Diffractive Optical Neural Networks with Arbitrary Spatial Coherence

Matthew Filipovich Aleksei Malyshev A. I. Lvovsky
Department of Atomic and Laser Physics
University of Oxford
{matthew.filipovich, aleksei.malyshev, alex.lvovsky}
@physics.ox.ac.uk

Abstract

Diffractive optical neural networks (DONNs) have emerged as a promising optical hardware platform for ultra-fast and energy-efficient signal processing. However, previous experimental demonstrations of DONNs have only been performed using coherent light, which is not present in the natural world. Here, we study the role of spatial optical coherence in DONN operation. We propose a numerical approach to efficiently simulate DONNs under input illumination with arbitrary spatial coherence and discuss the corresponding computational complexity using coherent, partially coherent, and incoherent light. We also investigate the expressive power of DONNs and examine how coherence affects their performance. We show that under fully incoherent illumination, the DONN performance cannot surpass that of a linear model. As a demonstration, we train and evaluate simulated DONNs on the MNIST dataset using light with varying spatial coherence.

1 Introduction

Machine learning models are currently executed using specialized electronic hardware, such as graphics processing units (GPUs) and tensor processing units (TPUs), which harness immense processing power and data parallelism. However, the growing compute requirements from advanced deep learning models are far outpacing hardware improvements anticipated by Moore's law scaling [1]. Given the constraints imposed by digital electronics, optics has gained recognition as a promising platform for machine learning applications with low latency, high bandwidth, and low energy consumption [2].

Diffractive optical neural networks (DONNs) are specialized hardware architectures that harness diffraction effects to process optical signals in free space [3; 4]. DONNs are generally composed of several successive modulation surfaces, denoted as diffractive layers, that modify the phase and/or amplitude of the incident optical signals through light-matter interactions, as shown in Fig. 1. The diffractive layers contain discrete pixels, each with an independent complex-valued transmittance coefficient. The output of the DONN corresponds to the total intensity of the optical field incident on designated detection regions in the output plane.

DONNs are usually trained *in silico*, i.e., the physical DONN is modeled using a computer to simulate the evolution of the optical signals through the system. The modulation patterns of the diffractive layers are optimized to achieve the desired transformation between the input and target output of the DONN, which is analogous to optimizing the weights in standard neural network models [5]. During training, the transmittance coefficient of each pixel in the diffractive layers is iteratively updated using an optimization algorithm to minimize the error in the model's output with respect to the training set. The backpropagation algorithm is used to efficiently calculate the gradient of the loss with respect to the transmittance coefficients [6].

37th First Workshop on Machine Learning with New Compute Paradigms at NeurIPS 2023 (MLNPCP 2023).

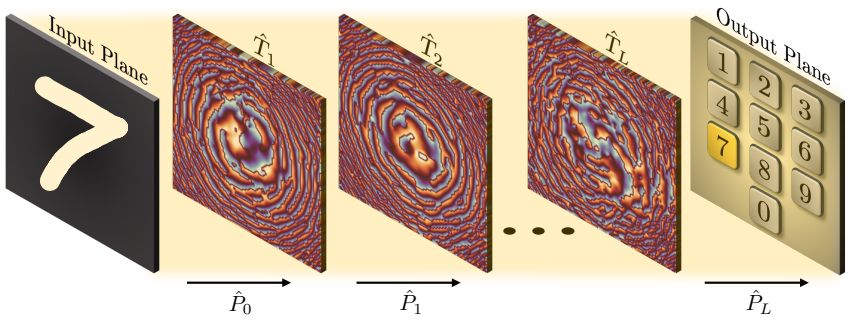

Figure 1: Illustration of a DONN trained to identify handwritten digits. The DONN is comprised of L diffractive layers that modulate the optical field as it propagates through the system. The output plane encompasses ten detection regions, which are each associated with a unique digit, and the predicted output corresponds to the region with the highest optical intensity. The transmission and propagation operators at the l-th layer are denoted by $\hat{T}_l$ and $\hat{P}_l$, respectively.

DONNs are particularly well-suited for use with real-world optical signals, as the optical fields can be directly fed into the system. However, such signals are typically incoherent. In contrast, most of the existing experimental work with DONNs is performed using fully-coherent illumination from laser sources. In this paper, we introduce a computationally efficient framework for simulating and training DONNs using incident illumination with arbitrary spatial coherence and evaluate the performance of simulated DONNs trained on the MNIST dataset of handwritten digits.

2 DONN operation with spatial coherence

2.1 Coherent illumination

In this section, we introduce a formalism for describing the evolution of coherent, monochromatic optical fields through DONNs using scalar diffraction theory [7]. We treat the optical field as a complex scalar quantity and employ Dirac notation to represent the transverse profile of the field at discrete spatial positions using ket-vectors. This discretization does not affect the generality of our treatment as long as the spatial sampling interval (i.e., pixel pitch) is much smaller than the characteristic transverse field feature size. The transformations applied to the field as it evolves through the DONN, which include free-space propagation and transmission through modulation surfaces, are expressed using linear operators.

At each layer in the DONN, the optical field is modulated by the diffractive surface and subsequently propagates through free space to the next layer. The incident field at the m-th discrete pixel of the l-th diffractive layer, before modulation, is represented by $\psi_l(m)$, where the time dependence of the signal is absent. The transverse profile of the field can be expressed using Dirac notation as $|\psi_l\rangle = \sum_m \psi_l(m)\,|m\rangle$, where the set $\{|m\rangle\}$ of all pixels forms an orthonormal basis. The mapping between the optical fields in the l-th and (l+1)-th layers can be expressed as $|\psi_{l+1}\rangle = \hat{P}_l\hat{T}_l\,|\psi_l\rangle$, where $\hat{T}_l$ and $\hat{P}_l$ are the transmission and free-space propagation operators, respectively.

The transmission operator $\hat{T}_l$ describes the phase and/or amplitude modulation applied to the optical field by each pixel in the l-th diffractive layer. The corresponding matrix is diagonal:

$$\hat{T}_l = \sum_m t_l(m) \cdot |m\rangle\langle m|\,, \tag{1}$$

where $t_l(m)$ is the complex-valued transmittance coefficient at the m-th pixel in the l-th diffractive layer, which satisfies $|t_l(m)| \leq 1$.

The operator $\hat{P}_l$ describes the free-space propagation of the field between the l-th and (l+1)-th diffractive layers using the Rayleigh-Sommerfeld solution [7]:

$$\hat{P}_l = \sum_{m,n} h(m,n) \cdot |n\rangle\langle m|\,, \tag{2}$$

with

$$h(m,n) = \frac{1}{i\lambda} \exp\left(\frac{i2\pi r(m,n)}{\lambda}\right) \frac{d}{r(m,n)^2} \tag{3}$$

being the point-spread function, i.e., the amplitude distribution in the (l+1)-th layer if only the m-th pixel of the l-th layer is illuminated. In the above equation, λ is the wavelength of the coherent optical signal (the central wavelength for quasimonochromatic light), d is the axial distance between the two diffractive layers, and $r(m,n)$ is the Euclidean distance between the m-th and n-th pixels in the l-th and (l+1)-th layers, respectively. The above expression is valid when the axial distance between layers is much greater than the (central) wavelength of light.

The input image processed by the DONN is encoded in the initial field $|\psi_{\text{in}}\rangle$. The output optical field, represented by $|\psi_{\text{out}}\rangle$, can then be expressed as

$$|\psi_{\text{out}}\rangle = \hat{U}\,|\psi_{\text{in}}\rangle\,, \tag{4}$$

where $\hat{U}$ is the evolution operator of the DONN that maps the input optical field onto the output field (i.e., the spatial impulse response of the system):

$$\hat{U} = \prod_{l=1}^{L} \left(\hat{P}_l \hat{T}_l\right) \hat{P}_0, \tag{5}$$

where the DONN has L diffractive layers. At the output plane of the DONN, the intensity of the evolved field is measured using image sensors:

$$I_{\text{out}}(n) = |\,\langle n|\psi_{\text{out}}\rangle\,|^2 = |\psi_{\text{out}}(n)|^2. \tag{6}$$

For classification tasks, the output of the DONN corresponding to each class c is defined as the total intensity incident on a specified spatial detection region $\mathcal{D}_c$ in the output plane:

$$o(c) = \sum_{n \in \mathcal{D}_c} I_{\text{out}}(n). \tag{7}$$

Training DONNs using a computer (i.e., *in silico*) requires simulating the evolution of coherent optical fields through the system. The calculated optical field at each layer is then used during the backward pass to compute the gradient of the loss function with respect to the diffractive layer transmittance coefficients. The propagation operator $\hat{P}_l$ can be evaluated in $\mathcal{O}(N \log N)$ time by utilizing the fast Fourier transform algorithm [8]. Additionally, the transmission operator $\hat{T}_l$ is described by a diagonal matrix and can be evaluated in $\mathcal{O}(N)$ time. Therefore, calculating the evolution of B different input fields through a DONN with L layers and N pixels per layer has a computational complexity of $\mathcal{O}(BLN \log N)$, and the backward pass has the same complexity.

2.2 Arbitrary spatial coherence illumination

Using DONNs for real-world applications requires the ability to process incoherent and partially coherent light. We assume quasimonochromatic illumination conditions, which is a good approximation for many cases. These conditions require that the input light is narrowband and its coherence length is much greater than the maximum path length difference between diffractive layers [9]. At the same time, we assume the coherence time to be much shorter than the inverse detection bandwidth, so the detection averages in time over the non-stationary interference pattern.

The spatial coherence of the optical field in the l-th layer is characterized by the mutual intensity function, which determines the time-averaged correlation of the field at two separate pixels [9; 10]:

$$J_l(m,m') = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} \psi_l(m;t)\,\psi_l^*(m';t)\,\mathrm{d}t,$$

where T is the detection time. This matrix represents an operator

$$\hat{J}_l = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} |\psi_l(t)\rangle\langle\psi_l(t)|\,\mathrm{d}t, \tag{8}$$

Table 1: Computational complexity of modeling the evolution of B input examples through a DONN with L layers and N pixels per layer under different illumination conditions.

Spatial Coherence	Computational Complexity
Coherent	$\mathcal{O}(BLN \log N)$
Arbitrary Coherence	$\mathcal{O}(BLN^2 \log N)$
Incoherent	$\mathcal{O}(BN^2) + \mathcal{O}(LN^2 \log N)$

where $J_l(m, m') = \langle m|\hat{J}_l|m'\rangle$. The time-averaged intensity of the field is given by the diagonal elements of the mutual intensity operator, such that

$$I_l(m) = J_l(m, m). \tag{9}$$

Similar to the evolution of coherent fields through DONNs, the evolution of the mutual intensity operator can be expressed using the transmission and propagation operators. The input mutual intensity operator $\hat{J}_{\text{in}}$ describes the spatial coherence of the initial field that encodes the input image to be processed by the DONN. The output mutual intensity operator is given by

$$\hat{J}_{\text{out}} = \hat{U}\, \hat{J}_{\text{in}}\, \hat{U}^{\dagger}, \tag{10}$$

where $\hat{U}$ is the evolution operator of the DONN defined in Eq. (5). The output of the DONN corresponds to the total time-averaged intensity, defined in Eq. (9), incident on the spatial detection regions along the output plane:

$$o(c) = \sum_{n \in \mathcal{D}_c} J_{\text{out}}(n, n). \tag{11}$$

The evolution of the mutual intensity operator and the corresponding DONN output can be simulated on a computer using Eqs. (10) and (11). Analogous to the previously discussed method, the fast Fourier transform can be leveraged to evaluate the propagation operator $\hat{P}_l$ applied to an arbitrary mutual intensity operator described by an $N \times N$ matrix, which scales as $\mathcal{O}(N^2 \log N)$. The transmission operator $\hat{T}_l$ can similarly be evaluated in $\mathcal{O}(N^2)$ time. Hence, simulating the evolution of B different input fields with arbitrary spatial coherence through a DONN with L layers and N pixels per layer has a computational complexity of $\mathcal{O}(BLN^2 \log N)$. The backward pass executed during training has the same computational complexity.

The computational cost of simulating DONNs with fully incoherent illumination can be amortised for multiple input examples using the impulse response that characterizes the system. A summary of the computational complexities of simulating DONNs under coherent, arbitrary coherence, and incoherent illumination is shown in Table 1.

3 Results

3.1 Expressivity of DONNs

The expressive power of DONNs is dependent on the spatial coherence of the input light. Under coherent illumination, DONNs have been shown to outperform linear models [3]. Since the coherent field evolves linearly through the system, this improvement in performance results from the nonlinear intensity measurement of the complex-valued field at the output plane (6), followed by the linear summation of the intensities over the detection regions (7). Thus, DONNs using coherent illumination can be understood as standard neural networks that consist of a complex-valued linear layer with a nonlinear activation function, followed by a real-valued linear layer.

In contrast, under incoherent illumination, the time-averaged output intensity is the sum of the intensity patterns from individual pixel sources in the input plane. Therefore, DONNs with incoherent input illumination cannot perform better than a linear model, as the time-averaged input and output intensity distributions are linearly related according to the intensity impulse response.

The improved expressive power of a coherently illuminated DONN arises from the off-diagonal elements in the input mutual intensity operator, which are absent for incoherent light. These elements

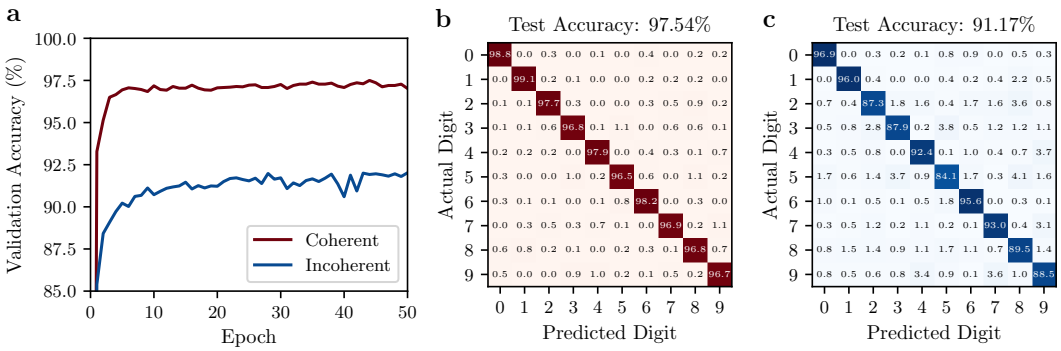

Figure 2: DONN training results on the MNIST dataset under coherent and incoherent illumination. **a** Validation accuracy attained by the models at each epoch during training. **b, c** Confusion matrix of the model trained using coherent (b) and incoherent (c) illumination evaluated on the test set.

represent the spatial coherence between two different pixels in the input image. For an arbitrary input mutual intensity operator J_{in}, the output intensity can be expressed, using Eqs. (9) and (10), as

$$I_{\text{out}}(n) = \sum_{m,m'} J_{\text{in}}(m,m') \cdot \langle n|\hat{U}|m\rangle \cdot \langle m'|\hat{U}^{\dagger}|n\rangle \,. \tag{12}$$

This summation includes off-diagonal elements of the mutual intensity operator, which depend nonlinearly on the field. Due to this nonlinearity, the performance of DONNs under partially coherent illumination can surpass that with incoherent light, as demonstrated in the following section.

3.2 Performance on MNIST dataset

Using the formalism introduced in the previous section, we trained simulated DONN models to identify handwritten digits from zero to nine using incoherent, partially coherent, and coherent illumination. The models were trained over 50 epochs using 55,000 images (5,000 images for validation) from the MNIST dataset, each consisting of 28×28 pixels [11]. The DONNs are composed of five diffractive layers, each with 100×100 pixels, which modulate the phase of the incident light and are spaced 5 cm apart. Each model was trained using a uniform, normalized optical field incident on the input image of the handwritten digit with a central wavelength of 700 nm. The cross-entropy loss function was used during training to calculate the output error of the model. Each pixel in the diffractive layers has a surface area of 10×10 μm^2, while each pixel in the input pattern is 30×30 μm^2. Each detection region in the output plane, which is associated with a unique digit, is 250×250 μm^2.

The spatial coherence of the input illumination is given by $J_{\text{in}}(m,m') = \sqrt{I_{\text{in}}(m) I_{\text{in}}^{*}(m')}\, \mu^{r_{\text{norm}}(m,m')}$, where $I_{\text{in}}(m)$ is the time-averaged intensity at the m-th pixel in the input image, $r_{\text{norm}}(m,m')$ is the Euclidean distance between the m-th and m'-th input pixels, normalized by the pixel pitch, and μ quantifies the degree of spatial coherence: $\mu = 1$ for fully coherent and $\mu = 0$ for fully incoherent light. We first trained two DONN models to process handwritten digits using fully coherent and incoherent illumination. During the training phase, we saved the model parameters that yielded the highest validation accuracy. We then evaluated the performance of these models using a test set of 10,000 images that were not shown during training. The models trained using coherent and incoherent light achieved test accuracies of 97.54% and 91.17%, respectively. The validation accuracy attained during training, as well as the performance of the models on the test set, are shown in Fig. 2.

We then trained DONNs using input illumination with partially coherent light, where each model was trained to process light with a different degree of spatial coherence. The performance of the models was evaluated, and the test accuracies are shown in Fig. 3a. We also evaluated the robustness of the models by testing their performance using input light with degrees of spatial coherence different from that used during training (Fig. 3b). As expected, the best performance is achieved when the model is evaluated under the same coherence conditions used during training. However, models trained using incoherent illumination are more robust against changes in spatial coherence.

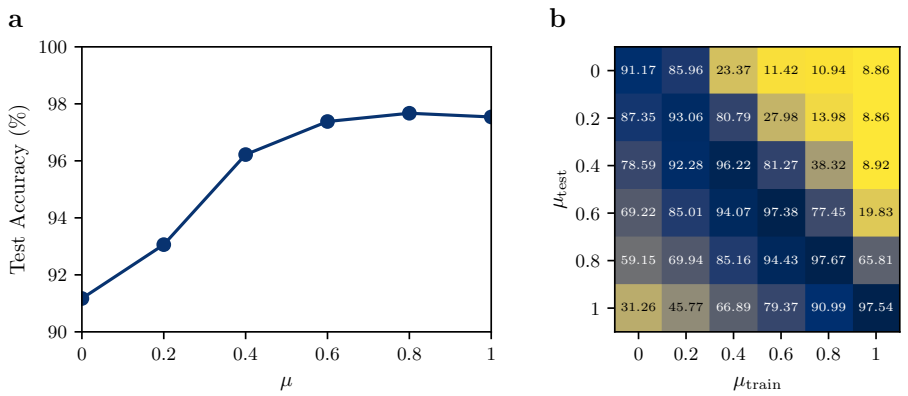

Figure 3: Performance of DONN models on the MNIST dataset with varied coherence. **a** Test accuracy achieved with input light of a specified degree of spatial coherence μ. **b** Accuracies attained by models trained with illumination of spatial coherence μ_{train} but tested using input illumination with spatial coherence μ_{test}. The test accuracies along the diagonal correspond to using the same illumination conditions for training and testing, as shown in (a).

4 Discussion

We have demonstrated that the performance of DONNs is dependent on the spatial coherence of the incident illumination. Models using incoherent illumination cannot outperform linear models for information processing tasks. However, as demonstrated in Fig. 3a, the degree of spatial coherence required to achieve optimal performance need not be high: $\mu \sim 0.6$ means that the mutual coherence between points separated by four pixels is reduced by a factor of 0.13. That is, performance almost at the fully coherent level can be reached even when the transverse coherence length is much less than the size of a whole MNIST digit. This implies that neighboring pixels contain more relevant information for pattern recognition compared to distant pixels. As a result, the DONN model can capture relevant nonlinear relationships without requiring full spatial coherence between all pixels.

A deep-learning based method was recently proposed for implementing linear transformations with DONNs under spatially incoherent illumination [12]. The method approximates incoherence for a single input example by averaging the output intensity patterns from numerous coherent input fields with random phase distributions. In contrast, our approach operates with the mutual intensity function (i.e., mutual coherence function), which is a compact and efficient way to represent the statistical properties of a partially coherent field [10]. This enables us to compute the evolution of light with arbitrary coherence more efficiently and accurately.

We emphasize that the above relation between the input coherence and DONN expressivity assumes that no further processing of the DONN data is implemented. If, for example, the DONN is followed by an electronic neural network with nonlinear activation layers, the DONN can surpass a linear model even if illuminated incoherently. For example, Rahman *et al.* trained a DONN to classify MNIST by associating two detection regions with each digit and then applying a rational function to compute the network prediction from the intensities of these regions. In this way, the accuracy reached was above that of a linear classifier [12].

Incoherently illuminated DONNs are more broadly applicable to real-world environments, as coherent illumination requires a laser source. However, some degree of coherence can also be achieved by illuminating the object with a distant incoherent source of narrow spatial extent according to the van Cittert – Zernike theorem [10]. As discussed above, illumination with even a short transverse coherence length can significantly enhance the DONN performance.

Acknowledgments

This work is supported by Innovate UK Smart Grant 10043476. M.J.F. is funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 956071.

References

[1] Dario Amodei and Danny Hernandez, "AI and Compute," May 2018.

[2] 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," *Nature Photonics*, vol. 15, pp. 102–114, Feb. 2021.

[3] 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*, vol. 361, pp. 1004–1008, Sept. 2018.

[4] D. Mengu, Y. Luo, Y. Rivenson, and A. Ozcan, "Analysis of diffractive optical neural networks and their integration with electronic neural networks," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 1, pp. 1–14, 2020.

[5] D. Mengu, M. S. Sakib Rahman, Y. Luo, J. Li, O. Kulce, and A. Ozcan, "At the intersection of optics and deep learning: statistical inference, computing, and inverse design," *Advances in Optics and Photonics*, vol. 14, no. 2, p. 209, 2022.

[6] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, Oct. 1986.

[7] J. W. Goodman, *Introduction to Fourier optics*. W. H. Freeman, 2017.

[8] F. Shen and A. Wang, "Fast-Fourier-transform based numerical integration method for the Rayleigh-Sommerfeld diffraction formula," *Applied Optics*, vol. 45, p. 1102, Feb. 2006.

[9] J. W. Goodman, *Statistical Optics*. John Wiley & Sons, May 2015.

[10] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics*. Cambridge: Cambridge University Press, 1995.

[11] L. Deng, "The MNIST database of handwritten digit images for machine learning research," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 141–142, 2012.

[12] M. S. S. Rahman, X. Yang, J. Li, B. Bai, and A. Ozcan, "Universal linear intensity transformations using spatially incoherent diffractive processors," *Light: Science & Applications*, vol. 12, p. 195, aug 2023.