

RESEARCH ARTICLE

Integrated Photonic Convolutional Neural Network Based on Silicon Metalines

OMID POORDASHTBAN^{1,*}, MAHMOOD REZA MARZABN^{1,2,*}, AND AMIN KHAVASI¹

¹Department of Electrical Engineering, Sharif University of Technology, Tehran 11155-4363, Iran

²School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

Corresponding author: Omid Poordashtban (omid.poordashtban@gmail.com)

*Omid Poordashtban and Mahmood Reza Marzabn contributed equally to this work.

ABSTRACT Compact and low-power CMOS-compatible hardware can be used for on-chip optical neural networks (ONNs), enabling affordable and portable image classification solutions for applications like autonomous vehicles, healthcare, and optical communication. In this work, we propose a novel one-dimensional Optical Convolutional Neural Network (OCNN) architecture that significantly reduces the number of learnable parameters required for an ONN. Our OCNN achieves an impressive accuracy of over 96% as a pattern classifier, utilizing only 90 learnable parameters, leading to a simpler structure compared to existing on-chip ONNs. Additionally, our OCNN demonstrates scalability and robustness, with an accuracy exceeding 89% in handwritten digit classification. The OCNN's convolutional layer employs a lenslet 4f system for convolving desired kernels on input images, while an on-chip lens facilitates the desired Fourier Transform effortlessly. The subsequent layer consists of a single metaline layer, implementing a fully connected layer. By parallelizing pre-trained OCNNs, an on-chip deep convolutional neural network (CNN) can be realized, where each OCNN functions as a separate kernel within a conventional CNN.

INDEX TERMS Photonic integrated circuits, silicon photonics, optical neural network, optical convolutional neural network, silicon metaline.

I. INTRODUCTION

The rapid development of deep learning [1] based on Convolutional Neural Networks has brought significant advances in a wide range of artificial intelligence applications, including image recognition [2], object classification [3], document analysis [4], autonomous driving [5], and more. Although 2D CNNs are widely used for analyzing grid-like 2D data, they have limitations that have led to the emergence of one-dimensional (1D) CNNs as a promising alternative in recent years [6], [7]. As demonstrated in a recent study [8], 1D convolution offers several advantages over 2D convolution. For one, 1D CNNs have significantly lower computational complexity than their 2D counterparts. Additionally, most 1D CNN applications use compact setups with only 1-2 buried CNN layers and networks containing less than 10,000 parameters. In contrast, nearly all 2D CNN applications employ “deep” architectures with over 1 million parameters, making them more difficult to train and implement.

The associate editor coordinating the review of this manuscript and approving it for publication was Sukhdev Roy.

Furthermore, compact 1D CNNs are well-suited for real-time and low-cost applications, particularly on mobile or handheld devices, due to their minimal processing needs.

To address the energy consumption bottleneck of CMOS electronics in machine learning [9], researchers are investigating optical computing as a potential alternative architecture for building neural networks. Optical parallelism has been used in ONN [10], [11], [12], [13] and Opto-Electronic Neural network implementations [14], [15], [16], [17], [18] to speed up computing, while optical passivity has been utilized to reduce energy costs and minimize latency. Several studies have reported the successful implementation of ONN designs utilizing multi-layer metasurfaces [10], [11], [12], [19]. These subwavelength structures are suitable for shaping the phase front of incident electromagnetic waves, allowing for a powerful optical analog signal processing system to be implemented with designed dispersion and diffraction [20], [21].

While being passive and operating at the speed of light, the realization of fully passive all-optical convolutional neural networks has seen limited progress. However, attempts

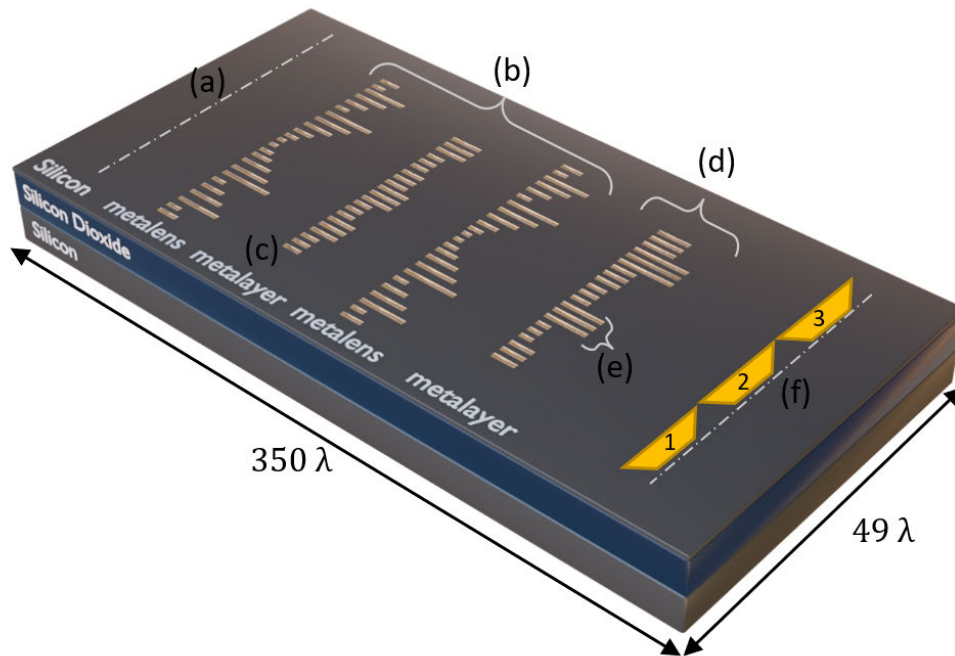


FIGURE 1. Illustration of the network architecture. The network comprises four metaline layers, where the first three layers function as a convolutional layer and the last layer acts as a fully connected layer. (a) depicts the inputs represented as optical pulse amplitudes propagating through the metalines. (b) showcases a $4f$ system implemented with two metalenses, enabling the optical implementation of the convolution operation. (c) illustrates the metaline associated with the desired phase mask of the convolution layer, applying the desired aperture transfer function (ATF) on the propagating optical wave. (d) displays the implemented optical fully connected layer. (e) showcases a phase shifter consisting of three meta-atoms. Finally, (f) presents three output regions, each connected to a detector with a tapered waveguide.

have been made to implement optical convolutional neural networks. Some existing implementations, such as [15] and [22], utilize a 2D $4f$ system for optical convolution. Nevertheless, these networks face challenges associated with misalignment errors, which hinder their integration into practical optical circuits and limit the utilization of the parallel processing power of optics. In contrast, other implementations of optical convolutional neural networks rely on optical delay lines. For example, a notable design utilizes a single convolution layer with repatching logic, employing 3dB splitters and variable-length delay lines [23]. While this network achieves high throughput of millions of inferences per second, it requires significant power consumption of 2mJ per inference, comparable to ASIC systems [24]. Moreover, recent advancements in optical neural networks have showcased the potential of utilizing broad optical bandwidths for accelerated computing. One noteworthy study demonstrated an impressive performance with a universal optical vector convolutional accelerator operating at over ten TOPS and achieving high accuracy in recognizing handwritten digit images [25]. However, the reliance on complex hardware, including a soliton crystal microcomb source, raises practical concerns for widespread adoption, particularly in common portable devices.

Our work presents the first demonstration of an all-optical integrated on-chip 1D convolutional neural network (OCNN) using metasurfaces. This innovative design improves energy

efficiency by using passive optical elements, eliminating the need for additional power-consuming components. The integration of our proposed OCNN is facilitated by utilizing $1.55\mu\text{m}$ silicon-compatible technology, enabling seamless connectivity to existing CMOS-compatible technology. Our OCNN achieves remarkable accuracy, with a performance of 96% on a binary letter image classification task using only 90 learnable parameters. This represents an impressive 80% reduction compared to the state-of-the-art ONN [10]. We also evaluated the scalability of our OCNN through tests on handwritten digit classification, where it achieved an accuracy exceeding 89%, demonstrating its robustness and adaptability. The compact CMOS-compatible structure enables seamless integration and parallelization, making our OCNN well-suited for low-cost and real-time applications, especially on portable devices. Furthermore, our OCNN leverages simple diffractive physics to achieve comparable accuracies in classification tasks without relying on complex hardware implementations like soliton crystal microcomb sources [25]. This eliminates the complexity associated with advanced sources, making our OCNN more efficient and compact. Importantly, our proposed OCNN exhibits significant advantages in mitigating misalignment errors compared to 2D optical CNNs [13], [26] that rely on 2D layered metasurfaces in a free space setup. Previous studies have demonstrated the superior alignment stability and performance of our OCNN architecture, further emphasizing its practical

viability [10], [27]. In summary, our all-optical integrated on-chip 1D convolutional neural network (OCNN) using metasurfaces offers improved energy efficiency, impressive accuracy, scalability, and compatibility with existing CMOS technology. These advantageous features position our OCNN as a promising solution for efficient and high-performance neural network applications. Looking towards the future, the integration of phase change materials [28], [29] holds great potential for realizing active optical neural networks on our platform, leveraging the low number of learnable parameters in our OCNN design to meet the energy demands of such networks.

II. OCNN ARCHITECTURE

The architecture of conventional 1D CNNs comprises 1D convolutional layers followed by a fully-connected network. In this work, we present a $4f$ architecture to implement the convolutional layer of our OCNN. Fig.1 shows the proposed design, where the input is first preprocessed and transformed into a 1-dimensional vector, which is then converted into optical pulse amplitudes and passes through the OCNN layers.

The optical architecture employed in this study utilizes 1D high-contrast transmit array metasurfaces [21], called “metalines,” which are composed of 1D etched rectangle slot arrays in a silicon-on-insulator (SOI) substrate, as shown in Fig. 1. Each rectangle slot is a meta-atom, with the silicon layer and SiO_2 insulator layer having thicknesses of 250nm and $2\mu\text{m}$, respectively. The meta-atom geometrical parameters include the lattice constant of the metalines (a), length (L), and width (w) of the rectangular slots [12]. The operational frequency is $1.55\mu\text{m}$, with the lattice constant of the metalines being 500nm which is approximately one-third of the operating wavelength, making the structure subwavelength.

Figs. 2(a) and 2(b) demonstrate the transmission amplitude and phase of the metalines for TE-polarized guided waves as the length and width of meta-atoms are swept. The width of the slots is fixed at 140nm. The proposed design parameters allow for adjustment of the propagation phase within the range of $-\pi$ to π by varying the length of the slots from $0.05\mu\text{m}$ to $2.5\mu\text{m}$ while maintaining transmission amplitudes greater than 0.96 for all phases, as shown in Figs. 2(a) and 2(b). The results presented in Fig. 2 are obtained using the commercial software package Lumerical FDTD.

According to the approach described in [12], each meta-atom in the proposed structure can be modeled as a complex transmission coefficient represented by ($T^l = t^l \exp(j\phi^l)$). Due to the near-unitary transmission amplitudes of the meta-atoms, the transmission loss is negligible, allowing for the modeling of the meta-atoms as phase shifters applying the transmission function ($T^l = \exp(j\phi^l)$) to the input wave propagating through each metaline. To achieve the desired phase profile, appropriate slot lengths must be selected at various positions for meta-atoms in each metaline layer. The transmission function is computed under

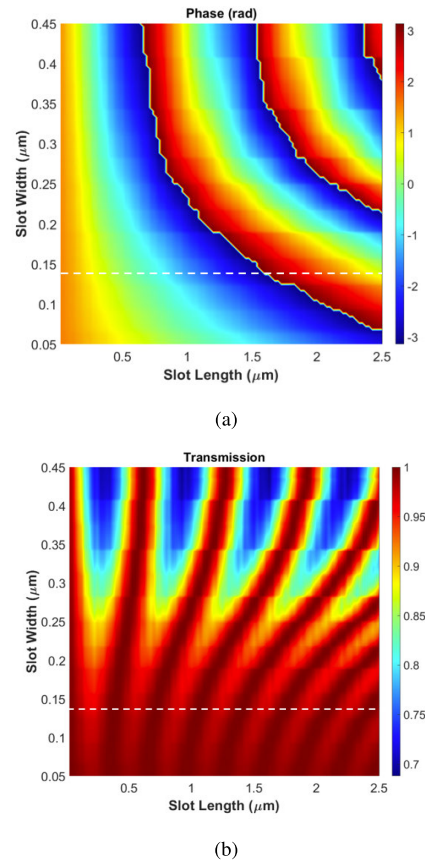


FIGURE 2. shows the phase (a) and transmission (b) amplitude of a meta-atom for TE-polarized guided wave at an operational wavelength of ($1.55\mu\text{m}$) versus slot length and width. The white dashed line in both (a) and (b) shows the phase and transmission amplitude of the slots with a fixed width of 140nm.

the assumption that the wave’s phase front is affected by a periodic array of meta-atoms, with every three slots (meta-atoms) assumed to act as a phase shifter to improve the accuracy of the locally periodic approximation [10].

III. OPTICAL CONVOLUTIONAL LAYER USING METASURFACE-BASED LENS

An interesting method to create an optical convolutional layer is using optical lenses that can freely perform a Fourier transform. Metasurface-based lens designs have been proposed in several works [20], [30], [31] that are useful in implementing the convolutional function. In our work, we control the wavefront on-chip using a metaline. The metaline along the y direction imposes a space-dependent phase shift on the incident wave (TE polarized) along the x direction. For achieving on-chip wave focusing, the phase shift of the transmitted wave is defined in the following equation [21]:

$$\phi(y) = \frac{2\pi}{\lambda_d} n_{eff} \left(f - \sqrt{f^2 + y^2} \right) \quad (1)$$

where λ_d is the design wavelength in free space, n_{eff} is the effective refractive index of the guided light confined in the silicon slab, and f is the focal length.

A. OPTICAL FOURIER TRANSFORM

According to [32], the output function of a two-dimensional metalens at its focal length, when the input is placed at a distance d from it, is given by:

$$g(x, y) = c' \exp \left[j\pi \frac{(x^2 + y^2)(f - d)}{\lambda f^2} \right] \mathcal{F} \left(\frac{x}{\lambda f}, \frac{y}{\lambda f} \right) \quad (2)$$

where \mathcal{F} stands for Fourier transform of the input signal, λ is the wavelength in the slab waveguide, f is the focal length of the lens, d is the distance between the input plane and the lens, j is the imaginary unit, (x, y) are the coordinates on the focal line, and c' is a constant. For one dimensional lens, Eq. 2 can be simplified as:

$$g(y) = c' \exp \left[j\pi \frac{y^2(f - d)}{\lambda f^2} \right] \mathcal{F} \left(\frac{y}{\lambda f} \right) \quad (3)$$

According to Eq. 3, when the object is placed at the focal length (i.e., $d = f$), the Fourier transform of the input appears in the output, and Eq. 3 simplifies to:

$$g(y) = c' \mathcal{F} \left(\frac{y}{\lambda f} \right) \quad (4)$$

Fig. 3 displays the electric field distribution in the z - y plane of an on-chip lens, which was obtained through finite-difference time-domain (FDTD) simulations. By utilizing the phase mask described in Eq. 1, the lengths of each slot can be determined using the method outlined in Section II.

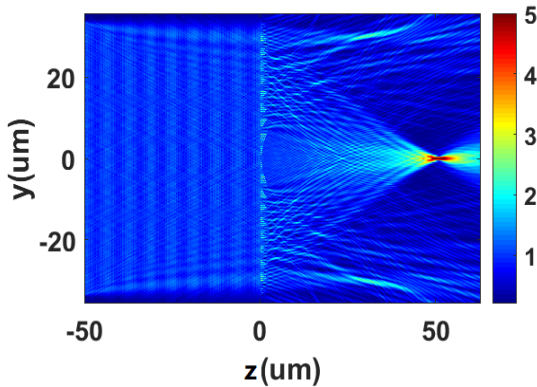


FIGURE 3. The in-plane distribution of $|E_y|$ in the middle of the silicon slab for the on-chip metalens based on the SOI platform, designed using Eq. 1 and the structure described in Sec. II to achieve a focal length of $50 \mu\text{m}$. The incident light is parallel to the optical axis of the lens.

IV. OPTICAL CONVOLUTIONAL LAYER

Convolutional neural networks usually contain multiple convolutional layers, where each layer performs pattern matching using various visual filters and must be trained during a learning process. The output of each convolutional layer in these networks is obtained by convolving input data with convolution kernels, as shown in Fig. 4(a). To perform the convolution operation optically, a 4f system is employed, as shown in Fig. 4(c).

In these linear optical systems, the output data is commonly modeled as a spatially invariant convolution of the input data with the point spread function (PSF) of the system.

The PSF can be described as:

$$PSF(y) = \sum_{a=1}^A W_a(y) \otimes \delta(y - a\Delta y) \quad (5)$$

where \otimes signifies a 1D convolution, W_a corresponds to a 'th convolutional kernel, and Δy is a shift to ensure these outputs are non-overlapping. As a result, the output formation is described as follows:

$$\begin{aligned} E_{out}(y) &= [E_{in} \otimes PSF](y) \\ &= E_{in}(y) \otimes \sum_{a=1}^A W_a(y) \otimes \delta(y - a\Delta y) \\ &= \sum_{a=1}^A [E_{in} \otimes W_a](y) \otimes \delta(y - a\Delta y) \end{aligned} \quad (6)$$

In order to define PSF correctly, we calculate it using the following formula based on the lens functionality:

$$PSF(y) = \mathcal{F}\{ATF(k_y)\} \quad (7)$$

where \mathcal{F} signifies 1D Fourier transfer, $k_y = \frac{y}{\lambda f}$ denotes spatial frequency, λ is the wavelength of light, and f is the focal length of lenses. The aperture transfer function (ATF) is a complex function that is decomposed into amplitude and phase as:

$$ATF(k_y) = A(k_y) \cdot \exp\{j\Delta\phi(k_y)\} \quad (8)$$

In this study, ATF is calculated directly, in contrast to previous hybrid networks [15], [18] in which PSF is first calculated, then ATF optimally is obtained. The metalines proposed in Section II are used to implement the ATF, and the amplitude of the ATF is considered approximately one. By utilizing available learning algorithms, it is possible to directly learn the phase mask of the ATF and the corresponding lengths of its slots.

V. MODELING

In this section, we adopt a similar approach to the electromagnetic modeling of the OCNN presented in [10]. The input images are first pixelated, and each pixelated image, comprising $N \times M$ pixels, is converted into a vector and then encoded into the amplitude of the input electric field to the OCNN. Thus, the input electric field (E^{in}) is a $(N * M) \times 1$ -dimensional vector. The input line is positioned at a distance of Δy from the first layer of the multi-layered metalines.

The electric field propagation from layer l with k neurons (meta-atoms) to the next layer with n neurons is similar to vector-matrix multiplication:

$$\begin{aligned} [m^{l+1}(1), \dots, m^{l+1}(n)] \\ = [m^l(1) * t^l(1), \dots, m^l(k) * t^l(k)] \cdot W \end{aligned} \quad (9)$$

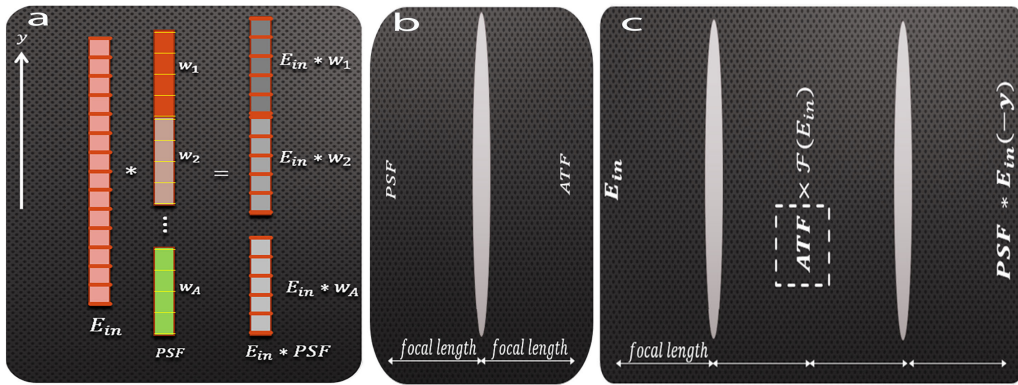


FIGURE 4. Illustration of the Convolution Layer and Optical Implementation using Fourier Transform and 4f System. The Convolution Layer involves convolving input data with a point spread function (PSF) that includes multiple kernels (W_A), as shown in (a). Optical implementation of convolution can be achieved by utilizing Fourier transform through lenses, as demonstrated in (b) where the ATF is the Fourier transform of the PSF and placed in the focal length of a lens. A 4f system structure can be used to perform the convolution operation in the spatial domain, as shown in (c).

In this equation, $t^l(p) = a * \exp(\phi^l(p))$ represents the transmission coefficient of the p -th neuron in the l -th layer, where the amplitude a is close to 1 for a slot width of 140 nm, and the phase shift $\phi^l(p)$ is proportional to the slot length. The amplitudes of the input electric field towards the p -th neuron in the l -th layer and the q -th neuron in the $(l + 1)$ -th layer are denoted as $m^l(p)$ and $m^{l+1}(q)$, respectively. The inter-layer connectivity matrix W is a $k \times n$ transfer matrix that represents wave propagation in the SOI slab waveguide, as given by the Rayleigh-Sommerfeld diffraction equation [33]. The (p, q) -th element of W is given by:

$$W(p, q) = \frac{\Delta y}{r^2} \left(\frac{1}{2\pi r} + \frac{1}{j\lambda} \right) \exp\left(\frac{j2\pi r}{\lambda}\right) \quad (10)$$

In the proposed neural network model, the distance between the p -th neuron in layer l and the q -th neuron in layer $l + 1$ is denoted by r and Δy represents the distance between these two layers. The effective wavelength in the planar waveguide is represented by λ . It has been observed in previous works [10], [11] that the angle-dependent transmission needs to be considered in each layer's output. As shown in Fig. 5(a), the transmission curve can be approximated using a simple exponential function, $\exp\left(\frac{\theta^2}{\sigma}\right)$, which can be fitted to the original curve. The angle can be calculated using Δy and r , as depicted in Fig. 5(b), and the value of σ depends on the meta-atom dimensions, which can be obtained by fitting the exponential function to the original transmission curve. As shown in Fig. 5(a), phase shifts do not vary substantially with angle, so they are not considered in this study.

A. AN ALTERNATIVE APPROACH FOR MODELING FORWARD PROPAGATION

In previous studies [12], [34], a discretized version of the angular spectrum wave propagator from [33] was utilized as a forward propagation model. This involves decomposing the sampled electric field at the output of each metaline plane into a superposition of plane wave components using the discrete Fourier transform. Each plane wave component is then

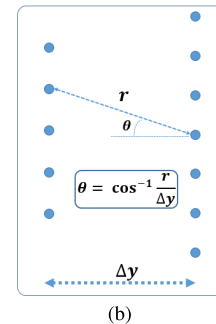
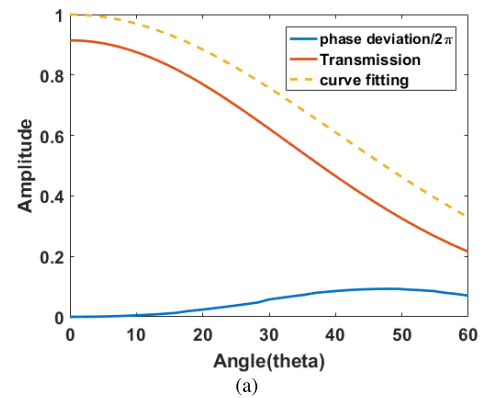


FIGURE 5. (a) Amplitude (red line) and phase (blue line) of the complex transmission coefficient versus the incident angle (θ indicated in (b)) with the slot width of 140 nm and length of 2 μm . The dashed line is an approximation of the amplitude using an exponential function. (b) Schematic of two sequential metaline layers.

multiplied by a proper phase factor to account for the accumulated phase as the wave propagates through the optical system. The resulting electric field is determined by an inverse Fourier transform operation. The vector w^m denotes the length of the slots in the m -th metaline plane, which determines a line of spatially-varying phase shifts experienced by the incident electric field at each metasurface plane. The output electric field E^{out} at the output line of the system is a function of the parameters w^1, w^2, \dots, w^M of the

M metalines, and can be calculated using a series of matrix-vector multiplications as follows:

$$E^{out} = \left(\prod_{m=1}^M \mathcal{F}^+ P^m \mathcal{F} \Phi^{w^m} \right) E^{in} \quad (11)$$

here, \mathcal{F} and \mathcal{F}^+ denote the discrete Fourier transform and its inverse, respectively. P^m is a diagonal matrix that takes into account the plane wave propagation from one metaline layer to the next (see [34]), and Φ^{w^m} is a diagonal matrix containing complex exponentials associated with the phase shifts induced by the meta-atoms on the m -th layer. The diagonal matrix Φ^{w^m} can be expressed as $\Phi^{w^m} = \exp(j\phi^{w^m})$, with ϕ^{w^m} being another diagonal matrix containing the phase shifts related to w^m .

For the purpose of comparing the mentioned model with the model proposed in Sec. V, we use a 4f system implemented by metalines. Fig. 6(a) shows two different electric field inputs imposed on the 4f system. Its outputs, created with 3D FDTD simulation and also with the two mathematical models, are compared as shown in Fig. 6(b) and Fig. 6(c). As is evident in Fig. 6, the first mathematical model follows the simulation more closely than the second.

VI. RESULTS

A. DESIGN OF OCNN FOR PATTERN CLASSIFICATION

In this study, the effectiveness of an OCNN for pattern recognition tasks is demonstrated. The network is trained to recognize letter patterns of ‘X’, ‘Y’, and ‘Z’ using a simple architecture consisting of one convolutional layer and one fully connected layer. The convolutional layer comprises two metalenses and one adjustable metaline, each of which has 135 phase shifters. To account for the pseudo-periodic assumption and improve the mathematical model’s accuracy, every three meta-atoms are considered as a single cell of equal length [11]. Each layer has 45 learnable parameters, and the inter-metaline distances are set to $100\mu m$.

To evaluate the performance of the OCNN, a dataset of binary letter images with amplitude flipping in random pixels is used. The two learnable metalines are pre-trained using 24,000 such matrices, and then the OCNN is trained using the TensorFlow framework. The network’s testing accuracy is evaluated using an additional 3,000 datasets, and it is found that the letter classifications are predicted with an accuracy of over 99%. The network achieves this level of accuracy in only two epochs, as shown in Fig.7(a).

After the phase masks for the two metalines have been learned, the length of each meta-atom is determined using the approach described in Sec. II. The performance of the network is then evaluated by detecting the output signal after it exits the final metaline and propagates a distance of $100\mu m$ until it reaches the output line of the network, which consists of three detector regions arranged in a linear configuration. Each detector is assigned to a specific letter, and the length of each detector is set to $15\mu m$. Our experimental results demonstrate the effectiveness of our OCNN for pattern recognition

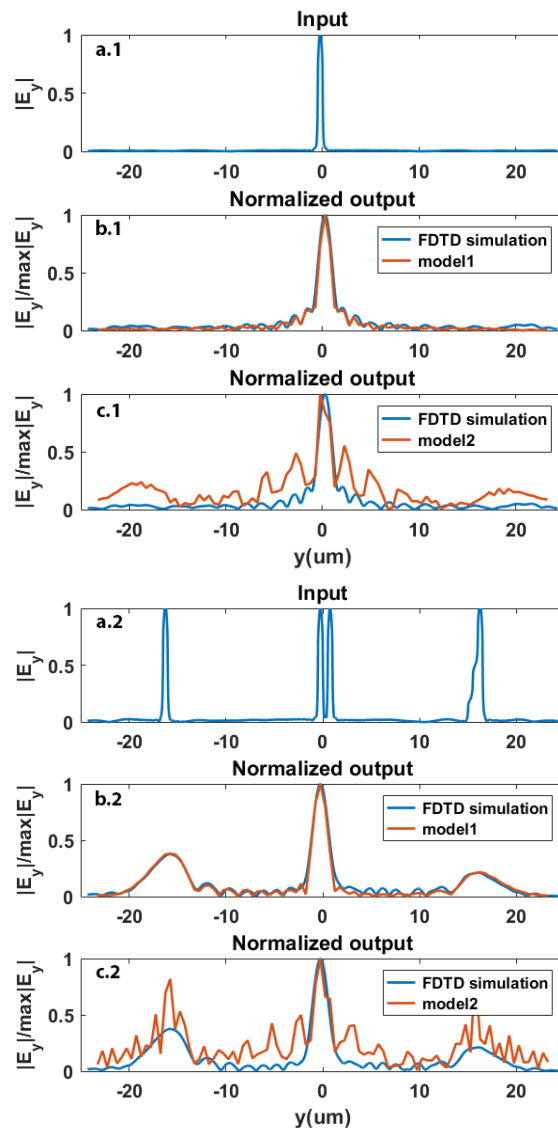


FIGURE 6. The figure depicts a comparison between two mathematical models based on metasurfaces. In panel (a), the input of a 4f optical system with a ($100\mu m$) focal length created by metalines based on SOL. In panel (b), the output of this system is calculated using the mathematical model presented in Sec. V and also compared with the results obtained from 3D FDTD simulations. Similarly, panel (c) shows the system’s output calculated using the model presented in Sec. V-A and compared it with the output obtained from 3D FDTD simulations.

tasks, and our approach shows promise for developing more advanced optical computing technologies in the future.

B. DESIGN VERIFICATION

We used numerical simulation with the Lumerical Mode Solution 2.5D variational FDTD solver to validate the electromagnetic model of the OCNN. The key characteristics of the OCNN structure are summarized in Table 1, and Fig. 9b shows three examples of the optical field intensity distribution generated by the implemented network. The output plane, located $100\mu m$ away from the last metaline, is divided into three regions corresponding to the three channels of classification results, namely up, middle, and down, which

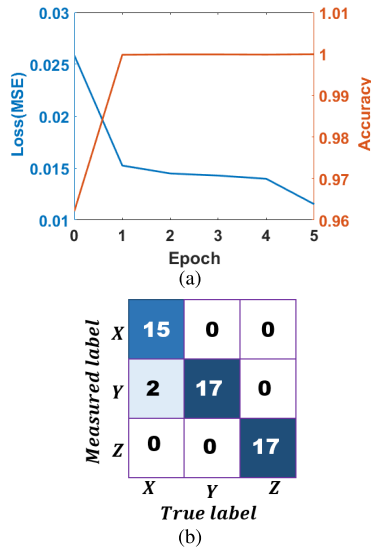


FIGURE 7. (a), the blue line represents the mean squared error (MSE) loss on the training set, while the red line represents the accuracy of the optical convolutional neural network (OCNN) at each epoch during the learning process for a pattern recognition task. (b) shows the confusion matrices obtained from finite-difference time-domain (FDTD) simulations for the pattern classification system.

TABLE 1. Characteristics of OCNN for pattern classification task.

Number of layers	2
Number of cells per layer	45
Number of meta-atoms per cell	3
Number of design variables	90
Number of input pixels	15
Lattice constant	500nm
Length of each metaline-layer	67.5μm
Distance between two neighboring metaline-layer	100μm
Total device size	350λ × 49λ
Number of detectors	3
Detectors arrangement	Linear

correspond to the input letter patterns of ‘X’, ‘Y’, and ‘Z’, respectively. When an input image of the letter ‘X’ is used, the light intensity is highest near the spatial position of the upper channel on the output plane. A detailed illustration of the light intensity distribution along the output plane can be found in Fig. 9c.

We evaluated the performance of the proposed OCNN through numerical simulation, and the resulting confusion matrix is presented in Fig. 7(b). The classification accuracy was found to be 96%, which demonstrates the effectiveness of the proposed design and its potential for various pattern classification applications.

C. METASYSTEM SCALABILITY AND ACCURACY

The scalability of the design algorithm is examined using a more complex system consisting of a modified National Institute of Standards and Technology (MNIST) handwritten digit database with 196-pixel inputs (original datasets are 784-pixel images). For this purpose, we numerically explored the 1D metasystem’s computing capability by designing one,

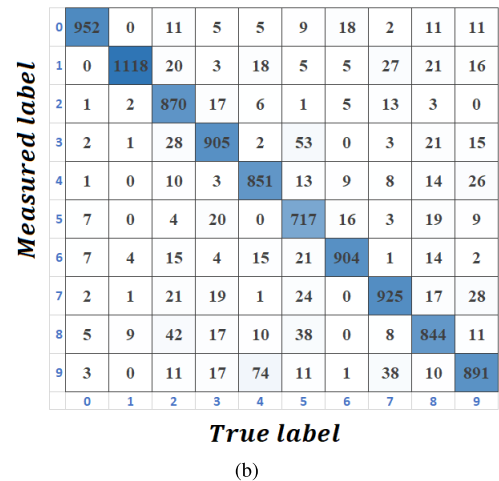
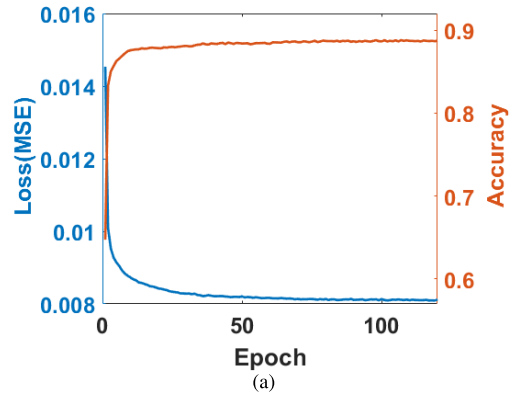


FIGURE 8. (a), the blue line represents the mean squared error (MSE) loss on the training set, while the red line represents the accuracy of the optical convolutional neural network (OCNN) at each epoch during the learning process for image classification on MNIST dataset. (b) shows the confusion matrices obtained from mathematical Python simulations for the handwritten digits classification system.

and Python simulations were used to assess the accuracy of the output. To classify MNIST inputs, we train OCNN with one convolution layer and two fully connected layers, each with 300 learnable parameters, achieving an accuracy of over 89%. Figure 8(a) shows the cost values for the training set and the accuracy values for the test set for each epoch throughout the learning procedure, while Figure 8(b) demonstrates the truth table for the MNIST classification task.

According to [35], the system’s accuracy dramatically improves with the depth of the system (the number of diffraction layers). In order to determine the sensitivity of system accuracy to the number of layers, we train OCNN with one fully connected layer and three ones, while the one convolution layer remains fixed. As a result, they achieve over 84% and 90% accuracy, respectively. As anticipated, OCNN requires fewer learning parameters compared with the current state-of-the-art metasurface-based optical neural network to achieve approximately the same accuracy. Table 2 provides a useful comparison between the OCCN and previously reported optical neural network schemes.

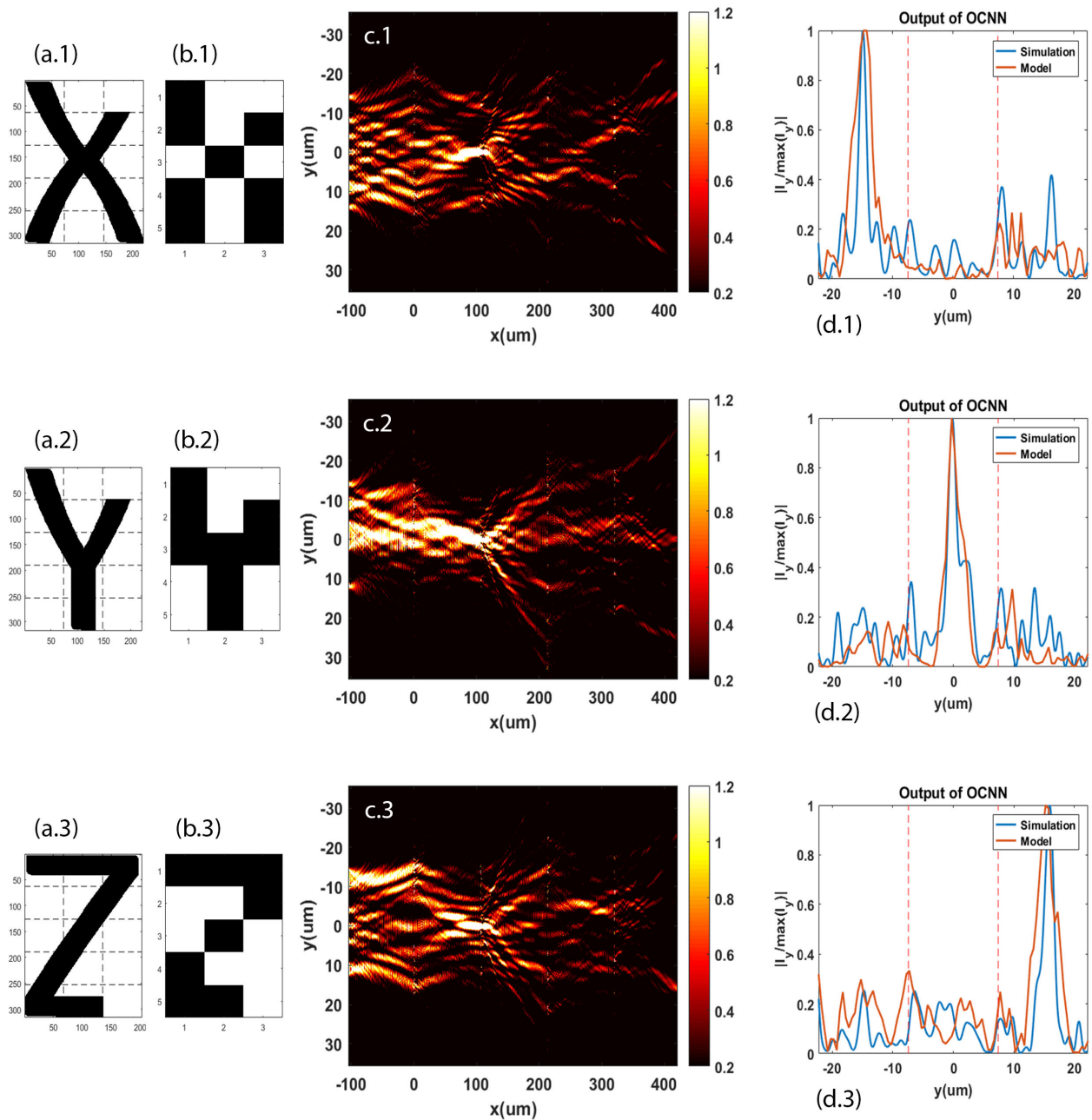


FIGURE 9. (a) Not exact “x”, “Y”, and “Z” in high resolution. To provide OCNN with test and train data, these letters are down-sampled to 15-pixel images. (b) shows the input images used by the designed OCNN to calculate its performance. (c) displays the $x - y$ electric field distribution of the simulated OCNN, based on 2.5D variational FDTD simulations for the input letter shown in (b). (d) shows the output field intensity obtained through electromagnetic modeling described in Section V. The red solid lines represent the output field intensity obtained through electromagnetic modeling, while the blue lines represent the output field intensity obtained through Lumerical Mode Solution. Both are shown at $x = x_{\text{output}}$.

VII. DISCUSSION

As shown in Table 2, in contrast to the proposed OCNN, which was implemented entirely optically, previous optoelectronic works implemented the convolution layer optically and the remaining parts of the designed networks electronically. In these networks, electronically implemented fully-connected layers with more degrees of freedom and nonlinear activation functions play a significant role in the

accuracy of the networks; however, OCNN achieves reasonable precision without these advantages.

Even though our structure is based on linear materials without the equivalent of a nonlinear activation function within the optical network, hybrid integration of active materials can incorporate reconfigurability and nonlinear activation functions into the metasystem platform. For instance, phase change materials with a high refractive index contrast can

TABLE 2. Comparison of neuron networks-based image classifiers.

	Convolution neural network [14]		Convolution neural network [15]	Convolution neural network [23]	Convolution neural network [16]	Diffraction neural network [10]	OCNN
Signal processing	Optoelectronic		Optoelectronic	Optoelectronic	Optoelectronic	All-optical	All-optical
Programmed layer(s)	One amplitude-only layer of DMD		One phase-only layer of diffractive optic elements	optical delay lines of MZI	One amplitude-only layer of optical mask	Phase-only layers of metasurface	Phase-only layers of metasurface
Reconfigurable	Yes		No	No	No	No	No
Technology	Free-space		Free-space	On-chip	Free-space	On-chip	On-chip
Weight matrix	16*228*208		16*32*32	[23]	16*3*3	1000*3	300*3
Dataset	MNIST	CIFAR	CIFAR-10	MNIST	MNIST	MNIST	MNIST
Input size	28*28	208*208	experiment	28*28	28*28	28*28	14*14
Accuracy	98%	63%	51%	97%	97%	92%	89.7%

fill those slots and offer sufficient phase tenability [36] for a fully programmable metasystem. Some active materials demonstrate highly nonlinear responses (like two-photon absorption-related free carrier absorption or absorption saturation) and are transparent at telecommunication wavelength ranges; these materials can be utilized as nano-scale activation functions in diffractive networks with solution processing [37], [38].

Even when using linear optical materials to create a diffractive deep neural network (D2NN), the optical network designed by deep learning exhibits a “depth” benefit, meaning a single diffractive layer lacks the degrees of freedom required for the same classification accuracy, power efficiency, and signal contrast at the output plane that multiple diffractive layers can in tandem achieve for the task at hand. For a linear diffractive optical network, the entire wave propagation and diffraction phenomena happening between the input and output planes can be compressed into a single matrix operation; however, this arbitrary mathematical operation defined by a number of learnable diffractive layers is unable to be carried out by just one diffractive layer placed between the same input and output planes. Therefore, multiple diffractive layers forming a D2NN demonstrate the depth advantage, statistically perform better than a single diffractive layer trained for the same classification task and achieve greater accuracy, as also discussed in the supplementary materials of [13].

In the mathematical model used in this study, there are some simplifications affecting accuracy. The assumption of the periodicity of slots in each metaline is the first and most significant factor that led to the model’s results not matching simulation results. As described in Section I, phase shifts of meta-atoms are measured where they are periodic, whereas the lengths of slots along a metaline vary in OCNN. In addition, transmission is presumed to be close to one, and the back reflection of incident waves on metalines is disregarded. In addition, the mathematical model neglects the effect of the edges in each metaline. In addition to the primary factor affecting the accuracy of the mathematical model, the approximation of the impact of the incident angle of an input wave on each metaline with an exponential transmission curve while its phase shift is ignored also has a minor effect.

VIII. CONCLUSION

In conclusion, our work presents a novel passive all-optical architecture for implementing a 1D OCNN using multi-layer metasurfaces. This design offers several advantages, including lower structural complexity and reduced implementation costs, enhancing its practicality for real-world applications. With a compact size and small footprint, our OCNN achieves an impressive accuracy of 96% on a similar image classification task while utilizing significantly fewer learnable parameters compared to state-of-the-art integrated photonic metasystems. Furthermore, our proposed OCNN exhibits scalability and robustness, demonstrated by its accuracy exceeding 89% in handwritten digit classification. The utilization of 1.55 μm silicon-compatible technology in our OCNN is particularly notable. This technology is well-developed and widely used in various optical devices, providing a mature platform for seamless integration with other optical components and facilitating the implementation of optical CNN kernels. This capability opens up exciting possibilities for developing more complex all-optical systems with applications in object recognition, natural language processing, and portable devices.

ACKNOWLEDGMENT

The datasets used during the current study were provided by Zenodo repository (<https://doi.org/10.5281/zenodo.6339743>). Numerical simulation support was provided by the Sharif University of Technology.

DISCLOSURES

The authors declare no conflicts of interest.

REFERENCES

- [1] E. Akleman, “Deep learning,” *Computer*, vol. 53, no. 9, p. 17, Sep. 2020.
- [2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012.
- [3] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [4] P. Y. Simard, D. Steinkraus, and J. C. Platt, “Best practices for convolutional neural networks applied to visual document analysis,” in *Proc. 7th Int. Conf. Document Anal. Recognit.*, Aug. 2003, pp. 1–6.
- [5] M. Al-Qizwini, I. Barjasteh, H. Al-Qassab, and H. Radha, “Deep learning algorithm for autonomous driving using GoogLeNet,” in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2017, pp. 89–96.

- [6] L. Eren, T. Ince, and S. Kiranyaz, "A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier," *J. Signal Process. Syst.*, vol. 91, no. 2, pp. 179–189, Feb. 2019.
- [7] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, Mar. 2016.
- [8] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mech. Syst. Signal Process.*, vol. 151, Apr. 2021, Art. no. 107398.
- [9] S. Bandyopadhyay, A. Sludds, S. Krastanov, R. Hamerly, N. Harris, D. Bunandar, M. Streshinsky, M. Hochberg, and D. Englund, "Single chip photonic deep neural network with accelerated training," 2022, *arXiv:2208.01623*.
- [10] Z. Wang, L. Chang, F. Wang, T. Li, and T. Gu, "Integrated photonic metasystem for image classifications at telecommunication wavelength," *Nature Commun.*, vol. 13, no. 1, pp. 1–8, Apr. 2022.
- [11] T. Fu, Y. Zang, H. Huang, Z. Du, C. Hu, M. Chen, S. Yang, and H. Chen, "On-chip photonic diffractive optical neural network based on a spatial domain electromagnetic propagation model," *Opt. Exp.*, vol. 29, no. 20, pp. 31924–31940, 2021.
- [12] S. Zarei, M. Marzban, and A. Khavasi, "Integrated photonic neural network based on silicon metalines," *Opt. Exp.*, vol. 28, no. 24, pp. 36668–36684, Nov. 2020.
- [13] 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, no. 6406, pp. 1004–1008, Sep. 2018.
- [14] M. Miscuglio, Z. Hu, S. Li, J. K. George, R. Capanna, H. Dalir, P. M. Bardet, P. Gupta, and V. J. Sorger, "Massively parallel amplitude-only Fourier neural network," *Optica*, vol. 7, no. 12, pp. 1812–1819, 2020.
- [15] J. Chang, V. Sitzmann, X. Dun, W. Heidrich, and G. Wetzstein, "Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification," *Sci. Rep.*, vol. 8, no. 1, pp. 1–10, Aug. 2018.
- [16] W. Shi, Z. Huang, H. Huang, C. Hu, M. Chen, S. Yang, and H. Chen, "LOEN: Lensless opto-electronic neural network empowered machine vision," *Light, Sci. Appl.*, vol. 11, no. 1, pp. 1–12, May 2022.
- [17] X. Xu, L. Zhu, and W. Zhuang, "Convolutional neural networks with coherent nanophotonic circuits," in *Proc. 10th Int. Symp. Adv. Opt. Manuf. Test. Technol., Intell. Sens. Technol. Appl.*, Dec. 2021, pp. 207–213.
- [18] C. M. V. Burgos, T. Yang, Y. Zhu, and A. N. Vamivakas, "Design framework for metasurface optics-based convolutional neural networks," *Appl. Opt.*, vol. 60, no. 15, pp. 4356–4365, 2021.
- [19] Z. Wu, M. Zhou, E. Khoram, B. Liu, and Z. Yu, "Neuromorphic metasurface," *Photon. Res.*, vol. 8, no. 1, pp. 46–50, 2020.
- [20] S. Abdollahramezani, O. Hemmatyar, and A. Adibi, "Meta-optics for spatial optical analog computing," *Nanophotonics*, vol. 9, no. 13, pp. 4075–4095, Sep. 2020.
- [21] Z. Wang, T. Li, A. Soman, D. Mao, T. Kananen, and T. Gu, "On-chip wavefront shaping with dielectric metasurface," *Nature Commun.*, vol. 10, no. 1, pp. 1–7, Aug. 2019.
- [22] S. Colburn, Y. Chu, E. Shilzerman, and A. Majumdar, "Optical frontend for a convolutional neural network," *Appl. Opt.*, vol. 58, no. 12, pp. 3179–3186, 2019.
- [23] H. Bagherian, S. Skirlo, Y. Shen, H. Meng, V. Ceperic, and M. Soljacic, "On-chip optical convolutional neural networks," 2018, *arXiv:1808.03303*.
- [24] Y. Chen, T. Krishna, J. S. Emer, and V. Sze, "Eyeriss: An energy-efficient reconfigurable accelerator for deep convolutional neural networks," *IEEE J. Solid-State Circuits*, vol. 52, no. 1, pp. 127–138, Jan. 2017.
- [25] X. Xu, M. Tan, B. Corcoran, J. Wu, A. Boes, T. G. Nguyen, S. T. Chu, B. E. Little, D. G. Hicks, R. Morandotti, A. Mitchell, and D. J. Moss, "11 Tops photonic convolutional accelerator for optical neural networks," *Nature*, vol. 589, no. 7840, pp. 44–51, Jan. 2021.
- [26] T. Yan, J. Wu, T. Zhou, H. Xie, F. Xu, J. Fan, L. Fang, X. Lin, and Q. Dai, "Fourier-space diffractive deep neural network," *Phys. Rev. Lett.*, vol. 123, no. 2, Jul. 2019, Art. no. 023901.
- [27] T. Fu, Y. Zang, Y. Huang, Z. Du, H. Huang, C. Hu, M. Chen, S. Yang, and H. Chen, "Photonic machine learning with on-chip diffractive optics," *Nature Commun.*, vol. 14, no. 1, p. 70, Jan. 2023.
- [28] S. Zolfaghary pour and K. Arik, "Ultra-broadband polarization-independent perfect absorber based on phase change material (Ge₂Sb₂Te₅ or GST) for the visible and infrared regions," *Opt. Quantum Electron.*, vol. 55, no. 2, p. 141, 2023.
- [29] S. Abdollahramezani, O. Hemmatyar, M. Taghinejad, H. Taghinejad, A. Krasnok, A. A. Eftekhari, C. Teichrib, S. Deshmukh, M. A. El-Sayed, E. Pop, M. Wuttig, A. Alù, W. Cai, and A. Adibi, "Electrically driven reprogrammable phase-change metasurface reaching 80% efficiency," *Nature Commun.*, vol. 13, no. 1, p. 1696, Mar. 2022.
- [30] S. Zarei and A. Khavasi, "Inverse design of on-chip thermally tunable varifocal metalens based on silicon metalines," *IEEE Access*, vol. 9, pp. 73453–73466, 2021.
- [31] W. T. Chen, A. Y. Zhu, V. Sanjeev, M. Khorasaninejad, Z. Shi, E. Lee, and F. Capasso, "A broadband achromatic metalens for focusing and imaging in the visible," *Nature Nanotechnol.*, vol. 13, no. 3, pp. 220–226, Mar. 2018.
- [32] B. E. Saleh and M. C. Teich, *Fundamentals of Photonics*. Hoboken, NJ, USA: Wiley, 2019.
- [33] J. Goodman, *Introduction to Fourier Optics* (McGraw-Hill Physical and Quantum Electronics Series). New York, NY, USA: W. H. Freeman, 2005.
- [34] A. Backer, "Computational inverse design for cascaded systems of metasurface optics," *Opt. Exp.*, vol. 27, no. 21, pp. 30308–30331, Oct. 2019.
- [35] D. Mengu, Y. Luo, Y. Rivenson, and A. Ozcan, "Analysis of diffractive optical neural networks and their integration with electronic neural networks," *IEEE J. Sel. Topics Quantum Electron.*, vol. 26, no. 1, pp. 1–14, Jan. 2020.
- [36] N. Wang, T. Li, B. Sun, Z. Wang, L. Zhou, and T. Gu, "Focusing and defocusing switching of an indium selenide-silicon photonic metalens," *Opt. Lett.*, vol. 46, no. 17, pp. 4088–4091, 2021.
- [37] F. Wang, L. Zhao, Y. Xiao, T. Li, Y. Wang, A. Soman, H. Lee, T. Kananen, X. Hu, B. P. Rand, and T. Gu, "Controlling microring resonator extinction ratio via Metal–Halide perovskite nonlinearity," *Adv. Opt. Mater.*, vol. 9, no. 22, Nov. 2021, Art. no. 2100783.
- [38] F. Wang, Z. Wang, D. Mao, M. Chen, Q. Li, T. Kananen, D. Fang, A. Soman, X. Hu, C. B. Arnold, and T. Gu, "Light emission from self-assembled and laser-crystallized chalcogenide metasurface," *Adv. Opt. Mater.*, vol. 8, no. 8, Apr. 2020, Art. no. 1901236.



OMID POORDASHTBAN was born in Bandar-Abbas, Hormozgan, Iran, in 1997. He received the B.S. degree in electrical engineering from the Isfahan University of Technology, Isfahan, Iran, in 2019. He is currently pursuing the M.S. degree in electrical engineering (fields and waves) with the Sharif University of Technology, Tehran, Iran. His research interests include metasurfaces, photonics, and analog optical computing.



MAHMOOD REZA MARZABN was born in Melbourne, Australia, in 1994. He received the bachelor's degree in electrical engineering from Shiraz University, in 2018, and the master's degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2021. He is currently pursuing the Ph.D. degree with the Georgia Institute of Technology. His research interests include nanophotonics, computational optics, optical neural networks, and artificial intelligence.



AMIN KHAVASI was born in Zanjan, Iran, in 1984. He received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2006, 2008, and 2012, respectively. Since 2013, he has been with the Department of Electrical Engineering, Sharif University of Technology, where he is currently an Associate Professor. His research interests include plasmonics, metamaterials, silicon photonics, and analog optical computing.