

EOSAM 2022

Guest editors: Patricia Segonds, Gilles Pauliat and Emiliano Descrovi

SHORT COMMUNICATION

OPEN ACCESS

Exploring the hidden dimensions of an optical extreme learning machine

Duarte Silva^{1,2,*}, Tiago Ferreira^{1,2}, Felipe C. Moreira^{1,2}, Carla C. Rosa^{1,2}, Ariel Guerreiro^{1,2}, and Nuno A. Silva^{1,2}

¹Departamento de Física e Astronomia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal

²INESC TEC, Centre of Applied Photonics, Rua do Campo Alegre 687, 4169-007 Porto, Portugal

Received: 7 December 2022 / Accepted: 10 January 2023

Abstract. Extreme Learning Machines (ELMs) are a versatile Machine Learning (ML) algorithm that features as the main advantage the possibility of a seamless implementation with physical systems. Yet, despite the success of the physical implementations of ELMs, there is still a lack of fundamental understanding in regard to their optical implementations. In this context, this work makes use of an optical complex media and wavefront shaping techniques to implement a versatile optical ELM playground to gain a deeper insight into these machines. In particular, we present experimental evidences on the correlation between the effective dimensionality of the hidden space and its generalization capability, thus bringing the inner workings of optical ELMs under a new light and opening paths toward future technological implementations of similar principles.

Keywords: Extreme Learning Machine, Optical Computing, Machine Learning, Optics.

1 Introduction

Over the last decades, Artificial Neural Networks (ANNs) were established as a powerful computing architecture across numerous fields of science and technology [1, 2]. Part of its success is linked to the scalability and versatility of the neuromorphic architecture, which with the impending plateau of Moore's law is now pushing towards the development of novel computing hardware, capable of bypassing the limits of electronics miniaturization [3]. Indeed, as the amount of different mathematical operations involved in such algorithms is not vast, mainly involving matrix multiplications and non-linear activation functions, the development of hardware accelerators for ANNs has become an attractive topic of research [4].

In this context, optical-based implementations appear particularly promising, offering non-trivial advantages when compared with electronic devices. Indeed, with the ability to handle information at the speed of light allied to multiplexing capabilities, optical information processing systems have the native potential for fast, massively parallelizable, and energy-efficient approaches. Nonetheless, realizing conventional ANNs with optics requires establishing precise neuron connections which can be quite hard to achieve, often limited by fabrication procedures, materials

or device imperfections. This, in turn, makes the already intensive training procedures largely ineffective.

For all these reasons, architectures that can bypass the tuning of all the weights have been increasingly explored for hardware development, from which we can highlight the implementations using Reservoir Computing (RC) [5] and Extreme Learning Machines (ELMs) [6]. In simple terms, the underlying concept of both models is to use a fixed *reservoir* to non-linearly project the input information onto a high-dimensional hidden space. The training process then occurs only between the hidden layer and the output layer, which strongly reduces the computation complexity and softens the requirements for hardware deployment.

In particular, optical ELMs have already been demonstrated through the use of complex optical media [7] and multimode fibers [8, 9], and in principle, many other optical phenomena can also be used to achieve such architectures. For instance, ELMs based on $\chi^{(3)}$ materials have been demonstrated numerically [10, 11], and experimentally [9]. Still, most of the works remain largely empirical, lacking a fundamental understanding of such machines. In this work, we study and implement an optical ELM based on strongly scattering media that is able to process information encoded either in the spatial distribution of the amplitude or the phase of a continuous wave optical beam. Introducing a simple model for the amplitude case, we study the dimensionality of the hidden space experimentally, as a function of different encoding schemes with linear and

* Corresponding author: duartejfs@hotmail.com

nonlinear intensity measurements. Benchmarking the device on standard ML regression and classification tasks, our results demonstrate the important role played by non-linearity in the deployment of effective optical extreme learning machines.

2 Theoretical framework

In simplified terms, the inner workings of an ELM consist in taking a N_I -dimensional input \mathbf{X} and feed it to an untrained hidden or reservoir layer, recording its output. Thus, for each $\mathbf{X}^{(i)}$ of the dataset, we obtain a N_o -dimensional vector $\mathbf{Y}^{(i)}$ given by

$$\mathbf{Y}^{(i)} = \begin{bmatrix} G_1(\mathbf{w}_1 \mathbf{X}^{(i)}) \\ \vdots \\ G_{N_o}(\mathbf{w}_{N_o} \mathbf{X}^{(i)}) \end{bmatrix}, \quad (1)$$

where G describes the dynamics of the hidden layer and is commonly referred to as the activation function, and \mathbf{w}_j a vector of weights for each output channel j . The ELM strategy is now to use this output and multiply it by an output weight vector $\boldsymbol{\beta} = [\beta_1, \dots, \beta_{N_o}]^T$ at the hidden layer to obtain a prediction for a given task as

$$\mathbf{P}(\mathbf{X}^{(i)}) = \sum_{j=1}^{N_o} \beta_j \mathbf{Y}_j^{(i)} = \mathbf{Y}^{(i)} \cdot \boldsymbol{\beta}. \quad (2)$$

Put in this way, it is straightforward to see that training an ELM to perform a task is reduced to simply computing a linear transformation performed by the output weight vector $\boldsymbol{\beta}$. One way to perform this while preventing the overfitting of the model is to fit the vector $\boldsymbol{\beta}$ via Ridge regression by minimizing a regularized loss function

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^{N_o \times N_T}} \|\mathbf{H}\boldsymbol{\beta} - \mathbf{T}\|^2 + \lambda \|\boldsymbol{\beta}\|^2, \quad (3)$$

where $\|\cdot\|$ denotes the Frobenius norm, and λ the regularization parameter. To perform this minimization, we take each element of input and associated target data of a training dataset, say pairs $\{\mathbf{X}^{(i)}, \mathbf{T}^{(i)}\}_{i=1}^N$, to construct the matrix \mathbf{H} by stacking all the output states as rows, i.e. $\mathbf{H}_{ij} = \mathbf{Y}_j^{(i)}$. Constructing the matrix \mathbf{T} with the associated targets stacked in the same way, equation (3) has then an analytical solution given by (for $N > N_o$)

$$\boldsymbol{\beta} = (\mathbf{H}^T \mathbf{H} + \lambda \mathbf{I})^{-1} \mathbf{H}^T \mathbf{T}, \quad (4)$$

where \mathbf{I} is the identity matrix.

In theory, and as it happens in neural network architectures, the performance of an ELM is intrinsically connected with the dimensionality of the hidden space and its activation function. Indeed, from the literature, it is mathematically shown that as long as (i) the weights are drawn from a random distribution and (ii) the activation function G is a nonlinear piecewise continuous function, the ELM will feature universal approximation capabilities on a

hidden space of dimensions equal or below the dimensionality of the training dataset [6]. Yet, we shall notice that fulfilling these conditions does not warrant by itself the deployment of a working algorithm that is able to generalize well for the task, nor being robust to external noise. As in most neural network architectures, the generalization performance is typically task-specific and shall be discussed for each case individually by taking into consideration the nature of the activation functions.

3 Implementation of an optical ELM

Our optical implementation of an ELM is based on wavefront shaping techniques and is schematically described in Figure 1, establishing the connection with the ELM framework. In short, we first make use of a Digital Micromirror Device (DMD), capable of both amplitude and phase modulation enabled by Lee holography [12], as the optical encoder to create the input state. The light is then coupled to a multimode fiber via a standard fiber collimator for our working wavelength, which works as the reservoir where the information is mixed. At the exit, the optical field is a speckle pattern that is known to possess Gaussian circular statistics [13] and guarantees the randomness required by an ELM. This pattern is then measured on a high-speed CMOS camera both in the linear and non-linear regime, which constitutes our hidden layer. Upon correct synchronization, the system can work within the kHz rate, limited by the detection and digital processing steps.

In particular, when using amplitude encoding with two distinct encoding regions, as depicted in Figure 1, we can make use of the properties of the optical transmission matrix \mathbf{M} to express the output field at the camera image plane as

$$G_i = \int dx_{\Delta_x(l-1)}^{\Delta_x(l)} \int dy_{\Delta_y(m-1)}^{\Delta_y(m)} \times F\left(|\mathbf{M}(E_{\text{ref}} + f_1(\mathbf{X}^{(i)})E_1 + f_2(\mathbf{X}^{(i)})E_2)|^2\right) \quad (5)$$

detected in the macropixel $i = (l, m)$ with $l \in \{1, \dots, N_x\}$, $m \in \{1, \dots, N_y\}$ and $N_x \times N_y = N_o$. Furthermore, the camera detection function F can be either linear $F(I) = I$ (no saturation, low exposure time) or nonlinear $F(I) = I/(I + I_{\text{sat}})$ (saturation, higher exposure time), thus corresponding to distinct activation functions.

4 Results and discussion

To understand the capabilities of our setup, we have tested it in standard regression and classification tasks. In specific, for the regression task we used a dataset of points randomly sampled from the function $f(x) = \frac{\sin(2\pi x)}{2\pi x}$. For the classification task we used a dataset of points based on the curves $x_1(\theta) = (2\theta + \pi) \cos(\theta)$, $x_2(\theta) = (-2\theta - \pi) \cos(\theta)$, $y_1(\theta) = (2\theta + \pi) \sin(\theta)$ and $y_2(\theta) = (-2\theta - \pi) \sin(\theta)$, where a

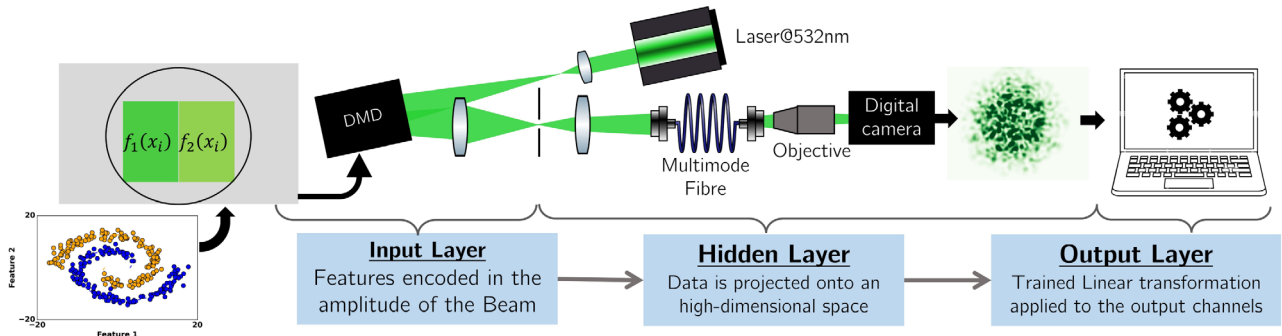


Fig. 1. Illustration of the setup for the implementation of an optical ELM. The information encoding is performed on the DMD, from where the optical signal follows to a multimode fiber producing a speckle pattern which is collected with a digital camera, constituting the hidden reservoir layer. The weights are then calculated digitally to be applied at the hidden layer to get a prediction.

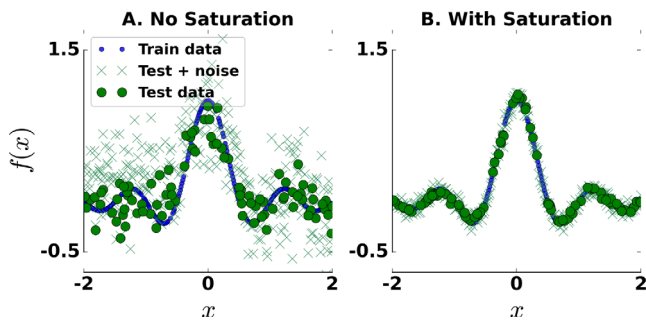


Fig. 2. Regression performance under amplitude modulation. In addition to the results of the 80–20% holdout strategy, we also represent a test for the robustness of the implementation by testing for a dataset with 5% of additional white noise at the end of the hidden layer.

sample j from class i consists of a pair of points $\{x_i(\theta_j) + \mathcal{N}_j, y_i(\theta_j) + \mathcal{N}_j\}$, where θ_j is sampled from a uniform distribution $\mathcal{U}(0, 2\pi)$ and \mathcal{N}_j is added random noise from a distribution $\mathcal{U}(0, 1)$. For both methods, we have used a total of 300 samples and trained with 80% of the whole dataset and tested the performance in the additional 20%. For both procedures, in order to encode the information in the optical domain, we have defined $A(q_i) = \frac{q_i - q_{\min}}{q_{\max} - q_{\min}}$, where q_i is a generalized coordinate, and q_{\max} and q_{\min} are the greatest and lowest coordinates within the dataset, respectively. For the scope of this manuscript we will only analyse the results for amplitude modulation, obtained by aggregating groups of DMD pixels resulting in various modulation levels.

In **Figure 2** we present the results for the regression task. First, it is straightforward to see that the saturation regime increases the performance both for the training and test datasets. This observation matches our empirical expectation and can be confirmed by making a connection with the dimensionality of the hidden space. To achieve this, we computed the rank of the output matrix \mathbf{H} by making use of the singular value spectrum. Still, we should take into consideration the effect of experimental noise, which can artificially increase the dimensionality of the hidden space. Anchored on Weyls inequality [14], we did this by counting

Table 1. Summary of the results for the Regression task, with the Rank and Mean Absolute Error (MAE) metric obtained through a 5-fold cross validation procedure.

Saturation	Rank (\mathbf{H})	MAE (Train)	MAE (Test)
Off	3	0.1	0.16
On	8	0.0009	0.02

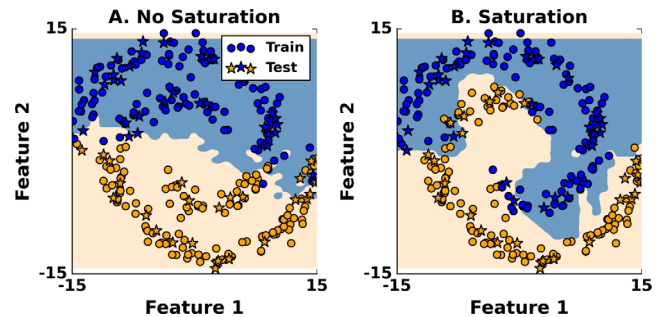


Fig. 3. Results for single fold for the classification task under amplitude modulation.

the number of singular values of \mathbf{H} above the highest singular value of the noise matrix of the i -th experiment $\mathbf{N}_i = \mathbf{H}_i - \langle \mathbf{H} \rangle$, where $\langle \mathbf{H} \rangle$ represents the average over 100 experiments.

As it can be seen in **Table 1**, the ELM performance increases with the rank. This happens because while both activation functions are nonlinear, the non-saturated regime only provides a second-degree polynomial while the saturation regime features a saturable response which can only be approximated by a higher order polynomial, effectively increasing the dimensionality of the hidden space.

Regarding the classification task, a benchmark result is found in **Figure 3**, together with a summary in **Table 2**. Again and as expected, the camera saturation results in increasing the dimensionality of the hidden space, allowing us to achieve higher accuracy. Also, it is interesting to see that the methodology provides a good generalization performance, separating the regions as intended.

Table 2. Summary of the Accuracy (Acc.) results for the classification task obtained through a 5-fold cross validation procedure.

Saturation	Rank (H)	Acc. (Train)	Acc. (Test)
Off	6	72%	71%
On	12	94%	92%

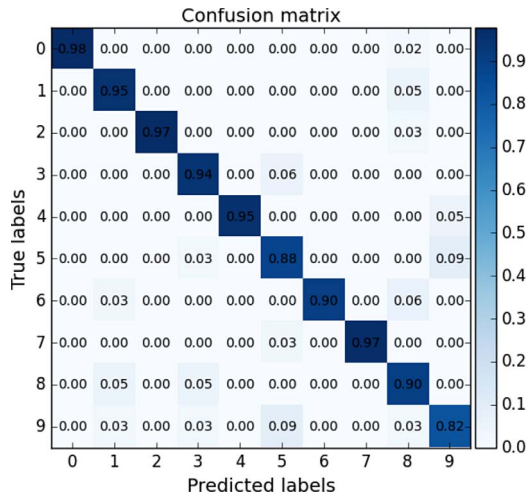


Fig. 4. Confusion matrix on the MNIST dataset. These results amount to a macro average accuracy and precision of 93% and a recall of 92%.

Finally, to test the performance of the optical ELM in more complex tasks such as processing and classifying images, we tested the setup on the classification of handwritten digits through the MNIST dataset (1797 images, with the same 80–20% holdout strategy) [15]. Overall, we obtained accuracies around 93%, with a confusion matrix depicted in Figure 4.

5 Final remarks

In this work, we demonstrated the implementation of an optical extreme learning machine that is able to process information encoded in the wavefront of an optical beam by making use of a multimode fiber and a camera detector. Using both standard regression and classification tasks, we have shown that the setup is capable of achieving good computing performances. Furthermore, by studying the dimensionality of the hidden space and comparing it against performance and generalisation capability, we have demonstrated a correlation between the two which aligns with the theoretical predictions. In particular, an increase of the performance can be obtained by including physical nonlinearities within the system, which is done using the saturation of the detection system. Put into perspective, the findings enclosed confirm the optical ELMs as a

promising platform for versatile non-Von Neumann analog computing, while simultaneously paving the way for a better understanding of such devices.

Acknowledgments. This work is financed by National Funds through the Portuguese funding agency, FCT – Fundação para a Ciência e a Tecnologia, within project UIDB/50014/2020. T.D.F. is supported by Fundação para a Ciência e a Tecnologia through Grant No. SFRH/BD/145119/2019.

References

- 1 LeCun Y., Bengio Y., Hinton G. (2015) Deep learning, *Nature* **521**, 7553, 436–444.
- 2 Barucci A., Cucci C., Franci M., Loschiavo M., Argenti F. (2021) A deep learning approach to ancient egyptian hieroglyphs classification, *IEEE Access* **9**, 123438–123447.
- 3 Shalf J. (2020) The future of computing beyond Moore’s law, *Philos. Trans. A Math. Phys. Eng. Sci.* **378**, 2166, 20190061.
- 4 Xingyuan X., Tan M., Corcoran B., Jiayang W., Boes A., Nguyen T.G., Chu S.T., Little B.E., Hicks D.G., Morandotti R., Mitchell A., Moss D.J. (2021) 11 TOPS photonic convolutional accelerator for optical neural networks, *Nature* **589**, 7840, 44–51.
- 5 Schrauwen B., Verstraeten D., Campenhout J. (2007) An overview of reservoir computing: theory, applications and implementations, in: Proceedings of the 15th European symposium on artificial neural networks, Bruges, Belgium, April 25–27, pp. 471–482. <https://www.esann.org/proceedings/2007>.
- 6 Huang G.-B., Zhou H., Ding X., Zhang R. (2012) Extreme learning machine for regression and multiclass classification, *IEEE Trans. Syst. Man. Cybern. B Cybern.* **42**, 2, 513–529.
- 7 Saade A., Caltagirone F., Carron I., Daudet L., Dremeau A., Gigan S., Krzakala F. (2016) Random projections through multiple optical scattering: Approximating kernels at the speed of light, in: 2016 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE.
- 8 Sunada S., Kanno K., Uchida A. (2020) Using multidimensional speckle dynamics for high-speed, large-scale, parallel photonic computing, *Opt. Express* **28**, 21, 30349.
- 9 Teğin U., Yıldırım M., Oğuz İ., Moser C., Psaltis D. (2021) Scalable optical learning operator, *Nat. Comput. Sci.* **1**, 8, 542–549.
- 10 Silva N.A., Ferreira T.D., Guerreiro A. (2021) Reservoir computing with solitons, *New J. Phys.* **23**, 2, 023013.
- 11 Marcucci G., Pierangeli D., Conti C. (2020) Theory of neuromorphic computing by waves: Machine learning by rogue waves, dispersive shocks, and solitons, *Phys. Rev. Lett.* **125**, 9, 093901.
- 12 Lee W.-H. (1979) Binary computer-generated holograms, *Appl. Opt.* **18**, 21, 3661–3669.
- 13 Goodman J.W. (2020) Speckle phenomena in optics: Theory and applications, 2nd ed., SPIE.
- 14 Horn R.A. (2012) *Matrix analysis*, 2nd ed., Cambridge University Press. <https://doi.org/10.1017/CBO9781139020411>.
- 15 Deng L. (2012) The mnist database of handwritten digit images for machine learning research, *IEEE Signal Process. Mag.* **29**, 6, 141–142.