ADVANCED PHOTONICS RESEARCH Open Access
www.adpr-journal.com

Neural Optimizer for Inverse Design of Complex-Modulated Hologram Implemented by Plasmonic Metasurfaces

*Huade Mao, Yue Yu, Yu-Xuan Ren, Ka Yan Chan, Jiqiang Kang, Xiankai Sun, Edmund Y. Lam, and Kenneth K. Y. Wong**

Inverse design of a metasurface involves searching parameters in a high-dimensional space, which needs huge computational power. To ease the computational burden, neural network, a well-researched computer science stream, has demonstrated its potential usage in the inverse design of a photonic device. Many studies primarily focused on the nanostructure's configuration. However, the near field of a metasurface requires further optimization to achieve a desired holographic image at relieved computational power. Here, a convolutional neural network is developed to optimize the complex field of a computer-generated hologram, which can be fabricated into a metasurface to generate the desired holographic image upon lensless image projection. The neural optimizer can accelerate the design speed 400 times faster than theoretical computation and reduce the time complexity from O(n^4) to O(n^2). The neural optimizer has been compared against three other methods, e.g., gradient-based optimization, global genetic algorithm, and coupled-mode theory, to demonstrate a lowered error rate from more than 10% to 1.38% for the benchmark testing and a reduced running time from hours to near 1 s. The neural optimizer is envisioned to play a key role in lensless image projection and real-time metasurface pattern design.

1. Introduction

Metasurface, as an integrated photonic device, has been used in many fields, including hologram,[1,2] polarization control,[3] color manipulation,[4] achromatic focusing,[5] etc. To achieve those goals, the metasurface needs to be meticulously designed to realize specific optical responses concerning both the nano-antenna configuration and the pattern arrangement. The design process usually consists of sweepings over different parameters, e.g., shape, height, length, etc., through an electrical-magnetic (EM) solver (e.g., Lumerical finite-difference time-domain (FDTD))[1] so as to find a suitable configuration.

Given the computation burden in the sweeping process, the deep learning model prevails in finalizing the nano-unit's geometry in a much faster way.[6,7] Apart from nano-antennas' design, the deep learning model has also been applied to optimize the arrangement of them.[8] A generative model is adopted to create a meta-grating with the best diffraction efficiency.[9] In addition, a convolutional neural network is trained to map from the holographic image to the phase mask.[10] Such inverse design problems could be solved readily through a deep learning model.[9,11–13] However, with metasurfaces taking on different forms, e.g., gratings, plasmonics, layers, etc., for different applications, it requires a tailored neural network to accommodate the various parameters.

In the field of hologram, the inverse problem from a holographic image to the phase mask is generally solved through the Gerchberg–Saxton algorithm[14] or inverse Fourier transform,[15] which could be easily achieved through a physical Fourier transform lens.[14] The underlying physics is about Fourier optics. Given a field transmission function $f(x,y)$, we could get its amplitude spectrum in the Fourier plane as $F(u,v) = \mathscr{F}(f(x,y))$, where $\mathscr{F}$ stands for Fourier transform.[15] Experimentally, it is common to use a spatial light modulator (SLM) to tune the transmission function $f(x,y)$ in a phase-only manner, thus generating the desired holographic image. The phase mask on SLM is usually termed as a

H. Mao, K. Y. Chan, J. Kang, E. Y. Lam, K. K. Y. Wong
Department of Electrical and Electronic Engineering
The University of Hong Kong
Pokfulam Road, Hong Kong 999077, China
E-mail: kywong@eee.hku.hk

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adpr.202200085.

DOI: 10.1002/adpr.202200085

Y. Yu, X. Sun
Department of Electronic Engineering
The Chinese University of Hong Kong
Shatin, New Territories, Hong Kong 999077, China

Y.-X. Ren
Institute for Translational Brain Research
Shanghai Medical College
Fudan University
Shanghai 200032, China

J. Kang, K. K. Y. Wong
Advanced Biomedical Instrumentation Centre
Hong Kong Science Park
Shatin, New Territories, Hong Kong 999077, China

computer-generated hologram (CGH).[16] Metasurface featuring the Pancharatnam–Berry phase could also be used for hologram with higher resolution and better performance.[17]

However, the phase-only modulation is apparently far worse than complex modulation in terms of sharpness and luminosity.[1] Moreover, complex modulation is free from the Fourier transform lens, as it incorporates the inverse propagation function into the complex design,[18] thus every point in the holographic image (object field) is the interference results of all the spherical waves from the CGH. This lensless modality, combined with metasurface, is of high interest in recent studies.[1,10,19] However, efficient inverse mapping from a desired holographic image to the original complex-modulated metasurface is still highly desirable in the community.

Here, to the best of our knowledge, we first presented a neural optimizer (NO) to design a complex-field CGH, which was 400 times faster than theoretical computation, to achieve a desired holographic image. The time complexity was reduced from $O(n^4)$ to $O(n^2)$, where n is the length of the CGH. The CGH was then fabricated into a metasurface within a lensless modality to deliver a compact optical system. The NO featured a dual-branch convolutional neural network to encode the complex pattern containing amplitude and phase. A similar architecture has been demonstrated in a classification work[20] to improve the accuracy. In our work, the NO tended to generalize from the holographic image field to the metasurface field. Our NO has outperformed three other methods, e.g., gradient-based optimization, global genetic algorithm, and coupled-mode theory, in terms of the running time and error rate in metasurface design. It is envisaged that such technique can promote lensless image projection and real-time metasurface pattern design.

2. Principle

This lensless holography could be simplified as the superposition of the spherical wave propagation from each point source on a metasurface. For a spherical wave, the free space propagation can be represented by the Green function $G_0(\vec{r}, \vec{r_1}) = \exp\left(\pm \frac{ik|\vec{r}-\vec{r_1}|}{4\pi|\vec{r}-\vec{r_1}|}\right)$,[18] where the plus (minus) sign denotes the forward (backward) propagation.

Within a metasurface, each unit performs as a secondary source according to the Huygens–Fresnel law.[19] The holographic image is, therefore, the superimposition of the secondary wave from all the units in the metasurface. Inversely, for a given holographic image, a metasurface's pattern is the summation of all contributions of backpropagation from the pixels, which is formulated as

$$M(x, y) = \sum_{i,j} \frac{\exp(-ik \cdot R_{ij}(x, y))}{R_{ij}(x, y)} \cdot H(i,j) \quad (1)$$

where $M(x,y)$ denotes the complex modulation at position (x,y) on a metasurface M, k represents the wavenumber, $H(i,j)$ is the intensity information at position (i,j) for a given holographic image H, and $R_{ij}(x,y)$ is the pixel-wise distance between M and H, which in mathematical form is $|\overrightarrow{\mathrm{OM}}(x, y) - \overrightarrow{\mathrm{OH}}(i, j)|$ with O being the coordinate origin.

Equation (1) is essentially an inverse design process. However, it involves, apparently, two loops to finalize the complex information of a metasurface. If the holographic image and metasurface are both in size of $n \times n$ pixels, the whole inverse design problem is of the complexity $O(n^4)$. Here, through a neural network, it is anticipated to generalize Equation (1) into

$$M(x, y) = \widehat{Q} \cdot H(x, y) \quad (2)$$

where the decoding process $\widehat{Q}$ is the transformation matrix which functions as NO. Hence, the problem is optimized to a complexity of $O(n^2)$. Through this process, we managed to accelerate the inverse design 400 times faster compared with the theoretical method in Equation (1) with a testing loss below 5%. It has the potential for lensless image projection on a faster basis.

3. Neural Network Architecture

Our NO aims at mapping a holographic image to the metasurface's complex pattern. It consists of U-Net[21] and residual network[22] architectures (**Figure 1**). Since the U-Net is commonly used in image segmentation and encoding,[23] here we extended the layer by adding a short-term memory within each block (Figure 1f). The NO model features two outputs to represent both the amplitude and phase of the complex field of a metasurface. The rationale behind this architecture is: 1) it is required to represent a complex number by two real numbers, therefore dual-output is adopted; 2) amplitude and phase are correlated to some extent, hence their skip connections share the same downsample block.

The model, similar to a prior report[10] but different in output branches, comprises five different blocks. Each block color represents: a) downsampling, b) dilation, c) upsampling, d) residual, and e) convolution, respectively (Figure 1a–e). For different blocks, its filters range from convolutions with a stride of (1,1) or (2,2), convolutions with a dilation rate of (2,2), to transpose convolutions with a stride of (2,2). The activation functions after each layer is Rectified Linear Unit (ReLU). And batch normalization (BN) is adopted to standardize the inputs for the next layer.[24] The skip connection inside each block is an Add function. The overall architecture of our NO model is illustrated in Figure 1f, within which the skip connection is concatenation and the dropout rate is set at 0.03 in the dropout layer. The kernel size of all the layers is (3, 3). The optimizer is Adam[25] with a learning rate of 0.001.

The input layer is the desired observation (holographic image) with a size of 416×416. Then it passes through downsample and dilation blocks. The channel size is specified on each block (Figure 1f). Within each downsample block, the size of the image is reduced by a factor of 2 along each dimension. The size reduces to (13, 13) at the final downsample block with a channel size of 192. Next, the block separates into two branches. The phase branch (top path in Figure 1f) contains five upsample blocks followed by a residual and a convolution block. The anterior four upsample blocks are concatenated with skip connection from the downsample path so as to solve the diminishing gradient issue and extract local and global features.[26] The amplitude branch, (lower path in Figure 1f) consists of the same structure as the phase branch.

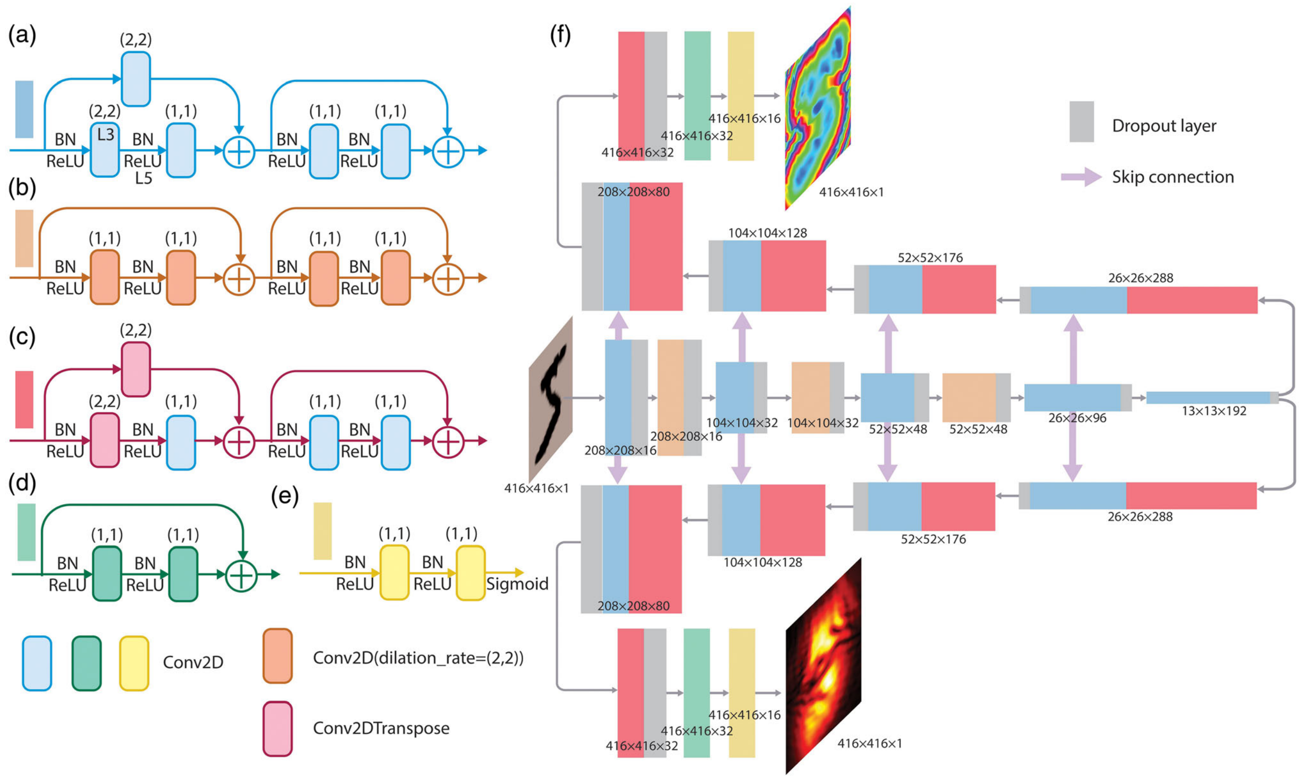

Figure 1. The architecture of the neural optimizer (NO) model consists of: a) downsample, b) dilation, c) upsample, d) residual, and e) convolution blocks. Different blocks are represented in a single rectangle of their corresponding color. f) The detailed structure of the NO model is built with all the above blocks with dropout layers denoted by a gray block and skip connections represented by a purple arrow. The numbers on the top or bottom of each block such as $a \times b \times c$ indicates the size of the feature map is $a \times b$ and the feature channel is c. In each block, the convolution filter types are specified by different colors in the legend at the bottom left corner, including Conv2D, Conv2D with a dilation rate of (2,2), and Conv2DTranspose. L3 and L5 mean layer 3 and layer 5 for the first block, whose maximum filter activation is shown in Figure 5.

The objective function to be minimized during training is the mean absolute error (MAE)

$$\text{Loss} = \sum \frac{1}{wd} ||P - T|| \tag{3}$$

where w and d denote the width and height of the output, P means the prediction output, and T represents the true output label, which is normalized in the range of [0,1]. The final MAE sums over the MAEs from two branches with a ratio of 1:1. The MAE is actually the pixel-wise average deviation for the prediction field. Therefore, the MAE is a raw value. To represent the percentage of deviation, we use a term, mean absolute deviation (MAD), i.e., MAD = MAE/span. The denominator is chosen as 1.0 given both amplitude and phase are normalized to a span of 1. Here, MAE is an absolute value whereas MAD is a percentage ratio.

We selected 8000 images from the Caltech101 dataset and resized them into 416×416 pixels. For each resized image H, we used Equation (1) to calculate the corresponding metasurface pattern M as its label, which led to 8000 pairs of datasets. The focusing distance from metasurface M to holographic image H was 750 μm. Each unit pixel had a size of $500 \times 500\ \text{nm}^2$. Hence, both the target holographic image and metasurface had a size of $213 \times 213\ \mu\text{m}^2$. The dataset was split into training and test datasets with a 9:1 ratio, which means 7200 pairs for training and 800 pairs for testing. Within the training data, we also took 10% of them as validation. Therefore, in every training epoch, 6480 pairs were used for training while 720 pairs were set aside for validation.

We trained the NO for 20 epochs with our dataset. The training history (**Figure 2**) suggested that the loss had already reached

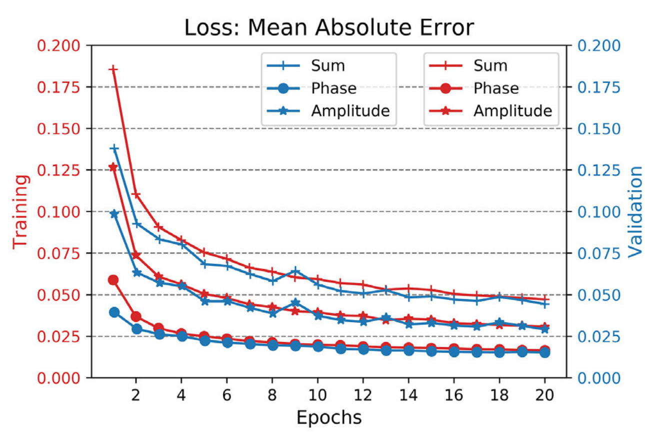

Figure 2. Training history over 20 epochs. Red line denotes training loss while blue line represents validation loss with respect to epochs.

a plateau from epoch 8 and onwards. At epoch 20, the training MAE reached 0.0471 (MAD of 4.71%) and the validation MAE was 0.0443 (MAD of 4.43%), implying that the NO was not overfitting. The test MAE was 0.0457 (MAD of 4.57%), suggesting a good bias and variance. Since the dropout layer will be deactivated in the evaluation process, the validation and test loss were slightly lower than the training loss.

To validate the novelty and impact of our NO, we compared it with three other methods, namely, gradient-based optimization, global genetic algorithm, and coupled-mode theory. The implementation of these three methods is detailed in S1, Supporting Information. We took the "HKU" letters as the benchmark testing sample with the visual comparison displayed in S2, Supporting Information. The comparison results were listed in **Table 1**, which shows a greater edge of our NO over other methods.

After nearly 30 h training, our NO could deliver a metasurface design in about 1 s with a 1.38% error in our benchmark testing of "HKU" letters. In comparison, the other three methods had to start from the beginning for each new input with an increased error. When the number of metasurface reaches 20, the total running time of gradient-based optimization is already much longer than our NO. Besides, in the three methods other than NO, we had to reduce the metasurface size or holographic image size due to the memory limit of our workstation. We acknowledged that the three methods other than NO may require hyperparameter tuning to increase efficiency and reduce error. But it was beyond the scope of this manuscript. And the benchmark testing dataset should contain a bundle of testing images rather than a single image of "HKU" letters. However, it would be too heavy a computation burden to prepare a large dataset for the community to duplicate while only one image would be easier to work with. We also validated our NO using a different open source dataset detailed in Section 4.

In conclusion, our NO has exceeded the other three methods in terms of running time, error rate, metasurface, and holographic image sizes.

4. Experimental Verification

The metasurface consisted of Au blocks coated on a glass substrate. The Au blocks were fixed with a height h of 80 nm and a width w_x of 200 nm. The length of the Au block ranged from 200 to 400 nm, and the orientation angle was tuned from 0 to π to achieve the 2π phase modulation according to the Pancharatnam–Berry phase.[2] The design of the Au block to achieve complex modulation was well presented previously.[27] In brief, the Au block illuminated by a left circularly polarized (LCP) light functioned as a birefringent crystal, and part of the LCP light changed into right circularly polarized (RCP) light (inset in **Figure 3**). The output can be formulated as

$$S_{\text{out}} = \left\langle R|\Gamma(-\alpha)\widehat{W}\Gamma(\alpha)|L\right\rangle \tag{4}$$

where R and L, respectively, denote RCP and LCP; $\Gamma(\alpha)$ represents the rotation matrix; α indicates the orientation angle of the Au block; $\widehat{W}$ is the transformation matrix.

The electric field of the filtered RCP component is

$$E_{\text{out}} = i \cdot \sin\left[\frac{kh(n_{\text{o}} - n_{\text{e}})}{2}\right] \cdot \exp\left(i\left[\frac{kh(n_{\text{o}} + n_{\text{e}})}{2} + 2\alpha\right]\right) \tag{5}$$

where k is the wavevector, n_{o} and n_{e} are the ordinary and extraordinary index of the Au block respectively. We have run a full-wave simulation for the Au antenna through Lumerical FDTD. The length of the antenna was specially picked to achieve an amplitude range of 0 to 1, while the orientation was tailored to exert a full range of 2π modulation.

To create the targeted observation pattern, the Au antenna with different lengths and orientations would be arranged to form a specific pattern. Accordingly, we adopted the NO to efficiently design such a pattern. The metasurface was then

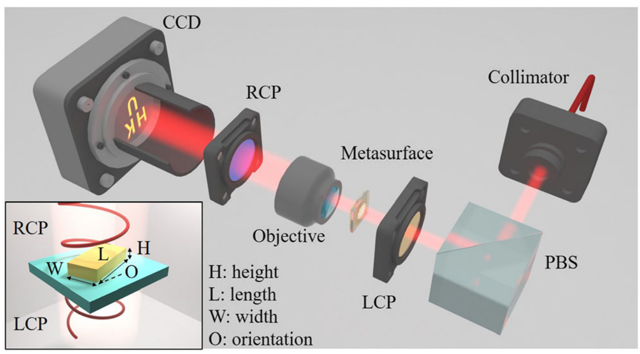

Figure 3. Experimental setup. PBS: polarization beam splitter. LCP/RCP: left/right circularly polarized light filter, containing a $\lambda/4$ waveplate and a polarizer. CCD: charge-coupled device camera. Inset: unit structure of the metasurface. It consists of an Au block on a glass substrate.

Table 1. Comparison table of four different methods to design a metasurface. All these methods were tested on the same CPU of AMD EPYC 7302P 3.0 GHz (16C/32T, 128M CACHE). The training time is not applicable to gradient-based optimization, global genetic algorithm, and coupled-mode theory. 1 iteration time denotes the time to run one forward simulation and one backward propagation to update the parameters. Running time records total forward propagation time to deliver a result. For example, the trained NO only requires one forward propagation to deliver the prediction, while the other three methods need multiple iterations to output the results. Error denotes the deviation percentage between simulated intensity and desired intensity averaged across pixels. Metasurface and holographic image size denote the number of units we adopted.

Method	Training time	1 iteration time	Running time	Error	Metasurface size	Holographic image size
Neural optimizer (NO)	≈30 h	N/A	≈1 s	1.38%	416 × 416	416 × 416
Gradient-based optimization	N/A	≈1146 s	≈1.6 h	10.69%	416 × 416	100 × 100
Global genetic algorithm	N/A	≈1013 s	≈5.6 h	19.90%	416 × 416	100 × 100
Coupled-mode theory	N/A	≈106 s	≈5.9 h	70.72%	52 × 52	25 × 25

fabricated through electron-beam lithography (EBL). We built an experimental platform to test the fabricated metasurface (Figure 3). The 1030 nm incident light was prepared with an LCP polarization, and then illuminated on the metasurface. Next, the modulated light passed through the RCP filter, finally casting a desired holographic image on the camera.

We selected two images to test the concept. None of them has been used to train the NO. They were "HKU" letters (**Figure 4**a), the Chinese character "光" (meaning "light") (Figure 4b). The theoretical complex field of the metasurfaces (row I in Figure 4) and the corresponding holographic image (row II in Figure 4) were calculated through Equation (1). It took ≈450 s to compute through MATLAB on a CPU of Intel i5 8500 4.1 GHz. In contrast, with the same CPU, the NO prediction only took about 1 s. Before training, the prediction of the neural network was purely noise without meaningful information (column a2, b2 in Figure 4). After training, the generated holographic image matched well with the theoretical one (column a3, b3 in Figure 4). Then we fabricated the metasurface featuring the complex modulation for testing. The experimental holographic image matched well with our theory (Figure 4).

Due to the ohmic loss and amplitude modulation, the transmitted power, in the case of "HKU" letters, was only 1.87% of the incoming power. However, the experimental results sufficed to demonstrate the concept of NO. And we will consider turning to dielectric metasurface in the future. The issue of the blurred letter "U" comes from the fabrication imperfection due to the unstable performance of the EBL machine.

The training process took roughly 30 h using Tensorflow on a CPU of AMD EPYC 7302P 3.0 GHz (16C/32T, 128M CACHE).

Apart from the Caltech101 database, 2000 images from Labeled Face in the Wild (LFW) were selected as a different test pool for cross-checking. The test MAE of the LFW database for the NO trained under Caltech101 was 0.0484 (MAD: 4.84%) on average, which was considerably good. The test images from a different database could verify that our NO generalized well across images in different datasets.

We further visualized the maximum filter activation of what has been learned by the neural network (**Figure 5**). The input size was 32 × 32. Starting from a randomly initialized input, we used a gradient ascent algorithm to maximize the filter response.[28] The chosen filter was maximally responsive to the finalized input (Figure 5). For a shallower layer (layer 3 in Figure 5a), it could capture simple features, including stripes or dots. When it came to a deeper layer, for example, layer 5 in Figure 5b, complex features like clouds or textures were learned.

To better reveal the neural network from the inside, we visualized the intermediate activation output in **Figure 6** for the input holographic image, a shape of omega from the lab logo of Kenneth K.Y. Wong's group at HKU. The channels are the random three channels of the last convolutional layer of each specific layer. The block Do-1&3&5, Aup-1&3 (Pup-1&3), and Ares1 (Pres1) represents downsample, upsample, and residual blocks (for amplitude and phase branches), respectively.

When a holographic image was fed into the neural network, the downsample blocks first extracted simple features (e.g., vertical/horizontal edges, contrast, etc.), then identified complex features (e.g., thickness of a stripe, etc.). After that, the upsample branches utilized such information to reconstruct the amplitude and phase pattern, which will lead to a desired holographic image.

5. Conclusion

In conclusion, we built an NO to intelligently design a CGH for an optical lensless 2D image projection. The CGH then was fabricated into a complex-modulated metasurface, which can generate a desired holographic image without a bulky lens system,

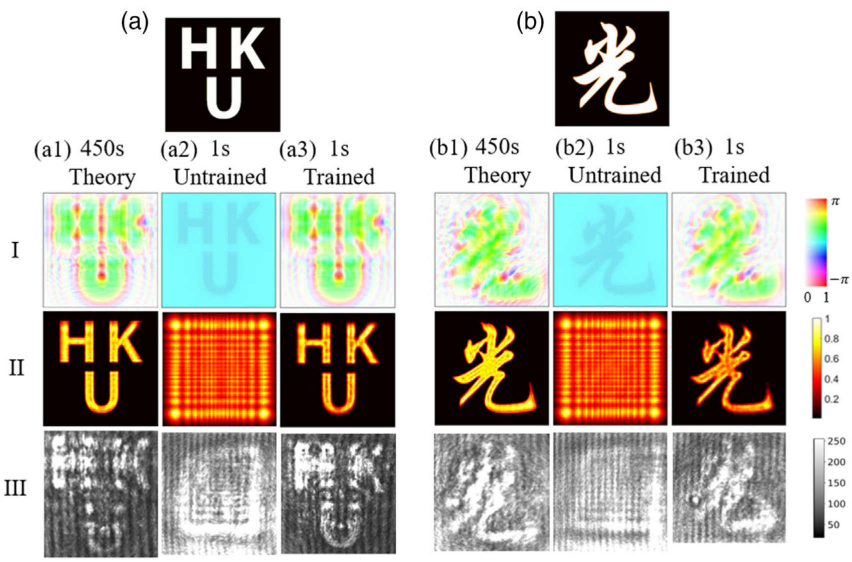

Figure 4. Experimental results of NO. Row I: complex field of metasurface's pattern. Row II: simulated holographic image with a distance of 750 μm from the metasurface. Row III: holographic image in experiments. The observed holographic images are: a) "HKU" letters, and b) "光" character. (Column a1, b1) Theoretical computation as in Equation (1) with a computation time of ≈450 s. (Column a2, b2) Prediction through the untrained NO with ≈1 s. (Column a3, b3) Prediction through the trained NO with ≈1 s. The training process takes ≈30 h to complete.

ADVANCED PHOTONICS RESEARCH Open Access
www.adpr-journal.com

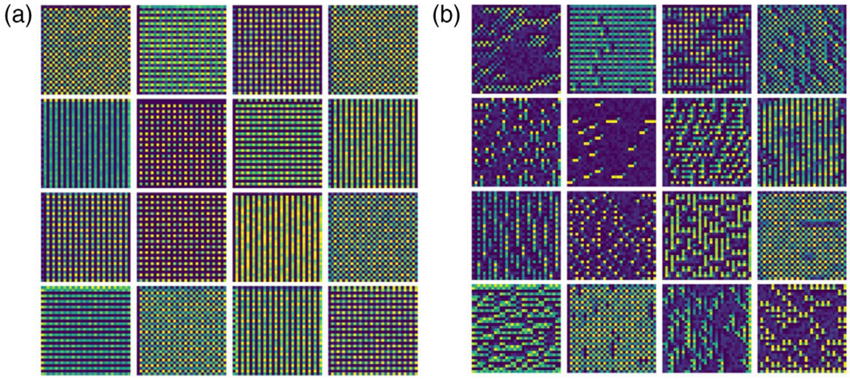

Figure 5. Maximum filter activation visualization. The patterns show the inputs that would maximize the 16 filter responses in: a) layer 3, the first convolutional layer in the first downsample block, and b) layer 5, the second Rectified Linear Unit (ReLU) layer in the first downsample block. The two layers are marked in Figure 1. The raw image is downsized by 2 in each dimension for both layers.

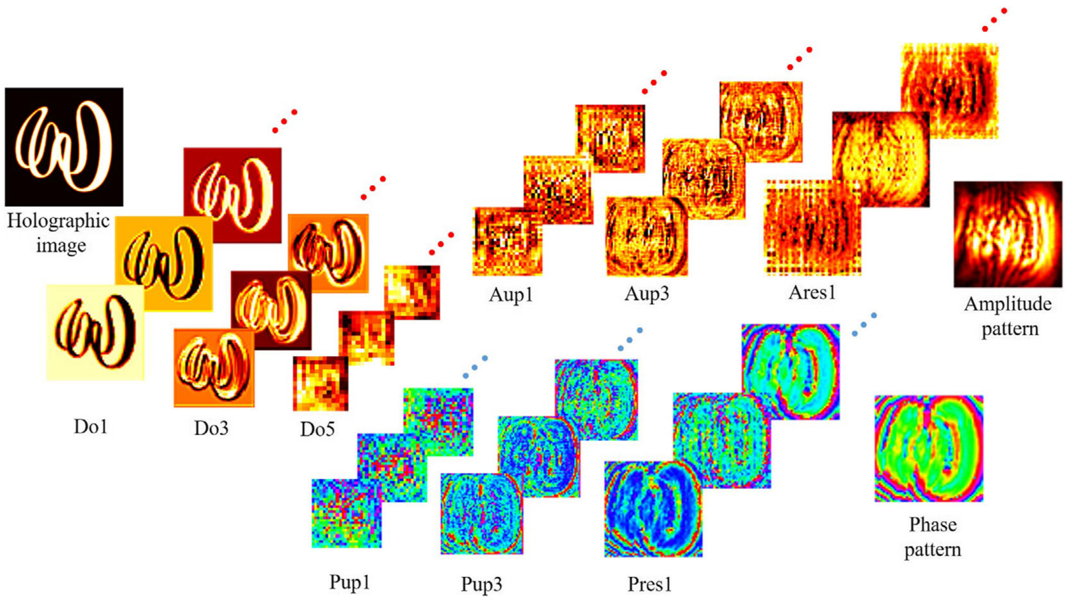

Figure 6. Intermediate activation map of NO when the logo of omega is inputted. Total number of channels is defined in the architecture in Figure 1. Only three random channels are displayed here. The chosen layer is the last convolution layer of a specific block. "Do" denotes downsample block. "Aup/Pup" represents the upsample block for amplitude and phase branch, while "Ares/Pres" means the residual block.

much more convenient than phase-only CGH, which needs a Fourier-transform lens. Though the experimental setup here contained other optical components, including LCP, RCP, and PBS, they can be replaced by a coating to achieve a handy modality. The NO tends to generalize the optical backpropagation process in free space from an object field to the CGH. The NO was trained under the Caltech101 database and had a test loss of 4.84% for the LFW database. The theoretical training results were well matched with experiments. Our NO can accelerate the inverse design process of the metasurface's pattern, from 450 s in a theoretically computing approach to ≈1 s enhanced by the prediction capacity. Although the time difference is negligible compared to the EBL process of nearly 10 h, the NO can readily be applied to a complex-modulated SLM to achieve a real-time design, which we will explore further. As for time complexity, it is reduced from $O(n^4)$ for theoretical computation to $O(n^2)$ for NO. When the number of metasurface reaches above 20, the total running time of NO is significantly reduced compared to other methods, e.g., gradient-based optimization, global genetic algorithm, and coupled-mode theory. Besides, the error can be lowered from more than 10% to only 1.38% for the selected testing image. Given that the complex-modulated metasurface has more sophisticated functions,[1] the proposed NO may assist in lensless image projection and real-time metasurface pattern design.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

H.M. and Y.Y. contributed equally to this work. The authors thank the supporting funding of the Research Grants Council of the Hong Kong Special Administrative Region of China (HKU 17205321, HKU 17200219, HKU 17209018, HKU C7047-16G, CityU T42-103/16-N) and the National Natural Science Foundation of China (NSFC) (N_HKU712/16).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

lensless modality, metasurfaces, neural networks

Received: March 23, 2022
Revised: August 10, 2022
Published online: August 28, 2022

[1] A. C. Overvig, S. Shrestha, S. C. Malek, M. Lu, A. Stein, C. Zheng, N. Yu, *Light: Sci. Appl.* **2019**, *8*, 92.
[2] G. Zheng, H. Mühlenbernd, M. Kenney, G. Li, T. Zentgraf, S. Zhang, *Nat. Nanotechnol.* **2015**, *10*, 308.
[3] J. Park, J.-H. Kang, S. J. Kim, X. Liu, M. L. Brongersma, *Nano Lett.* **2017**, *17*, 407.
[4] Y. Bao, Y. Yu, H. Xu, C. Guo, J. Li, S. Sun, Z.-K. Zhou, C.-W. Qiu, X.-H. Wang, *Light: Sci. Appl.* **2019**, *8*, 95.
[5] W. T. Chen, A. Y. Zhu, V. Sanjeev, M. Khorasaninejad, Z. Shi, E. Lee, F. Capasso, *Nat. Nanotechnol.* **2018**, *13*, 220.
[6] Z. Liu, D. Zhu, S. P. Rodrigues, K.-T. Lee, W. Cai, *Nano Lett.* **2018**, *18*, 6570.
[7] R. Lin, Y. Zhai, C. Xiong, X. Li, *Opt. Lett.* **2020**, *45*, 1362.
[8] W. Ma, Z. Liu, Z. A. Kudyshev, A. Boltasseva, W. Cai, Y. Liu, *Nat. Photonics* **2020**, *15*, 77.
[9] J. Jiang, J. A. Fan, *Nano Lett.* **2019**, *19*, 5366.
[10] A. Sinha, J. Lee, S. Li, G. Barbastathis, *Optica* **2017**, *4*, 1117.
[11] C. C. Nadell, B. Huang, J. M. Malof, W. J. Padilla, *Opt. Express* **2019**, *27*, 27523.
[12] X. Shi, T. Qiu, J. Wang, X. Zhao, S. Qu, *J. Phys. D: Appl. Phys.* **2020**, *53*, 275105.
[13] S. An, C. Fowler, B. Zheng, M. Y. Shalaginov, H. Tang, H. Li, L. Zhou, J. Ding, A. M. Agarwal, C. Rivero-Baleine, K. A. Richardson, T. Gu, J. Hu, H. Zhang, *ACS Photonics* **2019**, *6*, 3196.
[14] Z. Zalevsky, D. Mendlovic, R. G. Dorsch, *Opt. Lett.* **1996**, *21*, 842.
[15] J. W. Goodman, *Introduction to Fourier Optics*, Roberts and Company Publishers, Greenwood Village, CO **2005**.
[16] H. Zhang, N. Collings, J. Chen, B. Crossland, D. Chu, J. Xie, *Opt. Eng.* **2011**, *50*, 074003.
[17] X. Ding, F. Monticone, K. Zhang, L. Zhang, D. Gao, S. N. Burokur, A. de Lustrac, Q. Wu, C.-W. Qiu, A. Alù, *Adv. Mater.* **2015**, *27*, 1195.
[18] L. Novotny, B. Hecht, *Principles of Nano-Optics*, Cambridge University Press, Cambridge **2012**.
[19] W. Zhao, H. Jiang, B. Liu, J. Song, Y. Jiang, C. Tang, J. Li, *Sci. Rep.* **2016**, *6*, 30613.
[20] O. Stephen, U. J. Maduh, S. Ibrokhimov, K. L. Hui, A. A. Al-Absi, M. Sain, *2019 21st Int. Conf. on Advanced Communication Technology (ICACT)*, PyeongChang, Korea, **2019**.
[21] T. Falk, D. Mai, R. Bensch, Ö. Çiçek, A. Abdulkadir, Y. Marrakchi, A. Böhm, J. Deubner, Z. Jäckel, K. Seiwald, A. Dovzhenko, O. Tietz, C. Dal Bosco, S. Walsh, D. Saltukoglu, T. L. Tay, M. Prinz, K. Palme, M. Simons, I. Diester, T. Brox, O. Ronneberger, *Nat. Methods* **2019**, *16*, 67.
[22] K. He, X. Zhang, S. Ren, J. Sun, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA **2016**.
[23] O. Ronneberger, P. Fischer, T. Brox, *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany **2015**.
[24] S. Ioffe, C. Szegedy, *Proc. of the 32nd Int. Conf. on Machine Learning*, Lille, France, **2015**, p. 37.
[25] S. Mehta, C. Paunwala, B. Vaidya, *2019 Int. Conf. on Intelligent Computing and Control Systems (ICCS)*, Madurai, India, **2019**.
[26] J. Yamanaka, S. Kuwashima, T. Kurita, *Int. Conf. on Neural Information Processing*, Cham, Switzerland, **2017**.
[27] H. Mao, Y.-X. Ren, Y. Yu, Z. Yu, X. Sun, S. Zhang, K. K. Y. Wong, *Photonics Res* **2021**, *9*, 1689.
[28] F. Chollet, *Deep Learning with Python*, Manning, New York, **2018**.

