Research Article

Zhengyang Duan, Hang Chen and Xing Lin*

Optical multi-task learning using multi-wavelength diffractive deep neural networks

Abstract: Photonic neural networks are brain-inspired information processing technology using photons instead of electrons to perform artificial intelligence (AI) tasks. However, existing architectures are designed for a single task but fail to multiplex different tasks in parallel within a single monolithic system due to the task competition that deteriorates the model performance. This paper proposes a novel optical multitask learning system by designing multi-wavelength diffractive deep neural networks (D²NNs) with the joint optimization method. By encoding multi-task inputs into multi-wavelength channels, the system can increase the computing throughput and significantly alleviate the competition to perform multiple tasks in parallel with high accuracy. We design the two-task and four-task D²NNs with two and four spectral channels, respectively, for classifying different inputs from MNIST, FMNIST, KMNIST, and EMNIST databases. The numerical evaluations demonstrate that, under the same network size, multi-wavelength D²NNs achieve significantly higher classification accuracies for multitask learning than single-wavelength D²NNs. Furthermore, by increasing the network size, the multi-wavelength D²NNs for simultaneously performing multiple tasks achieve comparable classification accuracies with respect to the individual training of multiple single-wavelength D²NNs to perform tasks separately. Our work paves the way for developing the wavelength-division multiplexing technology to achieve high-throughput neuromorphic photonic computing and more general AI systems to perform multiple tasks in parallel.

Keywords: diffractive deep neural networks; multi-wavelength photonic neural networks; optical multi-task learning.

1 Introduction

Photonic computing utilizes photons instead of electrons for computation, which possesses the inherent advantages of light-speed processing, low-power consumption, and high-throughput capability [1–5]. Photonic neural networks (PNNs) [6–8], which implement the artificial neural network model based on photonic computing, can achieve leapfrog improvement in computing speed and energy efficiency. Therefore, it’s considered one of the most promising solutions to support the sustainable development of artificial intelligence (AI) in the post-Moore era [9, 10]. Among different PNN architectures, diffractive deep neural networks (D²NN) [11, 12] can achieve large-scale neural information processing and have attracted a vast amount of interest. D²NN consists of layers of diffractive neurons and their optical interconnections based on the diffraction of light, which can be trained with deep learning optimization methods to fit desired mapping functions between input optical fields and output detector measurements.

In the previous studies of D²NN architecture [11–16], monochromatic plane waves are used to encode the input data and propagate through modulation layers, performing a specific task. Inspired by biological intelligence [17], the multiplexing of different tasks in a single D²NN system is of great importance in improving its generalization and expanding its applications for different scenarios. However, performing multiple AI tasks in parallel with a monolithic D²NN system remains challenging. One of the major obstacles is the competition among tasks during the training, which leads to catastrophic forgetting [18, 19]. Catastrophic
forgetting occurs when systems are trained on multitasks, which leads to a tendency for knowledge of previously learned tasks to be abruptly lost while learning new tasks, resulting in the deterioration of performance on every single task. Previous solutions \cite{20, 21} require the mechanical movement of optical elements to switch between tasks one at a time or require the design of multiple different D2NN systems, one for each task, significantly increasing the hardware complexity.

Here, we propose an optical multitask learning monolithic system design that can simultaneously perform multiple classification tasks on different databases without mechanical movement by developing multi-wavelength D2NNs. Different from previous broadband D2NNs \cite{22, 23}, the wavelength dimension is exploited in this work to improve the computing throughput, which encodes different inputs into different wavelength channels and performs photonic computing in both spatial and spectral dimensions. We demonstrate that the multi-wavelength D2NNs allow for high-parallel processing of multiple inputs and significantly alleviate the competition among different tasks to preserve the high performance of each task.

In this design, the multi-wavelength D2NN has $N$ ($N \geq 2$) different parallel wavelength channels, encoding $N$ different inputs in parallel, and the detection area of each category is segmented into $N$ parts, where each part represents the category of input encoded at the corresponding wavelength channel. We use the multiwavelength joint optimization method with the loss functions of softmax cross-entropy and energy efficiency constraint to train the D2NN. We first verify the high-parallel characteristic of a three-wavelength D2NN by classifying three different inputs in parallel based on the MNIST database, where the accuracy at each wavelength is comparable to training three single-wavelength D2NN with sequential inputs. To perform multiple tasks in parallel, we encode the inputs from different databases into different wavelength channels. We utilize the two-wavelength and four-wavelength D2NNs for performing the two-task and four-task classifications, respectively, based on the databases of MNIST, FMNIST, KMNIST, and EMNIST. With the increase in task numbers, the multi-wavelength D2NNs achieve significantly higher classification accuracy than the single-wavelength D2NNs and maintain the model accuracy for each task with larger network sizes, demonstrating the great advantages of multiwavelength D2NNs in realizing optical multitask learning.

2 Methods

As shown in Figure 1, the proposed optical multitask learning system achieves multiple tasks in parallel by designing a multiwavelength D2NN. Each wavelength channel ($\lambda_i, i = 1, \ldots, N$) encodes the input targets of each task. By using the approximation theory of multiwavelength optical systems \cite{24–27}, the transformation of multiwavelength optical fields can be recognized as a combination of independent transformation of coherent optical fields at each wavelength, following the principle of superposition of optical intensities. The input optical fields $U_{\lambda_i}$ at the wavelength of $\lambda_i$, encoding the input targets of task $i$, are

![Figure 1: The architecture of multiwavelength D2NNs for optical multitask learning. The incident light comprises multiple wavelength channels, where each detection region is correspondingly segmented into multiple sub-regions. Each sub-region represents an input category type for each task.](image)
transformed by the D^2NN before the detection. We consider the linear D^2NN at the wavelength of \( \lambda_i \) with the complex transform function of \( M_{j_i}(\Phi) \), where \( \Phi \) represents the learnable phase modulation coefficient of the diffractive elements at multiple phase-only diffractive layers. The detailed derivation of complex transform function can be found in [11, 12, 23]. We first assume that the phase modulation coefficients of each diffractive layer are the same under different wavelengths with the multiwavelength diffractive optical element (DOE) design (see Discussion section). Then, the output optical fields at the \( i \)th wavelength \( \lambda_i \) can be formulated as: \( U_{j_i} = M_{j_i}(\Phi)U_{j_i} \), and the detector measures the intensity distribution of output optical fields that can be formulated as: \( I_{j_i} = \sum_{j_i} |U_{j_i}|^2 \). For the multiwavelength D^2NN, the total intensity distribution of different wavelengths can be formulated as the superposition of detected intensity distribution at each wavelength: \( I = \sum_{j_i} I_{j_i} = \sum_{j_i} |M_{j_i}(\Phi)U_{j_i}|^2 \).

We develop the joint optimization method to train multiwavelength D^2NNs for performing optical multitask learning. Each category detection area at the output plane is divided into multiple sub-regions \( \{D_i^j\} \), where \( i = 1, \ldots, N \) denotes the index of the task encoded at \( i \)th wavelength \( \lambda_i \); and \( j = 1, \ldots, M \) denotes the index of detection areas, representing the index of categories. We calculate the average intensity of the \( i \)th sub-region among the \( M \) category detection areas, i.e., \( P_i = \{ \text{avg}(U(D_i^j)), j = 1, \ldots, M \} \), where \( I(D_i^j) = \sum_{j_i} I_{j_i}(D_i^j) \) is for the broadband wavelength detection without using spectral filters on the detector. The information encoded in different spectral channels of multiwavelength D^2NNs is independent of each other during the diffraction propagation and modulation, where the presence of channel crosstalk only existed in the detection plane. The minimization of such channel crosstalk can be achieved by incorporating the additional wavelength selective filters or developing the multiwavelength D^2NN training method with a joint optimization process. The wavelength selective filter can be applied to each sub-region to completely eliminate the crosstalk during intensity detections among wavelength channels for further improving the task performance. In this case, each sub-region for each task only detects the optical signals at the corresponding wavelength channel, with which \( I(D_i^j) = I_{j_i}(D_i^j) \). To remove the filter for lower system complexity, the proposed joint optimization method during the training of multiwavelength D^2NNs can minimize the channel crosstalk during the detection. The category type of inputs for the \( i \)th task is determined by finding the sub-region of the detection area with maximum average intensity, i.e., the index of the maximum value in the vector \( P_i \). Besides, we further include the constraint to maximize the energy transmission efficiency of multiwavelength D^2NNs by minimizing the optical energy outside the category detection areas. Therefore, the joint optimization problem for multiwavelength D^2NNs training can be formulated as:

\[
\min_{\Phi} \left( \sum_{i} L(P_i, G_i) + \sum_{j_i} \text{MSE} \left( I_{j_i} - \sum_{j} I_{j_i}(D_i^j) \right) \right),
\]

where \( L(P_i, G_i) \) represents the softmax cross-entropy loss function of \( i \)th task at the wavelength of \( \lambda_i \) between the detection \( P_i \) and ground truth label \( G_i \); \( G_i \) is a one-hot vector with a length of \( M \); \( \text{MSE}(I_{j_i} - \sum_{j} I_{j_i}(D_i^j)) \) represents the total energy of optical intensity outside the sub-regions of category detection areas evaluated with the mean square error.

In the process of training, different wavelength channels share the same phase modulation coefficients at each diffractive layer that are iteratively updated to perform multitask functions by solving the joint optimization problem in Eq. (1). We use the stochastic gradient descent approach to train the multi-wavelength D^2NNs. The input targets of the training datasets of different tasks are encoded into the amplitude of optical fields at different wavelengths to feed into the network input layer. The residual error of network outputs with respect to ground truth labels and the total optical energy outside the category detection areas are calculated according to Eq. (1), which are used to perform the error back-propagation to optimize the network structure and the phase modulation coefficients of optical diffractive elements (DOEs).

To further improve the viability of multiwavelength DOEs, another design strategy can be incorporated to optimize the relative height map \( \Delta Z \) instead of the phase map \( \Phi \) of DOEs during the training, with which the complex transform function of multiwavelength D^2NNs can be changed from \( M_{j_i}(\Phi) \) to \( M_{j_i}(\Delta Z) \). As the relative height map of each diffractive layer is the same under different wavelength channels, the phase modulation value at each wavelength is determined by the wavelength-dependent material refractive index. The transformation between diffractive elements’ relative height map \( \Delta Z \) and the wavelength-dependent phase values \( \Phi_{j_i} \) under a certain wavelength channel \( \lambda_i \) can be formulated as: \( \Phi_{j_i} = 2\pi \Delta n_{j_i} \Delta Z / \lambda_i \), where \( \Delta n_{j_i} \) is the refractive index difference between the base material, e.g., the \( \text{SiO}_2 \) and air [11, 28]. With the same training process, the numerical analysis (see Table 1) validates that both design strategies, i.e., optimizing

### Table 1: Performance comparisons between the multiwavelength and single-wavelength D^2NNs for two-task classifications based on the design strategies of optimizing the phase maps or height maps of DOEs.

| Diffractive neural network model | Diffractive neural network size | Accuracy by optimizing phase maps (\( \Phi \)) | Accuracy by optimizing height maps (\( \Delta Z \)) |
|---------------------------------|---------------------------------|-----------------------------------------------|-----------------------------------------------|
| Single-task, single-wave        | 200 \( \times \) 200 \( \times \) 5 | 97.1% | 87.5% |
| Multitask, single-wave          | 200 \( \times \) 200 \( \times \) 5 | 92.4% | 83.1% |
| Multitask, multiwave (w/o filter) | 200 \( \times \) 200 \( \times \) 5 | 95.6% | 86.4% |
| Multitask, multiwave (w/Filter) | 200 \( \times \) 200 \( \times \) 5 | 95.9% | 87.0% |
| Multitask, multiwave (w/o Filter) | 400 \( \times \) 400 \( \times \) 5 | 97.5% | 91.7% |
| Multitask, multiwave (w/Filter) | 400 \( \times \) 400 \( \times \) 5 | 97.6% | 94.8% |
the phase maps or height maps, for training multiwavelength D²NNs achieve comparable performance, where optimizing the height maps has a slightly lower classification accuracy.

3 Results

3.1 Multiwavelength D²NNs for high-parallel classification

We first verify the application of multiwavelength D²NNs for high-parallel classification that can simultaneously classify multiple inputs in performing a single task. In this work, the multiwavelength D²NN models were implemented using Python version 3.8.13 and PyTorch framework version 1.11.0, using a desktop computer (GeForce GTX 3090 Ti Graphical Processing Unit, GPU, and Intel (R) Xeon (R) Gold 6226R CPU @2.90 GHz and 512 GB of RAM, running an Ubuntu 20.04 operating system). To demonstrate, we build a three-wavelength D²NN architecture with five diffractive modulation layers, which is applied for classifying the MNIST database that can recognize three handwritten digits at each instant of time. We consider the visible wavelength ranging from 400 nm to 700 nm, where the input light source was set as the combination of three wavelengths of 400 nm, 550 nm, and 700 nm, encoding three handwritten digits, respectively. Therefore, each category detection area at the output plane is correspondingly segmented into three sub-regions. The detection region and sub-region layout have minor effects on the performance, where we adopt a similar detection region layout in [28] and separate the detection regions into sub-regions. The Adam optimizer is used for network training to optimize the phase modulation coefficients \( \Phi \) or the relative height map \( \Delta Z \) of the optical diffractive elements. Each optical diffractive element size was set to \( 4 \mu m \times 4 \mu m \). We first evaluate the performance of multi-wavelength D²NNs by setting the modulation element number at each network layer to \( 200 \times 200 \), corresponding to the network layer size of \( 0.8 \text{mm} \times 0.8 \text{mm} \) (see Figure 2b). We further evaluate and compare the network performance under different modulation element numbers at each layer, i.e., \( K \times K \), where \( K = 200, 400, 600, \) and \( 800 \) (see Figure 2d). The layer number was set to 5, and the distance between successive layers was optimized according to the maximum half-cone diffraction angle theory [11, 28]. With a training batch size of 32, the initial learning rate is set to 0.01 and is reduced by half, i.e., multiplied by 0.5, after every epoch during the training. The network training converges after five epochs to achieve the desired mapping function for the multichannel inputs and output. The network was trained and blind-tested with 60,000 and 10,000 multiwavelength samples, respectively. The training and testing samples of each wavelength were constructed by randomly shuffling 60,000 and 10,000 handwritten digits from the MNIST training and testing sets, respectively. For the network layer with a modulation element number of \( K \times K \), each digit with a pixel number of \( 28 \times 28 \) is resized to \( K/2 \times K/2 \) and padded to \( K \times K \).

The numerical evaluation results are shown in Figure 2, where the performance of multiwavelength D²NNs is validated with and without wavelength selective filters on each category detection area. Figure 2a–c shows an exemplary result of simultaneously classifying three handwritten input digits, i.e., “7”, “2”, and “5”, encoded in the wavelengths of 700 nm, 550 nm, and 400 nm, respectively, under the phase modulation element numbers of \( 200 \times 200 \) at each layer. The classification result of each wavelength channel is determined by finding the maximum average intensity value among the corresponding sub-regions of category detection areas, indicated with three white arrows for three input digits, as shown in Figure 2b–c, left. The energy distributions of the classification results of three inputs at different wavelength channels in Figure 2b–c show that the proposed system could prominently identify the sub-region with maximum average intensity for the correct categorization. Due to the use of wavelength selective filters to completely eliminate the wavelength crosstalk during the detection, the classification accuracies of multiwavelength D²NNs with wavelength selective filters are 95.9%, 96.4%, and 96.9% for the wavelengths of 700 nm, 550 nm, and 400 nm, respectively, which are slightly higher than the classification accuracies of broadband wavelength detection without wavelength selective filters, i.e., 95.0%, 95.7%, and 96.4%, respectively. For both network settings, the classification accuracies of multi-wavelength D²NNs further improve at each wavelength with the increase of the modulation element numbers at each network layer, as shown in Figure 2d. Under the modulation element number of \( 800 \times 800 \) at each layer, the classification accuracies of multiwavelength D²NNs with wavelength selective filters, achieving 98.2%, 98.1%, and 98.1% for the wavelengths of 700 nm, 550 nm, and 400 nm, respectively, are comparable to training three single-wavelength D²NNs with the serial inputs, i.e., sequential input of digits. The results verify that multiwavelength D²NN can significantly increase the parallel computing capability. Using multiwavelength D²NNs for multitask learning by encoding different tasks into different channels enables different machine learning tasks to be implemented in parallel within a single system.
3.2 Optical multitask learning using multiwavelength D²NNs

To demonstrate the capability of multiwavelength D²NNs for optical multitask learning, we first construct a two-task classifier for classifying both the MNIST database (task I) and the fashion-MNIST (FMNIST) database (task II). Both databases include 60,000 training samples and 10,000 testing samples with 10 category numbers. For the multiwavelength D²NN, the training and testing samples of each wavelength were constructed by randomly shuffling the training and testing samples from each database, respectively. Therefore, the two-wavelength D²NN was constructed by dividing each of 10 detector areas into two sub-regions, where the upper and lower regions represent the classification results of tasks I and II, respectively, as shown in Figure 3. The handwritten digits of task I are encoded in the wavelength of 700 nm and the fashion products of Task II are encoded in the wavelength of 400 nm. With other network settings the same in Figure 2, we first set the two-wavelength D²NN to have five diffractive layers, each layer with a phase modulation element number of 200 × 200, without the wavelength selective filters on the detector. Figure 3a shows an exemplar result for simultaneously classifying a handwritten digit “7” with the category number of 7 from the MNIST database and a fashion product “pullover” with the category number of “2” from the FMNIST database. The maximum average intensity outputs of task I and task II were focused on the upper sub-regions of the No. 7 detector area and the lower sub-regions of the No. 2
Figure 3: Multiwavelength D²NNs working under two wavelengths for classifying both the MNIST and FMNIST databases. (a) Images from the MNIST and FMNIST datasets are encoded in wavelengths of 700 nm and 400 nm, respectively. The categories of two inputs are determined by finding the corresponding sub-regions with maximum average intensity as indicated by the white arrow. (b) Energy distributions of the classification results of two tasks demonstrated the success of the proposed approach for classifying two inputs. (c, left) Confusion matrices and (c, middle) energy matrices (percentage) of two tasks, corresponding to the classification accuracies of 95.6% and 86.8% for Task I and Task II, respectively. (c, right) Energy matrices (percentage) of two single-wavelength D²NNs for individually performing each of the two tasks, respectively. ST and SW, single-task using single-wavelength; MT and MW, multitask using multiwavelength.
were also shown in Figure 3b. The energy distributions of a fashion product “trouser” with the category number of “1” and a handwritten digit “6” with the category number of six and arrow in Figure 3a. The addition exemplar results, such as detector area, respectively, which were marked by the white of two tasks, are shown in Figure 3(c, left) and Figure 3(c, middle), respectively. The average energy percentages of correct categories are 20.8% and 21.8% for two tasks, respectively. We further compare the performance of two-wavelength D2NNs for performing two tasks in parallel model on the test datasets of MNIST and FMNIST achieves classification accuracies of 95.6% and 86.8%, respectively. The corresponding confusion matrices and energy distribution matrices, statistically summarizing the classification results of all samples and energy distribution percentages of two tasks, are shown in Figure 3(b) and Figure 3(c), respectively. The average energy percentages of correct categories are 20.8% and 21.8% for two tasks, respectively. We further compare the performance of two-wavelength D2NNs for performing two tasks in parallel with respect to the single-wavelength D2NNs for performing two tasks in parallel by overlapping to multiplex two images from two datasets, respectively, as the network input, as shown in Table 1. The classification accuracies of single-wavelength D2NNs for performing two tasks in parallel are 92.4% and 83.1%, respectively, which is much lower than the two-wavelength D2NNs. We also train two single-wavelength D2NNs for individually performing each of the two tasks, where the classification accuracies are 97.1% and 87.5% for tasks I and II, respectively. The average energy percentages of correct categories are 24.2% and 26.1% for two tasks with two energy distribution matrices shown in Figure 3(c, right), respectively. Although the two-wavelength D2NNs have lower average energy percentages of correct categories than two single-wavelength D2NNs, it has the comparable energy transmission rate at the detection regions. To improve the performance of two-wavelength D2NN, we can increase the modulation element numbers at each layer, which achieves the classification accuracies of 97.5% and 88.0% for two tasks with 400 × 400 modulation element numbers per layer. The performance can be further improved by using the wavelength selective filters on the category detection regions, which achieves the classification accuracies of 95.9% and 87.0% for two tasks with 200 × 200 modulation element numbers per each layer and 97.6% and 88.9% for two tasks with 400 × 400 modulation element numbers per each layer, showing comparable and even higher accuracy than individually training two single-wavelength D2NNs to perform two tasks separately. Besides, the design strategy of multiwavelength D2NNs by optimizing the relative height maps instead of phase maps of DOEs achieves comparable performance with slightly lower classification accuracies in the same tasks. The results are summarized in Table 1, which verifies that the designed two-wavelength D2NN with a joint training approach can successfully classify the targets from two tasks in parallel without any mechanical adjusting of diffractive layers for two tasks.

To demonstrate the capability of multiwavelength D2NNs for multi-task learning with more number of tasks, we constructed a four-wavelength D2NN for four-task classification that can simultaneously classify four targets from the databases of MNIST (task I), FMNIST (task II), Kuzushiji-MNIST (KMNNIST, task III), and extended-MNIST (EMNIST, task IV), respectively. The KMNNIST comprises images of ancient Japanese scripts with the same dataset size and category numbers as the MNIST and FMNIST databases. We randomly selected 10 categories of handwritten letters from the EMNIST database and kept the same dataset size as the other three tasks, i.e., 60,000 training samples and 10,000 testing samples. The databases of four tasks, from the task I to IV, are encoded in the wavelengths of 700 nm, 600 nm, 500 nm, and 400 nm, respectively. In this numerical experiment, the four-wavelength D2NNs are designed without using wavelength selective filters that have lower hardware complexity. With other network settings the same as Figures 2 and 3, we evaluated the classification accuracies of four-wavelength D2NNs in performing four tasks in parallel under different network sizes and compared the classification accuracies with the single-wavelength D2NNs, as shown in Figure 4. For the four-wavelength D2NN with five layers and the modulation element number of 200 × 200 at each layer, the classification accuracies of four tasks, from the task I to IV, are 92.8%, 83.0%, 81.0%, and 90.4% respectively, which are significantly higher than the single-wavelength D2NN of 64.6%, 68.7%, 52.5%, and 55.3%, under the same network size. The four-wavelength D2NNs for four-task classification consistently achieved much higher accuracies than the single-wavelength D2NNs with varying network sizes. As the number of tasks increases from two to four, the proposed multi-wavelength D2NN shows more advantages in realizing optical multi-task learning.

We further evaluated and compared the performance of the proposed four-wavelength D2NNs with respect to the individual training of four single-wavelength D2NNs to perform four tasks separately (see Figure 4 under different network sizes. Figure 4a increases the network size by increasing the layer numbers from one to 8 with the same element number of 200 × 200 at each modulation.
Figure 4: The performance of four-wavelength D^2NNs for four-task classification without wavelength selective filters. The classification accuracies of four-wavelength D^2NNs in performing the four-task classification are significantly higher than the single-wavelength D^2NNs. By increasing the layer number (a) and modulation element numbers (b) at each layer, the classification accuracies of four-wavelength D^2NNs increase at each of four tasks and approach the classification accuracies by individually training four single-wavelength D^2NNs to perform four tasks separately. ST and SW, single-task using single-wavelength; MT and SW, multitask using single-wavelength; MT and MW, multitask using multiwavelength.

layer. Figure 4b increases the network size by increasing the element number at each modulation layer with the same layer number of 5. Increasing the neural network size of multiwavelength D^2NNs for optical multitask learning can significantly improve its inference capability until the performance reaches a state of saturation. The performance of four-wavelength D^2NNs continues to improve with the increase of network size and approaches to the performance of training four single-wavelength D^2NNs. The classification accuracies of task I to task IV are 96.5%, 85.6%, 88.6%, and 93.8%, respectively, with the modulation layer number of five and the element number of 800 × 800 at each layer, which shows comparable performance with respect to the training of four single-wavelength D^2NNs with the same network size. The results demonstrate the effectiveness of the proposed approach for multi-task learning with a monolithic optical system and achieve much lower hardware complexity. The encoding of multitasks into multiwavelength channels alleviates the competition among different tasks and minimizes the performance reduction of each task.

4 Discussion

There are different methods that have been widely validated for fabricating multiwavelength DOEs. We have demonstrated two design strategies for training multiwavelength D^2NN by optimizing the height maps or phase maps of DOEs for each diffractive layer. With the height maps of DOEs, a multistep photolithography-etching process can be used as a viable fabrication method that has been extensively studied [28]. With the phase maps of DOEs, the same phase modulation characteristic under different wavelength channels can be achieved [29–34]. The geometry structure of each element can be determined to make the optical path length of each modulation element have the same phase value for each wavelength. This can be achieved by adding an integral multiphase delay (e.g., $2\pi$) to one wavelength until the other wavelength reaches the appropriate phase retardation [29, 30]. The overall physical height will be based on the actual accuracy requirements. The other method exploits the refractive index change of dispersive materials for different wavelengths. One can control the optical path length of each wavelength [31]. A multiwavelength DOE can also be designed by combining several aligned DOEs, made of different materials, similar to the polarization-selective DOEs [32, 33]. Furthermore, the flexibility of wavefront manipulation in different physical dimensions, e.g., phase, amplitude, wavelength, and polarization, in the metasurfaces makes it possible to encode multiple wavelength channels. For example, a designed metasurface consisting of different types of nanoblocks with
spatially varying rotation angles multiplexed in a subwavelength unit can make it resonant with different wavelengths [34].

We analyze and compare the impact of phase modulation noise and signal detection noise, modeled with the Gaussian noise, on the classification accuracies of the multiwavelength D²NNs and single-wavelength D²NNs. We evaluate with the models of two-wavelength D²NN and two single-wavelength D²NNs to perform two tasks of classifying MNIST and fashion-MNIST databases (see Figure 3). The plots of classification accuracy decrease with respect to the SNR are shown in Figure 5a and b for the phase noise and detection noise, respectively. Numerical analysis results demonstrate that although training single-wavelength D²NN for each task has higher robustness at the low SNR settings, both multiwavelength and single-wavelength D²NNs achieve a slight classification accuracy decrease at the high SNR settings. When the SNR of the phase modulation is greater than 10 dB, or the SNR of the signal detection is greater than 25 dB, the accuracy loss of the two-wavelength D²NN and the single-wavelength D²NNs is less than 1%.

We further compare the training time of 1, 2, 3, and 4-wavelength D²NNs for 1, 2, 3, and 4-task learning, respectively, running on a single GPU, with the other network settings the same as the model in Figure 3. With the training epoch number setting of five, the time consumptions of four models are 116 s, 227 s, 365 s, and 425 s, respectively, which show approximate linear time growth with the increase of the wavelength numbers. Multiple GPUs for parallel computing can be easily utilized to reduce training time. Notice that the optical fields of different wavelengths propagate simultaneously in the multiwavelength physical system, which demands no additional time consumption during the inference of the forward model.

The multiwavelength D²NN is an all-optical computing processor that simultaneously performs multiple tasks with extremely high computational throughput. The computing speed of multiwavelength D²NN is limited by the encoding speed of input objects. Considering the experimental system architecture of a single-wavelength D²NN, the high-speed spatial light modulation (SLM) that works at ~1000 fps in the visible spectrum typically can be used to encode the input images [8]. The SLM switching time is the bottleneck of the system, compared with the all-optical optical field propagation and detection rate; thus, the single-wavelength D²NN system can work at ~1000 fps with the system latency of 1 ms in theory. For N-wavelength D²NN, the number of frames and operations processed per second can be increased by N times. Taking the four-wavelength D²NN configured with five layers and 800 × 800 modulation elements at each layer in Figure 4 as an example, similar to the calculations in [14], the total operation number of the optical forward model with single and four wavelengths are 324.4 million and 1.3 billion, respectively. As the DOEs are passive optical elements, with 10 mW of input light source power, the four-wavelength D²NN system has a computing energy efficiency of 130.0 Tera operations per second per watt (TOPS/W). Besides, compared with the spatial multiplexing of multiple D²NNs, the optical signals of different wavelength channels are independent of each other without any crosstalk during the multiwavelength diffractive optical computing. Therefore, increasing the wavelength channels in a single monolithic system increases the computing throughput and facilitates more tasks.

![Figure 5: The comparisons of classification accuracy decrease between models of multiwavelength and multiple single-wavelength for multitask learning by incorporating different amounts of noise to the phase modulation (a) and the signal detection (b). MT and MW, multitask using multiwavelength; ST and SW, single-task using single-wavelength.](image-url)
Nowadays, artificial neural networks still cannot learn in a continuous manner like mammalian brains, which is a great hindrance to the development of general artificial intelligence. It is widely accepted that catastrophic forgetting is a necessary flaw in the connectionist model. Although machine learning algorithms, such as transfer learning, focus on storing knowledge gained while solving one problem and applying it to different but related problems, its essence is a mathematical process that ignores physical properties and can only achieve the serial processing of different tasks. Multiwavelength D2NN has the inherent advantages of parallel processing of multiple tasks with light-speed processing, low power consumption, and high throughput. By encoding different tasks into different wavelength channels, multiwavelength D2NN can significantly alleviate competition among different tasks and maintain high performance for each task. For each new task, a new wavelength channel can be added easily to implement the new task. The task expansion process is shown in Figure 3A, and the cost of the whole process is extremely low. Therefore, multiwavelength D2NN can take full advantage of each task. As the task number increases, the multi-wavelength D2NN shows greater advantages in realizing more general brain-inspired intelligence architecture in the future.

5 Conclusions

In this work, we have demonstrated the capability of multiwavelength D2NNs to achieve high-parallel classification and enable high-accuracy optical multitask learning with the joint optimization training method. By encoding multitasks into multiwavelength channels to exploit the wavelength dimension of the diffractive optical field, the proposed optical multitask learning approach can realize different tasks in parallel at the speed of light. The optical multitask function is implemented within a monolithic system and does not require the mechanical movement of diffractive modulation layers, significantly reducing the system’s complexity. Analysis reveals that the proposed method can significantly alleviate the competition between multitasks and maintain the performance of each task. As the task number increases, the multiwavelength D2NNs show greater advantages in realizing optical multitask learning. The proposed approach can be extended to other photonic neural network architectures by using the wavelength-division multiplexing technology to perform optical multitask learning that simultaneously achieves the capability of high-parallel, high-accuracy, and high-generality.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: This work is supported by the National Key Research and Development Program of China (No. 2021ZD0109902), and the National Natural Science Foundation of China (No. 62275139).

Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

References

[1] J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, and W. H. P. Pernice, “All-optical spiking neuromorphic networks with self-learning capabilities,” Nature, vol. 569, no. 7755, pp. 208–214, 2019.
[2] G. Barbastathis, A. Ozcan, and G. Situ, “On the use of deep learning for computational imaging,” Optica, vol. 6, no. 8, pp. 921–943, 2019.
[3] M. Gu, X. Fang, H. Ren, and E. Goi, “Optically digitalized holography: a perspective for all-optical machine learning,” Engineering, vol. 5, no. 3, pp. 363–365, 2019.
[4] Q. Zhang, H. Yu, M. Barbiero, B. Wang, and M. Gu, “Artificial neural networks enabled by nanophotonics,” Light: Sci. Appl., vol. 8, no. 1, pp. 1–14, 2019.
[5] K. Yao, R. Unni, and Y. Zheng, “Intelligent nanophotonics: merging photonics and artificial intelligence at the nanoscale,” Nanophotonics, vol. 8, no. 3, pp. 339–366, 2019.
[6] D. R. Solli and J. Bahram, “Analog optical computing,” Nat. Photonics, vol. 9, no. 11, pp. 704–706, 2015.
[7] Y. Shen, N. C. Harris, S. Skirlo, et al., “Deep learning with coherent nanophotonic circuits,” Nat. Photonics, vol. 11, no. 7 pp. 441–446, 2017.
[8] T. Zhou, X. Lin, J. Wu, et al., “Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit,” Nat. Photonics, vol. 15, no. 5, pp. 367–373, 2021.
[9] Y. Hao, S. Xiang, G. Han, et al., “Recent progress of integrated circuits and optoelectronic chips,” Sci. China Inf. Sci., vol. 64, no. 10, pp. 1–33, 2021.
[10] M. Nikdast, S. Pasricha, G. Nicolescu, A. Seyedi, and D. Liang, Eds. Silicon Photonics for High-Performance Computing and Beyond, Boca Raton, USA, CRC Press, 2021.
[11] X. Lin, Y. Rivenson, N. T. Yardimci, et al., “All-optical machine learning using diffractive deep neural networks,” Science, vol. 361, no. 6406, pp. 1004–1008, 2018.
[12] T. Yan, J. Wu, T. Zhou, et al., “Fourier-space diffractive deep neural network,” Phys. Rev. Lett., vol. 123, p. 2, 2019, Art. no. 023901.
[13] J. Li, D. Mengu, Y. Luo, Y. Rivenson, and A. Ozcan, “Class-specific differential detection in diffractive optical neural networks improves inference accuracy,” Adv. Photon., vol. 1, 2019, Art. no. 046001.
[14] T. Zhou, L. Fang, T. Yan, et al., “In situ optical backpropagation training of diffractive optical neural networks,” Photon. Res., vol. 8, no. 6, pp. 940–953, 2020.
[15] Y. Zuo, B. Li, Y. Zhao, et al., “All-optical neural network with nonlinear activation functions,” Optica, vol. 6, no. 9, pp. 1132–1137, 2019.

[16] X. Lin, “Artificial intelligence built on wireless signals,” Nat. Electron., vol. 5, no. 2, pp. 69–70, 2022.

[17] A. W. Rix, A. Bourret, and M. P. Hollier, “Models of human perception,” BT Technol. J., vol. 17, no. 1, pp. 24–34, 1999.

[18] R. M. French, “Catastrophic forgetting in connectionist networks,” Trends Cognit. Sci., vol. 3, no. 4, pp. 128–135, 1999.

[19] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, et al., “Overcoming catastrophic forgetting in neural networks,” Proc. Natl. Acad. Sci., vol. 114, no. 13, pp. 3521–3526, 2017.

[20] Y. Li, W. Gao, and C. Yu, “RubikONNs: multi-task learning with rubik’s diffractive optical neural networks,” in Preprint from Research Square, London, UK, Springer Nature, 2022.

[21] Yi Luo, D. Mengu, N. T. Yardimci, et al., “Design of task-specific optical systems using broadband diffractive neural networks,” Light: Sci. Appl., vol. 8, no. 1, pp. 11–14, 2019.

[22] J. Li, D. Mengu, N. T. Yardimci, et al., “Spectrally encoded single-pixel machine vision using diffractive networks,” Sci. Adv., vol. 7, p. 13, 2021, Art. no. eabd7690.

[23] J. Lin, X. C. Yuan, S. H. Tao, and R. E. Burge, “Collinear superposition of multiple helical beams generated by a single azimuthally modulated phase-only element,” Opt. Lett., vol. 30, no. 24, pp. 3266–3268, 2005.

[24] J. W. Goodman, Introduction to Fourier Optics. Goodman. New York, USA, McGraw-Hill, 1968.

[25] J. Perina, “Superposition of coherent and incoherent fields,” Phys. Lett. A, vol. 24, no. 6, pp. 333–334, 1967.

[26] C. Kai, Z. Feng, M. I. Dedo, et al., “The performances of different OAM encoding systems,” Opt. Commun., vol. 430, pp. 151–157, 2019.

[27] S. Noach, A. Lewis, Y. Arieli, and N. Eisenberg, “Integrated diffractive and refractive elements for spectrum shaping,” Appl. Opt., vol. 35, no. 19, pp. 3635–3639, 1996.

[28] J. E. Ford, F. Xu, and Y. Fainman, “Wavelength-selective planar holograms,” Opt. Lett., vol. 21, no. 1, pp. 80–82, 1996.

[29] Y. Arieli, S. Noach, S. Ozeri, and N. Eisenberg, “Design of diffractive optical elements for multiple wavelengths,” Appl. Opt., vol. 37, no. 26, pp. 6174–6177, 1998.

[30] F. Xu, J. E. Ford, and Y. Fainman, “Polarization-selective computer-generated holograms: design, fabrication, and applications,” Appl. Opt., vol. 34, no. 2, pp. 256–266, 1995.

[31] N. Nieuborg, A. Kirk, B. Morlion, H. Thienpont, and I. Veretennicoff, “Polarization-selective diffractive optical elements with an index-matching gap material,” Appl. Opt., vol. 36, no. 20, pp. 4681–4685, 1997.

[32] Z. Li, M. Premaratne, and W. Zhu, “Advanced encryption method realized by secret shared phase encoding scheme using a multi-wavelength metasurface,” Nanophotonics, vol. 91, no. 11, pp. 3687–3696, 2020.