Multi-wavelength diffractive photonic neural network for multi-task learning

Zhengyang Duan
Department of Electronic Engineering
Tsinghua University
Beijing, China
Co-first Author

Hang Chen
Department of Electronic Engineering
Tsinghua University
Beijing, China
Co-first Author

Xing Lin
Department of Electronic Engineering
Tsinghua University
Beijing, China
Corresponding author: lin-x@tsinghua.edu.cn

***Abstract*—Photonic neural networks use photons instead of electrons to perform artificial intelligence (AI) tasks, with the advantage of high-speed, low-power information processing. However, existing architectures are designed for a single task, which cannot reuse multiple tasks in parallel in a single system because the competition among different tasks will degrade the model performance. In this paper, a novel optical multi-task learning system is proposed by designing a multi-wavelength diffractive photonic neural network (DPNN) using a joint optimization method. By encoding the input of multiple tasks into multi-wavelength channels, the system can significantly reduce the competition to execute multi-tasks in parallel with high precision. We design the two-task and four-task DPNNs with two and four spectral channels respectively, for classifying different inputs from the EMNIST, KMNIST, FMNIST, and MNIST databases. Numerical evaluations show that for multi-task learning, multi-wavelength DPNNs achieve significantly higher classification accuracies than single-wavelength DPNNs under the same network size. Moreover, as the network size increases, the classification accuracy of multi-wavelength DPNNs is comparable to that of individually training multiple single-wavelength DPNNs to perform multiple tasks separately. Our work provides a proven technical solution for developing high-throughput neuromorphic photonic computing and more general artificial intelligence systems to perform multiple tasks in parallel.**

Keywords—multi-task learning, multi-wavelength photonic neural networks, diffractive photonic neural networks

I. INTRODUCTION

Photonic neural networks (PNNs) [1, 2] implement photonic computing-based artificial neural network models which can achieve a leapfrog improvement in computing speed and energy efficiency and are considered as one of the potential solutions to break Von Neumann bottleneck of electronic-based artificial intelligence (AI) systems in the post-Moore era. Among different PNNs, diffractive photonic neural networks (DPNNs) [3] control diffraction light from the input plane and perform advanced missions based on the desired output light intensity distribution, which can achieve large-scale neural information processing and attracted a vast amount of interest. However, in previous studies on DPNN architectures, monochromatic plane waves were used to encode input data and propagate through modulation layers to perform specific tasks. Executing multiple AI tasks in parallel using a monolithic DPNN system remains challenging. One of the major obstacles is the competition among different tasks during the training, which leads to catastrophic forgetting [4]. Catastrophic forgetting occurs when systems are trained on multi-tasks, which leads to a tendency for knowledge of previously learned tasks to be abruptly lost while learning new tasks, resulting in performance degradation on every single task. Previous solutions [5] required mechanical movement of optical elements to switch between tasks one at a time or require designing multiple different DPNN systems, one for each task, significantly increasing the complexity of the hardware.

Here, we propose an optical multi-task learning monolithic system design by developing multi-wavelength DPNNs that can simultaneously perform multiple classification tasks on different databases without mechanical motion. In this design, the multi-wavelength DPNN has N ($N \geqslant 2$) different parallel wavelength channels, encoding N different inputs in parallel. The detection area of each category is segmented into N parts, where each part represents the input category encoded at the corresponding wavelength channel. We train the DPNN using the multi-wavelength joint optimization method with the loss functions of softmax cross-entropy (SCE) and energy efficiency constraint. We utilize two-wavelength and four-wavelength DPNNs to perform two-task and four-task classifications, respectively, based on the MNIST, FMNIST, KMNIST, and EMNIST databases. Numerical evaluations demonstrate the great advantages of multi-wavelength DPNNs in realizing multi-task learning.

II. METHODS

The proposed optical multi-task learning system can achieve multiple tasks in parallel based on the multi-wavelength DPNN. As shown in Fig. 1, the input targets of each task i will be encoded into each wavelength channel ($\lambda_i, i = 1, \ldots, N$). According to the principle of light intensity superposition, multi-wavelength light field transformation can be regarded as the combination of independent transformation of coherent optical fields at each wavelength. We consider the linear DPNN with the complex transform function $\boldsymbol{M}_{\lambda_i}(\boldsymbol{\Phi})$, where $\boldsymbol{\Phi}$ represents the phase modulation coefficients of the diffractive elements in multiple phase-only diffractive layers. We assume that the phase modulation coefficients are the same at different wavelengths according to the multi-wavelength diffractive optical element (DOE) design strategies [6]. The output optical fields at the i-th wavelength λ_i can be formulated as $\boldsymbol{U}'_{\lambda_i} = \boldsymbol{M}_{\lambda_i}(\boldsymbol{\Phi})\boldsymbol{U}_{\lambda_i}$, while $\boldsymbol{U}_{\lambda_i}$ is the input optical fields at the wavelength of λ_i that encoding the input targets of task i. The detector measures the intensity distribution of the output optical fields, which can be expressed as $\boldsymbol{I}_{\lambda_i} = \left|\boldsymbol{U}'_{\lambda_i}\right|^2 = \left|\boldsymbol{M}_{\lambda_i}(\boldsymbol{\Phi})\boldsymbol{U}_{\lambda_i}\right|^2$. The total

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).

intensity distribution at different wavelengths in a multi-wavelength DPNN can be formulated as the superposition of the intensity distributions detected at each wavelength: $\boldsymbol{I} = \sum_{\lambda_i} \boldsymbol{I}_{\lambda_i} = \sum_{\lambda_i} \left|\boldsymbol{M}_{\lambda_i}(\boldsymbol{\Phi})\boldsymbol{U}_{\lambda_i}\right|^2$. Another training strategy is to assume the relative height map of each diffractive layer is the same under different wavelength channels. In this situation, DPNN's complex transform function will be changed from $\boldsymbol{M}_{\lambda_i}(\boldsymbol{\Phi})$ to $\boldsymbol{M}(\Delta\mathbf{Z})$ because of the diffractive elements' phase values $\boldsymbol{\Phi}$ under certain wavelength channel λ_i can be converted into a relative height map $\Delta\mathbf{Z}$ according to $\boldsymbol{\Phi}_{\lambda_i} = 2\pi\Delta n_{\lambda_i}\Delta\mathbf{Z}/\lambda_i$, where Δn_{λ_i} is the refractive index difference between the air and base material. The multi-step photolithography-etching process has been extensively studied as a practical fabrication method [7]. Under the same training parameters, two design strategies (optimizing the phase maps or height maps) achieve comparable performance, whereas optimizing the phase maps has a slightly higher performance [8].

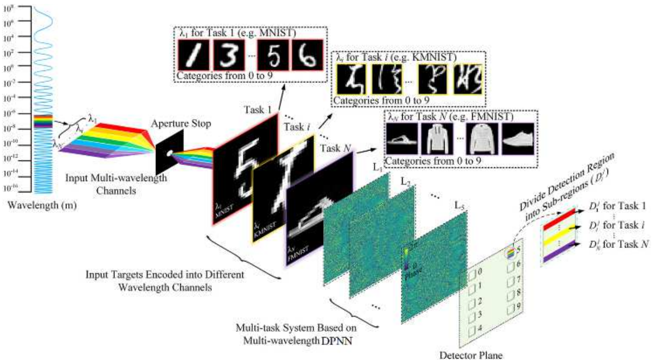

Figure 1 The multi-wavelength DPNNs framework enables multi-task learning.

To train multi-wavelength DPNNs for optical multi-task learning, each category detection area at the output plane will be divided into multiple sub-regions $\{D_i^j\}$, where $i = 1, \ldots, N$ denotes the index of the task encoded at i-th wavelength λ_i; and $j = 1, \ldots, M$ denotes detection areas' index, representing the index of categories. We calculate the average intensity of the i-th sub-region among the M category detection areas, i.e., $\boldsymbol{P}_i = \{avg(\boldsymbol{I}(D_i^j)), j = 1, \ldots, M\}$, where $\boldsymbol{I}(D_i^j) = \sum_{\lambda_i} \boldsymbol{I}_{\lambda_i}(D_i^j)$ is for the broadband wavelength detection without using spectral filters on the detector. If we use spectral filters on the detector, each sub-region for each task only detects the optical signals at the corresponding wavelength channel, with which $\boldsymbol{I}(D_i^j) = \boldsymbol{I}_{\lambda_i}(D_i^j)$. Using wavelength-selective filters can eliminate the wavelength crosstalk during detection and slightly improve the classification accuracies, but it increases hardware complexity [8]. To maximize the energy transmission efficiency by minimizing the optical energy outside the category detection areas, the joint optimization problem for multi-wavelength DPNNs training can be formulated as:

$$\min_{\boldsymbol{\Phi}}\left(\sum_i L(\boldsymbol{P}_i, \boldsymbol{G}_i) + \sum_{\lambda_i} MSE\left(\boldsymbol{I}_{\lambda_i} - \sum_j \boldsymbol{I}_{\lambda_i}(D_i^j)\right)\right), \quad (1)$$

where $L(\boldsymbol{P}_i, \boldsymbol{G}_i)$ represents the SCE loss function of the i-th task at the wavelength of λ_i between the detection $\boldsymbol{P}_i$ and ground truth label $\boldsymbol{G}_i$; $\boldsymbol{G}_i$ is a one-hot vector with a length of M; $MSE(\boldsymbol{I}_{\lambda_i} - \sum_j \boldsymbol{I}_{\lambda_i}(D_i^j))$ represents the total energy of optical intensity outside the sub-regions of category detection areas evaluated with the mean square error. We use the stochastic gradient descent approach and error back-propagation method to optimize the network structure.

III. Results

To demonstrate the capability, we first construct a two-task classifier for classifying both the MNIST database (task I) and the fashion-MNIST (FMNIST) database (task II). The handwritten digits of task I and the fashion products of Task II are encoded in the wavelength of 700 nm and 400 nm, respectively. The two-wavelength DPNN has five diffractive layers with a total phase modulation element number of 5 × 200 × 200, without the wavelength selective filters on the detector that have lower hardware complexity. The blind testing of the trained model achieves classification accuracies of 95.6% and 86.8% on the test datasets of MNIST and FMNIST, respectively. The classification accuracies of two-wavelength DPNNs are higher than that of the single-wavelength DPNNs for performing two tasks in parallel, which are 92.4% and 83.1%, respectively. Fig. 2(a) shows an exemplar result for simultaneously classifying a handwritten digit "7" with the category number 7 from the MNIST database and a fashion product "pullover" with the category number 2 from the FMNIST database. The maximum average intensity outputs were focused on the upper sub-regions (task I) of the No.7 detector area and the lower sub-regions (task II) of the No.2 detector area, respectively, as shown in Fig. 2(a) which were marked by the white arrow. Some other exemplar results, such as a handwritten digit "6" with the category number 6 and a fashion product "trouser" with the category number 1 were shown in Fig. 2(b). The energy distributions of the classification results in Fig. 2(b) show that the multi-wavelength DPNN can prominently identify the sub-region with maximum average intensity for different tasks' correct classification.

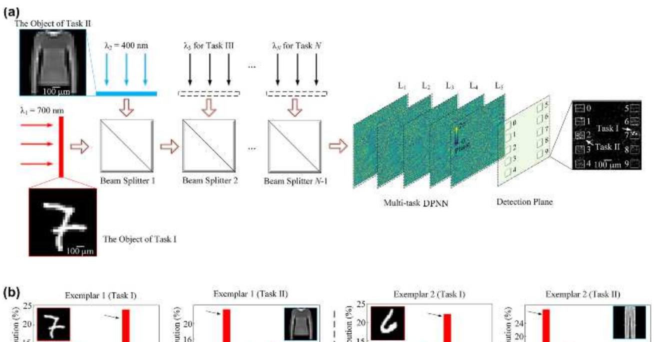

Figure 2: Multi-wavelength DPNNs operating at two wavelengths for classification on the MNIST and FMNIST databases. (a) The categories of two inputs from the MNIST and FMNIST datasets are determined by finding the corresponding sub-regions with maximum average intensity, as indicated by the white arrows. (b) The energy distribution of the classification results for the two tasks demonstrates the success of the proposed approach.

To further demonstrate the capability of multi-wavelength DPNNs for multi-task learning with more tasks, we constructed a four-wavelength DPNN for simultaneous four-task classification based on the datasets of MNIST (task I), FMNIST (task II), Kuzushiji-MNIST (KMNIST, task III), and Extended-MNIST (EMNIST, task IV), respectively. The databases for tasks I to IV are encoded at wavelengths of 700 nm, 600 nm, 500 nm, and 400 nm, respectively. With other network settings the same as Figs. 2, we evaluated the classification accuracies of four-wavelength DPNNs in performing four tasks in parallel under different network sizes and compared the classification accuracies with the single-wavelength DPNNs, as shown in Fig. 3. For the four-wavelength DPNN with the total modulation element number of 5×200×200, the classification accuracies are 92.8%, 83.0%, 81.0%, and 90.4% from the task I to IV, respectively. The classification accuracies are significantly higher than the single-wavelength DPNN of 64.6%, 68.7%, 52.5%, and 55.3%, under the same network size. The four-wavelength DPNNs for four-task classification consistently achieved much higher accuracies than the single-wavelength DPNNs. As the task number increases from two to four, the multi-wavelength DPNNs show more advantages in achieving multi-task learning.

We further evaluated and compared the performance of the proposed four-wavelength DPNNs against four single-wavelength DPNNs individually trained on four tasks at different network sizes. Fig. 3(a) increases the network size by increasing the number of layers from 1 to 8 with the same element number of 200×200 at each modulation layer. Fig. 3(b) increases the network size by increasing the number of elements in each modulation layer with the same layer number of 5. Increasing the neural network size can significantly improve its inference ability until the performance saturates. The performance of four-wavelength DPNNs keeps improving with increasing network size and approaches to the performance of training four single-wavelength DPNNs. The classification accuracies are 96.5%, 85.6%, 88.6%, and 93.8% from tasks I to IV, respectively, with the modulation layer number of five and the element number of 800×800 at each layer, which shows comparable performance concerning the training of four single-wavelength DPNNs 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. Encoding multi-tasks into multi-wavelength channels alleviates the competition between different tasks and minimizes the degradation of each task's performance.

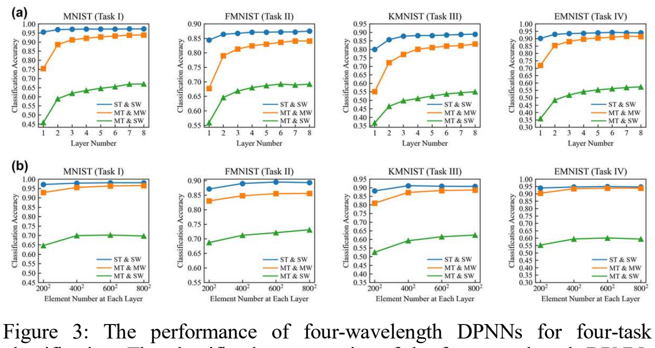

Figure 3: The performance of four-wavelength DPNNs for four-task classification. The classification accuracies of the four-wavelength DPNNs are significantly higher than that of the single-wavelength DPNNs when performing four-task classification. By increasing the layer number (a) and modulation element numbers (b) at each layer, the classification accuracies of four-wavelength DPNNs increase in each of the four tasks and approach the classification accuracies by individually training four single-wavelength DPNNs to perform four tasks separately. ST and SW, single-task using single-wavelength; MT and SW, multi-task using single-wavelength; MT and MW, multi-task using multi-wavelength.

IV. Conclusion

In this work, we demonstrate the ability of multi-wavelength DPNNs to achieve highly parallel classification and high-accuracy optical multi-task learning through a jointly optimized training method. By encoding multi-tasks into multi-wavelength channels to exploit the wavelength dimension of the diffractive light field, the proposed optical multi-task learning approach can realize different tasks in parallel at the speed of light. The optical multi-task function is implemented within a monolithic system without the need for mechanical movement of the diffractive modulation layer, significantly reducing system complexity. The Analysis shows that the proposed method can significantly alleviate the competition between multiple tasks and maintain the performance of each task. As the task number increases, the multi-wavelength DPNNs show greater advantages in realizing optical multi-task learning. The proposed approach can be extended to other photonic neural network architectures by using the wavelength-division multiplexing technology to perform optical multi-task learning that simultaneously achieves the capability of high parallelism, high accuracy, and high generality.

References

[1] D. R. Solli and J. Bahram, “Analog optical computing,” Nat. Photonics, vol. 9, no. 11, pp. 704–706, October 2015.

[2] 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, May 2019.

[3] X. Lin, Y. Rivenson, N. T. Yardimci, Y. Luo, M. Jarrahi and A. Ozcan, “All-optical machine learning using diffractive deep neural networks,” Science, vol. 361, no. 6406, pp. 1004–1008, July 2018.

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

[5] Y. Li, R. Chen, B. Sensale-Rodriguez, W. Gao, and C. Yu, “Real-time multi-task diffractive deep neural networks via hardware-software co-design,” Sci. Rep., vol. 11, no. 1, pp. 1–9, May 2021.

[6] 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, February 1998.

[7] H. Chen, J. Feng, M. Jiang, Y. Wang, J. Lin, J. Tan and P. Jin, “Diffractive deep neural networks at visible wavelengths,” Engineering, vol. 7, no. 10, pp. 1483–1491, October 2021.

[8] Z. Duan, H. Chen, and X. Lin, “Optical multi-task learning using multi-wavelength diffractive deep neural networks,” Nanophotonics, vol. 1358, pp. 1–11, January 2023.