots & \varphi(w_L \cdot x_H + b_L) \end{bmatrix}_{H \times L}, \tag{1}$$

where $\boldsymbol{w}_l = [w_{l1},\ w_{l2},\ \ldots\ ,\ w_{lm}]^T$, $b_l = [b_1,\ b_2,\ \ldots\ ,\ b_L]^T$, and $l = 1, \ldots, L$ are random values between the input nodes and the hidden layer. $\boldsymbol{\beta}$ is the weight matrix to train, connecting the

Optics EXPRESS

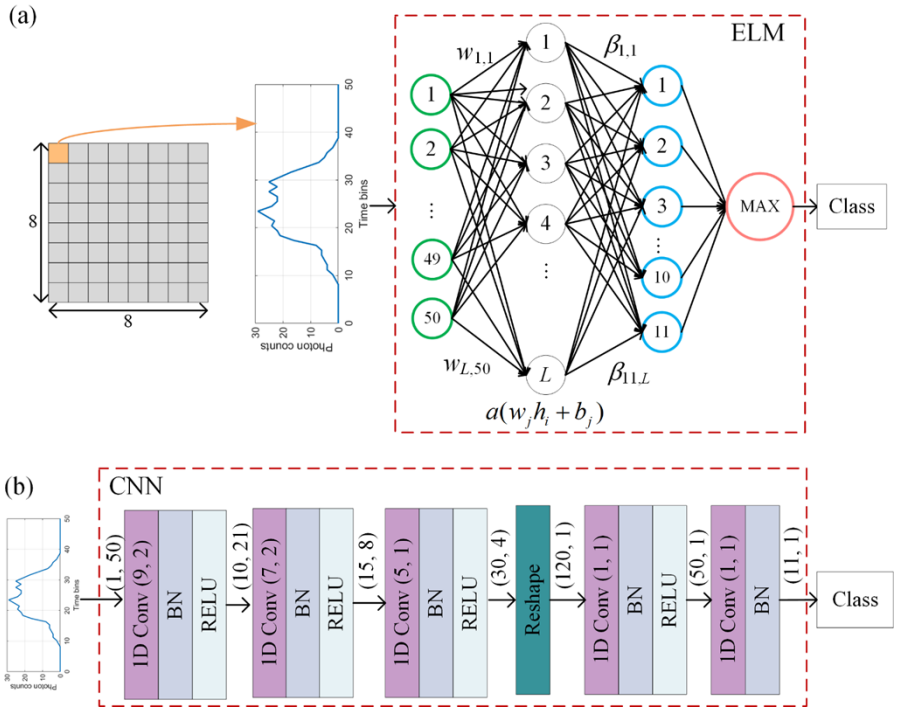

Fig. 3. Performing one-batch histogram inference using (a) ELM and (b) CNN architecture for pose inference: (a). Histograms from grey-coloured pixels are sub-sampled and merged with the yellow-coloured pixels. The aggregated histogram is then fed into the trained ELM for pose classification. (b). Numbers in 1D Conv blocks indicate kernel size and stride, the size of the feature map is presented after each RELU block.

hidden layer and the output layer, defined as:

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix} = \begin{bmatrix} \beta_{11} & \cdots & \beta_{1n} \\ \vdots & \ddots & \vdots \\ \beta_{L1} & \cdots & \beta_{Ln} \end{bmatrix}_{L\times n}. \tag{2}$$

We need to minimize the error in the ridge loss function to learn $\boldsymbol{\beta}$ with a dimension of $L \times n$:

$$\arg\min_{\beta \in \mathbb{R}^{L\times n}} ||\boldsymbol{A}\beta - \boldsymbol{Y}||^2, \tag{3}$$

where the $\boldsymbol{A}$ is the matrix composed of the activation functions with dimensions $H \times L$; $\boldsymbol{Y}$ is a matrix with dimensions $H \times n$ containing true class labels:

$$\boldsymbol{Y} = \begin{bmatrix} y_1^T \\ \vdots \\ y_H^T \end{bmatrix} = \begin{bmatrix} y_{11} & \cdots & y_{1n} \\ \vdots & \ddots & \vdots \\ y_{H1} & \cdots & y_{Hn} \end{bmatrix}_{H\times n}. \tag{4}$$

Through solving the loss function, we can obtain the matrix $\boldsymbol{\beta}$ by $\boldsymbol{\beta} = \boldsymbol{A}^{\dagger}\boldsymbol{Y}$. The pseudo matrix inversion $\boldsymbol{A}^{\dagger}$ can be resolved by singular value decomposition [20]. Once $\boldsymbol{\beta}$ is retrieved, we can

Optics EXPRESS

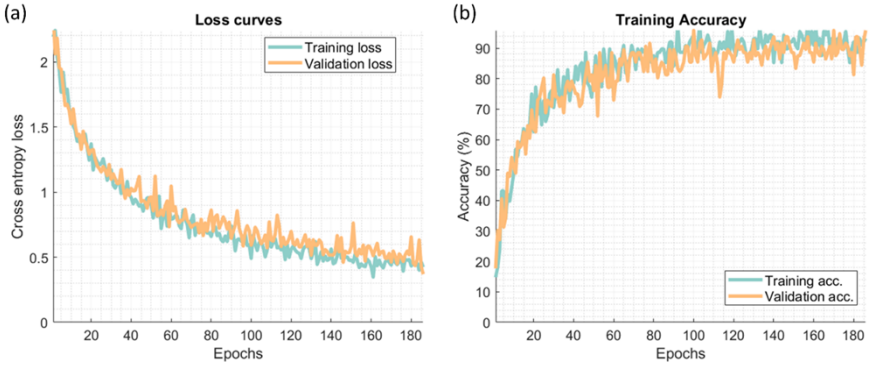

Fig. 4. CNN's (a) loss curves and (b) accuracy curves of training and validation

use it to execute forward inference and compare with true labels to obtain training accuracy. We trained and tested ELM on an Intel i7-13700 CPU, taking 14.08 s and 0.13 s without parallelism optimization. The training and test accuracy are 90.65% and 89.58%.

4.2. CNN training details

A CNN was trained to serve as a benchmark for the proposed ELM. It was implemented using PyTorch and trained on an NVIDIA RTX A1000 GPU with a batch size of 256, taking 803.72 seconds. The total number of epochs was pre-defined as 300. The training ceased at the 186th epoch. The learning rate was set to 0.0001, with cross-entropy as the loss function and SGD as the optimizer. The architecture of the CNN, depicted in Fig. 3(b), consists of five convolutional modules, each followed by batch normalization modules and RELU activation functions. Instead of fully connected layers, channel-wise convolution was employed after the Reshape module to reduce computational complexity. To prevent overfitting, early stopping with 20 patient epochs was applied. Figure 4 illustrates the accuracy curves and training and validation loss.

5. Performance evaluation

This section explains the reason for setting $L = 2{,}500$. Figure 5(a) and (b) present how L determines the accuracy and the latency. $L \in [100, 3{,}000]$ with an interval of 25. As L increases, the accuracy improves fast between 100 and 300 and improves slowly after 300. The training

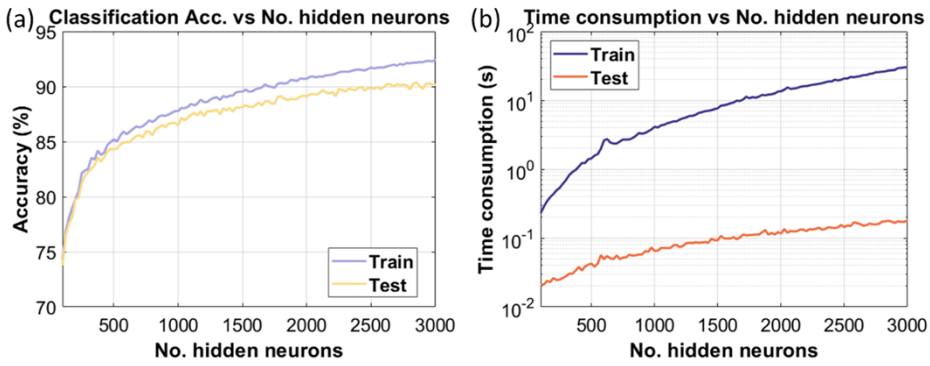

Fig. 5. Relations between L and (a) classification accuracy, and (b) time consumption in training and test datasets

Optics EXPRESS

and test accuracy is more than 90% when L exceeds 2,500. Figure 5(b) shows that L does not significantly affect the test speed, yet the training latency grows relatively noticeable. L can be flexibly selected according to latency or accuracy requirements. As overfitting of ELM only occurs when L exceeds the required amount [35], we confirm that there is no overfitting in the tested L because there is no accuracy degradation in Fig. 5(a).

For ELM, in Fig. 6(a), Class-1 and Class-5 show relatively lower accuracy than other classes. For CNN shown in Fig. 6(b), only Class-5 presents low accuracy. CNN exhibits a generally higher accuracy over 11 classes. Both ELM and CNN sometimes misclassify Class-5 to Class-6 and -8. Figure 5(a) indicates that ELM can approach CNN's accuracy (92.60%) when L further increases but sacrifice speed. Figure 6(c) and (d) illustrate how increasing frames (increasing fog density) influence accuracy. The x-axis indicates the acquired number of frames (representing increasing fog density), grouped by each set of 100 frames to calculate averaged accuracy within

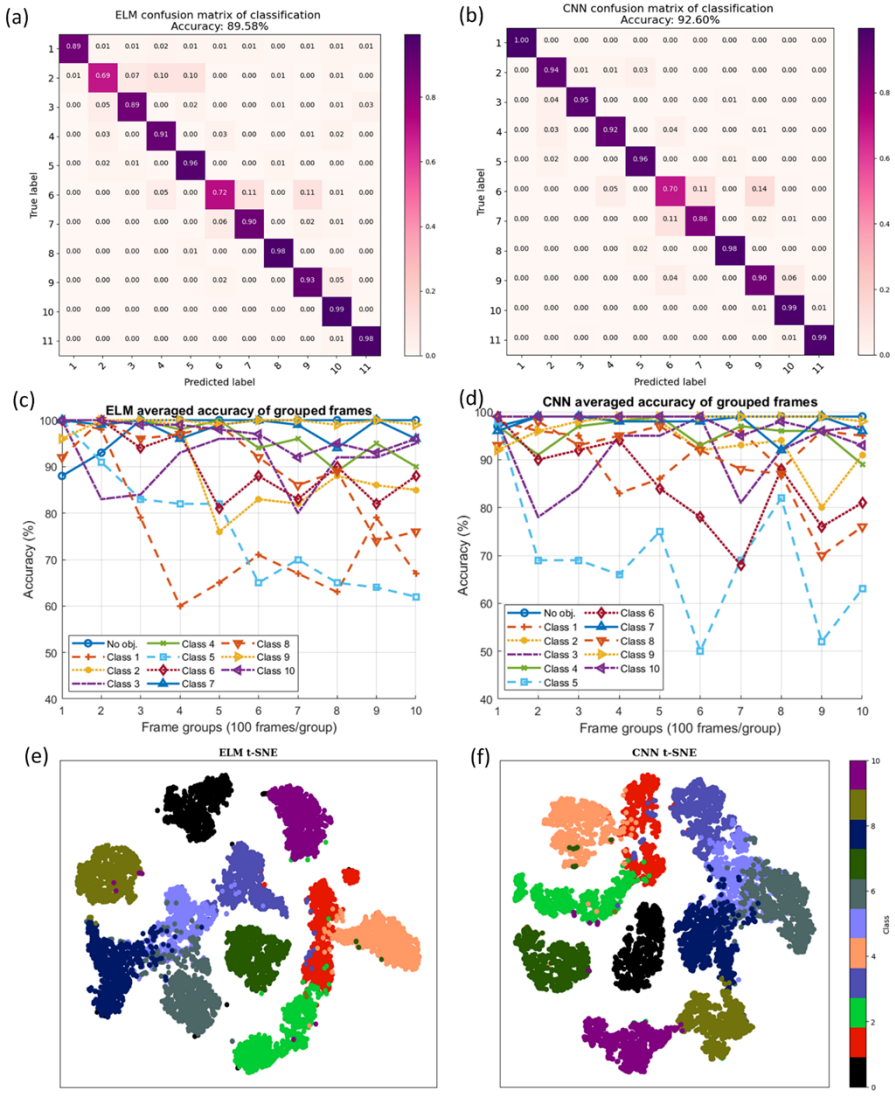

Fig. 6. Confusion matrix of classification of (a) ELM and (b) CNN. Accuracy of (c) ELM and (d) CNN while the frame increases (denser fog). t-SNE of (e) ELM and (f) CNN.

Optics EXPRESS

these 100 frames. Apart from Class-1 and Class-5, most classifications achieve more than 80% accuracy despite increasing the fog density. Similarly, in Fig. 6(b), most classifications are robust to the increased fog density apart from Class-5. We used t-distributed Stochastic Neighbour Embedding (t-SNE) to visualize the output clusters of ELM's and CNN's classification in test datasets, shown in Fig. 6(e) and (f). Both ELM and CNN correctly separate the 11 classes.

6. Discussion and conclusion

This work compares simple ELM and widely used CNN for classifying mannequin poses through dynamic heterogonous fog in a chamber. Eleven classes are accurately classified. We introduced an end-to-end data processing pipeline, ranging from front-end continuous data collection to neural networks' post-processing. Evaluation results reveal that ELM presents a slightly lower accuracy (89.58%) than CNN (92.60%), yet ELM's training and inference are 57.08× and 317.64× faster than CNN, providing two solutions for accuracy and speed requirements. ELM and CNN maintain high accuracy while fog density increases. ELM's fast training speed allows easy reconfiguration for other applications, such as human facial or pose recognition.

The limitation of the low-cost SPAD array is the severe signal attenuation to the range that is longer than 40 cm. The diffused flash laser cannot propagate to long distances due to its low power. Another limitation is the low spatial resolution and the limited number of time bins. The low spatial resolution hinders high-resolution depth or reflectance imaging without scanning modules. Jungerman *et al*. used another cheap SPAD array with 20×15 and deep learning to generate 640×458 depth images [27]. In comparison, our 8×8 SPAD array contains less spatial information, making the ill-posed super-resolution problem harder to solve. Besides, single-pixel SPAD with thousands of time bins have been proven feasible to generate high-resolution depth images using deep learning [36], [37]. However, the low-cost SPAD sensor we used only generates 128- and 18-time bins for an aggregated pixel or each pixel due to the memory constraint. Next-generation low-cost SPAD arrays might make this possible if on-chip memory is bigger to accommodate more time bins.

Besides pixel-wise histogram processing, ELM exhibits limitations when handling multi-channel 1D time-series data and higher-dimensional data such as images and tensors. In the case of multi-channel 1D data, dimensionality reduction techniques like principal or independent component analysis are typically required, introducing additional computational overhead. Reshaping and feeding 2D or 3D data into ELM may lead to the loss of spatial features and accuracy degradation. Therefore, for next-generation consumer SPAD arrays with higher spatial and temporal resolution, a hybrid deep neural network integrating 2D/3D CNN and ELM not only obtains fast training but also yields high accuracy without pre-processing [38].

Funding. Datalab; Photon Force; Xilinx; Innovate UK (10005391, HYDRI); Engineering and Physical Sciences Research Council (EP/T00097X/1).

Acknowledgements. We acknowledge the support from Datalab, Photon Force, Xilinx, the EPSRC (EP/T00097X/1): the Quantum Technology Hub in Quantum Imaging (QuantiC) and Innovate UK HYDRI (10005391).

Disclosures. The authors declare no conflicts of interest.

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

Reference

1. D. Shin, F. Xu, D. Venkatraman, *et al.*, "Photon-efficient imaging with a single-photon camera," Nat. Commun. **7**(1), 12046 (2016).
2. D. B. Lindell, M. O'Toole, and G. Wetzstein, "Single-photon 3D imaging with deep sensor fusion," ACM Trans. Graph. **37**(4), 1–12 (2018).
3. Z. Zang, D. Xiao, and D. D.-U. Li, "Non-fusion time-resolved depth image reconstruction using a highly efficient neural network architecture," Opt. Express **29**(13), 19278–19291 (2021).
4. X. Liu, J. Wang, L. Xiao, *et al.*, "Non-line-of-sight imaging with arbitrary illumination and detection pattern," Nat. Commun. **14**(1), 3230 (2023).

Optics EXPRESS

5. D. Huang, Q. Chen, Z. Wei, *et al.*, "Non-line-of-sight reconstruction via structure sparsity regularization," Opt. Lett. **48**(18), 4881–4884 (2023).
6. C. Pei, A. Zhang, Y. Deng, *et al.*, "Dynamic non-line-of-sight imaging system based on the optimization of point spread functions," Opt. Express **29**(20), 32349–32364 (2021).
7. Z.-P. Li, J.-T. Ye, X. Huang, *et al.*, "Single-photon imaging over 200 km," Optica **8**(3), 344–349 (2021).
8. A. M. Pawlikowska, A. Halimi, R. A. Lamb, *et al.*, "Single-photon three-dimensional imaging at up to 10 kilometers range," Opt. Express **25**(10), 11919–11931 (2017).
9. G. O. Williams, E. Williams, N. Finlayson, *et al.*, "Full spectrum fluorescence lifetime imaging with 0.5 nm spectral and 50 ps temporal resolution," Nat. Commun. **12**(1), 6616 (2021).
10. Z. Zang, D. Xiao, Q. Wang, *et al.*, "Compact and robust deep learning architecture for fluorescence lifetime imaging and FPGA implementation," Methods Appl. Fluoresc. **11**(2), 025002 (2023).
11. D. D.-U. Li, J. Arlt, D. Tyndall, *et al.*, "Video-rate fluorescence lifetime imaging camera with CMOS single-photon avalanche diode arrays and high-speed imaging algorithm," J. Biomed. Opt. **16**(09), 1–096012 (2011).
12. R. M. Field, S. Realov, and K. L. Shepard, "A 100 fps, time-correlated single-photon-counting-based fluorescence-lifetime imager in 130 nm CMOS," IEEE J. Solid-State Circuits **49**(4), 867–880 (2014).
13. W. Liu, R. Qian, S. Xu, *et al.*, "Fast and sensitive diffuse correlation spectroscopy with highly parallelized single photon detection," APL Photonics **6**(2), 1 (2021).
14. J. Giles, I. Bankman, R. Sova, *et al.*, "Lidar system model for use with path obscurants and experimental validation," Appl. Opt. **47**(22), 4085–4093 (2008).
15. T. G. Phillips, N. Guenther, and P. R. McAree, "When the dust settles: The four behaviors of lidar in the presence of fine airborne particulates," J. Field Robot. **34**(5), 985–1009 (2017).
16. R. Tobin, A. Halimi, A. McCarthy, *et al.*, "Robust real-time 3D imaging of moving scenes through atmospheric obscurant using single-photon LiDAR," Sci. Rep. **11**(1), 11236 (2021).
17. J. Mau, V. Devrelis, G. Day, *et al.*, "Through thick and thin: Imaging through obscurant using spad array," in *2020 IEEE SENSORS* (IEEE, 2020), pp. 1–4.
18. T.-H. Sang and C.-M. Tsai, "Histogram-based defogging techniques for LiDAR," in *2022 IEEE 16th International Conference on Solid-State & Integrated Circuit Technology (ICSICT)* (IEEE, 2022), pp. 1–3.
19. T.-H. Sang, Y.-C. Lin, and Y.-C. Hsiao, "On Training Deep-learning Models for Removing Airborne-Particle Points from SPAD LiDAR Multi-Echo Point Clouds," IEEE Sens. Lett. **1**, 1 (2023).
20. G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," Neurocomputing **70**(1-3), 489–501 (2006).
21. M. Tsukada, M. Kondo, and H. Matsutani, "A neural network-based on-device learning anomaly detector for edge devices," IEEE Trans. Comput. **69**(7), 1 (2020).
22. Ó. Mata-Carballeira, I. del Campo, V. Martínez, *et al.*, "A hardware/software extreme learning machine solution for improved ride comfort in automobiles," in *2019 International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2019), pp. 1–8.
23. G. Satat, M. Tancik, and R. Raskar, "Towards photography through realistic fog," in *2018 IEEE International Conference on Computational Photography (ICCP)* (IEEE, 2018), pp. 1–10.
24. G. Satat, M. Tancik, O. Gupta, *et al.*, "Object classification through scattering media with deep learning on time resolved measurement," Opt. Express **25**(15), 17466–17479 (2017).
25. A. Maccarone, G. Acconcia, U. Steinlehner, *et al.*, "Custom-technology single-photon avalanche diode linear detector array for underwater depth imaging," Sensors **21**(14), 4850 (2021).
26. R. Tobin, A. Halimi, A. McCarthy, *et al.*, "Three-dimensional single-photon imaging through obscurants," Opt. Express **27**(4), 4590–4611 (2019).
27. S. Jungerman, A. Ingle, Y. Li, *et al.*, "3D scene inference from transient histograms," in *European Conference on Computer Vision* (Springer, 2022), pp. 401–417.
28. C. Sifferman, Y. Wang, M. Gupta, *et al.*, "Unlocking the Performance of Proximity Sensors by Utilizing Transient Histograms," IEEE Robot. Autom. Lett. **8**(10), 6843–6850 (2023).
29. C. Callenberg, Z. Shi, F. Heide, *et al.*, "Low-cost SPAD sensing for non-line-of-sight tracking, material classification and depth imaging," ACM Trans. Graph. TOG **40**(4), 1–12 (2021).
30. A. Ruget, M. Tyler, G. Mora Martín, *et al.*, "Pixels2pose: Super-resolution time-of-flight imaging for 3d pose estimation," Sci. Adv. **8**(48), eade0123 (2022).
31. D. Palossi, N. Zimmerman, A. Burrello, *et al.*, "Fully onboard ai-powered human-drone pose estimation on ultralow-power autonomous flying nano-uavs," IEEE Internet Things J. **9**(3), 1913–1929 (2022).
32. Y. Zhang, Y. Wang, G. Zhou, *et al.*, "Multi-kernel extreme learning machine for EEG classification in brain-computer interfaces," Expert Syst. Appl. **96**, 302–310 (2018).
33. Y.-C. Chuang, Y.-T. Chen, H.-T. Li, *et al.*, "An arbitrarily reconfigurable extreme learning machine inference engine for robust ECG anomaly detection," IEEE Open J. Circuits Syst. **2**, 196–209 (2021).
34. Z. Zang, D. Xiao, Q. Wang, *et al.*, "Fast analysis of time-domain fluorescence lifetime imaging via extreme learning machine," Sensors **22**(10), 3758 (2022).
35. A. O. Abuassba, D. Zhang, X. Luo, *et al.*, "Improving classification performance through an advanced ensemble based heterogeneous extreme learning machines," Comput. Intell. Neurosci. **2017**(1), 1–11 (2017).
36. A. Turpin, G. Musarra, V. Kapitany, *et al.*, "Spatial images from temporal data," Optica **7**(8), 900–905 (2020).

Optics EXPRESS

37. P. Kirkland, V. Kapitany, A. Lyons, *et al.*, "Imaging from temporal data via spiking convolutional neural networks," in *Emerging Imaging and Sensing Technologies for Security and Defence V; and Advanced Manufacturing Technologies for Micro-and Nanosystems in Security and Defence III* (SPIE, 2020), 11540, pp. 66–85.
38. Y. Yoo and S.-Y. Oh, "Fast training of convolutional neural network classifiers through extreme learning machines," in *2016 International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2016), pp. 1702–1708.