Deep-learning-aided metasurface design for megapixel acoustic hologram Ⓕ

View Online   Export Citation   CrossMark

Xuan-Bo Miao,[1,2] (iD) Hao-Wen Dong,[3,a)] (iD) Sheng-Dong Zhao,[4] (iD) Shi-Wang Fan,[5] (iD) Guoliang Huang,[2] (iD) Chen Shen,[6,a)] (iD) and Yue-Sheng Wang[1,7,a)] (iD)

AFFILIATIONS

[1]Department of Mechanics, School of Mechanical Engineering, Tianjin University, Tianjin 300350, People's Republic of China
[2]Department of Mechanical and Aerospace Engineering, University of Missouri, Columbia, Missouri 65211, USA
[3]Institute of Advanced Structure Technology, Beijing Institute of Technology, Beijing 100081, People's Republic of China
[4]School of Mathematics and Statistics, Qingdao University, Qingdao 266071, People's Republic of China
[5]Department of Engineering Mechanics, Shijiazhuang Tiedao University, Shijiazhuang 050043, People's Republic of China
[6]Department of Mechanical Engineering, Rowan University, Glassboro, New Jersey 08028, USA
[7]Department of Mechanics, Beijing Jiaotong University, Beijing 100044, People's Republic of China

[a)]Authors to whom correspondence should be addressed: hwdong@bit.edu.cn; shenc@rowan.edu; and yswang@tju.edu.cn

ABSTRACT

Unlike the holography technique using active sound source arrays, metasurface-based holography can avoid cumbersome circuitry and only needs a single transducer. However, a large number of individually designed elements with unique amplitude and phase modulation capabilities are often required to obtain a high-quality holographic image, which is a non-trivial task. In this paper, the deep-learning-aided inverse design of an acoustic metasurface-based hologram with millions of elements to reconstruct megapixel pictures is reported. To improve the imaging quality, an iterative compensation algorithm is proposed to remove the interference fringes and unclear details of the images. A megapixel image of Mona Lisa's portrait is reconstructed by a 2000×2000 metasurface-based hologram. Finally, the design is experimentally validated by a metasurface consisting 30×30 three-dimensional printed elements that can reproduce the eye part of Mona Lisa's portrait. It is shown that the sparse arrangement of the elements can produce high-quality images even when the metasurface has fewer elements than the targeted image pixels.

INTRODUCTION

Holography is a promising technology for wave field reconstruction. It has many applications such as volumetric displays,[1] ultra-high-density data storage,[2] optical[3] or acoustic[4] tweezers, optical[5] or acoustic imaging,[6] acoustic suspension,[6,7] and biomedical engineering,[8,9] etc. Acoustic holography, as one representative holographic technique, has attracted considerable attention from both scientists and engineers in many fields. Conventional acoustic holography uses active phased arrays, which require many transducers with cumbersome phase-shifting circuits. Recently, metasurfaces, a two-dimensional equivalent of metamaterials with subwavelength microstructures, have emerged as a promising holographic technique in both fields of optics[10] and acoustics.[11,12] Metasurfaces can reduce system complexity as only a single source transducer is needed to achieve a desired scattered field based on either phase modulation (PM) or phase-amplitude modulation (PAM).

The PM method was applied by Melde et al.[7] and Xie et al.[13] in 2016 to design metasurface-based holograms for acoustic holographic imaging, and later by Bakhtiari-Nejad et al.[14] for generating a multifocal pressure pattern with waterborne ultrasound. The PAM technique can be realized through an elaborate design of a composite element which includes two parts modulating phase and amplitude independently.[11,15] Tian et al.,[16] Zhu et al.,[17] and Fan et al.[18] used this strategy to design transmissive[16] and reflective[17,18] metasurface-based holograms for acoustic imaging. Although it offers more freedom by controlling both the amplitude and the phase, the design of a metasurface-based hologram with decoupled (or quasi-decoupled) PAM is subject to its microstructure configuration and size with possible lack of precision and flexibility. Differently, Zhang et al.[19] introduced a modified weighted Gerchberg–Saxton algorithm to design monolithic acoustic holograms with consideration of PAM. A nine foci pattern and an image of the

letter "U" were shown by a 2 MHz transducer attached to the holograms with the designed thickness profiles. Brown[20] used two phase holograms to modulate both the phase and amplitude of an incident acoustic wave and generated letters "UCL" experimentally. The hologram thickness profiles were designed using an optimization method. The PAM technique was also applied to design multiplexed acoustic metasurface-based holograms.[21,22] Generally, the PAM can provide a larger design space for metasurface-based holograms than the PM. However, one difficulty is that one must find a series of functional elements, which can modulate phase and amplitude simultaneously. This issue is more prominent when many elements are involved.

Although some studies on metasurface-based holography have been reported, the merit has not been fully exploited due to the limitation of resolution and PAM precision of the proposed metasurfaces. Intuitively, the number of metasurface elements (pixels), which defines the resolution of the metasurface, is crucial to obtain complex holographic imaging with high resolution. As an example, we demonstrate the holographic images (Fig. 1) of a Quick Response code (QR-code) containing 100 words [as shown in Fig. 1(a)] by the metasurfaces with different numbers of elements. The images are displayed as the acoustic pressure distributions and the detailed calculations are presented in Fig. S3 in the supplementary text S2. It is noted that the image [Fig. 1(f)] by the metasurface with 500×500 elements [Fig. 1(d)] is recognizable (i.e., the QR-code image can be scanned and then translated into the presupposed characters), while the image [Fig. 1(e)] by the 120×120 metasurface [Fig. 1(c)] is unrecognizable. As shown in Fig. 1(b), a larger number of metasurface elements is required for more words. That is to say, a high-quality metasurface-based hologram needs a huge number of elements with precise PAM. This may be understood by considering the following facts. On the one hand, a higher resolution is required for compiling more words in a binary picture (as displayed in Fig. S4 in the supplementary text S2). On the other hand, realizing the binary holography image with $N \times N$ pixels

means that the number of elements in the acoustic metasurface needs to be increased to render a more complex image. Therefore, when the number of metasurface elements increases, the number of resulting recognizable words by the metasurface-based holography also becomes larger as well. It is worth noting that while a positive correlation exists between the number of metasurface elements and the length of recognizable words, there is no explicit relation (e.g., a formula or a limit) that can be found based on the current data.

However, it is usually more challenging to have a large quantity of elements while in the meantime achieving great precision. Conventional approaches either cannot achieve this or need a very tedious design process. For instance, optimization methods have been successfully applied to the inverse design of acoustic metasurfaces,[23,24] but usually require a lot of computing resources and time costs if thousands of elements are designed. Recently, deep learning (DL), a kind of machine learning method based on artificial neural networks,[25] has been applied in the design of optical metasurfaces,[26–29] acoustic metasurfaces,[30–33] non-photonic devices,[34] and even holographic image generation.[35,36] However, data-driven models for inverse problems remain largely unexplored in terms of efficiency and capability, especially in the context of acoustic metasurface-based holography with megapixel resolution. For instance, simultaneous amplitude and phase control is rarely studied in DL-based models.[37] Other inverse design approaches also suffer from poor efficiency and cannot achieve the rapid design of numerous elements with target responses.

On the other hand, the full potential of metasurface-based holography has yet to be realized. Generally, the loss of information during reconstruction is mainly caused by the limited metasurface size, discretization of the acoustic field, and diffraction limitations. If the metasurface is not sufficiently large, the image quality may be suppressed because of incomplete information. When the holographic data are discretized, the holography image will inevitably become suboptimal
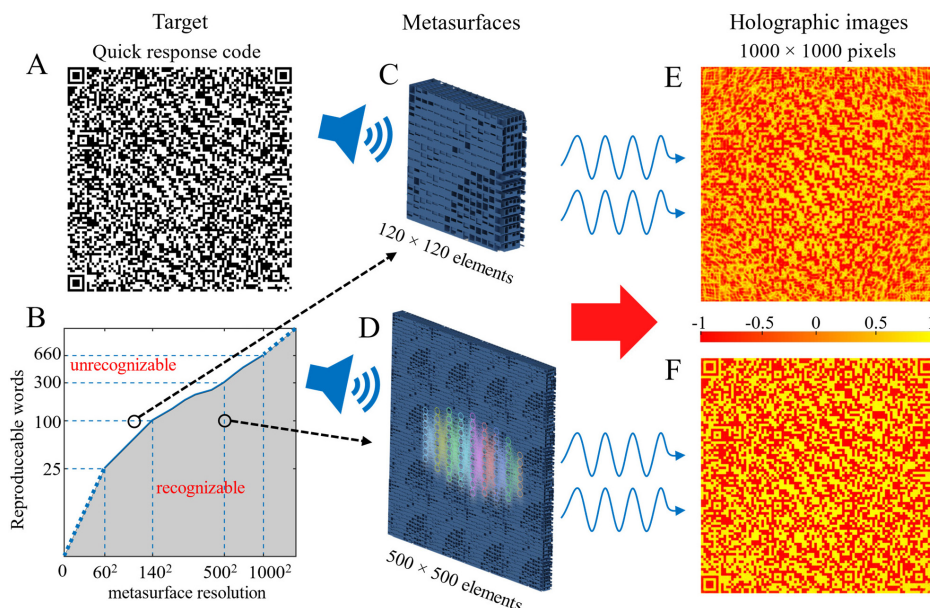
FIG. 1. Effects of the metasurface resolution on the hologram quality. (a) The target binary image of a QR-code containing information of 100 words. (b) Diagram illustrating the number of required elements to create a recognizable image as a function of reproducible words, demonstrating a positive correlation between metasurface elements and recognizable words. The 100-word target image cannot be created by a 120×120 metasurface but can be created by a 500×500 metasurface. Using the metasurface with different-scale elements, different numbers of recognizable words can be obtained through the holography image. (c), (d) Diagrams of the metasurfaces designed using the DL method with (c) 120×120 elements and (d) 500×500 elements. (e), (f) Corresponding holographic images from a metasurface with (c) unrecognizable and (d) recognizable resolutions, respectively. The details about the scanning device are shown in the supplementary text S2.

compared to that from continuous data. In addition, the intrinsic diffraction limitations can restrict the resulting imaging resolution. The ability to rapidly construct unit cells with unique PAM and compensate for the information loss during metasurface design is desired to promote the application of metasurface-based holograms.

This paper proposes a DL-based high-definition acoustic metasurface optimization method to achieve high-quality holographic reconstructions. The capacity of the approach is manifested by a megapixel image where it is shown the deep neural network (DNN) method combined with the genetic algorithm (GA) can accelerate the convergence of the inverse design of elements. This enables one to design millions of highly customized elements with high precision of PAM in a few hours. The element topology exhibits a bi-anisotropic feature that can address the requirement of fully controllable phase and amplitude.[38] An iterative algorithm of the phase-amplitude distribution optimization is developed to make more efficient use of millions of bi-anisotropic elements. Several demonstrations are presented to show how DL aids the rapid custom designs of acoustic metasurface-based holograms. In addition to the aforementioned holograms for QR-code images to show how the number of pixels affects the hologram performance, designs for imaging of Tianjin University logo are also presented to discuss proper spatial relationship between the hologram and image (incl. focal-length, and metasurface size and resolution). Finally, a 2000×2000 metasurface-based hologram for the high-quality megapixel image of Mona Lisa's portrait is designed by the proposed method and the eye part of Mona Lisa's portrait reconstructed by a three-dimensional (3D) printed metasurface is experimentally demonstrated.

RESULTS

Bi-anisotropic element topology

To construct a metasurface-based hologram, we should first develop an element topology that has good potential to realize PAM with the required acoustic properties (i.e., transmitted phase and amplitude). The proposed 3D asymmetric cuboid element topology with a length of 10.08 mm and a square cross-section of 2.88×2.88 mm^2 is shown schematically in Fig. 2(a). The element consists of four Helmholtz resonators with two inlet baffle plates and two outlet baffle plates and is separated with the plates in the front and back sides. The inner structural geometry of the element is described by 11 pending parameters, $l_1 \sim l_{11}$ [see Fig. 2(b)], which will be optimized in their allowable value ranges (Table S1) to meet the requirement of the hologram. Apart from these 11 parameters, the other geometric parameters are fixed as shown in Fig. 2(a). Theoretically, it is possible to use more geometric parameters and a larger design space (e.g., topology optimization), which can potentially provide more abundant topologies. This, however, will inevitably increase the computational load and lower the overall efficiency of the design process. As will be shown in the following, the choice of these 11 parameters provides sufficient freedom for the optimization of the microstructures by integrating the machine learning approach. All geometric dimensions are carefully selected to make them suitable for the stack laser 3D-printer. A 3D printed sample made of epoxy resin is shown in Fig. 2(c), which shows the good quality of fabrication. Our analysis shows that the thermal-viscous loss has a very slight influence on the sound propagation, especially at the considered frequency of 12.5 kHz, as given in Fig. S1.
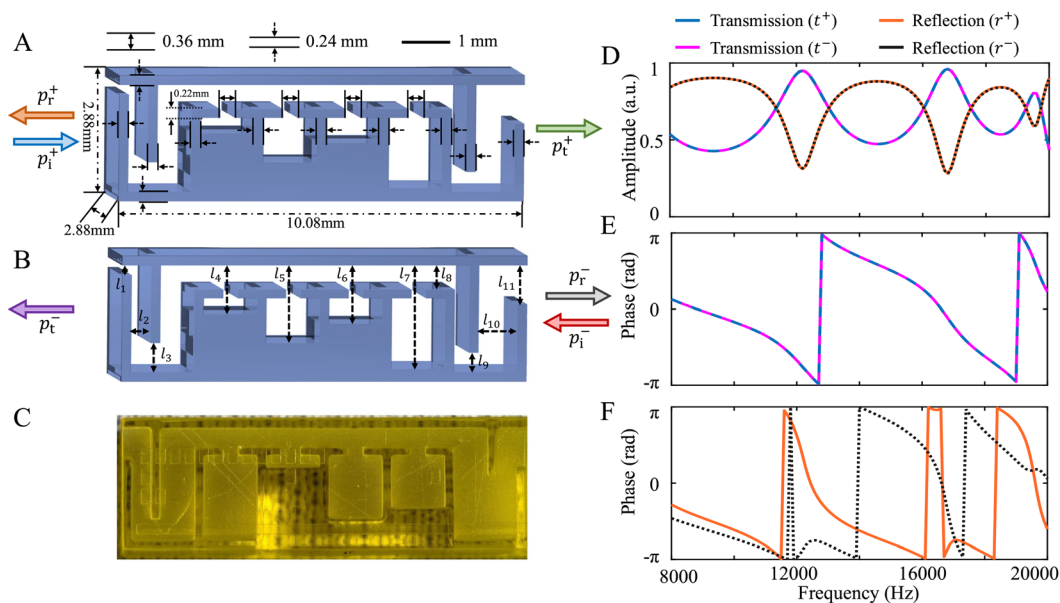
FIG. 2. Schematics of the bi-anisotropy element and its acoustic properties. (a), (b) Topological definition of the element, with fixed parameters shown in (a) and adjustable parameters $l_1 \sim l_{11}$ shown in (b). (c) Photograph of a 3D-printed element sample. (d)–(f) Bi-anisotropic wave properties of a typical element in the frequency range of 8–20 kHz. The transmission and reflection coefficients (d), transmissive phase variation (e), and reflective phase variation (f) from both sides of the element are illustrated, respectively. The reflection and transmission coefficients are the same for forward or backward incidence, i.e., the $= t^-$, $r^+ = r^-$, and the transmissive phase shifts are also the same. Different reflective phase shifts are observed from the structure, demonstrating the bi-anisotropic properties.

The designed element possesses bi-anisotropic property, which has been proven to be helpful to effectively control both amplitude and phase.[38] As shown in Figs. 2(a) and 2(b), we consider the response of the element subject to forward [Fig. 2(a)] and backward [Fig. 2(b)] incident waves. The following scattering relation can be obtained:[38]

$$\begin{bmatrix} p_s^+ \\ p_s^- \end{bmatrix} = \begin{bmatrix} r^+ & t^- \\ t^+ & r^- \end{bmatrix} \begin{bmatrix} p_i^+ \\ p_i^- \end{bmatrix}, \tag{1}$$

where $p_s^\pm = p_r^\pm + p_t^\mp$ represent the amplitudes of the scattered wavefields at both sides; $p_r^\pm$, $p_t^\pm$, and $p_i^\pm$ represent the amplitudes of the reflected, transmitted, and incident waves in forward (+) and backward (−) directions, respectively; and $r^\pm$ and $t^\pm$ are the reflection and transmission coefficients, respectively. Figures 2(d) and 2(f) illustrate the amplitudes and phases of $r^\pm$ and $t^\pm$ varying with frequency for a particular element with the detailed geometric parameters presented in the supplementary text S1. It can be seen that only the phases of the reflected waves (r^+ and r^-) are different, while the other properties are the same. This reflection phase asymmetry is a clear sign of the bi-anisotropy phenomenon.[38] As shown in Fig. S2 in the supplementary text S1, the energy conservation holds true for the present passive system.

Notably, the proposed element shows a strong bi-anisotropic response in the range of 8–20 kHz. In this case, bi-anisotropic structures are adopted to broaden the design space of our proposed metasurface by enabling more geometric parameters. Indeed, it has been shown that even for a purely transmissive design, bi-anisotropy can enhance multiple scattering and interaction among different components within the unit cell, which together lead to improved performance.[23]

Training DNN for acoustic property prediction of elements

The design of elements with the required phase and amplitude is crucial to constructing a metasurface hologram. As demonstrated previously (Fig. 1), tens of thousands or even millions of elements are needed to obtain a high-quality hologram with megapixels. This is obviously difficult to be implemented by the traditional optimization method[23] because of limited computational efficiency and large time cost. In this paper, we propose a design strategy for individual elements by combining the DNN with GA. The DNN is a generalized data-driven method that is used to realize accurate prediction[39] or feature extraction.[40] We first develop a DNN with the ability to reproduce the relationship between the element's geometry and its acoustic properties. The DNN model takes $\boldsymbol{l} = (l_1, \ldots, l_{11})$ as input and the amplitude (y_{p^+}) and phase (y_{θ^+}) of the forward transmission coefficient as output, as shown in Fig. 3(a). The relationship between the input and output can be expressed as $[y_{p^+}, y_{\theta^+}] = \mathcal{D}(\boldsymbol{l})$, where $\mathcal{D}$ represents the predicting method of DNN. The neural network contains seven hidden layers (layers 1–7), as shown in Fig. 3(a). Layers 0–8 contain 11, 400, 750, 750, 1000, 750, 750, 300, and 2 neurons, respectively. The ReLU activation function is employed in layers 1–4 to accelerate convergence, and the softplus activation function is employed in layers 5–7 to increase the performance of the nonlinear regression.

The dataset is generated by randomly selecting 11 parameters $l_1 \sim l_{11}$ in their allowable value ranges and then calculating the training-required characteristic outputs (i.e., the amplitudes and phases) by the finite element method (FEM). It is noticed that the selection of $l_4 \sim l_7$ is based on the values of l_8, and therefore $l_1 \sim l_3$ and $l_8 \sim l_{11}$ should be first identified.
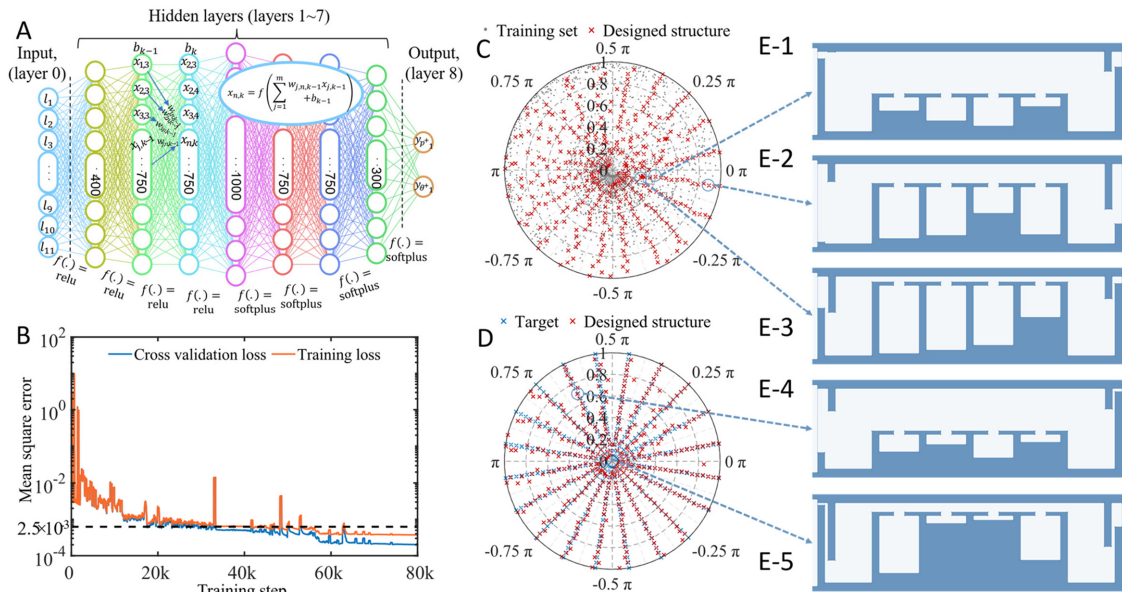
FIG. 3. Illustration of the neural network. (a) DNN structure map, including the size and activation function of each layer. (b) The gradient-descent process and model performance of the employed DNN. (c), (d) The training and prediction sets for the phase and amplitude values. The radial and angular coordinates of the polar plot correspond to the normalized transmission amplitude and phase of the element response, respectively. A batch of training set and a batch of tested structure (c), demonstration of DNN predictive ability (d), and representative elements (e) inversely designed by the DNN-based GA are shown as well.

The dataset contains 4000 individuals, which are divided into ten batches with 400 individuals per batch, one batch of them is shown in Fig. 3(c). Among the ten batches, eight are randomly selected as the training sets, one as the cross-validation set; and the rest one is for estimating the DNN. The gradient-decent process shown in Fig. 3(b) demonstrates that the DNN achieved a mean squared error of approximately 0.000376 on the training set and 0.000196 on the cross-validation set. These results imply that the DNN can accurately predict output values for both datasets without overfitting the training data, indicating that the DNN-based GA can predict structures that are not present in the training set in any desired responses, as shown in Figs. 3(c) and 3(d). The mean relative error of the DNN can reach over 97% for both y_{p^+} and y_{θ^+} in the frequency range from 8 kHz to 35 kHz.

The DNN can predict the properties of 1×10^6 elements in 0.23 s. The prediction speed of the DNN method is seven orders of magnitude faster than that of the FEM method (about 30 seconds per element), confirming the efficiency of the DL model. The proposed method exhibits good scalability and computational efficiency, which are key factors in addressing the challenges of metasurface design. By leveraging deep learning's capabilities in handling high-dimensional data and complex relationships, along with techniques such as parallelization, model compression, and transfer learning, the scalability of the approach is enhanced.

Inverse design of elements by DNN-based genetic algorithm

Although DL is fast in predicting the acoustic properties of elements (a forward problem), it may encounter difficulties of convergence with inverse problems. In order to accelerate convergence and obtain optimal results, the inverse design of elements for the prescribed amplitudes and phases is formulated as an optimization problem which is tackled by using GA. It is known that a larger population size can generate an optimized solution more easily but is time-consuming. The traditional FEM-based GA is generally unable to afford such a huge time cost. Thanks to the extreme calculation speed of DNN, GA combined with the above DNN model will be developed to accelerate the design progress of a huge number of elements.

To realize the targeted acoustic wave responses of an element, the Euclidean distance is utilized to characterize the difference between the desired and actual responses. Therefore, the objective function of the optimization problem is formulated as

$$
\begin{aligned}
Maximize: \ &\psi(\Omega) \\
&= 1 - \frac{\sqrt{(r^* \cos \theta^* - r_\Omega \cos \theta_\Omega)^2 + (r^* \sin \theta^* - r_\Omega \sin \theta_\Omega)^2}}{2},
\end{aligned} \quad (2)
$$

where ψ denotes the objective function of a metasurface element in optimization; Ω represents the design domain, i.e., an arbitrary metasurface element with eleven typical topological parameters ($l_1 \sim l_{11}$); θ_Ω and r_Ω are the phase shift and amplitude of an element involved in the current generation, respectively; θ^* and r^* are the corresponding targeted responses. Note that θ_Ω and r_Ω are retrieved by the DNN directly instead of the FEM. When the Euclidean distance varies from 2 to 0, the actual response becomes more and more similar to the targeted response. After a certain number of generations, the algorithm generates a near-optimal metasurface element with the optimized

$l_1 \sim l_{11}$ for the wave response of (θ^*, r^*) on demand. The parameters of GA are set as the population size $N_p = 300$, the maximum number of generations $E_n = 500$, the crossover probability $P_c = 0.1$, and the mutation probability $P_m = 0.05$.

The DNN-based GA with a large population size has an extremely high speed with sufficient accuracy and thus can realize the online inverse design of elements on demand. The present design method can inversely generate large-scale metasurface with millions of elements very quickly (i.e., $\sim$25 min for 1×10^6 elements and 1.5 ms for one element) in contrast to the traditional time-consuming forward simulations and optimization process. In this way, it could be used as a real-time, rapid-feedback approach to design the elements. The design of 1×10^6 elements all exhibiting unique phase and amplitude combinations costs about 25 min. As shown in Fig. 3(c), the results predicted by the DNN-based GA (red crosses) can reach those areas that the training set (gray dots) does not cover. Some of those areas are marked in the blue circles. The DNN-based GA is powerful to predict various elements to fit the requirement of the perfect PAM holography, as shown in Fig. 3(d). Some representatives inversely designed elements are shown in Fig. 3(e).

Iterative compensation algorithm for improving holographic images

To improve the quality of the holographic image, an iterative compensation algorithm is further developed to diminish the deviation of the reconstructed holographic image from the target image as far as possible.

According to Huygens–Fresnel principle, each point on the outlet plane of the metasurface can be considered as a secondary source; and all waves generated by these secondary sources are superimposed through interference to form a holographic image on a plane. For further details and illustration, as shown in the supplementary text S5 and Fig. S9. Let us consider the case of acoustic waves with an angular frequency of ω and omit the time-harmonic term $e^{-i\omega t}$. For a point acoustic source placed at $\boldsymbol{x}_p$ with unit intensity, the generated acoustic wave field at $\boldsymbol{x}_s$ can be expressed as[17]

$$
\varphi(\boldsymbol{x}_s) = \frac{1}{|\boldsymbol{x}_p - \boldsymbol{x}_s|} e^{ik|\boldsymbol{x}_p - \boldsymbol{x}_s|}, \quad (3)
$$

where $k = 2\pi/\lambda = 2\pi c/f$ with $\lambda = c/f$ and $f = 2\pi\omega$ being the wavelength and frequency, respectively. When trying to reproduce this point source, a PAM sound source array should generate acoustic wave satisfying $\varphi(\boldsymbol{x}_s)$. Therefore, to reconstruct an image using an metasurface-based acoustic hologram, each element should generate $\varphi_e(\boldsymbol{x}_e)$ as the output by treating them as a point source,

$$
\varphi_e(\boldsymbol{x}_e) = \int_T \frac{A(\boldsymbol{x}_t)}{|\boldsymbol{x}_t - \boldsymbol{x}_s|} e^{ik|\boldsymbol{x}_t - \boldsymbol{x}_e|} \, d\boldsymbol{x}_t, \quad (4)
$$

where $\varphi_e(\boldsymbol{x}_e)$ is the complex pressure at the element position $\boldsymbol{x}_e$ and its amplitude $|\varphi_e|$ is not necessarily equal to 1; $\boldsymbol{x}_e$ is the 3D position vector of the element; and $A(\boldsymbol{x}_t)$ is the gray value of the pixel at $\boldsymbol{x}_t$ on the target image T and is normalized from 0 to 1. After integrating $\boldsymbol{x}_t$ over all pixels on the target image T, we can obtain $\varphi_e(\boldsymbol{x}_e)$. The normalized amplitude $\alpha(\boldsymbol{x}_e)$ and phase $\theta(\boldsymbol{x}_e)$ generated by the metasurface elements are then obtained as

$$\alpha(\boldsymbol{x}_{\mathrm{e}}) = \frac{|\varphi_{\mathrm{e}}(\boldsymbol{x}_{\mathrm{e}})|}{\max\limits_{\boldsymbol{x}_{\mathrm{e}} \in T}\left(|\varphi_{\mathrm{e}}(\boldsymbol{x}_{\mathrm{e}})|\right)}, \tag{5}$$

$$\theta(\boldsymbol{x}_{\mathrm{e}}) = \arctan\left(\frac{\mathrm{Im}(\varphi_{\mathrm{e}}(\boldsymbol{x}_{\mathrm{e}}))}{\mathrm{Re}(\varphi_{\mathrm{e}}(\boldsymbol{x}_{\mathrm{e}}))}\right). \tag{6}$$

As the metasurface is ensonified by an incident wave, a pressure with amplitude $\boldsymbol{\alpha}(\boldsymbol{x}_e)$ and phase $\theta(\boldsymbol{x}_e)$ is generated at the outlet of element e in the metasurface, and the sound pressure field $\varphi_h(\boldsymbol{x}_h)$ in the holographic image plane H can be expressed as[17]

$$\varphi_{\mathrm{h}}(\boldsymbol{x}_{\mathrm{h}}) = \int\limits_{M} \frac{\alpha(\boldsymbol{x}_{\mathrm{e}})}{|\boldsymbol{x}_{\mathrm{h}} - \boldsymbol{x}_{\mathrm{e}}|} \mathrm{e}^{\mathrm{i}\left(k|\boldsymbol{x}_{\mathrm{h}} - \boldsymbol{x}_{\mathrm{e}}| - \theta(\boldsymbol{x}_{\mathrm{e}})\right)} \mathrm{d}\boldsymbol{x}_{\mathrm{e}}. \tag{7}$$

It is known that $|\varphi_{\mathrm{h}}(\boldsymbol{x}_{\mathrm{h}})|$ is the amplitude of the sound pressure at the reconstructed pixel $\boldsymbol{x}_{\mathrm{h}}$ of the holographic image. Its normalized value (denoted as $\overline{|\varphi_{\mathrm{h}}|}$) can be displayed as a gray-white image (with $\overline{|\varphi_{\mathrm{h}}|}$ = 0 to 1 corresponding to black to white). For an ideal lossless holographic image, $\overline{|\varphi_{\mathrm{h}}|}$ should be identical to $A(\boldsymbol{x}_{\mathrm{t}})$. In practice, however, the metasurface is finite and only placed on one side of the image. Furthermore, the metasurface is discrete with finite elements. All these facts will cause missing of the information, resulting in the deviation of the reconstructed holographic image from the target image. To improve the quality of the holographic image, we propose the following iterative compensation algorithm to compensate for the missed information.

To perform the compensation algorithm, the following matrices are introduced: $\boldsymbol{A}_{\mathrm{T}}$, a real matrix whose elements are the normalized gray values at all pixels of the target image T; $\boldsymbol{\Phi}_{\mathrm{M}}$, a complex matrix whose elements are the sound pressure at all element outlets of the metasurface-based hologram; $\boldsymbol{\Phi}_{\mathrm{H}}$, a complex matrix whose elements are the sound pressure at all pixels of the holographic image H; $\overline{\boldsymbol{\Phi}}_{\mathrm{H}}$, a real matrix whose elements are the normalized sound pressure amplitudes at all pixels of the holographic image H (i.e., the absolute values of the elements of $\boldsymbol{\Phi}_{\mathrm{H}}$); and a system error matrix $\boldsymbol{\Delta} = \boldsymbol{A}_{\mathrm{T}} - \overline{\boldsymbol{\Phi}}_{\mathrm{H}}$ (real), which represents the deviation of the reconstructed holographic image from the target image. Furthermore, we introduce two mappings,

$$\boldsymbol{\Phi}_{\mathrm{M}} = g(\boldsymbol{A}_{\mathrm{T}}), \tag{8}$$

$$\boldsymbol{\Phi}_{\mathrm{H}} = h(\boldsymbol{\Phi}_{\mathrm{M}}), \tag{9}$$

where the first one is obtained from Eq. (4) and is the mapping from the target image to a metasurface-based hologram; the second one, obtained from Eq. (7), is the mapping from the hologram to the holographic image, as illustrated in Fig. 4. The iterative compensation process may be summarized as follows (refer to Fig. 4 for the basic concept):

(1) At the beginning (the 0th-iteration), $\boldsymbol{\Phi}_{\mathrm{H}} = \boldsymbol{0}$, and thus $\boldsymbol{\Delta}^{(0)} = \boldsymbol{A}_{\mathrm{T}}$. Then, from Eqs. (8) and (9), we can calculate the 1st-iteration values: $\boldsymbol{\Phi}_{\mathrm{M}}^{(1)} = g(\boldsymbol{A}_{\mathrm{T}})$, $\boldsymbol{\Phi}_{\mathrm{H}}^{(1)} = h(\boldsymbol{\Phi}_{\mathrm{M}}^{(1)}) = h(g(\boldsymbol{A}_{\mathrm{T}}))$, and

$$\boldsymbol{\Delta}^{(1)} = \boldsymbol{A}_{\mathrm{T}} - \overline{\boldsymbol{\Phi}}_{\mathrm{H}}^{(1)}. \tag{10}$$

(2) For the 2nd-iteration, replacing $\boldsymbol{A}_{\mathrm{T}}$ with $\boldsymbol{\Delta}^{(1)}$, and then we obtain the compensations of $\boldsymbol{\Phi}_{\mathrm{M}}^{(1)}$ and $\boldsymbol{\Phi}_{\mathrm{H}}^{(1)}$: $g(\boldsymbol{\Delta}^{(1)})$ and $h(g(\boldsymbol{\Delta}^{(1)}))$. Therefore, we have the 2nd-iteration values,
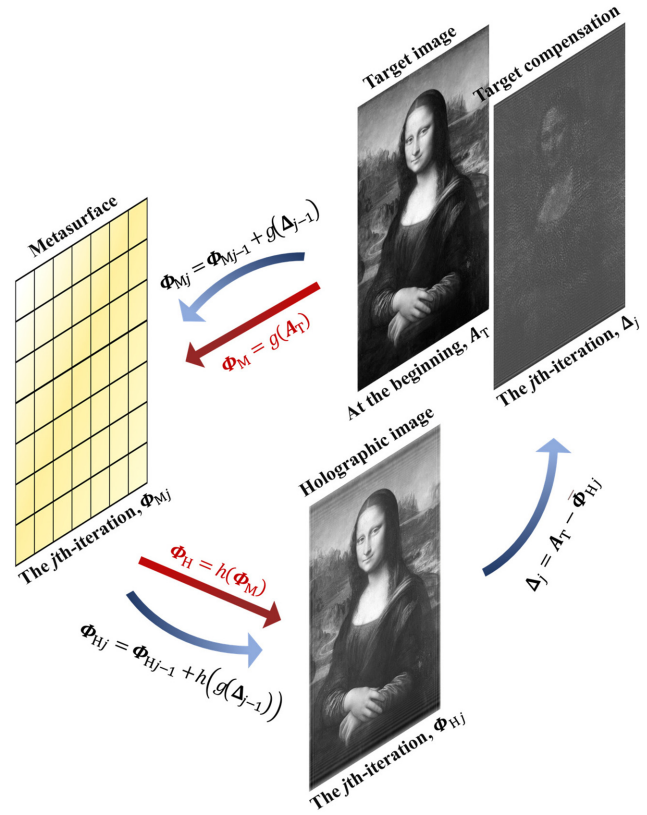

FIG. 4. Schematic illustration of iterative compensation algorithm and mapping relations between metasurface. The algorithm initially obtain metasurface $\boldsymbol{\Phi}_{\mathrm{M}}^{(1)}$ from the target image $\boldsymbol{A}_{\mathrm{T}}$ and calculates holographic image $\boldsymbol{\Phi}_{\mathrm{H}}^{(1)}$ using acquired $\boldsymbol{\Phi}_{\mathrm{M}}^{(1)}$ (following the straight red arrows). The algorithm then computes the difference as $\boldsymbol{\Delta}^{(1)} = \boldsymbol{A}_{\mathrm{T}} - \overline{\boldsymbol{\Phi}}_{\mathrm{H}}^{(1)}$. Subsequently, the algorithm replaces the target $\boldsymbol{A}_{\mathrm{T}}$ with the target compensation $\boldsymbol{\Delta}^{(j)}$ and continues iterating among $\boldsymbol{\Delta}^{(j)}$, $\overline{\boldsymbol{\Phi}}_{\mathrm{M}}^{(j)}$, and $\overline{\boldsymbol{\Phi}}_{\mathrm{H}}^{(j)}$ by following the curved blue arrows. The straight arrows represent the mappings, while the curved arrows indicate the iterative relations between the different stages. Figure credit: "Mona Lisa" by Leonardo da Vinci is in the public domain.

$$\boldsymbol{\Phi}_{\mathrm{M}}^{(2)} = \boldsymbol{\Phi}_{\mathrm{M}}^{(1)} + g\left(\boldsymbol{\Delta}^{(1)}\right) = g\left(\boldsymbol{\Delta}^{(0)}\right) + g\left(\boldsymbol{\Delta}^{(1)}\right), \tag{11}$$

$$\boldsymbol{\Phi}_{\mathrm{H}}^{(2)} = \boldsymbol{\Phi}_{\mathrm{H}}^{(1)} + h\left(g\left(\boldsymbol{\Delta}^{(1)}\right)\right) = h\left(g\left(\boldsymbol{\Delta}^{(0)}\right)\right) + h\left(g\left(\boldsymbol{\Delta}^{(1)}\right)\right), \tag{12}$$

$$\boldsymbol{\Delta}^{(1)} = \boldsymbol{A}_{\mathrm{T}} - \overline{\boldsymbol{\Phi}}_{\mathrm{H}}^{(1)}, \tag{13}$$

where we have denoted $\boldsymbol{\Delta}^{(0)} = \boldsymbol{A}_{\mathrm{T}}$.

(3) Following the above procedure, we can obtain the jth-iteration values as

$$\boldsymbol{\Phi}_{\mathrm{M}}^{(j)} = \boldsymbol{\Phi}_{\mathrm{M}}^{(j-1)} + g\left(\boldsymbol{\Delta}^{(j-1)}\right) = g\left(\boldsymbol{\Delta}^{(0)}\right) + g\left(\boldsymbol{\Delta}^{(1)}\right) + \cdots + g\left(\boldsymbol{\Delta}^{(j-1)}\right)$$
$$= \sum_{i=0}^{j-1} g\left(\boldsymbol{\Delta}^{(i)}\right), \tag{14}$$

$$\begin{aligned}
\boldsymbol{\Phi}_{\mathrm{H}}^{(j)} &= \boldsymbol{\Phi}_{\mathrm{H}}^{(j-1)} + h\left(g\left(\boldsymbol{\Delta}^{(j-1)}\right)\right) \\
&= h\left(g\left(\boldsymbol{\Delta}^{(0)}\right)\right) + h\left(g\left(\boldsymbol{\Delta}^{(1)}\right)\right) + \cdots + h\left(g\left(\boldsymbol{\Delta}^{(j-1)}\right)\right) \\
&= \sum_{i=0}^{j-1} h\left(g\left(\boldsymbol{\Delta}^{(i)}\right)\right),
\end{aligned} \tag{15}$$

$$\boldsymbol{\Delta}^{(j)} = \boldsymbol{A}_{\mathrm{T}} - \overline{\boldsymbol{\Phi}}_{\mathrm{H}}^{(j)}. \tag{16}$$

It is noted that $\boldsymbol{\Phi}_{\mathrm{M}}^{(\infty)} = \boldsymbol{A}_{\mathrm{T}}$.

Since each iteration is based on the finite discrete metasurface arranged on one side of the image and will not yield an exact solution, the oscillation near the optimal solution will appear in the iteration process. Therefore, a diminishing compensation rate is introduced into the algorithm to avoid oscillation. Equations (15) and (16) are, hence, replaced by

$$\hat{\boldsymbol{\Phi}}_{\mathrm{M}}^{(j)} = \sum_{i=0}^{j-1} \mu^{(i)} g\left(\boldsymbol{\Delta}^{(i)}\right), \tag{17}$$

$$\hat{\boldsymbol{\Phi}}_{\mathrm{H}}^{(j)} = \sum_{i=0}^{j-1} \mu^{(i)} h\left(g\left(\boldsymbol{\Delta}^{(i)}\right)\right), \tag{18}$$

where $\mu^{(i)}$ is the diminishing compensation rate of the ith iteration, it is similar to learning rates in machine learning and balances convergence speed and stability. In general, a smaller compensation rate is beneficial for higher-resolution metasurfaces. In the present algorithm, the compensation rates are selected based on extensive simulation tests, which are set as 0.5–0.3 for the first three iterations, and then are reduced to 0.2–0.05 for the subsequent iterations.

Inverse-designed metasurface-based holograms

To demonstrate the effectiveness of the proposed method, we next present some examples. The entire process of designing a high-quality metasurface-based hologram is summarized as follows:

(1) Train a DNN model to estimate the elements' properties (i.e., transmitted phases and amplitudes) at a designated frequency.
(2) Employ the iterative compensation method to determine and optimize the phase and amplitude distributions of the element array of the acoustic metasurface-based hologram according to the target image.
(3) Inversely design the metasurface with customized elements obtained by DNN-based GA.

To characterize the holographic image quality, we define the following parameter $\mathcal{F}$ to estimate the fitness between the target image and the designed sound pressure distribution

$$\mathcal{F} = \int_T \left(1 - \frac{\left|A(\boldsymbol{x}_{\mathrm{t}}) - \overline{|\varphi_h(\boldsymbol{x}_{\mathrm{h}})|}\right|}{A(\boldsymbol{x}_{\mathrm{t}})}\right) \mathrm{d}\boldsymbol{x}_{\mathrm{t}}, \tag{19}$$

where $\boldsymbol{x}_{\mathrm{h}} = \boldsymbol{x}_{\mathrm{t}}$ as the holographic image and target image share the same spatial region. The integral is performed over the target image area T (the same as the holographic image area H). In calculation, both A and $\overline{|\varphi_{\mathrm{h}}|}$ are, respectively, normalized by their own mean values in the entire pressure field.

Proper spatial relationship between hologram and image

In addition to the meticulous design of elements to obtain precise modulation of phases and amplitudes, the spatial relationship between the metasurface and image also has significant effects on the quality of holographic imaging. Such spatial relations include the sizes of the metasurface and image, the distance between them (i.e., focal-length F_0), and the elements per unit length R (i.e., pixel density measured by pixels per meter, PPM). In addition, the wavelength is also an important parameter to be considered. The hologram performance is closely tied to these spatial parameters and therefore they need to be carefully selected within specific ranges and follow certain rules. Herein, we will discuss this issue by considering the imaging of Tianjin University logo. Since the speed of sound in air is approximately 343 m/s at room temperature, the corresponding wavelength is 27.4 mm when the operating frequency is selected as 12.5 kHz. Clearly, the wavelength is much larger than the size of an element [Fig. 2(a)]. Here, the frequency is chosen so that it ensures a good spatial resolution while keeping the thermal-viscous loss small. The metasurface is constructed by inserting the elements uniformly and discretized to form a square lattice while other portions are filled with soundproof baffles. The number of the elements (i.e., pixels) is assumed to be $N \times N = 250 \times 250$, which is the same as the pixels of the target image. Both the metasurface and image are square with the same side length (S).

Figure 5(a) demonstrates images of the logo obtained after 10 iterations of the compensation algorithm by the customized metasurfaces with different sizes and focal lengths. The background color in the figure represents the fitness of the images. The abscissa (K) in the figure is the ratio of side-length S to focal-length F_0. The equal side-length curves are hyperbolas given by $S = K \cdot F_0 = const.$ The metasurface pixel-density is $R = N/S = N/(K \cdot F_0)$, thus the above hyperbolas also define the equal pixel-density curves. From the upper-left to the lower-right in the figure, the side-length increases as the pixel-density decreases. This is also shown in Fig. 5(b). The curve in either Fig. 5(a) or Fig. 5(b) is the equal side-length curve with $S = K \cdot F_0 = \lambda N$ or the equal pixel-density curve with $R = 1/\lambda$. The images near this curve have high fitness, showing a preferred area of spatial parameters for the best results. The one as shown in Fig. 5(c) with $K = 1.2$ and $F_0 = 6$ m has the highest fitness of 9.345. The fitness becomes lower as the image is farther from this curve. Specifically, the images below this curve with large S ($S > \lambda N$) or small R ($R < 1/\lambda$) have a lot of blurred lines and interference patterns and are of very low definition. More iterations of compensation are necessary to improve the quality of these images. Fortunately, most images above this curve have satisfactory legibility.

Figure 5 also indicates the best choices regarding metasurface size and focal length for a fixed resolution ($N \times N = 250 \times 250$). To show the influence of the metasurface resolution ($N \times N$) on the image quality, we present the results in Fig. 6 with different values of N while fixing the focal-length $F_0 = 11$ m and metasurface side-length $S = 11$ m (the same as the side-length of image). The target image has 500×500 pixels. The improvement of the image quality by the iterative compensation algorithm is also demonstrated in Fig. 6(a) which illustrates the fitness improvement with the iterative order for the cases of $N \times N = 350 \times 350$, 500×500, 650×650, and 800×800. It is shown that more elements of the metasurface can achieve higher fitness. The fitness can be improved gradually by the iterative compensation
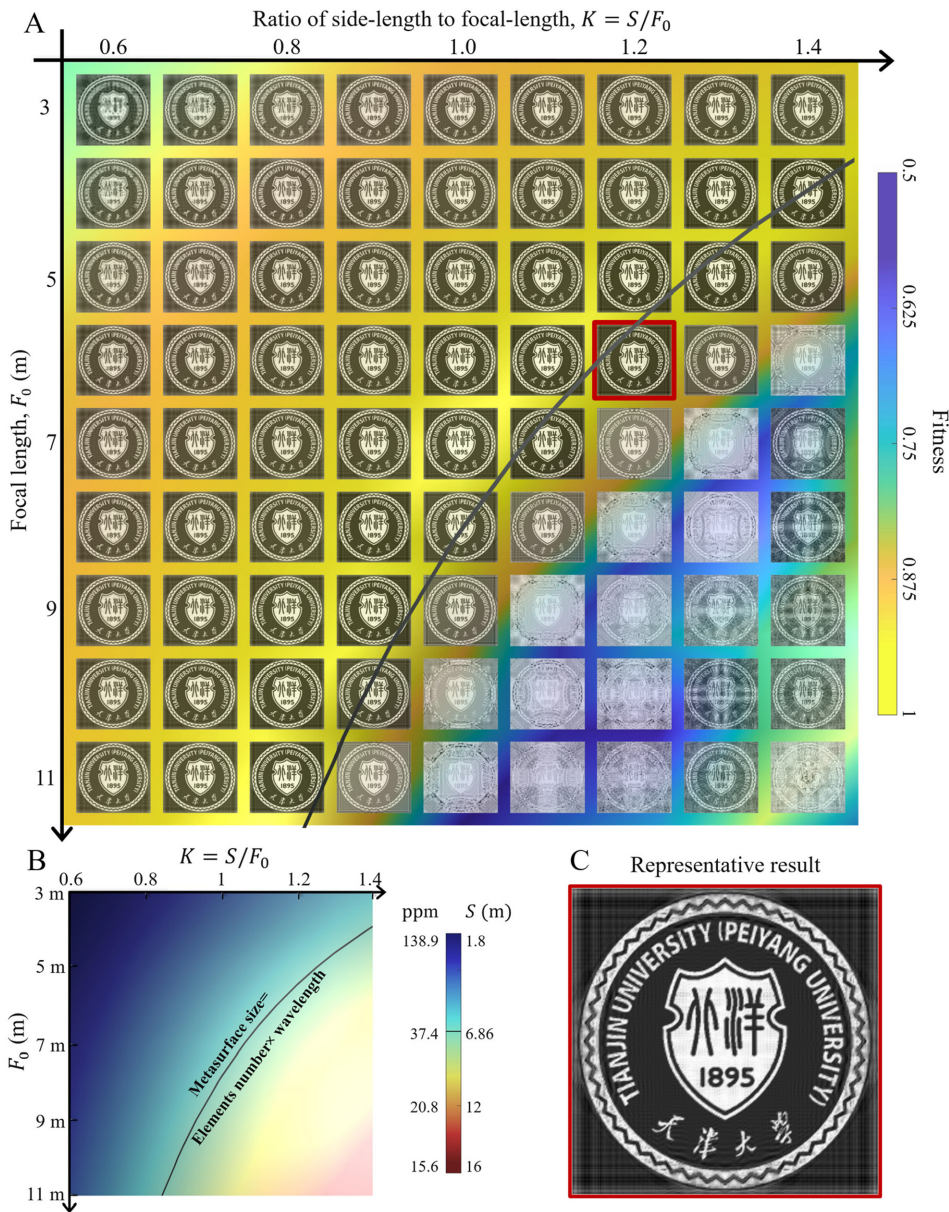
FIG. 5. Influence of spatial relationship between the metasurface and image on the holographic images of Tianjin University logo. (a) The holographic images after 10 iterations of the compensation algorithm with the fitness of results varying with focal length and metasurface size. (b) Contour maps for metasurface size (S) and pixel density (R). (c) Representative image with the highest fitness which is marked by the red square in (a). The area near the equal side-length curves shown in (a) and (b) yields optimal choices of the spatial parameters.

algorithm especially for lower resolution metasurfaces. For instance, the fitness of the image generated by the 800×800 metasurface is improved by 0.0258 from 0.9360 to 0.9619; while that by the 350×350 metasurface is improved by 0.0348 from 0.9137 to 0.9486. The increase in the fitness is fast in the first several iterations but slows down in the subsequent iterations. This is mainly due to the choice of the gradually reduced compensation rate μ_i to avoid oscillation. The fitness cannot reach 1, implying that the missed information cannot be fully compensated. Notably, some example images may turn gray as their fitness continues to increase. For instance, as shown in the 500×500 resolution logo image in Fig. 6(b), the logo text part in the 2nd-iteration image appears darker than in the 1st-iteration. This is caused by the normalization process applied during figure drawing.

The images are normalized according to the maximum intensity value among all the pixels. The drawing algorithm normalizes a few pixels at the top of the logo in the 2nd-iteration image to 1, which has large deviations in intensity compared to other pixels (pure white). This can make other regions appear "darker."

The images obtained with different iterative orders are shown in Fig. 6(b) from which one can observe that the interference fringes and unresolved details appearing at the pattern edges and junctions in the images are gradually removed with the iterative compensation progressing. The data-driven brightness variation over each iteration is caused by the normalization process. The images are normalized according to the maximum intensity value among all the pixels, which could vary from one iteration to another. An enhancement of the
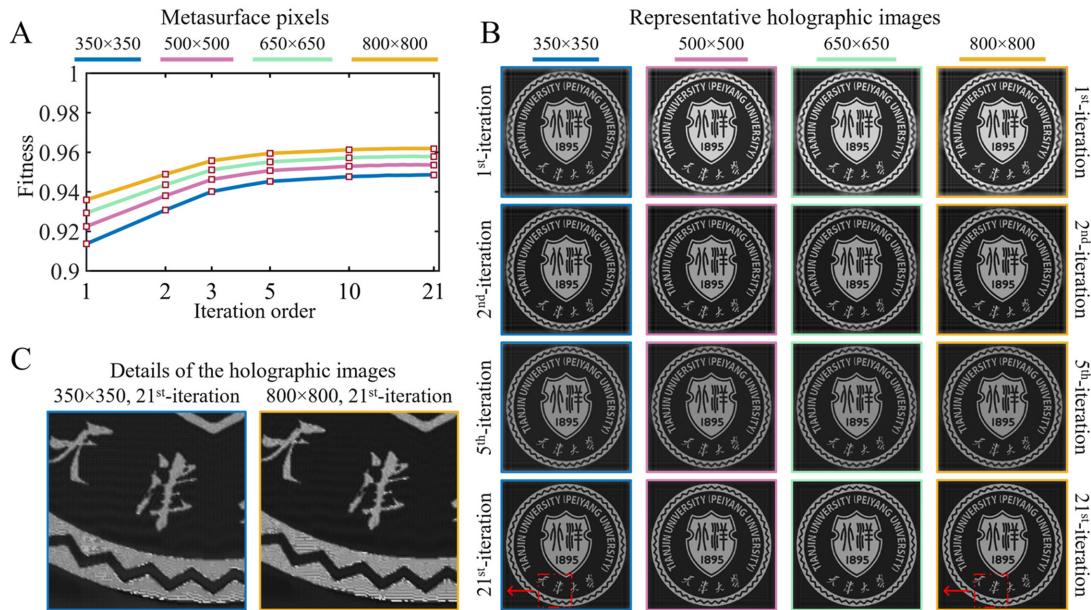
FIG. 6. Impact of metasurface pixel resolution on holographic images of Tianjin University logo. (a) Fitness curve depicting the quality of representative holographic images as a function of iterations. (b) Holographic images at the 1st, 2nd, 5th, and 21st iteration for different metasurface resolutions (pixels) with corresponding fitness values marked by hollow squares shown in (a). (c) Comparison of specific image features marked in the bottom line at (b) and generated by metasurfaces with 350 × 350 and 800 × 800 pixels after the 21st iteration, illustrating the differences in sharpness, contrast, and overall image quality.

details is presented in Fig. 6(c), which compares the zoomed-in images obtained after the 21st iteration with the 350×350 and 800×800 metasurfaces. More details about the quantitative evaluation are shown in the supplementary text S3.

The regular patterns revealed in Figs. 5 and 6 provide a guideline to design a metasurface-based hologram. Generally, more elements of a metasurface can yield a higher-quality image for a particular target picture. In the case of the same size and pixel numbers of both metasurface and image, the best choice of the element number (N) is near the ratio of side length (S) to wavelength (λ) and the ratio of side length to focal length (K) is better to be close to 1 (around 0.8–1.2). On the other hand, the developed iterative compensation method can improve the definition of images effectively in most cases.

High-quality imaging of the Mona Lisa's portrait

To demonstrate the potential of the developed method in designing a megapixel hologram, we present a holographic reconstruction of the Mona Lisa's portrait with 2000×2000 pixels by a high-resolution metasurface with 2000×2000 elements at 12.5 kHz as shown in Fig. 7(a). The focal length is 22 m and the ratio of side-length to focal length is 1. The target image is shown in Fig. 7(b). The design of 4×10^6 elements was completed within 2 h. The images at the representative steps of the iterative compensation algorithm and the corresponding fitness are illustrated in Figs. 7(c1)–7(c7) and 7(d), respectively. Benefiting from the large pixel number of the metasurface, the fitness of the 1st step is already as high as $\mathcal{F}^{(1)} = 0.9668$. It further increases rapidly to $\mathcal{F}^{(3)} = 0.9808$ in the first three iterations and then slowly to $\mathcal{F}^{(51)} = 0.9880$ at the 51st iteration as shown in Fig. 7(d). The image quality is improved gradually with the iterations as shown

in Figs. 7(c1)–7(c7). Compared with the first iteration result [Fig. 7(c1)], the final optimized image [Fig. 7(c7)] is significantly improved in definition, contrast, and sharpness. The stripes at the edges are also suppressed after the iterations. A detailed comparison among the target image, the final optimized image, and the first iteration image using zoomed-in view of some parts of the images [Figs. 7(e1)–7(e3)] reveals that the final optimized image has almost no difference from the target image. The texture of the oil painting is captured, and the interference fringes are removed. The comparison demonstrates that the iterative compensation algorithm can bring much more detail to the image along with the increasing fitness especially considering that only a quarter metasurface is filled with elements (the remaining portions are filled with soundproof baffles).

Experimental measurements were carried out to verify the proposed design. Due to the limitation of the cost and experimental condition, only the left eye part of Mona Lisa was reconstructed and measured. The test setup is described in detail in the supplementary text S5. A metasurface with 30×30 elements was designed to produce the eye part image with 100×100 pixels in the focal plane 0.12 m away at 12.5 kHz. The metasurface and image are of the same size of $0.288 \times 0.288 \ \mathrm{m}^2$. The metasurface was fabricated by 3D laser printing [Fig. 8(a)]. A 6×6 speaker array [Fig. 8(b)] is used as the sound source. The measurements were performed by scanning the sound field in the focal plane. The measured holographic image (amplitude) is demonstrated in Fig. 8(c) with the phase distribution shown in Fig. 8(d). For comparison, the simulated images after the 1st ($\mathcal{F}^{(1)} = 0.9668$), 3rd ($\mathcal{F}^{(3)} = 0.9808$), and 51st ($\mathcal{F}^{(51)} = 0.9880$) iterations are presented in Figs. 8(e)–8(g), and the phase distribution of 51st iterations is shown in Fig. 8(h). For the details of the simulation, we refer to supplementary text S5. The fitness of the measured image is
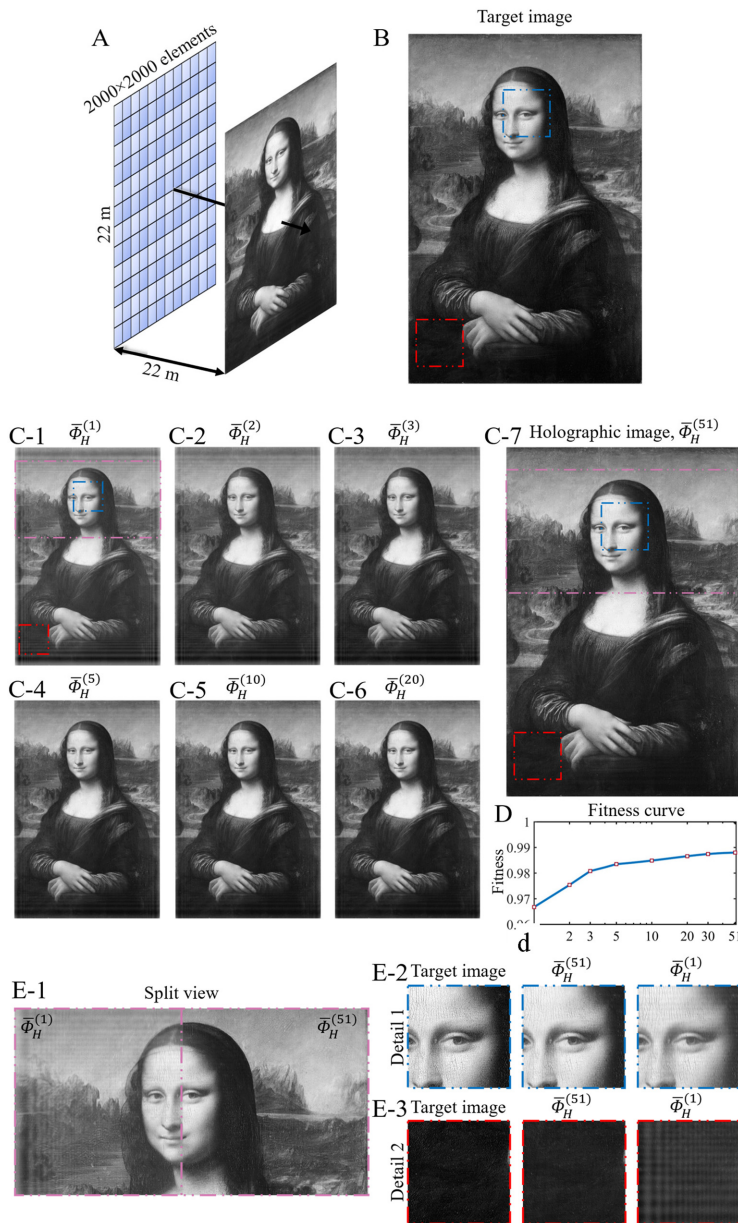
FIG. 7. High-quality holographic imaging of Mona Lisa's portrait. (a) 3D schematic view of the hologram and the generated image. (b) The target image of the Mona Lisa's portrait. (c1)–(c7) Reconstructed images with different iteration orders, the last one is the 51st-iteration result $\overline{\Phi}_H^{(51)}$. (d) Fitness curves correspond to iterative process. (e1) Split view comparison of the face part of "Mona Lisa's portrait" between the 51st-iteration result $\overline{\Phi}_H^{(51)}$ and the 1st-iteration result $\overline{\Phi}_H^{(1)}$. (e2), (e3) The comparison of eye part and left bottom part of Mona Lisa's portrait between target image, the 51st-iteration result $\overline{\Phi}_H^{(51)}$, and the 1st-iteration result $\overline{\Phi}_H^{(1)}$.

$\mathcal{F} = 0.944$ which is close to the simulation results. It can be observed that the measurements are generally in good agreement with the simulations, despite the presence of aliasing and speckles in holographic images. These result from factors like unwanted reflections, noise, inability to simulate ideal plane waves, and manufacturing imperfections in metasurface elements. The slight difference may stem from the inhomogeneity of the excited plane wave and the unavoidable errors during sample fabrication and assembly. An analysis of how noise influences results can be found in the supplementary text S6. In addition to the Mona Lisa's portrait, several other examples are demonstrated and summarized in the supplementary text S4, which confirm

the effectiveness of the proposed method. Due to its ability to reconstruct high-resolution complex images with precise phase-amplitude modulation, the presented high-resolution metasurface can offer numerous potential practical applications. These include volumetric displays, acoustic tweezers, nondestructive testing/inspection, medical imaging, and acoustic communication, among others.

DISCUSSION

In summary, a high-resolution metasurface holography optimization method is developed to achieve high-quality megapixel holographic images. The proposed element topology exhibits bi-anisotropic
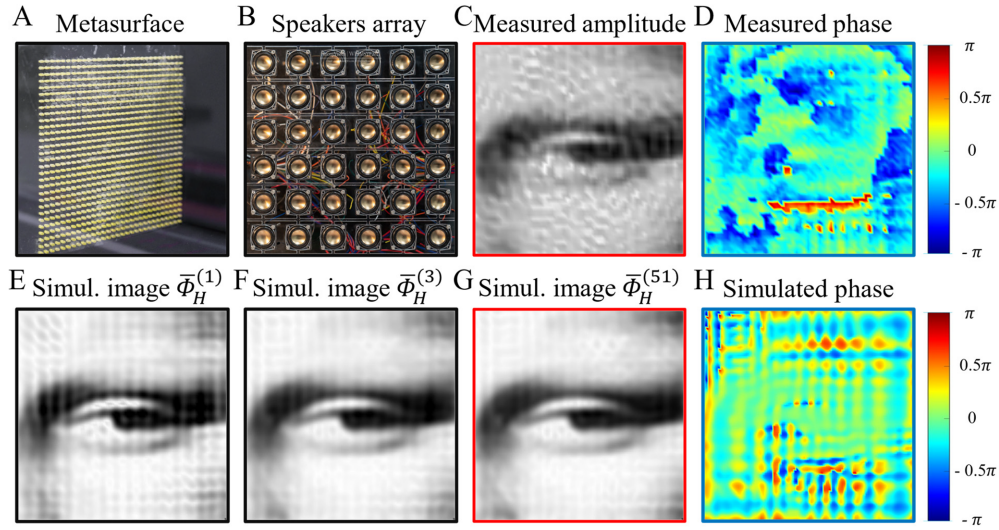
FIG. 8. Experimental measurements for the holographic image eye part of Mona Lisa's portrait. (a) The fabricated metasurface sample with 30×30 3D printed elements. (b) The 6×6 speaker array used in the measurement for the generation of a plane wave. (c), (d) Experimentally measured amplitude distribution and phase distribution on the target image plane, respectively. (e)–(g) Computational holographic images corresponding to 1st-, 3rd-, and 51st-iteration. (h) Simulated phase distribution on the target image plane.

properties and allows for flexible and precise PAM. The developed DNN-based GA can design millions of bi-anisotropic elements with customized phases and amplitudes in a fast and efficient manner, which is difficult to achieve using empirical approaches or conventional optimization methods. The proposed iterative compensation method can systematically improve the quality of the hologram by reducing the inconsistencies between target and projected images. As an example, the interference fringes are removed, and the unclear details are improved by adopting the compensation method.

Overall, the proposed methodology significantly enhances metasurface-based hologram designs, promoting efficiency, precision, and real-world applications. The results also demonstrate the possibility of realizing frequency-selective imaging, which opens a way to achieve frequency-selective hologram in multiple focal planes by designing elements at different frequencies. The approach can be extended for other wave-based metamaterials and devices by designing and optimizing the microstructures, with potential applications ranging from impedance matching, wavefront engineering, and polarization control. Future work could address challenges such as computation complexity, model generalization, and fabrication constraints. Advancements in scalable deep learning algorithms, such as transfer learning and neural architecture search, can reduce training time and resource requirements while maintaining performance and generality.

MATERIALS AND METHODS
M1. Deep neural network (DNN)

Here a fully connected neural network is employed to predict element properties.[41] Adjacent layers are fully interconnected: all neurons of the previous layer contribute to the next layer and every neuron derives its value from all previous layer's neurons. Fully connected DNNs deployed in this paper can be decomposed as perceptrons.[42]

The basic theory of perceptron and DNNs are given as follows. The fully connected neural network architecture represents an easy-to-train approach that is sufficient to give rise to the arbitrary phase and amplitude of an element. More sophisticated network architectures may provide enhanced accuracy and capability for some tasks; however, it is important to balance the complexity of the network architecture with the available computational resources.

A DNN (or perceptron) contains input, output, and hidden layers. An n-dimensional perceptron with m-dimensional input $\{x_j\}_{j=1}^{m}$ can be represented as

$$y_n = p\left(\sum_{j=1}^{m} w_{j,n} x_j + b\right), \tag{20}$$

where y_n represents the nth output neuron of the layer; $p(.)$ is the activation function of the neural network and we employed ReLU and softplus in the current work; $w_{j,n}$ is the weighting factor of the jth input neuron x_j; and b is the bias term. In DNN, each neuron can be represented as

$$x_{n,k} = p\left(\sum_{j=1}^{m} w_{j,n,k-1} x_{j,k-1} + b_{k-1}\right), \tag{21}$$

where $x_{n,k}$ represents the nth neuron of the kth layer for $k > 1$; $x_{j,k-1}$ represents the jth neuron of the previous $(k-1)$th layer; b_{k-1} denotes the bias of the previous $(k-1)$th layer; and $w_{j,n,k-1}$ is the nth weighting factor of the neuron $x_{j,k-1}$. When $k = 1$, $x_{n,1}$ represents the nth input neuron; and the $b_k = 0$ for the output layer.

The eleven structural parameters, $l_1 \sim l_{11}$, are inserted into $x_{n,1}$ ($n = 1–11$), then DNNs inference is processed through Eq. (21). The model will generate biased predictive values after the training set is input into $x_{n,1}$ ($n = 1–11$), the mean square error, **MSE**, can be estimated as[43]

$$MSE = \frac{1}{N} \sum_{i=1}^{N} \sum_{n=1}^{q} (l_{i,n} - y_{i,n})^2, \qquad (22)$$

where N is the number of the output neurons; q is the dimension of the output layer; $y_{i,n}$ is the nth output property of the ith element; and $l_{i,n}$ is the corresponding property calculated by FEM. The mean relative error, **MRE**, can be estimated as

$$MRE = \frac{1}{N} \sqrt{\sum_{i=1}^{N} \sum_{n=1}^{q} \left(\frac{l_{i,n} - y_{i,n}}{y_{i,n}} \right)^2}. \qquad (23)$$

Note that some prominent aspects of data dependency, model interpretability and robustness to fabrication errors need to be solved for further improving the present DNN method. In addition, the cutting-edge generative pre-trained transformer can also be applied here if the required data collection and preprocessing and model adaptation are solved.

M2. Numerical simulation

All wave simulations and calculations of the elements and metasurface are carried out by the finite element analysis software COMSOL Multiphysics. For the calculation of elements, we use thermo-viscous acoustics module in pressure acoustic module. We use pressure acoustic module for the wave response based on the whole metasurface. Perfectly matched layers are applied in the numerical simulation to reduce the reflection on the boundaries, corresponding to the sound-absorbing sponge used in the experiment. The loss in the air is considered by the viscous fluid model with bulk viscosity of 1.849×10^{-5} Pa s (air, 25°) in the Pressure Acoustic Module. The phase shift and transmission rate are extracted from the scattered fields of sound. To improve the accuracy, the transmission is calculated by extracting the integrated amplitude of strength at the cross-section two wavelengths away from the element outlet in the radial direction to avoid near-field effects.

M3. Sample preparation and experimental measurement

All metasurface elements were fabricated by 3D laser printing with a resolution of 10 μm. The printed material is hard UV-curable resin with a density of 1220 kg/m^3 and a longitudinal wave speed of 2650 m/s. The elements were secured in a laser-cut PMMA plate. The PMMA has a density of 1180 kg/m^3 and a longitudinal wave speed of 2690 m/s. The characteristic impedance of the elements and PMMA is much larger than the air, and therefore all solids are considered acoustically rigid. The pressure distribution of the hologram was acquired by an acoustic measurement system. In the experiments, an array of 36 speakers was used to generate the incident plane wave one meter away from the metasurface. The sound field at the focal plane was scanned by a moving sound sensor with a step of 3 mm in an area of 0.288×0.288 m^2. The sound sensor can collect acoustic pressure with a sensitivity of 1 mV/Pa from 6.5 Hz to 140 kHz. We measured the sound pressure at each point five times, performed the Fourier transform, and then took the average value to reduce the background noise. The experiment was carried out in an anechoic environment (see supplementary text S5 for details).

SUPPLEMENTARY MATERIAL

See the supplementary material for (S1) element topological definition and thermal-viscous loss effect; (S2) QR code generation and detailed calculation; (S3) analysis on holographic image detail level improvement; (S4) additional holographic images with detailed information; (S5) experiment setup; and (S6) experiment result analysis.

AUTHOR DECLARATIONS
Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Xuanbo Miao: Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Resources (equal); Supervision (equal); Validation (equal); Writing – original draft (equal). **Hao-Wen Dong:** Conceptualization (lead); Formal analysis (equal); Funding acquisition (equal); Methodology (equal); Project administration (equal); Resources (equal); Supervision (lead); Writing – original draft (equal); Writing – review & editing (equal). **Sheng-Dong Zhao:** Validation (equal). **Shiwang Fan:** Validation (equal); Visualization (equal). **Guoliang Huang:** Data curation (equal); Validation (equal). **Chen Shen:** Conceptualization (equal); Data curation (equal); Software (equal); Supervision (equal); Validation (equal); Writing – original draft (equal); Writing – review & editing (equal). **Yue-Sheng Wang:** Conceptualization (equal); Project administration (lead); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

All data needed to evaluate the conclusions in the paper are present in the paper and/or the supplementary material.

REFERENCES

[1]S. Tay, P. A. Blanche, R. Voorakaranam, A. V. Tunc, W. Lin, S. Rokutanda, T. Gu, D. Flores, P. Wang, G. Li, P. St. Hilaire, J. Thomas, R. A. Norwood, M. Yamamoto, and N. Peyghambarian, "An updatable holographic three-dimensional display," Nature **451**, 694–698 (2008).

[2]D. G. Grier, "A revolution in optical manipulation," Nature **424**, 810–816 (2003).

[3]M. L. Juan, M. Righini, and R. Quidant, "Plasmon nano-optical tweezers," Nat. Photonics **5**, 349–356 (2011).

[4]A. Marzo and B. W. Drinkwater, "Holographic acoustic tweezers," Proc. Natl. Acad. Sci. U.S.A. **116**, 84–89 (2019).

[5]X. Ni, A. V. Kildishev, and V. M. Shalaev, "Metasurface holograms for visible light," Nat. Commun. **4**, 2807 (2013).

[6]A. Marzo, S. A. Seah, B. W. Drinkwater, D. R. Sahoo, B. Long, and S. Subramanian, "Holographic acoustic elements for manipulation of levitated objects," Nat. Commun. **6**, 8661 (2015).

[7]K. Melde, A. G. Mark, T. Qiu, and P. Fischer, "Holograms for acoustics," Nature **537**, 518–522 (2016).

[8]Y. Hertzberg and G. Navon, "Bypassing absorbing objects in focused ultrasound using computer generated holographic technique," Med. Phys. **38**, 6407–6415 (2011).

[9]Z. C. Ma, A. W. Holle, K. Melde, T. Qiu, K. Poeppel, V. M. Kadiri, and P. Fischer, "Acoustic holographic cell patterning in a biocompatible hydrogel," Adv. Mater. **32**, 1904181 (2020).

[10]Z. L. Deng, X. P. Li, and G. X. Li, *Metasurface Holography*, Synthesis Lectures on Materials and Optics Vol. **1** (Morgan & Claypool, 2020), pp. 1–76.

[11]A. L. Chen, Y. S. Wang, Y. F. Wang, H. T. Zhou, and S. M. Yuan, "Design of acoustic/elastic phase gradient metasurfaces: Principles, functional elements, tunability and coding," Appl. Mech. Rev. **74**, 020801 (2022).

[12]B. Assouar, B. Liang, Y. Wu, Y. Li, J. C. Cheng, and Y. Jing, "Acoustic metasurfaces," Nat. Rev. Mater. **3**, 460–472 (2018).

[13]Y. B. Xie, C. Shen, W. Q. Wang, J. F. Li, D. J. Suo, B. I. Popa, Y. Jing, and S. A. Cummer, "Acoustic holographic rendering with two-dimensional metamaterial-based passive phased array," Sci. Rep. **6**, 35437 (2016).

[14]M. Bakhtiari-Nejad, A. Elnahhas, M. R. Hajj, and S. Shahab, "Acoustic holograms in contactless ultrasonic power transfer systems: Modeling and experiment," J. Appl. Phys. **124**, 244901 (2018).

[15]J. Zhang, Y. Tian, Y. Cheng, and X. J. Liu, "Acoustic holography using composite metasurfaces," Appl. Phys. Lett. **116**, 030501 (2020).

[16]Y. Tian, Q. Wei, Y. Cheng, and X. J. Liu, "Acoustic holography based on composite metasurface with decoupled modulation of phase and amplitude," Appl. Phys. Lett. **110**, 191901 (2017).

[17]Y. F. Zhu, J. Hu, X. D. Fan, J. Yang, B. Liang, X. F. Zhu, and J. C. Cheng, "Fine manipulation of sound via lossy metamaterials with independent and arbitrary reflection amplitude and phase," Nat. Commun. **9**, 1632 (2018).

[18]S. W. Fan, Y. F. Zhu, L. Y. Cao, Y. F. Wang, A. L. Chen, B. Vincent, Y. S. Wang, and B. Assouar, "Broadband tunable lossy metasurface with independent amplitude and phase modulations for acoustic holography," Smart Mater. Struct. **29**, 105038 (2020).

[19]J. Zhang, Y. Yang, B. P. Zhu, X. J. Li, J. Jin, Z. Y. Chen, Y. Chen, and Q. F. Zhou, "Multifocal point beam forming by a single ultrasonic transducer with 3D printed holograms," Appl. Phys. Lett. **113**, 243502 (2018).

[20]M. D. Brown, "Phase and amplitude modulation with acoustic holograms," Appl. Phys. Lett. **115**, 053701 (2019).

[21]Y. F. Zhu and B. Assouar, "Systematic design of multiplexed-acoustic-metasurface hologram with simultaneous amplitude and phase modulations," Phys. Rev. Mater. **3**, 045201 (2019).

[22]Y. F. Zhu, N. J. Gerard, X. X. Xia, G. C. Stevenson, L. Y. Cao, S. W. Fan, C. M. Spadaccini, Y. Jing, and B. Assouar, "Systematic design and experimental demonstration of transmission-type multiplexed acoustic meta-holograms," Adv. Funct. Mater. **31**, 2101947 (2021).

[23]H. W. Dong, C. Shen, S. D. Zhao, W. Qiu, H. Zheng, C. Zhang, S. A. Cummer, Y. S. Wang, D. Fang, and L. Cheng, "Achromatic metasurfaces with inversely customized dispersion for ultra-broadband acoustic beam engineering," Natl. Sci. Rev. **9**, nwac030 (2022).

[24]H. T. Zhou, W. X. Fu, X. S. Li, Y. F. Wang, and Y. S. Wang, "Loosely coupled reflective impedance metasurfaces: Precise manipulation of waterborne sound by topology optimization," Mech. Syst. Signal Process. **177**, 109228 (2022).

[25]Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature **521**, 436–444 (2015).

[26]S. S. An, C. Fowler, B. W. Zheng, M. Y. Shalaginov, H. Tang, H. Li, L. Zhou, J. Ding, A. M. Agarwal, C. Rivero-Baleine, K. A. Richardson, T. Gu, J. J. Hu, and H. L. Zhang, "A deep learning approach for objective-driven all-dielectric metasurface design," ACS Photonics **6**, 3196–3207 (2019).

[27]C. C. Nadell, B. H. Huang, J. M. Malof, and W. J. Padilla, "Deep learning for accelerated all-dielectric metasurface design," Opt. Express **27**, 27523 (2019).

[28]L. L. Li, H. X. Ruan, C. Liu, Y. Li, Y. Shuang, A. Alù, C. W. Qiu, and T. J. Cui, "Machine-learning reprogrammable metasurface imager," Nat. Commun. **10**, 1082 (2019).

[29]I. Tanriover, W. Hadibrata, and K. Aydin, "A physics based approach for neural networks enabled design of all-dielectric metasurfaces," ACS Photonics **7**, 1957–1964 (2020).

[30]J. Weng, Y. J. Ding, C. B. Hu, X. F. Zhu, B. Liang, J. Yang, and J. C. Cheng, "Meta-neural-network for real-time and passive deep-learning-based object recognition," Nat. Commun. **11**, 6309 (2020).

[31]Q. Lin, J. Q. Wang, F. Y. Cai, R. J. Zhang, D. G. Zhao, X. X. Xia, J. P. Wang, and H. R. Zheng, "A deep learning approach for the fast generation of acoustic holograms," J. Acoust. Soc. Am. **149**, 2312 (2021).

[32]K. Donda, Y. F. Zhu, A. Merkel, S. W. Fan, L. Y. Cao, S. Wan, and B. Assouar, "Ultrathin acoustic absorbing metasurface based on deep learning approach," Smart Mater. Struct. **30**, 085003 (2021).

[33]H. Ding, X. S. Fang, B. Jia, N. Y. Wang, Q. Cheng, and Y. Li, "Deep learning enables accurate sound redistribution via nonlocal metasurfaces," Phys. Rev. Appl. **16**, 064035 (2021).

[34]S. So, T. Badloe, J. Noh, J. Rho, and J. Bravo-Abad, "Deep learning enabled inverse design in nanophotonics," Nanophotonics **9**, 1041–1057 (2020).

[35]L. Shi, B. C. Li, C. Kim, P. Kellnhofer, and W. Matusik, "Towards real-time photorealistic 3D holography with deep neural networks," Nature **591**, 234–239 (2021).

[36]A. Maimone, A. Georgiou, and J. S. Kollin, "Holographic near-eye displays for virtual and augmented reality," ACM Trans. Graphics **36**(4), 1–16 (2017).

[37]Q. Zhang, C. Liu, X. Wan, L. Zhang, S. Liu, Y. Yang, and T. J. Cui, "Machine-learning designs of anisotropic digital coding metasurfaces," Adv. Theory Simul. **2**, 1800132 (2019).

[38]J. Li, C. Shen, A. Díaz-Rubio, S. A. Tretyakov, and S. A. Cummer, "Systematic design and experimental demonstration of bianisotropic metasurfaces for scattering-free manipulation of acoustic wavefronts," Nat. Commun. **9**, 1342 (2018).

[39]W. Ye, C. Chen, Z. Wang, I. H. Chu, and S. P. Ong, "Deep neural networks for accurate predictions of crystal stability," Nat. Commun. **9**, 3800 (2018).

[40]H. Larochelle, Y. Bengio, J. Louradour, and P. Lamblin, "Exploring strategies for training deep neural networks," J. Mach. Learn. Res. **10**, 1–40 (2009).

[41]M. H. Hassoun, *Fundamentals of Artificial Neural Networks* (MIT Press, 1995).

[42]I. Stephen, "Perceptron-based learning algorithms," IEEE Trans. Neural Networks **50**, 179–191 (1990).

[43]R. Hecht-Nielsen, "Theory of the backpropagation neural network," in *Neural Networks for Perception* (Academic Press, 1992), pp. 65–93.