On-chip 4F-System Based on Concave Mirrors for Optical Neural Networks

Jun Dai*[1], Xiaowen Dong[2], Chong Li[2], Jian-Jun He[1]
[1]Center for Integrated Optoelectronics, State Key Laboratory of Modern Optical Instrumentation, College of Optical Science and Engineering, Zhejiang University, Hangzhou, China, 310027;
[2] Huawei Technologies Co., Ltd., Bantian, Longgang, Shenzhen, Guangdong, China, 518000
*jundai@zju.edu.cn; phone 17612289359

ABSTRACT

Optical neural networks (ONNs) have the potential for accelerating the inference of AI models, since ONNs have the advantage of high compute speed, and high parallelism. Integrated diffractive optical network for implementing parallel Fourier transforms has been proved efficient and is promising for large scale ONNs. We propose a novel on-chip Fourier transform implementation based on etched concave mirrors, enabling the construction of a photonic integrated 4F system to perform the convolution computation in the Convolutional Neural Networks (CNNs). One of the input vectors is encoded in a modulator array at the object plane, and the Fourier transform of the other vector is encoded in another modulator array at the spectrum plane. We simulated the computing process by the diffractive propagation of the optical field from the object plane to the image plane according to the Kirchhoff's diffraction formula. Finally, we used our simulation system to replace the traditional convolution layers in the electronic system to implement CNNs on three different datasets, Iris-Flower, MNIST and Fashion-MNIST, and obtained 96.67%, 95.6% and 89.4% classification accuracies, respectively, demonstrating comparable performance with the electronic counterpart.
Keywords: Optical Neural Networks, Diffractive Optics, 4F-System

1. INTRODUCTION

In the past decade, deep learning-based Artificial Intelligence (AI) has achieved significant success in the fields of computer vision/graphics, speech recognition, and natural language processing[1], among others. However, as models become increasingly complex with more layers and parameters, both the inference and training speeds of AI models have slowed down, accompanied by a rapid increase in energy consumption[2]. Optical neural networks (ONNs) are considered promising solutions to address the current limitations in AI computing power[3]. They excel in high parallelism, rapid computational speed, and ultra-high computational bandwidth. The Convolutional Neural Network (CNN) stands as a fundamental cornerstone in modern AI, serving as one of its most vital components[4]. Its significance is especially pronounced in the realm of computer vision, where it plays a role of paramount importance[5]. However, it's essential to note that traditional electronic computing systems, constrained by their inherent sequential processing nature, result in substantial computational and energy costs during convolutional operations[6]. Recently, several studies have been dedicated to implementing CNNs using optical methods. Notably, Zhu et al.[7] introduced the integrated diffractive neural network (IDNN), which leverages integrated optical techniques to execute CNNs, achieving comparable performance to their electronic counterparts. Nonetheless, their approach requires additional design considerations to create optical structures suitable for Fourier transformation. This necessary enhancement, in turn, leads to a significant rise in both the design and fabrication costs associated with implementing optical convolution operations. In response to these challenges, we introduce a simplified optical 4F-system based on concave reflective mirrors. This innovative configuration (illustrated in Figure 1) eliminates the requirement for additional design complexities, providing an efficient approach for performing optical convolution operations while minimizing cost increases. We simulated the entire on-chip 4F-system according to the Kirchhoff's diffraction formula, and the two vectors involved in the convolutional operations are encoded by the modulators in both the phase and amplitude of the optical field. Finally, we evaluate the performance of our ONN by three classification tasks, including Iris-Flower classification, MNIST and Fashion-MNIST classification.

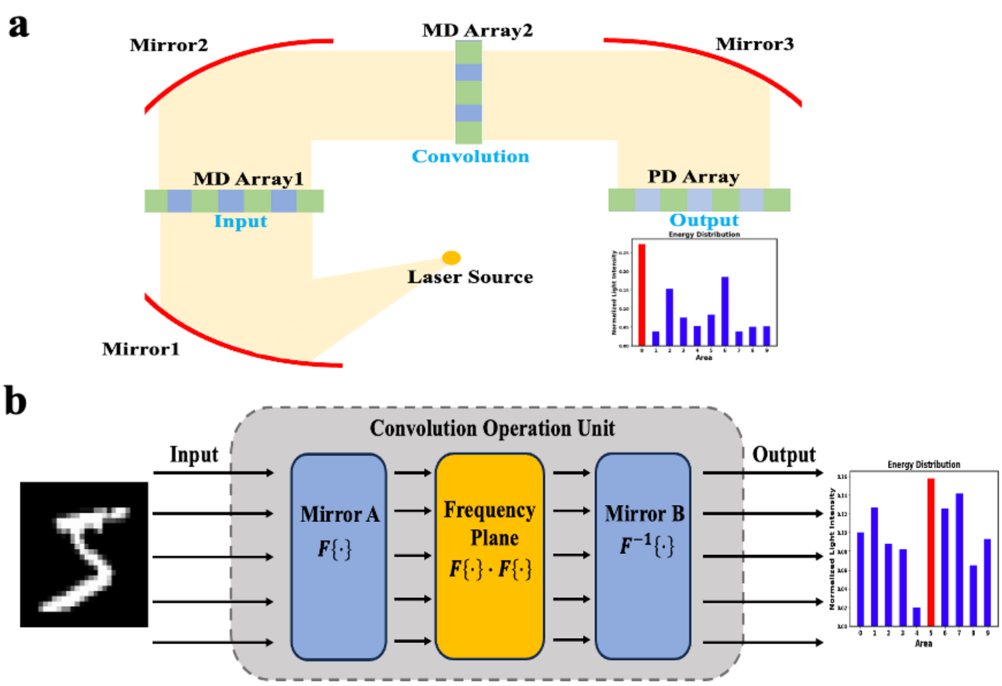

Figure 1. a. The 4F-system proposed overall diagram and b. Schematic diagram of the convolution operation principle.

2. PRINCIPLE AND ARCHITECTURE

The ultimate goal of ONNs is to imitate the calculation of convolutional operation in CNNs, we will discuss the principles and methods involved in our on-chip 4F-system ONN that we proposed in this section.

2.1 Optical 4F-System for Convolution Operation

In practice, the convolutional operations utilized within neural networks diverge from the strict definition of mathematical convolution calculation. Within the context of CNNs, the convolutional operation takes on the character of a sliding, weighted summation procedure, expressed by equation (1).

$$f[x,y] * g[x,y] = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} f[n_1,n_2]\, g[x-n_1, y-n_2] \qquad (1)$$

$$F\{f(x,y) * g(x,y)\} = F\{f(x,y)\} F\{g(x,y)\} \qquad (2)$$

Following the convolution theorem of Fourier transform, convolution in the spatial domain is equivalent to the Fourier inverse transform of the element-wise multiplication result in the frequency domain, as demonstrated in Equation (2). The 4F-system represents a classic optical information processing approach that exploits the convolution theorem to facilitate the convenient optical implementation of Fourier transform operations. As a result, we embarked on the construction of a comprehensive on-chip 4F-system for optical convolution operations, with concave parabolic mirrors serving as its foundation. The incorporation of the concave parabolic mirror structure eliminates the need for introducing any additional design or fabrication complexities while maintaining equivalent performance characteristics.

2.2 On-chip 4F-System ONN's architecture

The on-chip ONN that we propose primarily consists of a parabolic reflective mirror structure, an array of modulators, an array of detectors, and a laser light source (depicted in Figure 1). Further details are provided in Table 1. Initially, a laser light source undergoes collimation as it passes through the initial parabolic mirror. The input vector used in the convolutional operation is modulated to encode amplitude information within the optical field. Subsequently, the second parabolic mirror performs a Fourier transform on this modulated optical field. The resulting spectrum of the input vector is then subjected to element-wise multiplication with the spectrum of the convolutional kernel vector, carried out via the second array of modulators. Finally, the outcome of the multiplication undergoes an inverse Fourier transform using the third reflective mirror structure. The convolution result is ultimately retrieved through the array of detectors, as shown in Figure 1b.

Table 1. The parameters used in our simulation system.

Wavelength λ	1550nm
Focal Length of mirrors, f	2000μm
Refractive Index, n	1.50
Gap between input/output waveguides, d_{gap}	5μm
Width of input/output waveguides, d	5μm

3. SIMULATION EXPERIMENTS AND RESULTS

We subsequently utilized the Kirchhoff diffraction equation (Equation 3) and conducted end-to-end simulations for the envisioned ONN. In a sequential manner, we applied the resulting simulation architecture to a series of standardized recognition tasks (Iris-Flower, MNIST, and Fashion-MNIST). The primary objectives of these tasks were twofold: to validate the effectiveness of our proposed on-chip optical convolutional neural network and to evaluate its performance.

$$E_{out} = \frac{1}{2}\sqrt{\frac{n}{\lambda}}\int \frac{E_{in}}{\sqrt{r}} \cdot (1 + cos\, \alpha)e^{-jk|r|}ds \tag{3}$$

where α is the diffraction angle, r is the propagation distance, n is the material's refractive index and λ is the wavelength.

3.1 Iris-flower classifier

We demonstrate the performance of our system in classifying the Iris-flower dataset. In this task, every sample have four input parameters (the lengths and widths of the sepals and petals of a candidate flower) and the goal of classifier is to determine which of the three possible subspecies the flower belongs to (setosa, versicolor, and virginica). The classification task is realized by employing a single layer neural network comprising eight neurons. Within the group of eight neurons, four of them carry the four specific features from the dataset. The other four neurons act as placeholders, kind of like zeros padding. This arrangement is made to match the real physical setup, where nearby input arrays have gaps in between. A complex, circulant weight matrix links the input and output, governed by a trained convolutional kernel with eight real-value parameters. The modulation of input field amplitude within the system encodes input signals. In the Fourier domain, both amplitude and phase modulation are used to transform inputs into optical energy distributed across three output waveguides. These outputs are then captured by the final PD array and the highest optical intensity measurement outcome is used to indicate the classified result.

The complete dataset containing 150 instances is divided into a training set and a testing set, following an 8:2 ratio. The weights are exclusively trained using the instances within the training set. Initially, we establish the identical architecture on an electronic computer and train a convolutional layer to accomplish the same task. Upon obtaining trained convolutional kernels, these kernels parameters are transformed into the frequency domain (complex-value) and subsequently loaded onto the second layer's array of modulators. This process culminates in the execution of frequency-domain element-wise multiplication. The kernel parameters are optimized by back-propagation algorithm[8] and the final testing accuracy (on electronic PC) is 97.3%.

After the training phase, we assess the performance of our simulation system with complex modulation on the Iris flower classification testing dataset. The experimentally obtained energy distribution from the three final output PDs is shown in Figure 2b for a specific testing data point (setosa) as an example, with the red bar stand for the ground truth label. The entire 30 testing samples are experimentally evaluated using our simulation system and the final result is depicted in a confusion matrix in Figure 2c. The only one-layer simulation model achieved 96.7% testing accuracy, close to its electronic counterpart (97.3%), which is the same as the conventional one-layer fully connected neural network[7], manifesto that our scheme has the potential to get the similar performance with less extra design and fabrication.

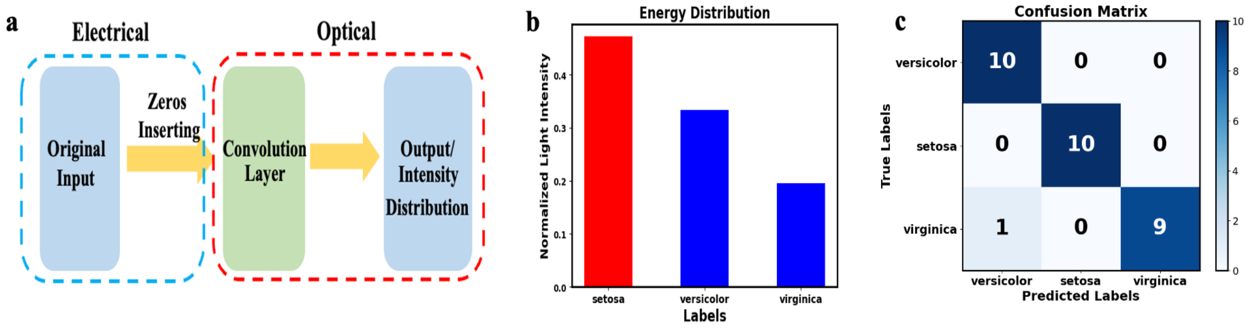

Figure 2. a. The overall experimental network diagram, b. The normalized light intensity distribution among final PD arrays when the sample's label is setosa and c. The confusion matrix for total 30 test samples.

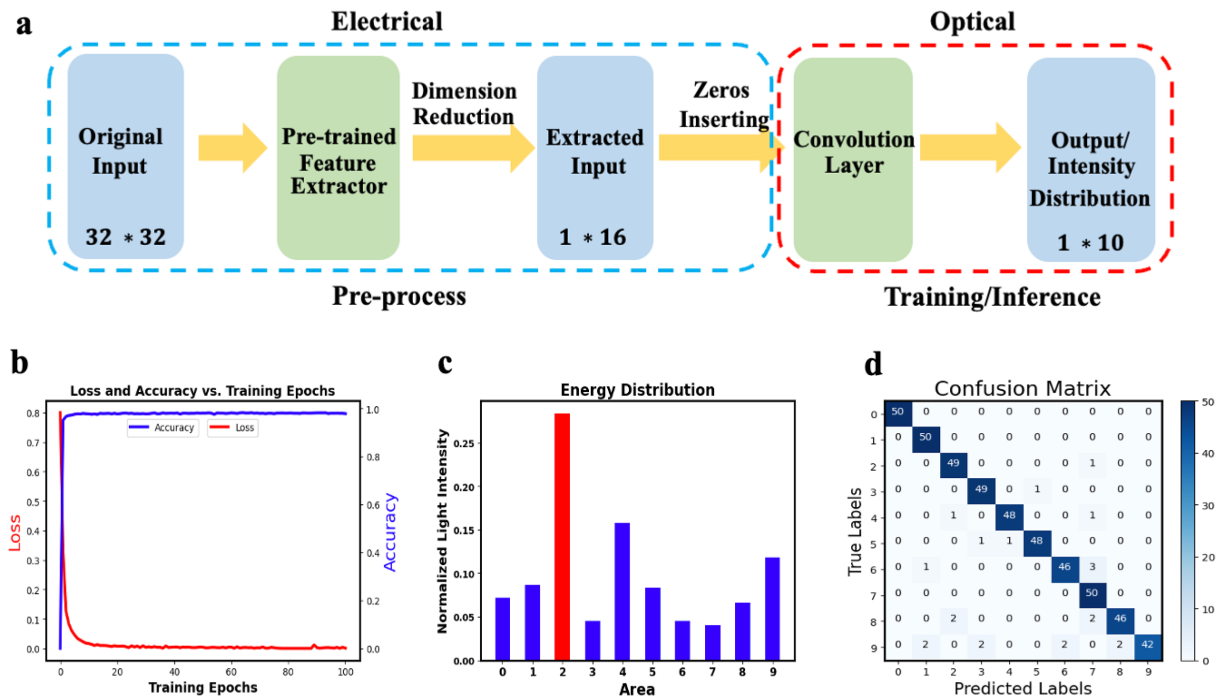

Figure 3. a. The extractor and the entire architecture of our experiment, b. The Loss and Validation Accuracy curve with the training epochs, c. The normalized light intensity distribution of the final PD arrays when the sample is digit 2 and d. The confusion matrix for 500 test images.

3.2 Handwritten digit and Fashion product classifier

More complex datasets, MNIST (handwrite digits images), and Fashion-MNIST (clothing images) datasets are used to further validate the functionality of our on-chip mirrors-based ONN. The two datasets are both split into training datasets (with 60,000 images) and test datasets (with 10, 000 images). We trained our convolutional kernels on the entire training datasets respectively and uniformly choose 50 samples for every class (namely 500 test samples in total). Limited by the amount of input channels in our system, we firstly put an extractor before our convolutional layers, which will decrease the size of input from 784 into 32, and the whole model architecture is shown in Figure3a. The extractor could be a pre-trained layers (we used in here) or any others methods that can downsize the input, like using the frequency information. The final optical intensities distribution captured by the final PDs indicate the recognition result. The numerical validation accuracy and loss vary with the epoch number as shown in Figure3b with a classification accuracy of 97.9%. Furthermore, the energy distribution simulated based on the intensity detection outcomes from ten outputs is illustrated in Figure, effectively demonstrating the classification performance. This visualization underscores that, in line with our numerical simulations, the waveguide exhibiting the highest energy typically corresponds to the accurate identification of the handwritten digit, the red bar stand for the ground truth label as well. Besides, we experimentally tested all the 500 images and the confusion matrix in Figure3c shows an accuracy of 95.6% in the generated predictions, in contrast to 97.9% in the electronical computer. These results show that we have successfully implemented the classification task on the 4F-system based mirrors ONN we proposed. We further utilize our simulation system to perform the Fashion-MNIST task with the same model architecture as we do in MNIST task. The results are shown in the same way as MNIST task's results in Figure4. Our ONN simulation model achieved a test accuracy of 88.8%, which is almost same as the result obtained in electronical computer (89.8%).

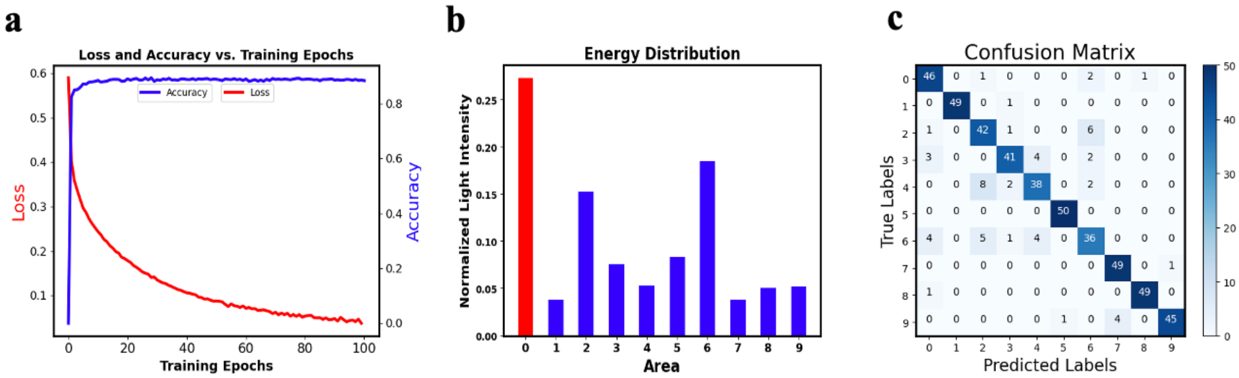

Figure 4. (Fashion-MNIST) a. The Loss and Validation Accuracy curve with the training epochs, b. The normalized light intensity distribution of the final PD arrays when the sample is a pullover and c. The confusion matrix for 500 test images.

4. CONCLUSION AND DISCUSSION

We propose and conduct an end-to-end simulation of an on-chip 4F-system, encompassing the entire integrated framework. Through this simulated system, we successfully execute several canonical recognition tasks. These outcomes theoretically substantiate that our novel on-chip 4F-system, while simplifying design and fabrication processes, achieves comparable computational and processing performance. In contrast to previous on-chip ONNs, which are predominantly based on Mach-Zehnder interferometer (MZI) configurations, often resulting in larger spatial footprints[9], some research endeavors to tackle the challenge posed by the quadratic relationship between the necessary number of modulators and the dimensions of vectors involved in computations. Leveraging diffractive-based methodologies, this issue has been previously shown to transition the quadratic relationship into a linear one[7]. Building upon this advantageous aspect, our research ensures the retention of this benefit while simultaneously mitigating the complexities associated with the design and fabrication of specific compositional structures. Simultaneously, achieving performance levels comparable to those of electronic computers across the three prominent recognition tasks serves as a strong confirmation of the competence of our proposed framework in effectively performing the tasks associated with the convolutional layer. This substantiates the potential of advancing our approach in a forthcoming manner—by integrating diffractive methodologies onto extant architectures—to further extend this competence towards encompassing fully connected layers. Such an entire framework, which allows the entire inference of AI to run on, bears paramount significance in expediting the inferential velocity of AI models necessitating rapid inference capabilities. Our proposition significantly favors the establishment of this comprehensive framework on a chip, thereby contributing to the realization of a more extensive, fully integrated on-chip optical artificial intelligence computing platform.

REFERENCES

[1] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." nature 521.7553 (2015): 436-444.
[2] Strubell, Emma, Ananya Ganesh, and Andrew McCallum. "Energy and policy considerations for modern deep learning research." Proceedings of the AAAI conference on artificial intelligence. Vol. 34. No. 09. 2020.
[3] Ambs, Pierre. "Optical Computing: A 60-Year Adventure." Advances in Optical Technologies (2010).
[4] LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324.
[5] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
[6] Huang, Wenjin, et al. "FPGA-based high-throughput CNN hardware accelerator with high computing resource utilization ratio." IEEE Transactions on Neural Networks and Learning Systems 33.8 (2021): 4069-4083.
[7] Zhu, H. H., et al. "Space-efficient optical computing with an integrated chip diffractive neural network." Nature communications 13.1 (2022): 1044.
[8] Wetzstein, Gordon, et al. "Inference in artificial intelligence with deep optics and photonics." Nature 588.7836 (2020): 39-47.
[9] Shen, Yichen, et al. "Deep learning with coherent nanophotonic circuits." Nature photonics 11.7 (2017): 441-446.