Free-Space Deep Optical CNN Simulation Based on 4F Optical Correlator

Maryam Dehbashizadeh Chehreghan, Ripalta Stabile

Technische Universiteit Eindhoven, EHCI, Den Dolech 2, 5612 AZ, Eindhoven

Abstract– *Deep Convolutional Neural Networks (DCNNs) are renowned for their effectiveness in image classification and pattern recognition tasks. However, their computational demands, lead to substantial resource consumption and extended execution time. For instance, a single forward-pass require up to 15 GFLOPs and take several milliseconds per image in ResNet-50 model on high-end GPUs. Here, we present the design and simulation of Deep Optical CNN (Deep OCNN) utilizing a 4F optical-correlator to perform multiple convolution operations in an optical environment by leveraging free-space Fourier transforms and incorporating electro-optical nonlinearity. Our simulation of the optical system is implemented using TensorFlow in Python, using a filter applied via kernel convolution in Fourier space, modeling the light propagation and amplitude modulation for simulating the system output, followed by the application of a threshold-like behavior (as non-linear function). In the last step, in order to extend the system to a DCNN, we incorporate an electronic conversion. Our calculations demonstrate 10^4 latency reduction and 10^5 power efficiency improvement for Deep OCNN compared to DCNN on state-of-the-art electronics.*

Introduction

A Convolutional Neural Network (CNN or ConvNet) is a deep learning model designed for tasks like object recognition, surpassing human performance in image classification [1, 2]. CNNs are now applied across various fields, including object tracking [3], text recognition [4], speech recognition [5], and autonomous driving [6]. Inspired by the human visual cortex, CNNs autonomously extract scalable features, eliminating manual feature engineering [4]. However, computational demands scale poorly with image size and convolution layers, with electronic convolutions exhibiting complexity $O(n^2 \times k^2)$ for an $n \times n$ image and a $k \times k$ kernel [7]. This becomes even more challenging with multiple convolution layers, p, reaching $O(p(n^2 \times k^2))$, as needed in the case of Deep CNNs .To manage complexity, electronic systems often incorporate dropout and pooling layers. While these techniques can reduce the amount of information from feature maps, they can also impact accuracy. Optical implementations offer a promising alternative, as they can process identical kernel sizes on large matrices without repetitive scanning, significantly reducing computational load. Optical neural networks support high-resolution inputs at high speeds with low power consumption. A 4F correlator, leveraging convex lenses and the Fourier convolution theorem, can reduce convolution complexity from $O(n^2 \times k^2)$ to $O(1)$ [8]. Here, we simulate optical convolution layers to measure execution time and power consumption, comparing an Optical Deep CNN (ODCNN) using a 4F Fourier optical correlator with an NVIDIA A100 GPU in speed, power and accuracy across datasets. Initially trained electronically, our CNN uses a Fourier convolution layer for optical validation.

Method

In OCNNs, a 4F amplitude-only Fourier optical correlator classifies datasets by generating feature maps via dot products between kernel and input images. In DOCNNs with multiple convolution layers and no down-sampling, spatial information is preserved, and this is expected to enhance detection of high-level features important in tasks like semantic segmentation and object detection [9]. Fig. 1a illustrates a free-space optical engine: a collimated laser beam, modulated by a spatial light modulator (SLM), undergoes an optical Fourier Transform (FT) through the first lens, and is projected onto a second SLM, where the kernel FT is centered in the Fourier plane. The results of the multiplication in the Fourier domain is followed by an Inverse Fourier Transform (IFT) through the second lens, bringing the outcomes into the spatial domain.

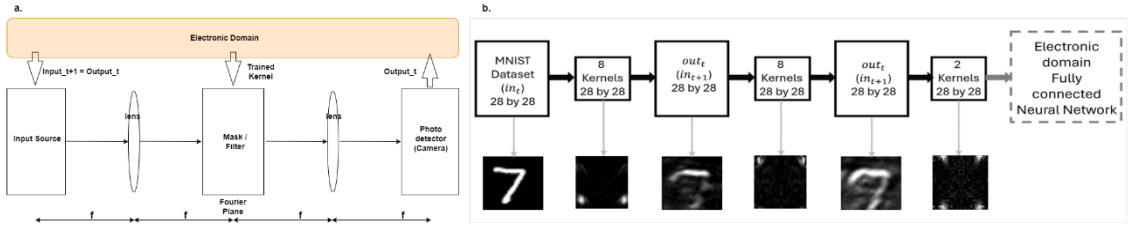

Figure 1: a. Illustration of 4F optical correlator as a deep convolution operation in free-space and post-processing in the electronic domain. b. Flow-chart of the implemented DCNN by using not-down-sampled convolution layers

To extend the optical correlator from a single to multiple convolution layers, the output images (denoted as out_t) are saved and re-used as inputs for the subsequent layers, such that $in_{t+1} = out_t$. Reloading outputs onto the first light modulator enables deeper feature extraction from complex inputs. To evaluate this setup, we benchmarked a DCNN on datasets MNIST and F-MNIST. Fig. 1b presents the DCNN flowchart. Input images and kernels, both in 8-bit grayscale, are rescaled to 250×250 pixels. The output images serve as inputs for further optical convolution layers, and after the final convolution, outputs are converted to tensors for the fully connected layer and classification.

Simulation results and computations

We first examined the relationship between point-wise multiplication count and kernel size in digital convolutions to identify trade-off (Fig. 2). The blue curve shows that number of operations increases with kernel size, raising computational complexity. The red curve identifies an optimal accuracy range at kernel sizes between 3×3 and 5×5 for the case of F-MNIST (Fig. 2b). In contrast, optical convolution, green curve, can use larger kernels, up to the input image size, without raising computational complexity and accuracy loss, highlighting its flexibility and efficiency.

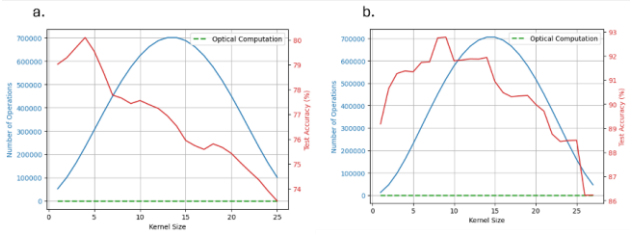

Figure 2: The blue curve represents the correlation between operation count and kernel size, the red curve highlights the accuracy peaks observed at specific kernel sizes. The green curve represents optical computing, demonstrating that it does not increase computational complexity. a. MNIST dataset. b. F-MNIST dataset

The accuracy of the DOCNN was evaluated across various datasets. Table 1, presents OCNN classification accuracies across datasets at various convolutional depths.

Data sets	Number of convolution layers	Accuracy
MNIST	1	98.13%
F-MNIST	3	93.17%

Table 1: Classification accuracy achieved on different datasets using different numbers of convolution layers.

These findings suggest that non-down sampled, sequential convolution layers enhance the performance of OCNNs, particularly for more complex image datasets. Latency and power are crucial in network performance assessment.

Figure 4a compares the power consumption of DOCNN and DCNN on an A100 GPU across various convolution layers, using identical input and kernel sizes. The results demonstrate a 10^5-fold improvement in power efficiency for DOCNN compared to DCNN on state-of-the-art electronic hardware. Figure 4b presents a comparison of latency between DOCNN and DCNN across different convolution layers, also with identical input and kernel sizes. DOCNN latency in optical domain calculated by:

$$T_{total(optical)} = (T_{source} + T_{lm2} \times k) \sum_{i=0}^{l} \left[\frac{T_{lm1}}{k} + T_c \right] in_1 \times k^i + T_{transf} \qquad (1)$$

Here in_1 is number of input images in first layer, T_{source} is laser source delay, T_{lm} light modulator setting time, k kernel count, T_c camera acquisition time, l convolution layers count, and T_{transf} is the time needed for data transfer between layers. Time execution for the DCNN in digital domain calculated by:

$$T_{total(digital)} = \left[\frac{M^2 \times N^2 \times in_1}{FLOPS} \sum_{i=0}^{l} k^i \right] + T_{transf} \qquad (2)$$

M and N represent the dimensions of the input images and kernels, respectively, and $FLOPS$ is $3.2 \times 10^{-12} s^{-1}$. The total computation time for digital-domain convolutions depends on both kernel count k^i in each layer, and image/kernel sizes and there is 10^4 improvement for DOCNN compared to DCNN.

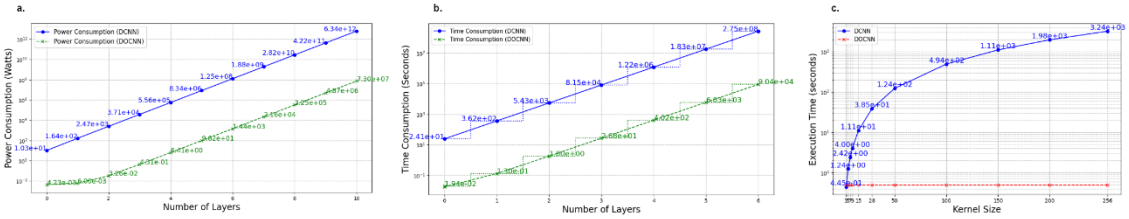

Figure 3: a. Comparison of power consumption for optical and digital deep convolution operations versus the number of layers, with same image and kernel sizes for both approaches. b. Comparison of the execution time of the convolution operation in an optical 4f correlator with digital b. Execution time in the digital domain and the optical domain (red graph) versus kernel size.

Figure 3c shows that digital execution time rises exponentially with kernel size, while optical time remains constant, demonstrating optical efficiency. This paper also compares power of DOCNN and DCNN, in optical domain:

$$P_{total(optical)} = (P_{source} + P_{lm2} \times k) \sum_{i=0}^{l} \left[\frac{P_{lm1}}{k} + P_c \right] in_1 \times k^i + P_{transf} \qquad (1)$$

where P_{source} represents light source power, P_{lm} light modulator power, and P_c is detector power.

The power consumption in DCNN in digital domain calculated by:

$$P_{total(digital)} = \left[\frac{GPU_{power} \times M^2 \times N^2 \times in_1}{FLOPs} \sum_{i=0}^{l} k^i\right] + P_{transf} \qquad (2)$$

GPU_{power} is 400 Watt. In the optical domain, power depends on the number of inputs and outputs per layer, but in digital domain, it relies on kernel count k^i in each layer and image/kernel sizes.

Conclusion

This study proposes and simulates a free-space optical convolution system for executing DCNNs, evaluating the accuracy and execution time of ODCNNs across various datasets and convolution layer setups. Results show that multi-layer ODCNNs improve classification accuracy, especially with additional layers without down-sampling. We analyzed how kernel size affects computational load and accuracy in both digital and optical convolutions and compared latency and power usage between electronic DCNNs and multi-layer ODCNNs. Findings reveal that ODCNNs offer significant reductions in execution time and power consumption, underscoring their potential for high-accuracy DCNNs with ultra-low latency.

Acknowledgment

This work has been supported by the EC KDT JU CLEVER (grant no. 101097560) project.

References

[1] Krishnan SG P Subathra Ramachandran R, Rajeev DC. Deep learning an overview. IJAER, 2015.
[2] Y. Wu J. Fan, W. Xu and Y. Gong. Human tracking using convolutional neural networks, neural networks. IEEE Transactions, 2010.
[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.
[4] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. IEEE conference 2016.
[5] Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, and Gerald Penn. Applying convolutional neural networks concepts to hybrid nn-hmm model for speech recognition. IEEE, 2012.
[6] Yaodong Cui, Ren Chen, Wenbo Chu, Long Chen, Daxin Tian, Ying Li, and Dongpu Cao. Deep learning for image and point cloud fusion in autonomous driving: A review. IEEE Transactionss, 2021.
[7] Edward Cottle, Florent Michel, Joseph Wilson, Nick New, and Iman Kundu. Optical convolutional neural networks–combining silicon photonics and fourier optics for computer vision., 2020.
[8] Mario Miscuglio, Zibo Hu, Shurui Li, Jonathan K George, Roberto Capanna, Hamed Dalir, Philippe M Bardet, Puneet Gupta, J Sorger. Massively parallel amplitude-only fourier neural network. Optica, 2020.
[9] A Ting He, Ying Liu, Yabin Yu, Qian Zhao, and Zhongkang Hu. Application of deep convolutional neural network on feature extraction and detection of wood defects. Measurement, 152:107357, 2020.