

---

# Numerical simulations on optoelectronic deep neural network hardware based on self-referential holography

Rio Tomioka<sup>1</sup> · Masanori Takabayashi<sup>2,3</sup>

**Abstract** We propose a novel optoelectronic deep neural network (OE-DNN) hardware called the self-referential holographic deep neural network (SR-HDNN). The SR-HDNN features a combination of an optical computing part utilizing a volume hologram and an electronic part connecting the optical elements virtually. Since the shape of a volume hologram, which is a 3-dimensional (3D) refractive index distribution in this case, can be changed by its recording conditions, it is expected to realize the flexible design of optical computing functions by coupling between specific nodes. In addition, the electronic part enables the construction of multi-layer networks without extending the optical system and enabling arbitrary signal processing, including nonlinear operations. By integrating flexible optical and electronic parts, the SR-HDNN consisting of both flexible optical and electronic parts has the potential to maximize the performance of OE-DNN. In this study, we numerically simulate image classification tasks to investigate the feasibility and potential of the SR-HDNN.

**Keywords** Optical neural network · Self-referential holography · Holographic data storage · Optical information processing

---

1. Graduate School of Computer Science and Systems Engineering,  
Kyushu Institute of Technology 680-4 Kawazu, Iizuka, Fukuoka, 820-8502, Japan  
E-mail: tomioka.rio288@mail.kyutech.jp ·

2. Faculty of Computer Science and Systems Engineering,  
Kyushu Institute of Technology 680-4 Kawazu, Iizuka, Fukuoka, 820-8502, Japan  
E-mail: takabayashi@phys.kyutech.ac.jp

3. Research Center for Neuromorphic AI Hardware  
Kyushu Institute of Technology 2-4 Hibikino, Wakamatsu, Kitakyushu, 808-0196, Japan

## 1 Introduction

An artificial neural network (ANN), one of the machine learning methods inspired by the biological neural networks of the brain, is widely known as a powerful tool for solving complicated tasks with high accuracy. An ANN with deep layers is referred to as a deep neural network (DNN), which has been applied in various research and industrial fields to solve advanced tasks such as natural language processing and self-driving cars. However, to achieve sustainable development of artificial intelligence (AI) technologies such as ANN, reducing the computational cost for training and inference is essential, which consumes significant energy. In particular, reducing the inference cost is important because it is continuously incurred in the operation of the ANN system. The structure of the ANN is a network with many tunable weights, and most of the computational cost in the inference process comes from large-scale parallel operations to compute signal propagation through weighted edges. Therefore, the effective implementation method of the parallel operation of the inference process of the ANN is expected to be established.

An optical neural network (ONN) is one of the candidates to achieve parallel computation with lower power consumption compared to general-purpose computers. The spatial light modulation of 2-dimensional (2D) patterns has received much attention among the various types of ONN technology [1–11]. This method enables an optical implementation of a large-scale ANN by regarding each pixel of the modulated 2D pattern as an ANN node and inducing interactions between them through diffraction and/or interference. A diffractive deep neural network (D<sup>2</sup>NN) [1] is the representative method of the spatial-modulation-based ONN. A typical D<sup>2</sup>NN consists of phase plates designed by in-silico training and fabricated by a 3D printer. When the input light enters these plates, the intensity distribution of the light is changed via diffraction. The main purpose of the D<sup>2</sup>NN is to achieve the designed input-output conversion by coupling between nodes through free-space propagation of the optical wave weighted by modulation at each point of the phase plates. For example, for an image classification task, the light modulated by the image to be classified is illuminated to the phase plates and focused onto a target area. Here, the phase plates are designed through in-silico training to ensure that the light modulated by an image is focused onto the area corresponding to the correct label. Although the D<sup>2</sup>NN has great potential for being energy-effective AI hardware, there is concern that the performance improvement based on the deep-learning principle is limited by the lack of nonlinearity and the increasing size associated with multilayer.

One solution to this challenge is fusion with electronic computing, which is referred to as an optoelectronic deep neural network (OE-DNN). OE-DNN aims to complement the advantages of energy-efficient parallel processing with optical computing and high expressiveness with electronic computing. Several OE-DNN methods have been proposed thus far. One of the convolution layers in a convolutional neural network (CNN) is optically implemented based on the optical Fourier transformation using lenses in the hybrid optical-electronic

convolutional neural network, whereas the other layers are electronically implemented [3]. In the optical convolution layer, the phase plate placed on the focal plane of the lenses is designed and used such that multiple images resulting from convolutional calculations with multiple kernels can be obtained simultaneously. It is shown that this approach significantly reduces the computational cost, even when the number of optically implemented convolutional layers is only one. As another method, an optronic convolutional neural network (OPCNN) in which both convolution and affine layers are optically implemented and electronically connected to realize a CNN has been proposed [10]. The OPCNN can realize several optical computing functions by individual optical systems and construct the DNN by combining these optical systems electronically. In other words, The OPCNN connects various optical computing parts with simple electronic connections to implement various types of ANNs. Furthermore, the method using a simple optical part referred to as the diffractive processing unit (DPU) is widely known [11]. Various networks, such as a CNN and a recurrent neural network (RNN) can be realized by electronic connections of the simple optical units DPU. It is the opposite of OPCNN, a simple optical computing part named DPU is connected by various ingenious electronic connections to implement various types of ANNs. Consequently, there is concern about the system enlargement in the hybrid optical-electronic convolutional neural network and the OPCNN and increasing the burden of electronic computing in DPU-based methods.

In this paper, we propose a self-referential holographic deep neural network (SR-HDNN) which consists of an optical computing part using holograms and an electronic part connecting the optical parts virtually. Since both computing parts of SR-HDNN are flexible to design, it can be understood SR-HDNN is OE-DNN with properties intermediate between the two types described above. In other words, it is expected to maximize the potential of OE-DNN by providing flexibility in both the optical and electronic computing parts. Furthermore, the base technology of SR-HDNN, self-referential holography (SRH), can be applied to holographic data storage (HDS), which is referred to as self-referential holographic data storage (SR-HDS) [12], and SR-HDNN is highly compatible with SR-HDS. Therefore, there is enough possibility to realize SR-HDS systems, including AI-based signal detection using SR-HDNN, for example. The purpose of this work is to investigate the feasibility of the SR-HDNN as OE-DNN, which is a first step toward the concept. Specifically, we numerically demonstrate an image classification task using SR-HDNN to investigate its trainability and generalization ability. Finally, we discuss the feasibility and potential of SR-HDNN.

## 2 SR-HDNN

### 2.1 SRH

The SRH is a technology to control the spatial intensity distribution of a light wave by a volume hologram recorded without using reference light. The SRH has two processes: the writing and reading process. During the writing process, a writing pattern (WP) is phase modulated into a light wave by a spatial light modulator (SLM). Then, the light wave is referred to as the writing light. When a lens focuses the writing light, self-interference between pixels of WP occurs near the focal plane. The self-interference pattern is recorded as a hologram by placing a recording medium near the focal plane, such as a photopolymer. During the reading process, a reading light to which the reading pattern (RP) is phase-modulated illuminates the hologram using the same optical system as the writing process. When the reading light illuminates the hologram, the energy coupling between the pixels of RP occurs. Specifically, when focusing on a specific pair of pixels on the SLM, energy transfer is induced if the phase difference of these two pixels is different in the writing and reading processes. This energy interaction occurs among all pixels, even when there are more than three pixels involved. As a result, the intensity distribution of the reading light becomes non-uniform and is decided by the relationship between WP and RP. In particular, when the difference pattern WP and RP satisfy special conditions [12], the intensity pattern obtained in the reading process corresponds to the difference between WP and RP. Then, the principle of SRH can be applied to data storage, i.e., SR-HDS. The typical patterns used for SR-HDS are shown in Fig.1. WP and RP are determined such that the recorded signal pattern (SP) corresponds to the difference between WP and RP using an additional pattern (AP) is an arbitrary pattern.

### 2.2 SR-HDNN

The SRH reflects the spatial intensity changes of the reading light, resulting from the energy coupling between the pixels of the RP derived from the phase difference between the writing and reading light. Based on the intensity change principle of SRH, it is expected that the desired intensity distribution can be achieved by changing the phase distribution of RP. Therefore, similar to the principle of ANNs, there is a possibility of obtaining a versatile phase distribution for a specific task by optimizing the phase distribution for multiple inputs, such that the difference between the output and target distribution is minimized. In other words, SR-HDNN is expected to be a feasible OE-DNN based on SRH.

Figure 2 shows a conceptual diagram of SR-HDNN for image classification. The RPs in the SR-HDNN are the summation of an input of layer and a controlling pattern (CP). In the first layer, the light to which the image to be recognized and CP<sub>1</sub> is to be modulated is illuminated in the hologram.

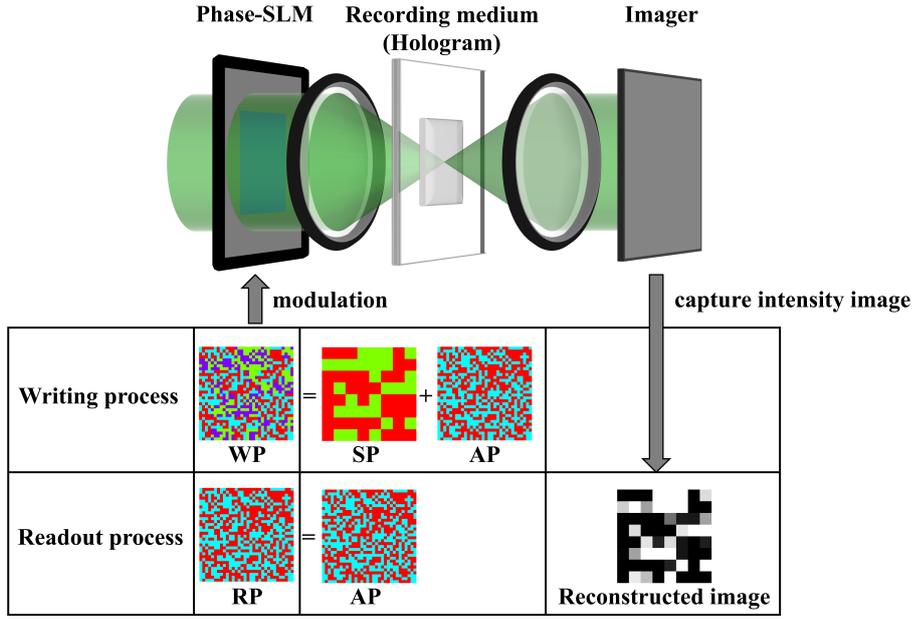


Fig. 1 Conceptual diagram of SR-HDS

Subsequently, pattern  $S_1$  is acquired from the intensity distribution of the reading light on the camera plane and is normalized and converted to  $P_1$ . In the second layer, the light in which  $P_1$  and  $CP_2$  are phase-modulated illuminates the hologram, and  $S_2$  is captured. Subsequently, these steps are repeated. Consequently, the desired intensity distribution, in which the intensity of the area indicates the image label, becomes the largest. To recognize arbitrary images, the  $CP_1, CP_2, \dots, CP_n$  must be designed via the training process, which is explained as follows.

In the training process of the SR-HDNN, CPs were designed using a dataset, along with the procedure shown in Fig. 3. First, we prepared a dataset consisting of pairs of the input and desired output images and defined a loss function to quantitatively evaluate the difference between the system output and desired output image. For example, the loss function is defined as the intensity ratio of the desired area, indicating the correct label for other areas. Subsequently, an optimization algorithm is applied to determine each pixel value of the CPs to minimize loss. Among the optimization methods, the gradient method is well known. In this method, the values of the parameters are updated to a gradient direction calculated by the differential. According to the general conversion rules extracted from the dataset, this optimization process enables the optical system to infer various data, whether included in the dataset or not. Consequently, the general CPs that can be used for the classification of arbitrary images can be obtained. In this way, the features of SR-HDNN are to design CPs such that the output intensity distribution becomes desired one, to use a hologram instead of the free space used in  $D^2NN$

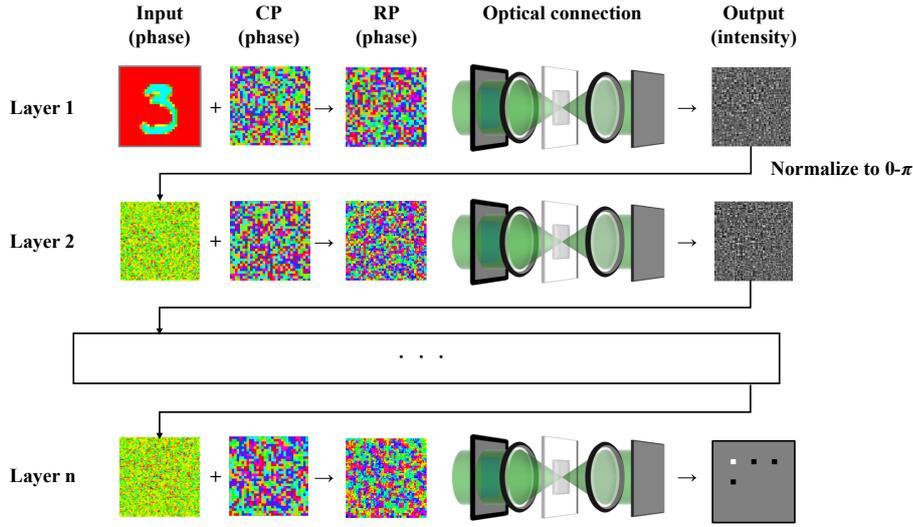


Fig. 2 Procedure of SR-HDNN

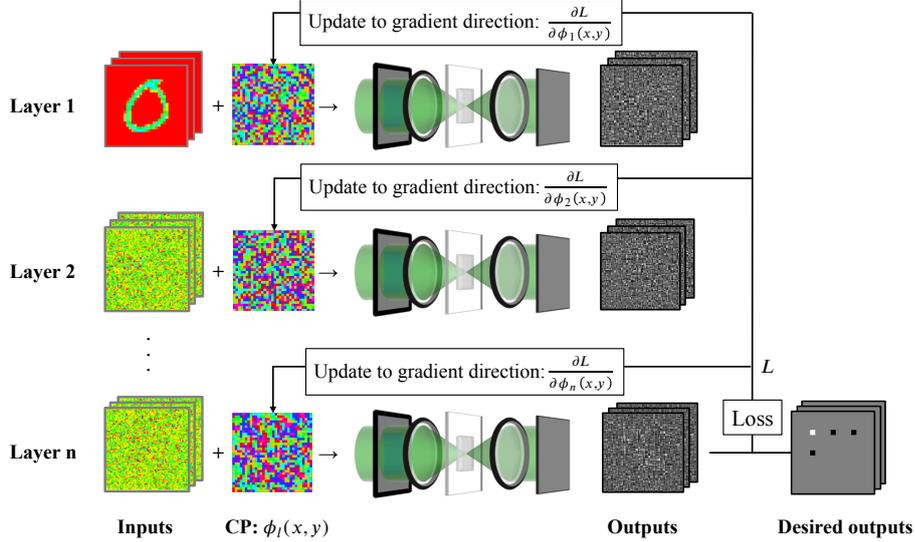


Fig. 3 Optimization of CPs in the training process

and DPU to change the wavefront shape, and to realize nonlinear functions and multilayer using electronic computing and feedback.

### 3 Numerical simulations

We simulated a 4-class image classification using SR-HDNN to demonstrate its feasibility. In Section 3.1, we describe the numerical simulation model. Section

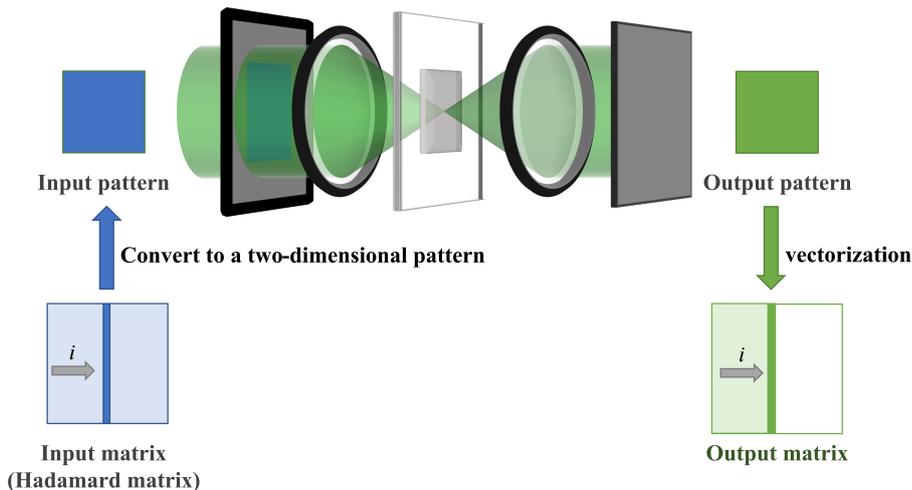
3.2, demonstrates the feasibility of an optical computing part based on SRH by simulating a single-layer network. In Section 3.3, we demonstrated the effectiveness of the electronic multi-layering through a simulation of a 3-layer SR-HDNN model.

### 3.1 Numerical simulation method and conditions

We built a simulation environment for the SR-HDNN based on the fast Fourier transform beam propagation method (FFT-BPM). Huge computational costs and time are required because the calculations of the reading process of the SRH must be repeated in the training process of the SR-HDNN. However, because the recorded hologram is fixed in the SR-HDNN, we can instantly obtain the output complex amplitude distribution from the arbitrary input distribution using a transmission matrix and drastically reduce the computation cost and time. Here, the transmission matrix is the matrix that links the complex amplitude distributions of the two planes in a linear optical system [13, 14]. When the complex amplitude distribution is in vector form on the input and output planes,  $v_{in}$  and  $v_{out}$  satisfy the condition  $v_{out} = M_{tm}v_{in}$ , where  $M_{tm}$  is the transmission matrix of the system.

In this simulation, we derived the transmission matrix for a hologram recording of the randomly generated WP using FFT-BPM. Here, WP plays a crucial role in the distribution of holograms, which in terms determines how RP pixels are coupled in the optical computing part. While one of the attractive features of SR-HDNN is the flexible design of the optical computing part through the hologram recording conditions such as the distribution of WP, we used a random WP for proof-of-principle of SR-HDNN which was generated using a Mersenne twister and has the pixel number of (32,32) and the gradation number of 8. The random WP is expected to reproduce the coupling between all pixels of RP, i.e., the affine layer. The laser wavelength, intensity, and recording time were 532 nm, 1.0 mW, and 1.0 s, respectively. The thickness, maximum refractive index, and sensitivity of the recording medium were 400.0  $\mu\text{m}$ ,  $4.0 \times 10^{-3}$ , and 40.0  $\text{cm}^2/\text{J}$ , respectively. The pixel size and pixel pitch of the imager and SLM were (128,128) and (22.5  $\mu\text{m}$ , 22.5  $\mu\text{m}$ ), respectively.

Figure 4 shows the measurement process for deriving the transmission matrix using the input matrix  $M_{in}$  and output matrix  $M_{out}$ . The input matrix  $M_{in}$  is a Hadamard matrix, where each column is used as an input vector  $v_{in}$ . Each column  $v_{out}$  of the output matrix represents the response of the optical system to the input vector  $v_{in}$ , corresponding to the same column of the input matrix. Because the Hadamard matrix is an orthogonal matrix with elements 1 and -1, its transpose matrix can be used as its inverse and can be represented in the optical system of the SR-HDNN as a phase modulation pattern of 0 and  $\pi$ . The input matrix  $M_{in}$  and output matrix  $M_{out}$  are related as:  $M_{out} = M_{tm}M_{in}$ . By taking the transpose of  $M_{in}$  on both sides, the transmission matrix can be derived as  $M_{tm} = M_{out}M_{in}^T$ .

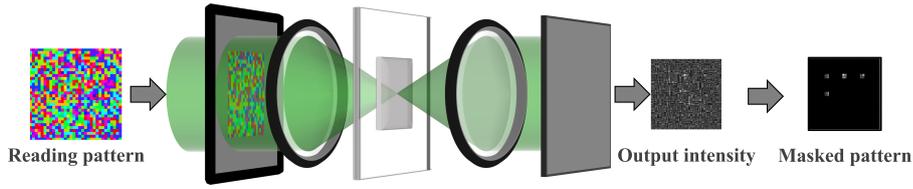


**Fig. 4** Procedure of the measurement process for the derivation of the transmission matrix

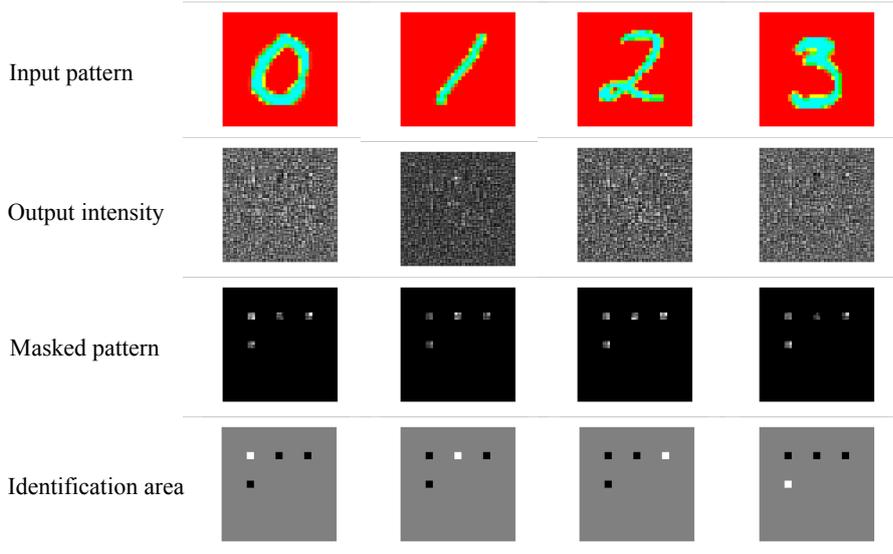
In the image classification simulation, we used input images and CPs of size  $(32,32)$ . The input images were processed using normalization and zero-padding. The Modified National Institute of Standards and Technology database (MNIST) dataset [15] consists of grayscale images of handwritten digits with a size of  $(28,28)$ . These images were normalized to  $0-\pi$  and zero-padded to expand the size to  $(32,32)$ . For the camera plane of the output layer, we prepared four identification areas that corresponded to each class. Subsequently, we chose the class corresponding to the area with the highest intensity signal as the inference result. The CPs were trained using a search-based optimization algorithm that minimized the loss by selecting each pixel value with a lower loss than the current loss from a predefined set of discrete values and raster scanning each pixel. Although the optimization algorithm requires many iterations, it can be implemented with a small number of computational resources to control the SLM and camera in future experiments. The CPs were uniformly discretized into eight levels in the range  $0$  to  $2\pi$ . The loss function used the light-focusing ratio for the correct area among the four identification areas. In particular, the ratio  $I_c/I_i$  is used, where  $I_c$  is the average intensity of the correct area and  $I_i$  is the average intensity of the incorrect area. We randomly selected 50 and 500 images from each class as the training and test datasets, respectively.

### 3.2 Optical computing unit based on SRH

In this section, we present the results of the image classification simulation with a single layer of SR-HDNN to investigate the feasibility of the optical computing part based on SRH. We trained a single layer of the SR-HDNN. The system configuration and information-processing procedure are shown in



**Fig. 5** System configuration and procedure of information processing of a single layer of SR-HDNN.



**Fig. 6** Output of a single layer of SR-HDNN for part of the test dataset image.

Fig. 5. The reading pattern was phase-modulated into the reading light on the SLM. Subsequently, the output intensity distribution was measured and masked to obtain the output signal intensity of each identification area.

Figure 6 shows an example of the input–output pairs and outputs of the system from the part of the test dataset with the correct prediction. Masked patterns were created to show this result clearly and were not used in the calculation process of this simulation. The white and black areas in the image of the identification area correspond to the correct and incorrect labels, respectively. The confusion matrix is presented in Fig. 7. Most of the outputs are on the diagonal, where the true and predicted labels correspond to this matrix. These results demonstrate that the optical system learns the characteristics required to realize the desired function from the dataset during the training process.

		Predicted label				total	accuracy
		label 0	label 1	label 2	label 3		
True label	label 0	298	180	12	10	500	59.60%
	label 1	0	385	2	113	500	77.00%
	label 2	53	148	215	84	500	43.00%
	label 3	4	123	23	350	500	70.00%
total		355	836	252	557	2000	62.40%

**Fig. 7** Confusion matrix of a single layer of SR-HDNN.

### 3.3 Multi-layering using electrical feedback

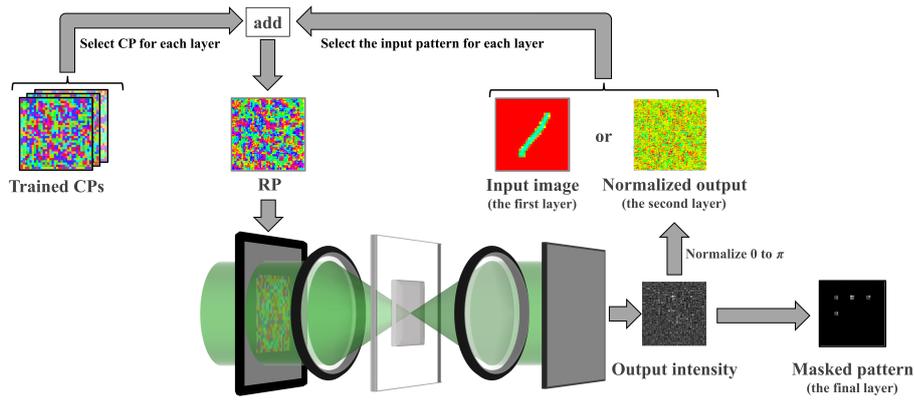
In this section, we investigate the feasibility and effectiveness of the implementation method of DNNs using an electronic computing part connecting each optical computing part. We trained a 3-layer SR-HDNN model using an optoelectronic multi-layering method.

Figure 8 shows the system configuration and information-processing procedure. Handwritten digit images were used as the input images for the first layer. The output of the previous layer was normalized to  $0-\pi$  and used as the input image for the second and subsequent layers. The nonlinear activation function is defined as the squared absolute value of the complex amplitude distribution with intensity measurement. Finally, in the output layer, the output intensity distribution was measured and masked to obtain the output signal intensity of each identification area. Additionally, we used the same hologram in all the layers in this simulation.

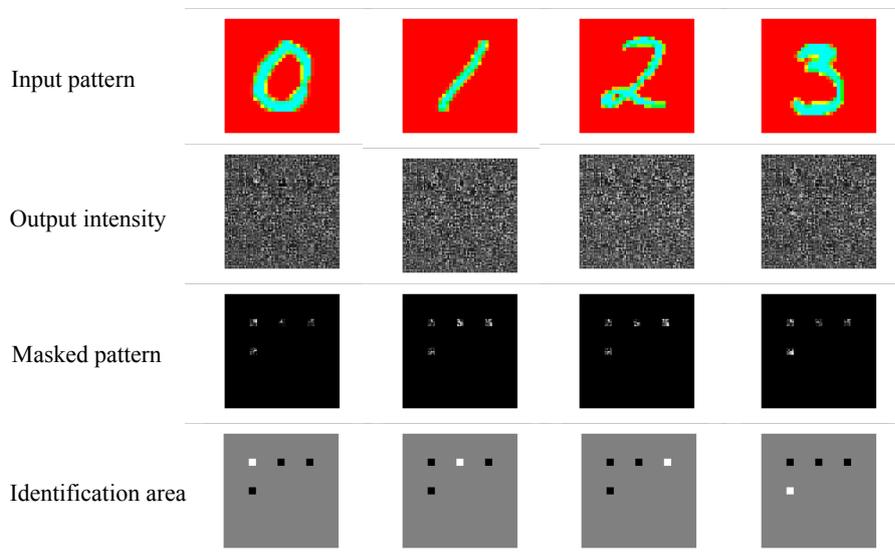
Figure 9 shows the input images, output intensity, output intensity masked with the identification area, and mask for the identification areas from the part of the test dataset with the correct predictions. These results indicate that the trained optical system identifies the input images by focusing on the energy in the correct area. Figure 10 shows the confusion matrix for the simulation. The accuracy increased by 18.45% with the introduction of electronic multi-layering from the single-layer SR-HDNN. These results demonstrate the adaptability of the optoelectronic DNN implementation method for SR-HDNN.

## 4 Discussions

We have proposed the SR-HDNN, which combines an optical computing part using holograms with an electronic part. The flexible design of optical computing functions will be achieved because the pixel-to-pixel coupling characteristics of holograms can be controlled by the recording conditions, i.e., the pattern of WP. For example, the optical computing part possibly allows the realization of the coupling of all of the pixels and only neighboring pixels, which mimic the affine and convolution layers, respectively. In addition, by



**Fig. 8** System configuration and procedure of information processing of a three-layered SR-HDNN.



**Fig. 9** Output of a three-layer SR-HDNN for part of the dataset image.

		Predicted label				total	accuracy
		label 0	label 1	label 2	label 3		
True label	label 0	392	32	55	21	500	78.40%
	label 1	0	479	16	5	500	95.80%
	label 2	28	46	395	31	500	79.00%
	label 3	15	58	76	351	500	70.20%
total		435	615	542	408	2000	80.85%

**Fig. 10** Confusion matrix of a three-layer SR-HDNN with nonlinear function.

recording different holograms achieving various functions on a disc, it is expected to realize a DNN which requires different computations for each layer, such as a CNN. In this way, SR-HDNNs, which can design both the optical and electronic computing parts, are expected to utilize the potential of the OE-DNN by appropriately combining each other. This study demonstrates the feasibility of SR-HDNN, then paves the way approach, which enables OE-DNN to balance the burden between optical and electronic computing. The result of this study enables researchers to realize various computing functions with a hologram disc.

Another advantage of SR-HDNN is its high compatibility with SR-HDS, which share the same optical system. In this case, holograms recorded for SR-HDNN, which are different from holograms recorded with information, are used. For instance, when a signal is read out, the recording medium is rotated or shifted to change the hologram to one for SR-HDNN while the read-out pattern is fed back to the SLM. This enables the application of SR-HDNN processes to readout signals, as demonstrated in this paper. Specific processing enabled by SR-HDNN in SR-HDS includes denoising of the readout patterns [16] and decoding the block-coded signals [17–19]. In the denoising application, CPs are trained to make the SR-HDS readout image closer to the ground truth signal pattern. Using the versatile phase pattern obtained in the training phase, an arbitrary readout signal is expected to be denoised [20, 21]. On the other hand, in the block-coded signal decoding application, the block codes, such as the 3/16 code, are classified in the same way as the handwritten digit image classification demonstrated in this study. However, the method for block-coded signal decoding requires local detection of the page data read by HDS and is a topic for our future works. Thus, SR-HDNN is expected to realize SR-HDS including AI-like processing with a small computational cost by an electronic computer.

In addition, scaling the training parameters can improve the performance of SR-HDNN. In Section 3, the SR-HDNN model has 1024 training parameters in the optical-connection layer, which is only one-third of the 3136 parameters in the minimum configuration of the software-based fully-connected layer for the 4-class MNIST image classification task, with a 784-nodes input layer and 4-nodes output layer, as shown in Fig. 11. In regards to DPU [11], software-based DNNs are surpassed using approximately 2.2 million training parameters in each layer, which is the total number of pixels in the SLM. Therefore, the performance can be improved by using all the pixels in the SLM as training parameters in the SR-HDNN.

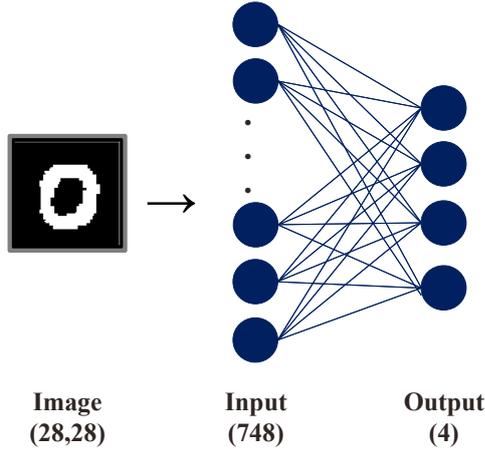
Scaling of the training dataset can also improve the performance of the SR-HDNN. In Section 3, we limited the dataset size because of the computational time required for training. To investigate the effect of the size of the dataset, we tested it with the simple software-based network shown in Fig. 11. Consequently, the accuracies of the network trained by the entire MNIST and limited datasets, as presented in Section 3, were 97.65% and 84.82%, respectively. This result shows that there is room for improvement regarding the performance of the SR-HDNN in terms of the dataset size. Therefore, it is nec-

---

essary to modify the training environment to overcome the limitations with respect to the limited number of training parameters and dataset size. The main issue in overcoming these limitations is the computational time derived from the computational cost for simulating the optical propagation process and the inefficiency of the training algorithm. There are two approaches that can address these issues: utilizing a fast propagation process on the physical optical system and changing the training algorithm.

The first approach uses a physical optical system for the propagation process rather than employing a numerical simulation. This approach enables the use of larger datasets and parameters by the time reduction of each forward-propagation process by utilizing the high speed of light. The second approach is to introduce a back-propagation algorithm, which is a fast optimization algorithm that updates all parameters through one forward and backward propagation process. Although this algorithm requires large-scale numerical differentiation using an electrical computer for the back-propagation process, indicating that high computing power is required for the training process, it allows the SR-HDNN models to use larger datasets and parameters by fast optimization. However, as the accuracy decreases because of errors from disturbance and device construction, it is necessary to find methods, such as adaptive training methods, that can eliminate errors [11].

In addition, improving the representation of a single optical-connection layer is also a valid approach. Specifically, there are two approaches in this regard: designing a hologram and utilizing complex amplitude fields in the SLM and camera plane. The first approach is to design a propagation medium (i.e., a hologram) as previously stated. The second approach uses complex amplitude fields. It has been shown that extending the modulation pattern to complex amplitude fields improves the performance of diffractive neural networks [11]. As shown in Fig. 12, we verified the improvement in accuracy by 5.95% by introducing complex amplitude modulation on the three-layer model under the same conditions, as described in Section 3.2, where an amplitude pattern was used as the input information and a phase pattern was used as CPs. We plan to investigate the implementation of complex amplitude modulation methods, such as using two SLMs or holography. Furthermore, using interferometry instead of intensity measurement in the camera plane with complex electrical activation functions also enables the implementation of complex-valued neural networks that excel in periodic signal processing.



**Fig. 11** The minimum configuration of the fully-connected layer for the 4-class MNIST image classification task.

		Predicted label				total	accuracy
		label 0	label 1	label 2	label 3		
True label	label 0	445	7	44	4	500	89.00%
	label 1	0	494	2	4	500	98.80%
	label 2	40	53	380	27	500	76.00%
	label 3	8	43	32	417	500	83.40%
total		493	597	458	452	2000	86.80%

**Fig. 12** Confusion matrix of a three-layer SR-HDNN with complex amplitude modulation.

## 5 Conclusion

We proposed the SR-HDNN as a novel OE-DNN based on the principle of SRH. The SR-HDNN has an optical computing part using holograms and an electronic computing part for multi-layering and nonlinear processing. By using the method of HDS to change propagation features, it is possible to bring flexibility to node-to-node couplings, such as by connecting only specific nodes. Furthermore, the electronic computing part enables the representation of various networks without changing the size of the optical system. The SR-HDNN is expected to enable more efficient network design than the conventional methods by providing flexibility in both optical and electronic computing. We numerically simulated of image classification task using SR-HDNNs to show its feasibility. First, we demonstrated a single-layer SR-HDNN model to investigate the behavior of the optical computing part based on SRH. As a result, the single-layer model achieved an accuracy of 62.40% for test data. By this, we confirmed that the optical system of SRH trained with RP could behave as a layer of an ANN. Next, we demonstrated a 3-layer SR-HDNN

model to investigate the behavior of a multi-layered model using both optical and electronic computing parts. As a result, the image classification accuracy improved, achieving 80.85%. This confirms the effectiveness of the DNN implementation with an electronic computing part and the feasibility of SR-HDNN.

The results obtained in this study are expected to pave the way for a novel OE-DNN framework with flexible optical and electronic computing parts and high compatibility with SR-HDS. In the future, we will investigate methods for designing holograms for various optical computing. In addition, we plan to scale up the training process and utilize the complex amplitude fields to the maximum potential of SR-HDNNs. Furthermore, the applicability to SR-HDS will also be investigated.

**Acknowledgements** We would like to thank Prof. Atsushi Shibukawa at Hokkaido University for the fruitful comments on the transmission matrix. We also acknowledge thank Editage (www.editage.com) for English language editing.

### Conflict of interest

The authors declare that they have no conflict of interest.

### References

1. Lin, X., Rivenson, Y., Yardimci, N.T., Veli, M., Luo, Y., Jarrahi, M., Ozcan, A.: All-optical machine learning using diffractive deep neural networks. *Science* **361**(6406), 1004–1008 (2018). <https://doi.org/10.1126/science.aat8084>
2. Wu, Z., Zhou, M., Khoram, E., Liu, B., Yu, Z.: Neuromorphic metasurface. *Photon. Res.* **8**(1), 46–50 (2020). <https://doi.org/10.1364/PRJ.8.000046>
3. Chang, J., Sitzmann, V., Dun, X., Heidrich, W., Wetzstein, G.: Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. *Scientific Reports* **8**(1), 12324 (2018). <https://doi.org/10.1038/s41598-018-30619-y>
4. Miscuglio, M., Hu, Z., Li, S., George, J.K., Capanna, R., Dalir, H., Bardet, P.M., Gupta, P., Sorger, V.J.: Massively parallel amplitude-only fourier neural network. *Optica* **7**(12), 1812–1819 (2020). <https://doi.org/10.1364/OPTICA.408659>
5. Shi, W., Huang, Z., Huang, H., Hu, C., Chen, M., Yang, S., Chen, H.: Loen: Lensless opto-electronic neural network empowered machine vision. *Light: Science & Applications* **11**(1), 121 (2022). <https://doi.org/10.1038/s41377-022-00809-5>
6. Sadeghzadeh, H., Koohi, S.: Translation-invariant optical neural network for image classification. *Scientific Reports* **12**(1), 17232 (2022). <https://doi.org/10.1038/s41598-022-22291-0>
7. Rafayelyan, M., Dong, J., Tan, Y., Krzakala, F., Gigan, S.: Large-scale optical reservoir computing for spatiotemporal chaotic systems prediction. *Phys. Rev. X* **10**, 041037 (2020). <https://doi.org/10.1103/PhysRevX.10.041037>
8. Dou, H., Deng, Y., Yan, T., Wu, H., Lin, X., Dai, Q.: Residual d2nn: training diffractive deep neural networks via learnable light shortcuts. *Opt. Lett.* **45**(10), 2688–2691 (2020). <https://doi.org/10.1364/OL.389696>
9. Rahman, M.S.S., Li, J., Mengu, D., Rivenson, Y., Ozcan, A.: Ensemble learning of diffractive optical networks. *Light: Science & Applications* **10**(1), 14 (2021). <https://doi.org/10.1038/s41377-020-00446-w>
10. Gu, Z., Gao, Y., Liu, X.: Optronic convolutional neural networks of multi-layers with different functions executed in optics for image classification. *Opt. Express* **29**(4), 5877–5889 (2021). <https://doi.org/10.1364/OE.415542>

11. Zhou, T., Lin, X., Wu, J., Chen, Y., Xie, H., Li, Y., Fan, J., Wu, H., Fang, L., Dai, Q.: Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit. *Nature Photonics* **15**(5), 367–373 (2021). <https://doi.org/10.1038/s41566-021-00796-w>
12. Takabayashi, M., Okamoto, A.: Self-referential holography and its applications to data storage and phase-to-intensity conversion. *Opt. Express* **21**(3), 3669–3681 (2013). <https://doi.org/10.1364/OE.21.003669>
13. Popoff, S.M., Lerosey, G., Carminati, R., Fink, M., Boccarda, A.C., Gigan, S.: Measuring the transmission matrix in optics: An approach to the study and control of light propagation in disordered media. *Phys. Rev. Lett.* **104**, 100601 (2010). <https://doi.org/10.1103/PhysRevLett.104.100601>
14. Popoff, S., Lerosey, G., Fink, M., Boccarda, A.C., Gigan, S.: Image transmission through an opaque material. *Nature Communications* **1**(1), 81 (2010). <https://doi.org/10.1038/ncomms1078>
15. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **86**(11), 2278–2324 (1998). <https://doi.org/10.1109/5.726791>
16. Hao, J., Lin, X., Chen, R., Lin, Y., Liu, H., Song, H., Lin, D., Tan, X.: Phase retrieval combined with the deep learning denoising method in holographic data storage. *Opt. Continuum* **1**(1), 51–62 (2022). <https://doi.org/10.1364/OPTCON.444882>
17. Katano, Y., Nobukawa, T., Muroi, T., Kinoshita, N., Ishii, N.: [paper] efficient decoding method for holographic data storage combining convolutional neural network and spatially coupled low-density parity-check code. *ITE Transactions on Media Technology and Applications* **9**(3), 161–168 (2021). <https://doi.org/10.3169/mta.9.161>
18. Katano, Y., Muroi, T., Kinoshita, N., Ishii, N., Hayashi, N.: Data demodulation using convolutional neural networks for holographic data storage. *Japanese Journal of Applied Physics* **57**(9S1), 09SC01 (2018). <https://doi.org/10.7567/JJAP.57.09SC01>
19. Shimobaba, T., Kuwata, N., Homma, M., Takahashi, T., Nagahama, Y., Sano, M., Hasegawa, S., Hirayama, R., Kakue, T., Shiraki, A., Takada, N., Ito, T.: Convolutional neural network-based data page classification for holographic memory. *Appl. Opt.* **56**(26), 7327–7330 (2017). <https://doi.org/10.1364/AO.56.007327>
20. Sakib Rahman, M.S., Ozcan, A.: Computer-free, all-optical reconstruction of holograms using diffractive networks. *ACS Photonics* **8**(11), 3375–3384 (2021). <https://doi.org/10.1021/acsp Photonics.1c01365>
21. Li, Y., Luo, Y., Bai, B., Ozcan, A.: Analysis of diffractive neural networks for seeing through random diffusers. *IEEE Journal of Selected Topics in Quantum Electronics* pp. 1–20 (2022). <https://doi.org/10.1109/JSTQE.2022.3194574>