Supporting Material

Gradient-descent Optimization of Metasurface based on One Deep-enhanced Resnet

Yi Xu[1], Fu Li[1], Jianqiang Gu[1*], Quan Xu[1], Zhen Tian[1], Jiaguang Han[1,3], and Weili Zhang[2*]

1. Simulation details for the meta-atom library

We employ commercial electromagnetic simulation software to characterize the transmission spectrum of our meta-atom. The period of the meta-atom is 200 μm, the thickness of the substrate is 2800 μm, and the height of the pillars is uniformly set as 200 μm. All pillars within the lattice are rectangular. To ensure the fabricability of the assembled pillars, the minimum size of the pillars as well as the minimum gap between adjacent pillars is not less than 20 μm, to ensure that the depth-to-width ratio is not greater than 10. A periodic boundary is set in both *x*- and *y*-directions, and a perfect match layer is set in the *z*-direction. The meta-atoms are made of lossless silicon material with a dielectric constant of 11.9. We use *x*- and *y*- polarized plane waves to illuminate the meta-atoms, respectively.

We edit Python scripts to control the simulation software to perform the modeling and calculation automatically. The structural parameters of each pillar were generated by Python random functions sampled randomly over a uniform distribution, in which 9 classes of meta-molecules containing multiple pillars (ranging from 1×1 to 3×3) are also randomly generated. To ensure that the transmission of all the structures has adequate spectral accuracy, we uniformly set the simulation duration to 110 ps. The working frequency band is from 0.3 to 1.2 THz, located in the range with the best performance of the experimental system. One loop is completed for each simulation run, and the number of loops is the initial size of the dataset. Then we augment the dataset based on the intrinsic symmetry of the meta-atoms by mirroring or rotating the pillars, leading to a total of 170,747 samples, which are evenly divided into training, validation, and testing sets in a ratio of 1:1:1.

2. Numerical encoding details for different classes of meta-atoms

We employ the one-hot encoding method to represent different classes of meta-atoms. One-hot encoding is used to deal with categorical data to convert all variables into a unified numerical format, which can be fed into the deep learning algorithm for further processing. This approach distinguishes among different classes while maintaining the disorder and randomness of data, rendering it suitable for regression problems. Here, each meta-atom is represented as a tensor of size 27, including 9 category parameters and 18 structure parameters, as listed in **Table S1**. The letters *x* and *y* represent the dimensions of the dielectric column within the lattice in the *x* and *y* directions, respectively. The order of the structure parameters is from left to right and from top to bottom.

Class	Category Parameters	Structure Parameters
1×1	100000000	00000000xy00000000
1×2	010000000	00xy0000000000xy00
1×3	001000000	00xy0000xy0000xy00
2×1	000100000	000000xy00xy000000
2×2	000010000	xy00xy000000xy00xy
2×3	000001000	xy00xyxy00xyxy00xy
3×1	000000100	000000xyxyxy000000
3×2	000000010	xyxyxy000000xyxyxy
3×3	000000001	xyxyxyxyxyxyxyxyxy

Table S1. Encoding for category and structure parameters of 9 classes of meta-atoms.

In each epoch, we use the Softmax function to judge the category of meta-atoms, which is mathematically expressed as:

$$\mathrm{Softmax}\left(z_i\right)=\frac{\exp(z_i)}{\sum_j z_j}. \tag{S1}$$

Equation S1 is fundamentally employed to convert numerical vectors into normalized probability distribution where $\sum\mathrm{Softmax}\left(z_i\right)=1$, which is suitable for processing multi-class classification problems. In our study, we feed the category parameters into the Softmax function

to determine the classification of meta-atoms, which corresponds to the maximum of the probability value. Following class assignment, the structure parameters undergo a dual-stage processing protocol: First, we set the structure parameters that do not belong to this category as zero, to eliminate the effect from extraneous terms in the spectrum prediction of the meta-atoms. Second, we re-assign the structure parameters within the valid design space that belong to this category but are out of the range through random sampling, to ensure the optimized meta-atoms meet design requirements. This dual-step regularization ensures both categorical specificity and physical realizability of the structural configurations. These adjusted parameters are then input to the neural network for the next optimization cycle. Through repeated iterations, this process gradually improves the metasurface design until it meets the target performance requirements.

3. Hardware configuration and software environment

We provide the hardware configuration and software environment along with other details here. The workstation is equipped with an Intel Xeon Platinum 8373C CPU and two NVIDIA GeForce RTX 3090 GPUs. It utilizes eight Samsung server-grade 32 GB Registered ECC memory modules, totaling 256 GB of RAM. Storage includes a combined capacity of 9 TB across SSDs and HDDs, allocated for system installation and data storage respectively. To enable efficient neural network training, a high-performance Linux workstation running Ubuntu 22.04.3 LTS was configured. After installing the Ubuntu OS, critical dependencies such as GPU drivers, CUDA, and Cudnn were deployed. Anaconda 22.9.0 was then installed to manage Python environments, followed by the setup of PyTorch within a dedicated Conda environment. All hardware and software components were rigorously validated before initiating model training. The entire training framework operates under a Python 3.9 environment with version-controlled dependencies. During the training process, data loading and preprocessing utilized 2 parallel worker threads, and input batches were randomized through shuffling to prevent ordering bias.

4. Prediction results of the deep-enhanced ResNet

Several samples in the test set are randomly presented here in **Figure S1**, as the complement to the main text, in which the brown and orange lines are the truth in the amplitude and phase of

the transmission under *x*- and *y*-polarized incidence, respectively, while the dots represent the corresponding results obtained by the prediction of the network.

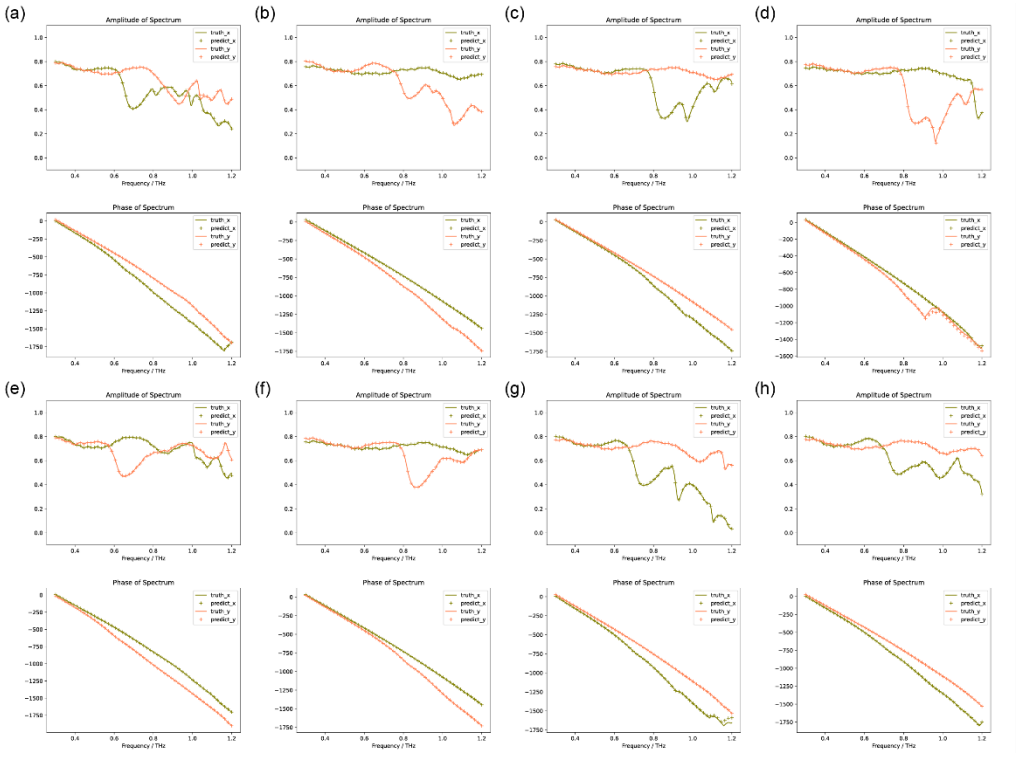

Figure S1. Examples demonstrating the performance of the deep-enhanced ResNet.

5. Details for paradigm-optimized configuration

In addition to the disparities in the flow between the training of the forward prediction network and the metasurface design, as illustrated in **Figure S2**, their configurations are also different. During the training process, the label is the full-spectrum data output by the electromagnetic simulation software, and the weights and biases of the network are trained to make the network converge. Whereas, when the forward prediction network is well-trained, we embed it into the end-to-end design framework and freeze its weights and biases. The design requirements, i.e., the target amplitudes and phases under different polarizations at arbitrary frequencies, are applied as the label, and the category and structure parameters are optimized as variables until the convergence criterion is reached. Specifically, we adopt a learning rate of 0.01 and a stair-

step learning rate decay with a γ of 0.9 every 1000 epochs. The stop threshold is set to 1×10^{-5}, that is, the optimization loop is stopped when the loss is less than this value otherwise the paradigm will output the result after 100000 epochs as the optimal solution. It is worth noting that the duration for each epoch depends on the type and function of the metasurface, and the number of iterations to reach the stop threshold is also different, hence leading to a varying total time of the optimization process. To avoid getting trapped in a locally optimal solution due to the gradient descent algorithm, every 20 epochs we re-perform random sampling for the meta-atom with the largest loss. The batch size equals the number of meta-atoms on the metasurface. We utilize Adam as the optimizer the same as the network training process.

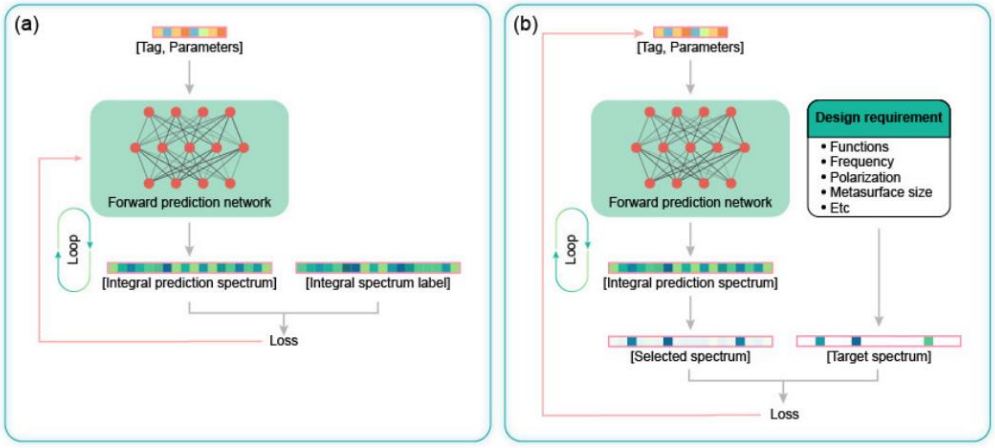

Figure S2. Comparison between the flow of (a) forward prediction network training and (b) metasurface design.

6. Discussion on the computational efficiency of the proposed design paradigm

To further clarify the computational efficiency of our method, we will elaborate mainly on three aspects:

1) Network training time. With the same hardware configuration and software environment specified in Section 3, the forward neural networks converged after 4×10^6 iterations. The training for Base, Medium, and Large models took about 12 days, 13 days, and 17 days, respectively.
2) Network vs. conventional simulations. The network training process is a one-time effort. Once well-trained, the network can predict the spectrum of a meta-atom in milliseconds

(ms), whereas full-wave simulations require 2-3 minutes. Moreover, tensor-based computation rules enable parallel calculation. In contrast, full-wave simulations typically allow a maximum of 3 concurrent runs on the same device, and simultaneous execution significantly degrades computational efficiency.

3) Optimization Convergence. As illustrated in **Figure S3a**, the gradient-descent optimization converged after ~34,000 iterations with a loss of 6.805×10^{-3}. And it can be notably seen that the patterns generated by the network are very close to the target ones after around 10,000 iterations in Fig. S3b and S3c. After 20,000 iterations, the background noise is almost negligible.

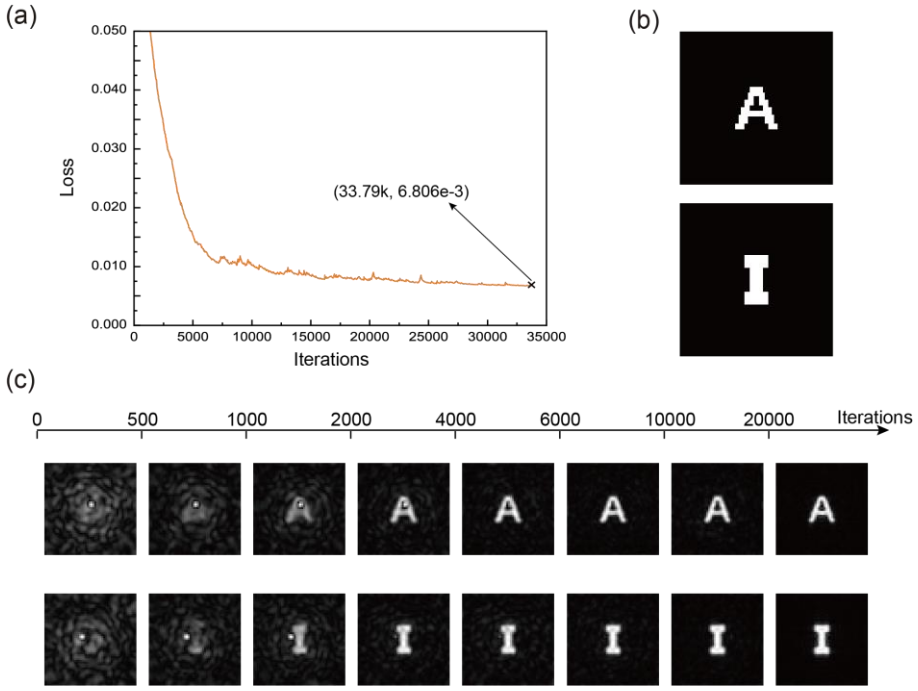

Figure S3. Gradient-descent optimization results for the designed holographic metasurface. (a) Convergence plot that takes the loss value as the function of the iteration. (b) Target images. (c) Details of the reconstructed hologram images at the target planes during the optimization.

7. Design and experimental demonstration of metalens

As an initial validation for the proposed paradigm, we designed and fabricated a metalens, the most fundamental metasurface device. The ideal phase of each meta-atom that constitutes the

metalens can be derived according to **Equation S2**:

$$\varphi_{lens}(x,y)=\frac{\omega}{c}\left(\sqrt{x^2+y^2+f^2}-f\right), \tag{S2}$$

where c is the light speed in vacuum, ω is the angular frequency, (x, y) is the Cartesian coordinates of the meta-atom, and f is the focal length of the metalens. A metalens is aimed to operate at 0.7 THz with a focal length of 10 mm for x-polarized incidence and a focal length of 12 mm for y-polarized incidence, which consists of 51 × 51 meta-atoms with a lattice size of 200 μm. The proposed end-to-end paradigm automatically optimizes each meta-atom and arranges them into the whole metasurface. Then we fabricated the designed metalens according to the optimized parameters and characterized it by using a home-built terahertz time-domain spectroscopic near-field scanning imaging system as shown in **Figure S4a**, and the whole and local magnification of the SEM image of the sample are shown in Fig. S4b. The amplitude and the phase of the incident terahertz beam at 0.7 THz are shown in Fig. S4c, which indicates a typical Gaussian-like beam. The simulated calculation and measured results are illustrated in Fig. S4d and S4e. The simulation results show that the transmitted x-polarized terahertz wave converges to the foci at z = 10 mm in both the *xoz*- and *yoz*- sections, and the transmitted y-polarized terahertz wave converges to the foci at z = 12 mm in both the *xoz*- and *yoz* sections, conforming to the intended design function, which manifest the effectiveness of the proposed paradigm. While, the measured focal lengths are 7.5 mm and 9.0 mm for x- and y- polarization, respectively. Though there is a focal shift of 2~3 mm compared with the expected target, which is due to the unavoidable imperfect incidence and the defects of the sample, the metalens still exhibits polarization multiplexing characteristics and has good focal spot shapes. Also, the metalens achieved efficiencies of 35% and 34% under x- and y- polarized illumination, respectively. These quantified results confirm that our meta-devices maintain competitive efficiency while enabling multifunctional performance through polarization multiplexing, which further demonstrate the effectiveness and universality of our end-to-end design paradigm. The variation in efficiency values between different devices can be attributed to their distinct phase modulation requirements and inherent diffraction limitations in each operational mode, as well as the measurement and fabrication errors.

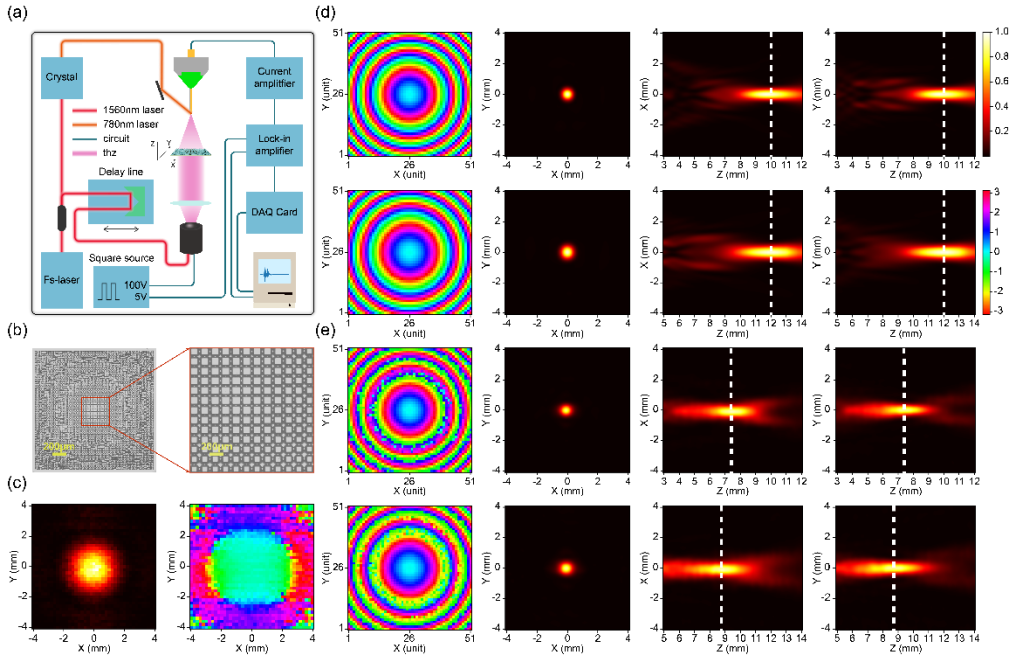

Figure S4. Design and characterization results of the designed metalens. (a) Schematic diagram of the terahertz near-field time-domain spectroscopy system used to characterize the sample. (b) The SEM image of the whole sample and magnification of the selected area. (c) The intensity and phase of the incident terahertz beams scanned without samples at 0.7 THz. (d) The phase and intensity of the focal plane as well as the intensity of *xoz*- and *yoz*- sections calculated by electromagnetic simulation software for *x*-polarization at z = 10 mm (top row) and *y*-polarization at z = 12 mm (bottom row), respectively. (e) The phase and intensity of the focal plane, as well as the intensity of *xoz*- and *yoz*- sections measured experimentally for *x*-polarization at z = 7.5 mm (top row) and *y*-polarization at z = 9.0 mm (bottom row), respectively. The white dashed lines indicate the focal planes.

8. Device fidelity analysis of the holographic and vortex metasurfaces.

1) For holographic metasurfaces, the fidelity primarily evaluates the similarity between its reconstructed image and the objective image, with deviations between the actual and theoretical intensity distributions being a key metric. We calculate the fidelity based on the following equations:

$$\Delta I^{(n)} = I_{\text{norm}}{}^{(n)} - \hat{I}_{\text{norm}}{}^{(n)}, \tag{S3}$$

$$f = \frac{1}{N}\sum_{n=1}^{N} |\Delta I^{(n)}|, \tag{S4}$$

where $I_{\text{norm}}^{(n)}$ and $\hat{I}_{\text{norm}}^{(n)}$ are the calculated/measured and target intensity, and are normalized separately. N is the total number of pixels in the image plane. In our work, the theoretically calculated fidelities for patterns "A" and "I" were determined to be 0.013 and 0.010, respectively, while experimental measurements yielded corresponding values of 0.054 and 0.050. This close correspondence between theoretical predictions and experimental outcomes demonstrates excellent agreement with the predetermined design specifications.

2) For vortex metasurfaces, the fidelity typically refers to vortex purity, which is a core metric for assessing the quality of the orbital angular momentum (OAM) pattern of an optical vortex beam by quantifying the similarity between the actual and the ideal OAM pattern.[1-2] It is defined as the ratio of the power of the target OAM mode to the total power of the actual light field. The higher the vortex purity, the fewer the stray modes, indicating that the beam quality is closer to the ideal vortex state. We employ the mode decomposition method to calculate the vortex purity of the proposed metasurface. The principle involves projecting the actual light field onto the orthogonal OAM mode basis vectors and calculating the weight coefficients C_l of each mode:

$$C_l = \frac{1}{2\pi}\int_0^{2\pi} \tilde{E}(\theta)\exp(-il\theta)d\theta, \tag{S5}$$

where $\tilde{E}(r,\theta)$ is the measured complex light field, $\tilde{E}_l(r,\theta) = \exp(il\theta)$ is the ideal OAM pattern, $\tilde{E}_l^*(r,\theta)$ is the complex conjugate to $\tilde{E}_l(r,\theta)$. By taking a ring-shaped electric field centered at the phase singularity of the vortex beam and selecting the radius r_0 where the electric field amplitude is maximum, the above equation can be simplified to:

$$C_l = \frac{1}{2\pi}\int_0^{2\pi} \tilde{E}(\theta)\exp(-il\theta)d\theta, \tag{S6}$$

where the linking Fourier relationship is given by:

$$\tilde{E}(\theta) = \sum_{l=-\infty}^{+\infty} C_l \exp(il\theta), \tag{S7}$$

and the vortex purity can be denoted as the normalized power:

$$P_l = \frac{|C_l|^2}{\sum_{l'}|C_{l'}|^2}. \tag{S8}$$

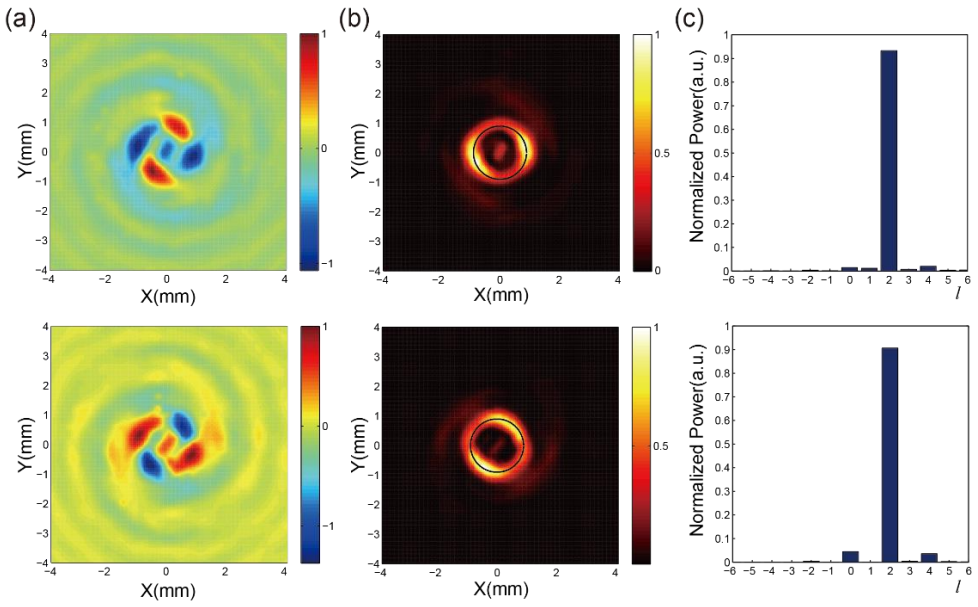

Figure S5. Measured vortex purity for the designed vortex metasurface. (a) The experiment results for (a) the real parts, (b) the normalized intensity distributions, and (c) the vortex purities calculated along the black circle in (b) under *x*- and *y*- polarized incidence, respectively.

In our work, the target OAM mode (i.e. the target topological charge number of the vortex beam) is l = +2, and we only select the l' from –6 to 6. The measured vortex purities are illustrated in **Figure S5**. The weight of the scheduled +2 OAM mode achieves 93.3% for *x* polarization and 90.6% for *y* polarization, remarkably exceeding other OAM modes, which demonstrates that our proposed vortex beam generator metasurface possesses high fidelity.