Convolutional Free-space Optical Neural Networks for Image Recognition

用于图像识别的卷积自由空间光学神经网络

Qiuhao Wu[1], Xiubao Sui[1]*, Liping Wang[1], and Qian Chen[1]

[1]*School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China*

E-mail: sxbhandsome@163.com

With its unique parallel processing capability, optical neural network has shown low-power consumption in image recognition and speech processing. At present, the manufacturing technology of programmable photonic chip is not mature, and the realization of optical neural network in free-space is still a hot spot of photonic AI. In this letter, based on MNIST datasets and 4f system, one- and three-layer optical neural networks are constructed, whose recognition accuracy can reach 86.06% and 93.66%, respectively. Our network is better than the existing free-space optical neural network in terms of spatial complexity, and the three-layer's accuracy.

光学神经网络以其独特的并行处理能力，在图像识别和语音处理中表现出低功耗。
目前，可编程光子芯片的制造技术还不成熟，在自由空间中实现光学神经网络仍然是光子人工智能的一个热点。
本文基于MNIST数据集和4f系统构建了一层和三层光学神经网络，其识别准确率分别达到86.06%和93.66%。
我们的网络在空间复杂度和三层精度上都优于现有的自由空间光学神经网络。

In the era of big data, more requirements are put forward for the information processing capacity of computers, which gives rise to the research of quantum computing,[1] optical computing,[2] and other emerging fields of supercomputing.[3-4] After half a century of development, artificial neural network (ANN) has shown excellent performance in image processing,[5-6] speech recognition,[7] three-dimensional imaging,[8-9] target detection,[10-12] etc. However, the demand of computing power is becoming more and more stringent. Therefore, the mode of signal transmission in neural network is transformed from electrical to optical. It is well known that using light as the signal transmission medium enables the system to have the ability of large-scale parallel processing and high bandwidth.[13-15] In recent years, the research on ONN has become more and more popular. At present, it has achieved success in natural language processing,[16] handwriting digit recognition,[17-21] pedestrian detection,[22] intelligent sensor design and other fields.[23] However, their architectures are complex and the recognition accuracy needs to be improved. As one of the most widely used architectures in computer vision, convolutional neural network (CNN) utilizes filters to extract image features.[24] Here, we use optical signal convolution theory,[25] based on the MNIST datasets in the free-space, all-optical convolutional neural network is constructed, and the accuracy of identification can reach 93.66%. Compared to the convolutional opto-electrical hybrid neural network relied on digital neural network,[26-27] our architecture makes full use of the parallel processing ability of the light, the whole system has lower spatial complexity and low power consumption.

As shown in Fig.1(a), the neural network realizes its function through the cross connection of neurons and by means of the back-propagation algorithm.[28] The core of its work includes matrix multiplication, activation and objective function. The process of constructing ONN can be regarded as finding suitable optical devices to transplant the function of digital neural network into optical system. It is known from physical optics that light has wave-particle duality. Based on the superposition principle of diffractive wavelet,[29] X. Lin *et al.* analogized the cross interconnection of neurons to realize information transmission, and thus constructed an all-optical diffractive neural network,[18] as shown in Fig.1(b). Different from the superposition principle of wavelet, we use 4f system to realize matrix multiplication, as we mentioned in the review of ONN.[30]

Inspired by the 1×1 full CNN and to improve the accuracy of the model, we initialized a weight matrix W of the same size as the input image (512×512), multiply the corresponding elements of the input, and add a bias matrix B to obtain the pre-processed image. This can be realized by MZI device in optical system, as shown in Fig.1(c). We know that for the superposition of two vector waves E_1 and E_2 which have an angle α between the direction of vibration. It can be described as

$$E(r,t) = E_1(r,t) + E_2(r,t) \tag{1}$$

The intensity of a vector field is the time average of its conjugate dot product

$$I = \langle E \cdot E^* \rangle = \langle [E_1(r,t) + E_2(r,t)] \cdot [E_1^*(r,t) + E_2(r,t)] \rangle \tag{2}$$

It can be further written as

$$\begin{aligned} I &= E_1 \cdot E_1^* + E_2 \cdot E_2^* + \langle \mathrm{Re}\{2E_1 \cdot E_2^*\} \rangle \\ &= A_1^2 + A_2^2 + 2A_1A_2 \cos\alpha \langle \cos\psi \rangle \\ &= I_1 + I_2 + I_{12} \end{aligned} \tag{3}$$

It can be seen from formula (3) that when α=90°, namely, when the vibration directions of two waves are perpendicular to each other, the coherent term I_{12} is zero, and the intensity effect of superposition is shown as the summation of their respective intensities. This is just enough to optically realized WX+B, what we called preprocessing.

The pre-processed image is optical convolved with the initial phase mask on the Fourier plane by the optical 4f system. The output light field after Fourier inversion is received by the detector placed on the focal plane of the lens. This enables the forward propagation of optical signal in our proposed architecture. As shown in Fig.1(d).

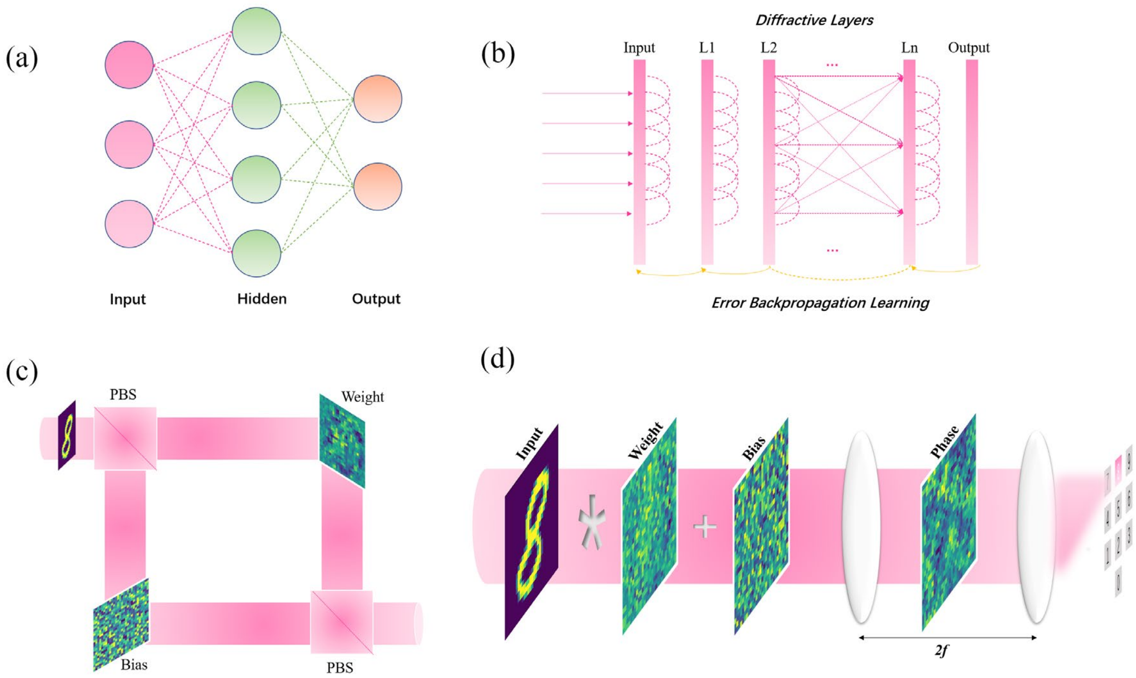

Fig.1. (a) the structure of nanophotonic neural network: $X_{n+1} = F(X_n \cdot (W_{n+1} + B_n))$, F is the nonlinear activation function. (b) All-optical Diffractive Neural Network: $X_{n+1} = W_{n+1} \cdot (X_n \circ B_n)$. (c) Preprocessing: use MZI to realize WX+B. (d) the schematic diagram of our convolutional ONN. The pre-processed input image is received by the detector

through the optical 4F system, and the process can be expressed as $O = \mathcal{F}^{-1}\{\mathcal{F}(X \circ W + B) \circ \phi)\}$. Here, $\mathcal{F}$means Fourier transform, $\circ$ means elements multiplication. O: the output field of the last layer. In addition, the symbol * in figure equals to $\circ$ in equation and the symbol + can realized by MZI based on superposition of non-interference.

Our goal is to identify the MNIST datasets to be tested, which contains 10 categories, so we divided 10 detected areas on the detector. If the input image is a handwritten digit 8 and the light intensity of the eighth detection area is the largest compared to other areas, it is considered that our architecture correctly recognizes the image to be tested. We know that any iterative algorithm can be trained and converged by a neural network model. In our algorithm, initialized weights W, bias B, and phase masks are all learnable variables. Finally, the ideal model was obtained by using 55,000 training sets, 5000 validation sets, and the generalization ability of the model was verified by 10,000 test sets. The algorithm includes forward propagation and back propagation, adopts stochastic gradient descent, cross entropy loss function, and α represents learning rate.

$$Pre = W \circ X + B \tag{4}$$

$$O = \mathcal{F}^{-1}\{\mathcal{F}(Pre) \circ \phi)\} \tag{5}$$

$$I = O \cdot O^{*} \tag{6}$$

$$LOSS = -\sum_{i=1}^{n} \hat{I} \log{(I)} \tag{7}$$

$$W = W - \alpha \frac{\partial LOSS}{\partial W} \tag{8}$$

$$B = B - \alpha \frac{\partial LOSS}{\partial B} \tag{9}$$

$$\phi = \phi - \alpha \frac{\partial LOSS}{\partial \phi} \tag{10}$$

Based on the above algorithm, we set up the all-optical neural network with layer 1 and layer 3 respectively. The so-called 3-layer ONN is achieved by adding two additional 4f systems and two high reflectance dielectric mirror (HRDM), as shown in Fig.2(a). As we discussed in ref.30 and ref.31, all-optical nonlinear activation plays key role in the multi-layer ONN. In the training on PC, we use the Relu activation and HRDM (99.9% reflectivity) is used to realize this function in the optical system. Considering the light intensity is non-negativity, we choose shift the turning point of the function to fit it, as shown in Fig.2(b).

正如文献30和31所讨论的，全光非线性激活在多层ONN中起着关键作用。
在PC机上的训练中，我们使用了Relu激活，并在光学系统中使用了HRDM(99.9%反射率)来实现这一功能。
考虑到光强是非负的，我们选择平移函数的转折点来拟合，如图2(b)所示。

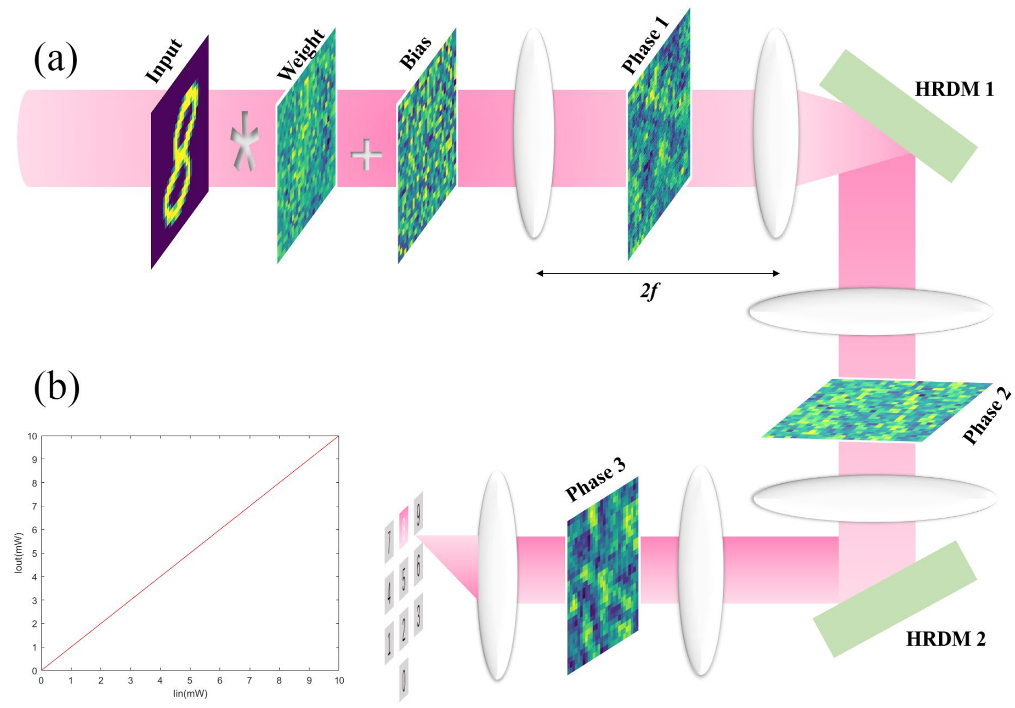

Fig.2. (a) the schematic diagram of 3 layers ONN. HRDM is high reflectance dielectric mirror with 99.9% reflectivity. (b) the shift Relu activation. Light intensity is non-negativity, so we don't have to worry about the negative half axis.

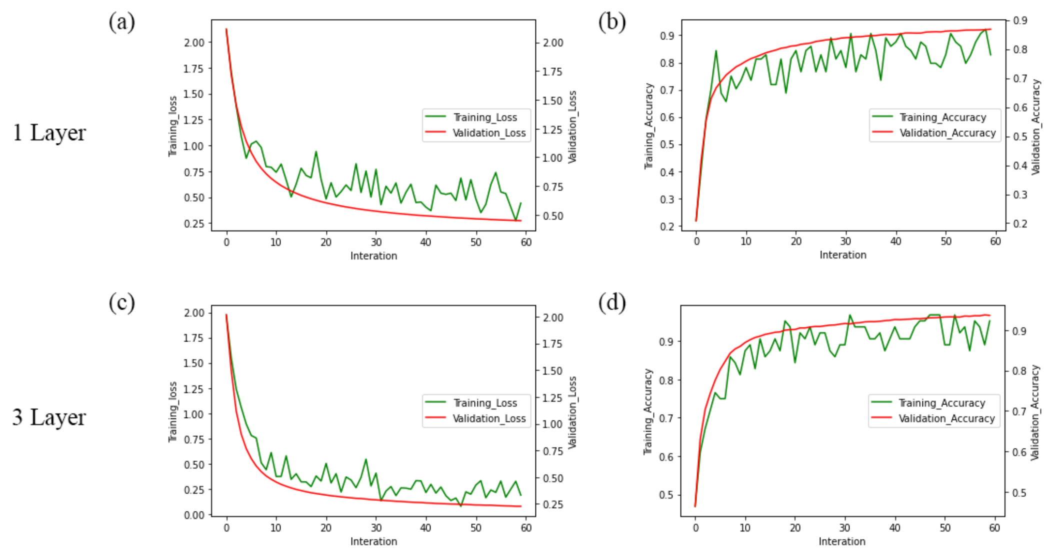

Fig.3. The results of the training for the one- and three-layer ONN with the size of 55000 training datasets and 5000 validation datasets. (a), (c) describe the training loss and validation loss of the ONN. (b), (d) show the training accuracy and validation accuracy of the ONN.

In view of the fact that we used optical 4f system and convolution theory to realize matrix multiplication in free space, we called this architecture as the free-space convolutional optical neural network (CONN). Different from our other research,[31)] this algorithm is a crucial step for input image preprocessing by learning the optimal weight W and bias B, and ultimately reduces the number of layers of the network and optimizes the spatial complexity of the model. Similar to the previous studies on ONN, we also used MNIST data set to design the optical neural network system. The performance of training is shown in Fig.3. In our architecture, the recognition accuracy of one- and three-layers reached 86.06% and 93.66%, respectively. In particular, the accuracy of 3-layer ONN exceeds that of any previous free-space ONN's performance on the MNIST data set, which is shown in Table 1. In addition, Fig.4 shows the recognition accuracy of MNIST data.

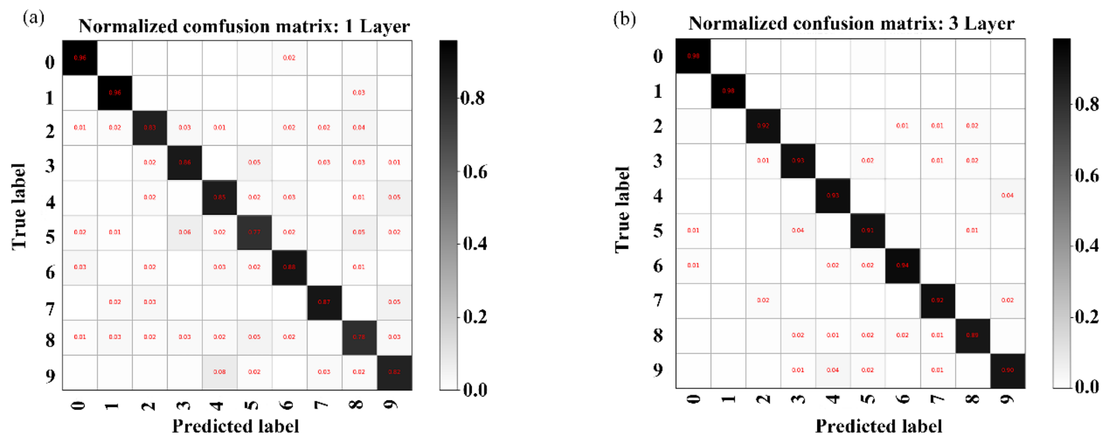

Fig.4. (a)-(b) show the test performance of the 10000 MNIST datasets, and the accuracy for one-and three-layer ONN is 86.06% and 93.66%, respectively.

It can be seen from Table I. that our 3-layer CONN has the lowest spatial complexity and the highest recognition accuracy. We have reason to believe that by increasing the number of layers of the network and adopting Regularization, Batch Norm, Shortcut connection, our architecture will have higher recognition accuracy.

Table I. The test accuracy of MNIST in all-optical Neural Networks.

Research	Accuracy
Ref.17 (Nano ONN)	79%
Ref.18 (5-layers D^2NN)	91.75%
Ref.18 (7-layers D^2NN)	93.39%
Ref.19 (5-layers ONN)	88.2%
Ref.21 (10-layers ONN)	91.96%
Ref.31 (6-layers FONN)	92.51%
Ours (1-layer CONN)	86.06%
Ours (3-layer CONN)	93.66%

We know the concept of using phase-masks in 4f system is well known and has also extensively been studied in the context of neural networks, such as Ref.26 and Ref.27. However, they do not use the 4f system to build all-optical NN. Their optical system is hybrid optoelectronic based on diffraction or convolution, our architecture is an all-optical implementation that makes full use of the parallelism of photons and the ability for large-scale interconnections. In our opinions, the main method of Chang et al. is based on CNN, they use multi-kernal to extract features of the object and tail them in a 2D plane. Although they use an optical correlator similar to our 4f system, but their one optical correlator must be processed by the digit CNN or full-connected NN and then they obtain a good performance in MNIST (their accuracy lower than ours). In addition, we can optically implement the nonlinear activation function by using the sCMOS as discussed in ref.32.[32] It also proves that our idea is feasible and we first submit this paper in 18-Dec-2020.

The biggest difference of our architecture is that we use polarization to preprocess the input image, utilize HRDM to implement optical nonlinear activation and then pass through three spatial optical convolutional layers to obtain excellent recognition accuracy. The three 4f systems have a lower complexity than the multi-layer diffractive neural network. Because the number of neurons in each layer is 512×512, it can be seen as the nodes of the 1-layer ONN. Total nodes of our network are 3×512×512, ~786000 nodes. In addition, a lens is a passive device whose Fourier transform takes about 1ps to perform. If the parameters of our ONN have been learned, we then use the amplitude and phase spatial light modulator (SLM) to load the weight, bias and phase, whose high response rate (1 kHz, different from frame rate) also ensures that our optical testing system responds very quickly. The response time of the CCD (80Hz) is 8 ms, and the latency of our single-layer convolution is only a few milliseconds. In our implementation, maintaining the SLM requires some power of ~10mW per modulator, on average. Considering our CONNs have 3 layers and the 10^7 HZ clock rate, which correspond to 7.86×10^{12} dimensional multiplications per second. The number of operations (FLOPs) can be given by

$$N = 2 \times 786000^2 \times 10^7\, FLOPs \qquad (11)$$

The power consumption of our CONN is dominated by the laser and SLMs, whose total power needed is approximately 100mW. Therefore, the energy per FLOP of the CONN is $2 \times 786000^2 \times 10^8\, FLOPs / J$ ~1.23×10^{20}FLOPs/J.[16] As a result, our CONN has higher speed, much lower complexity and power consumption.

In summary, we introduce a convolutional optical neural network in free space with only three layers. This constructure means our model have the lowest spatial complexity compared with other researches in free space ONN. The aim of our research is to design a better ONN model to be integrated into the photonic platform such as photonic integrated circuits. Our study will contribute to this futural AI.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 11773018 and 61727802), in part by the Key Research and Development programs in Jiangsu China (Grant No. BE2018126), and in part by the Fundamental Research Funds for the Central Universities (Grant Nos. 30919011401 and 30920010001).

References

1) Stefanie Barz, J. Phys. B: At. Mol. Opt. Phy*s.* **48**, 083001 (2014).
2) C. Ríos, N. Youngblood, Z. Cheng, M. Le Gallo, W. H. P. Pernice, C. D. Wright, A. Sebastian, H. Bhaskaran, Sci. Adv. **5**, eaau5759 (2019).
3) F. Wang, H. Lv, Q. Li, J. Li, X. Zhang, J. Shi, L. Wang, C. Fan, Nat. Commun. **11**, 121

(2020).
4) B. Leonard, M. Fan, J. S. Welling, PEARC '20. Pra. & Exp. in Adv. Res.Com. 2020, p. 473.
5) M. I. Waris, J. Mir, V. Plevris, A. Ahmad, 2020 IOP Conf. Ser. Mater. Sci. Eng. **899**, 012014 (2020).
6) K. H. Jin, M. T. McCann, E. Froustey and M. Unser, IEEE Trans Image Process. **26**, 4509 (2017).
7) Q. Wang, B. Zheng, Q. Li, C. Shen and Z. Ba, IEEE Trans Inf Foren Sec. **16**, 896 (2020).
8) J. Dong, W. Cong, D. Ai, Y. Chu, Y. Huang, H. Song, Y. Jiang, Y. Wang, M. Li, J. Yang, IEEE Trans Compu Imaging. **5**, 251 (2019).
9) A. Kulikajevas, R. Maskeliūnas, R. Damaševičius, S. Misra, Sensors. **19**, 1553 (2019).
10) L. Long, F. Jiang, J. Phys.: Conf. Ser. **1098**, 012030 (2018).
11) E. Lashnits, P. Neupane, R. G. Maggi, K. E. Linder, J. M. Bradley, N. Balakrishman, B. L. Southern, G. P. McKeon, R. Chandrashekar, E. B. Breitschwerdt, J Vet Intern Med. **34**, 145 (2020).
12) Q. Liu, W. Chen, L. Zhao, C. Ma, IOP Conf. Ser.: Mater. Sci. Eng. **631**, 052021 (2019).
13) H. J. Caulfield, S. Dolev, Nat. Photonics. **4**, 261 (2010).
14) J. Bueno, S. Maktoobi, L. Froehly, I. Fischer, M. Jacquot, L. Larger, D. Brunner, Optica. **5**, 756 (2018).
15) Y. Luo, D. Mengu, N. T. Yardimci, Y. Rivenson, M. Veli, M. Jarrahi, A. Ozcan, Light Sci. Appl. **8**, 112 (2019).
16) Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, M. Soljacic, Nat. Photonics. **11**, 441 (2017).
17) E. Khoram, A. Chen, D. Liu, L. Ying, Q. Wang, M. Yuan, Z. Yu, Photonics Res. **7**, 823 (2019).
18) X. Lin, Y. Rivenson, N. T. Yardimei, M. Veli, Y. Luo, M. Jarrahi, A. Ozcan, Science, **361**,1004 (2018).
19) L. Liu, L. Zhu, Q. Zhang, B. Zhu, Q. Yao, M. Yu, H. Niu, M. Dong, G. Zhong, Z. Zeng, IEEE Photon. Technol. Lett. **31**, 1952 (2019).
20) D. Mengu, Y. Luo, Y. Rivenson, A. Ozcan, IEEE J Sel Top Quantum Electron. **26**, 3700114 (2020).
21) T. Zhou, L. Fang, T. Yan, J. Wu, Y. Li, J. Fan, H. Wu, X. Lin, Q. Dai, Photonics Res. **8**, 940 (2020).
22) P. Antonik, N. Marsal, D. Brunner, D. Rontani, Nat. Machine Intelligence. **1**, 530 (2019).
23) L. Mennel, J. Symonowicz, S. Wachter, D. K. Polyushkin, A. J. Molina-Mendoza, T.

Mueller, Nature. **579**, 62 (2020).

24) A. Krizhevsky, I. Sutskever, G. E. Hinton, NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems. **1**, 1097 (2012).
25) G. C. Sherman, J. Opt. Soc. Am. **57**, 546 (1967).
26) J. Chang, V. Sitzmann, X. Dunn, W. Heidrich, G. Wetzstein, Sci. Rep. **8**, 12324 (2018).
27) S. Colburn, Y. Chu, E. Shilzerman, A. Majumdar, Appl. Opt. **58**, 3179 (2019).
28) Y. Hirose, K. Yamashita, S. Hijiya, Neural Netw. **4**, 61 (1991).
29) A. S. Marathay, J. F. McCalmont, J. Opt. Soc. Am. A. **18**, 2585 (2001).
30) X. Sui, Q. Wu, J. Liu, Q. Chen, G. Gu, IEEE Access. **8**, 70773 (2020).
31) Q. Wu, X. Sui, Y. Fei, C. Xu, J. Liu, G. Gu, Q. Chen, submitted to Appl. Phys. Lett.
32) Z. Gu, Y. Gao, X. Liu, Opt. Express. **29**, 5877-5889 (2021).