Deep Diffractive Neural Networks - An investigation of new approaches and evaluation of Fourier optics as design method

Christian Eder*, Andreas Heinrich*

Center for Optical Technologies, University of Aalen

mailto:christian.eder@hs-aalen.de

Optical computing promises advantages over conventional electronic computers, in terms of cross-talk, efficiency, and computational speed. This proceeding discusses diffractive optical neural networks, their physical operating principle, and the digital computational methods for determining network parameters.

1 Introduction

Artificial neural networks (ANN) become increasingly important as they provide the means to solve complex tasks without exact knowledge about the task itself. Such a network must be trained using exemplary data. Once trained an ANN is capable of solving defined tasks by itself and even apply its methods to unknown input data. This abstraction ability is one of the reasons why ANNs are widely used in the fields of object detection, image recognition, prediction of complex systems and many more.

Deep diffractive neural networks (D²NNs) are physical implementations of such networks, using light as computation medium instead of transistors and electrons in computers. D²NNs promise the capability for extensive parallelism and very short calculation times. [1]

2 Concept of D²NNs

The input for D²NNs is a coherent wave front. Information is modulated in amplitude or phase of this input wave front. Like their digital counterparts, D²NNs are structured in layers, whereby each layer consists of a complex modulating medium and a free-space between each layer. The modulation can be realized by varying the wave's phase or amplitude by a diffractive optical element (DOE). The output of the network is measured as an interference intensity at a detector layer. The structural layout is also sketched in Fig. 1

A single neuron of a D²NN is a single change of modulation in each DOE. Formally, the neuron's complex transmission t becomes:

$$t_j^l = T_j^l \cdot e^{2\pi i \Delta\varphi_j^l} \tag{1}$$

Whereby T_j^l is the transmission and $\Delta\varphi_j^l$ the phase modulation of the neuron j at the layer l. The transfer function of one layer in matrix notation is:

$$\boldsymbol{x}^l = \boldsymbol{x}^{l-1} * \boldsymbol{w}^l \circ \boldsymbol{t}^l \tag{2}$$

Where $\boldsymbol{x}^l$ and $\boldsymbol{x}^{l-1}$ are the complex optical fields of the layer l and the preceding layer, respectively. The ∘- and *-operator are the element-wise multiplication and convolution, respectively. The matrix $\boldsymbol{w}^l$ describes the propagation of light between the previous layer to the current layer.

The output intensity in the detector plane $l = L$ is:

$$y = \|x^L\|^2 \equiv x^L x^{L\dagger}$$

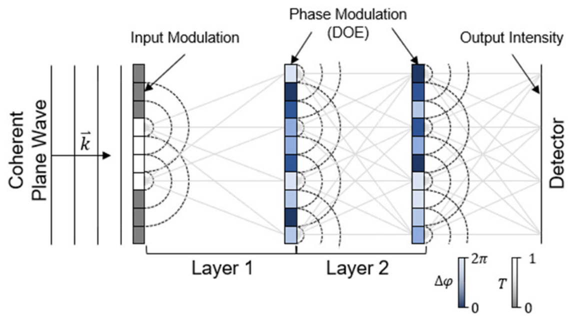

Fig. 1 *A two-dimensional sketch of the structural layout of a two-layer phase-only D²NNs.*

3 Training Model

The goal of training D²NNs (or ANNs in general) is to minimize the network's loss. The loss $\mathcal{L}$ is a function describing the difference of the networks actual and the desired output. Typically, the stochastic gradient descent (SGD) method is used to minimize $\mathcal{L}$. A diagram of the complete training process is shown in Fig. 2. In general, the training consists of two steps: the forward propagation of the optical field through the network and backpropagation of the error to find the gradients of the loss with respect to every network value. The gradient matrices, e.g. $d\boldsymbol{\mathcal{L}}/d\boldsymbol{\Delta\varphi}^l$ for phase-only modulation, can be determined by:

$$\frac{d\mathcal{L}}{d\boldsymbol{\Delta\varphi}^l} = \frac{d\mathcal{L}}{dx^l}\frac{dx^l}{d\boldsymbol{\Delta\varphi}^l} + \frac{d\mathcal{L}}{dx^{l\dagger}}\frac{dx^{l\dagger}}{d\boldsymbol{\Delta\varphi}^l} = 2\Re\left\{-i\frac{d\mathcal{L}}{dx^l} \circ \boldsymbol{x}^l\right\} \tag{3}$$

Whereby $\Re\{\dots\}$ denotes the real part of a complex value and † the conjugate complex. The error gradient field for backpropagation is determined by using the chain rule:

DGaO Proceedings 2021 – http://www.dgao-proceedings.de – ISSN: 1614-8436 – urn:nbn:de:0287-2021-A037-5

submitted: 26.10.2021 published: 17.11.2021

$$\frac{d\mathcal{L}}{dx^l} = \begin{cases} \left(\frac{d\mathcal{L}}{dx^{l+1}} \circ \boldsymbol{t}^l\right) * \boldsymbol{w}^l & ,for\ l < L \\ \frac{d\mathcal{L}}{dy} \circ 2\Re\{(\boldsymbol{x}^{l-1})^{\dagger}\} & ,for\ l = L \end{cases} \quad (4)$$

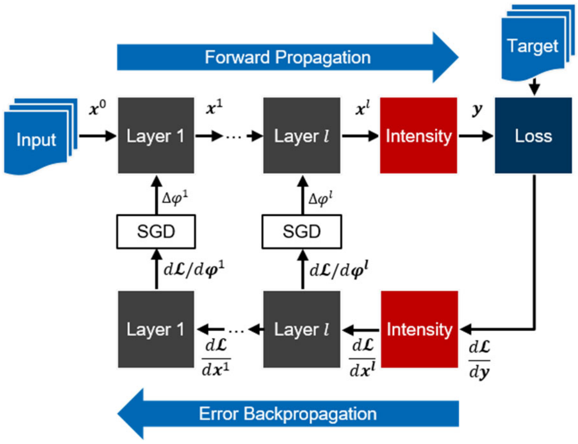

***Fig. 2** Process of training a D²NN using the SGD method.*

4 Field Propagation Model

When training a D²NN, the optical fields and error gradients at every layer must be calculated. Here for one might use a scalar diffraction model based on the first Rayleigh-Sommerfeld (RS) integral. Note that the matrix $\boldsymbol{w}^l$ of (2) and (4) can be expressed using the RS integral.

It is necessary to optimize the simulation for computational speed, because for successful training a large amount of training inputs must be used. An efficient way of calculating the optical field is using the Angular Spectrum (AS) method with the Fast Fourier Transform (FFT). Formally, the propagation of an optical field for a distance d can be described as:

$$\boldsymbol{u}^l = IFFT\left(\boldsymbol{H}^l \circ FFT(\boldsymbol{x}^{l-1})\right) \equiv \boldsymbol{x}^{l-1} * \boldsymbol{w}^l \quad (5)$$

Whereby $\boldsymbol{H}^l$ is called the propagation kernel matrix and can be calculated by [2]:

$$\boldsymbol{H}^l = e^{i2\pi d^l \sqrt{\lambda^{-2} - \boldsymbol{v}_x^2 - \boldsymbol{v}_y^2}} \quad (6)$$

The matrices $\boldsymbol{v}_x$ and $\boldsymbol{v}_y$ are spatial frequency coordinates and λ is the operating wavelength.

Note that like the forward pass, the error backpropagating fields can also be calculated using the AS method.

5 Field Sampling & Computational Speed

When using the AS method, one must take two sampling conditions into account. First, the FFT is a form of the Discrete Fourier Transform and therefore is only valid for periodic signals. As the aperture field is non-periodic, one must add a zero-padding region around the calculation field, so that the resulting field size doubled. Second, the propagation kernel of (6) is an oscillating radial-symmetric function which must be sampled correctly, so that the highest occurring frequency is resolved with at least two points, which is the case when [3]:

$$ds_{x,y} \leq \frac{\lambda}{2}\sqrt{4 \cdot S_{x,y}^{-2} d^2 + 1} \quad (7)$$

Whereby $ds_{x,y}$ is the spatial sampling distance and $S_{x,y}$ total length of the aperture in x- and y-direction. The resulting number of computation points in each direction is $N_{x,y} = S_{x,y} / \, ds_{x,y}$.

To estimate the training time for a given network and training data set, we measured the calculation time for one field propagation. The results for different numbers of total field points when using the AS method in comparison to the direct calculation of the RS integral is shown in Fig. 3. The resulting times are generated using an intel core i9-10900X processor.

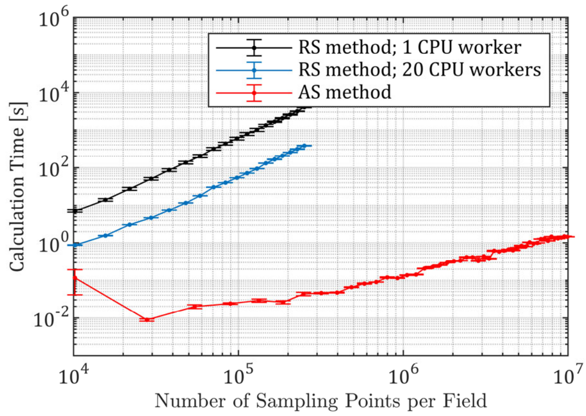

***Fig. 3** Calculation times for different numbers of sampling points ($N = N_x \cdot N_y$) using the AS and RS (with 1 & 20 parallel processes) methods.*

The total time then can be estimated multiplying the respective time with each field propagation for the forward and backward pass.

6 Summary

We showed the concept of how D²NNs operate and the basic principles of training such a network. Further we gave a physical model for calculating the optical fields inside a D²NN and measured the corresponding calculation times using our reference system.

References

[1] X. Lin et al., “All-optical machine learning using diffractive deep neural networks,” Science (New York, N.Y.), vol. 361, no. 6406, pp. 1004–1008, 2018, doi: 10.1126/science.aat8084

[2] J. W. Goodman, Introduction to Fourier optics, 3rd ed. Englewood, Colo.: Roberts & Co, 2005.

[3] K. Matsushima and T. Shimobaba, “Band-limited angular spectrum method for numerical simulation of free-space propagation in far and near fields,” Optics express, vol. 17, no. 22, pp. 19662–19673, 2009, doi: 10.1364/OE.17.019662

DGaO Proceedings 2021 – *http://www.dgao-proceedings.de* – ISSN: 1614-8436 – urn:nbn:de:0287-2021-A037-5

submitted: 26.10.2021 published: 17.11.2021