DTU Library

Modeling of Optical Matrix Multipliers Using Transposed Convolutional Neural Networks

Cem, Ali; Yan, Siqi; Moura, Uiara Celine De; Ding, Yunhong; Zibar, Darko; Ros, Francesco Da

Published in:
Proceedings of 2022 IEEE Photonics Conference

Link to article, DOI:
10.1109/IPC53466.2022.9975676

Publication date:
2022

Document Version
Peer reviewed version

Modeling of Optical Matrix Multipliers Using Transposed Convolutional Neural Networks

Ali Cem
DTU Electro
Kongens Lyngby, Denmark
alice@dtu.dk

Siqi Yan
School of Optical & Electrical Information
Huazhong Univ. of Science and Technology
Wuhan, China
siqya@hust.edu.cn

Uiara Celine de Moura
DTU Electro
Kongens Lyngby, Denmark
uiamo@dtu.dk

Yunhong Ding
DTU Electro
Kongens Lyngby, Denmark
yudin@dtu.dk

Darko Zibar
DTU Electro
Kongens Lyngby, Denmark
dazi@dtu.dk

Francesco Da Ros
DTU Electro
Kongens Lyngby, Denmark
fdro@dtu.dk

***Abstract*—We demonstrate a data-driven model for optical matrix multipliers utilizing Mach-Zehnder interferometer meshes. For a fabricated chip, a transposed convolutional neural network model learns from experimental measurements offline and predicts the weights across 100 frequency channels in the C-band with high precision (RMSE<0.8 dB).**

Keywords—machine learning, neuromorphic photonics, optical matrix multiplication, transposed convolutional neural network

I. Introduction

Photonic integrated circuits (PICs) are well suited as hardware for implementing machine learning algorithms with low power consumption and high speed. This has resulted in an increased interest in neuromorphic photonics over the last decade [1]. In particular, photonic hardware for accelerating artificial neural networks (ANNs) is of great interest. Photonic ANNs necessitate the implementation of nonlinear activation functions and both linear transformations [2]. Impressive results have been reported in the literature for the latter using various optical implementations of matrix-vector multiplication, including, but not limited to, [3] and [4], where microring resonators and Mach-Zehnder interferometers (MZIs) have been used for weighting, respectively.

For the case of MZI meshes, accurate models for fabricated PICs are required to tune the implemented linear weights. This is commonly achieved by optimizing a set of voltages for phase tuning the interferometers. While carefully calibrated analytical models have been used for this task, they only perform well for input light within a narrow spectral band around the frequency used for training and training a new model for each desired input frequency is a very time-consuming approach. As a result, such models are not suitable for applications involving frequency multiplexing, or where frequency drifts may impact the system. Recently, it was shown that ANNs can also be used to model such PICs and even outperform simple analytical models in the presence of fabrication errors and effects such as crosstalk [5]. The analysis of [5], however, did not consider multiple frequency channels but frequency-averaged performance.

In this work, we demonstrate a machine learning model based on a transposed convolutional neural network (TCNN) [6] for a fabricated PIC that can predict the implemented matrix weights for 100 different frequency bands simultaneously. Experimental measurements obtained from the PIC were reshaped into a 2D form well-suited for convolutional operations given the spatial/spectral correlations, resulting in low-error modeling of the chip across the spectrum despite fabrication tolerances.

II. Experimental Setup for Optical Matrix Multiplication Using Mach-Zehnder Interferometer Meshes

The experimental measurement setup for the silicon PIC implementing optical multiplication by a 3x3 matrix is shown on Fig. 1(a). The optical switches at the input and the output together with the optical spectrum analyzer (OSA) are used to measure the spectral responses of the 9 weights consecutively. The captured responses are then downsampled to the 100 frequency channels corresponding to the central frequencies of the ITU DWDM grid for the C-band with 50 GHz spacing. Further details on the experimental setup and the PIC can be found in [5] and [7], respectively. A sample measurement of the 9 matrix weights for the 100

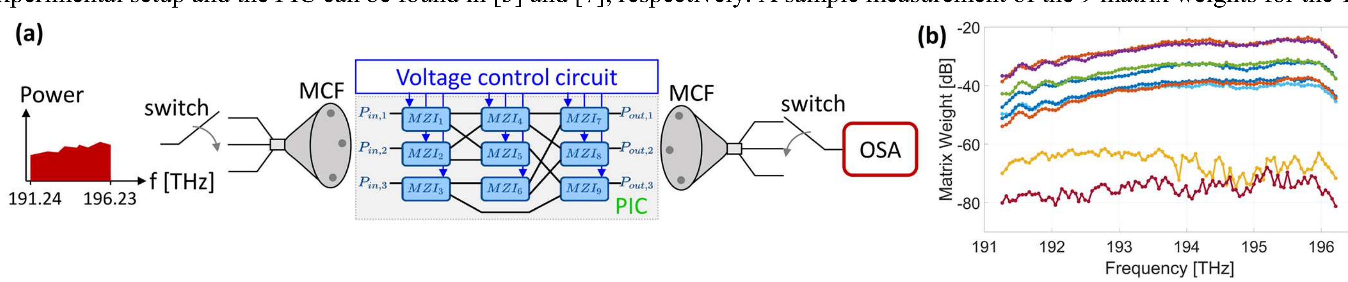

Fig. 1. (a) Experimental setup (MCF: multi-core fiber). (b) Sample spectral measurement results, each of the 9 curve represents a different matrix entry.

frequency channels and a fixed set of input voltages is shown in Fig. 1(b). In order to obtain a measurement dataset for modeling the PIC, more than 5100 random voltage sets were applied, each sampled from a uniform distribution [0, 2V], corresponding to one half-period of the MZI responses.

III. Transposed Convolutional Neural Network Model

In order to better utilize the fact that the matrix weights for neighboring frequency channels are correlated, the flattened matrix weights $\mathbf{W}$ were reorganized in the shape of a 2D matrix $\mathbf{W_{ij}} = \mathbf{W_i(f_j)}$, where each row i corresponds to a different matrix weight and each column j corresponds to a different frequency. Under this formulation, the TCNN can make use of the spatial structure of the data. Fig. 2a shows the architecture for the TCNN which was used to model the mapping between the input voltages $\mathbf{V}$ and the implemented matrix weights $\mathbf{W}$ for the 100 frequency bands. The input layer consists of the 9 heater voltages along with their squares, as physically the phase shifts are directly proportional to squared voltages for MZIs with thermos-optic phase shifters. The output of the TCNN is the 9x100 spectral weight matrix $\mathbf{W_{ij}}$ in decibels and both the inputs and the outputs are normalized between -1 and +1.

A splitting ratio of 70:15:15 was chosen for dividing the measured data into training, validation, and testing sets. TCNN hyperparameters for both the fully-connected hidden layers and the transposed convolutional layers are shown on Fig. 2(a) and are optimized such that the root-mean squared error (RMSE) between the experimentally measured and the predicted outputs is minimized for the validation set. The TCNN was trained on PyTorch using the L-BFGS optimizer with the default parameters.

After training the model, the RMSE between the measured and the predicted matrix weights in the testing set (averaging over all 9×100 2D weight profiles) was found to be 0.79 dB, which is comparable to the results in [5] obtained for a frequency-averaged model. All weights across all frequencies within the testing set are shown using a scatter plot in Fig. 2(b). The model performs especially well for matrix weights closer to 0 dB, but the error is higher for smaller matrix weights. This is most likely due to the higher relative measurement noise. In practice, errors when implementing matrix weights that are close to 0 (e.g. ≤ -30 dB) may not be as impactful on task performance [5]. The probability distribution function (PDF) of the difference between predicted and measured matrix weights over the testing set is plotted in Fig. 2(c). While the absolute prediction error can be as high as 16 dB when implementing particularly lower matrix weight values, more than 92% of the predicted weights are within 1 dB of the experimental measurements.

IV. Conclusion

We present and experimentally evaluate a TCNN-based modelling approach for optical matrix multipliers with the MZI mesh architecture. Given a set of input voltages, our model accurately predicts the implemented matrix weights across 100 frequencies in the C-band for a fabricated PIC. Such models can especially be useful in future applications where multiple spectral bands are utilized to accelerate multiple independent tasks by using the same PIC as a matrix multiplier.

References

[1] B. J. Shastri et al., "Photonics for artificial intelligence and neuromorphic computing," Nature Photonics, vol. 15, no. 2, pp. 102–114, 2021.

[2] C. Huang et al., "Prospects and applications of photonic neural networks," Advances in Physics: X, vol. 7, no. 1, 2021.

[3] A. N. Tait et al., "Neuromorphic photonic networks using silicon photonic weight banks," Scientific Reports, vol. 7, no. 1, 2017.

[4] Y. Shen et al., "Deep learning with coherent nanophotonic circuits," Nature Photonics, vol. 11, no. 7, pp. 441–446, 2017.

[5] A. Cem et al., "Comparison of models for training optical matrix multipliers in Neuromorphic PICs," Optical Fiber Communication Conference (OFC), 2022.

[6] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010.

[7] Y. Ding et al., "Reconfigurable SDM switching using novel silicon photonic integrated circuit," Scientific Reports, vol. 6, no. 1, 2016.

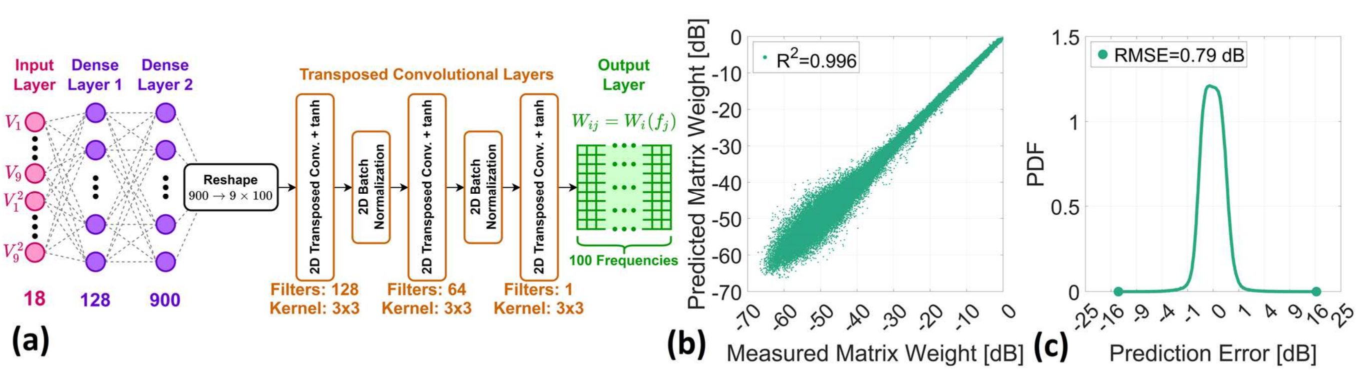

Fig. 2. (a) Architecture of the TCNN model. (b) Scatter plot of the predicted and measured matrix weights for all 100 frequencies. (c) Probability density function (PDF) obtained by normalizing the error histogram, where the error is defined to be the difference between the predicted and measured matrix weights.

This work has received funding by Villum Foundations, Villum YI, OPTIC-AI, grant n. 29344, and ERC CoG FRECOM, grant n. 771878.