ISSN 1998-0124 CN 11-5974/O4
https://doi.org/10.26599/NR.2026.94909057

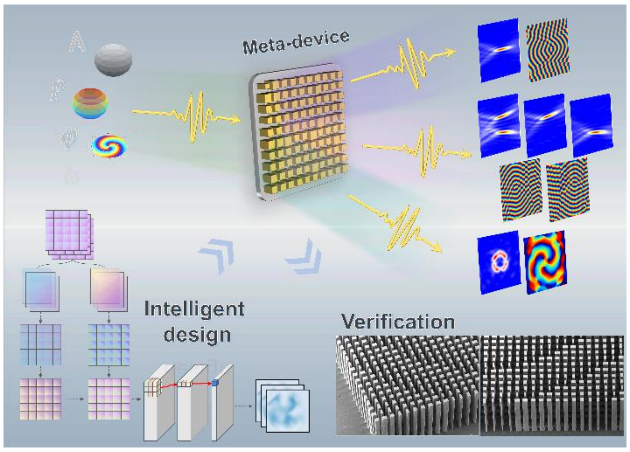
Research Article

Lightweight deep neural network empowers multifunctional metasurface design for arbitrary wavefront manipulations

Nan Zhang, Donghai Han(✉), Zhonglei Shen, Yuqing Cui, Liuyang Zhang(✉), Ruqiang Yan, Xuefeng Chen
School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China

Nano Res., **Just Accepted Manuscript** • https://doi.org/10.26599/NR.2026.94909057
https://www.sciopen.com/journal/1998-0124 on Jul. 28, 2026

TABLE OF CONTENTS (TOC)



A task specific multifunctional lightweight deep neural network integrating forward and inverse design is developed for efficient and accurate metasurface design for arbitrary wavefront manipulation. Combined with data distillation, the framework reduces the dataset requirement to 20% while improving design precision, as demonstrated by polarization multiplexed metalens and vortex generation metasurfaces.

Nano Research

Open Access

Research Article

Lightweight deep neural network empowers multifunctional metasurface design for arbitrary wavefront manipulations

Nan Zhang, Donghai Han, Zhonglei Shen, Yuqing Cui, Liuyang Zhang, Ruqiang Yan, and Xuefeng Chen

School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China

Received: 1 April 2026; **Revised:** 2 July 2026; **Accepted:** 28 July 2026

Address correspondence to Donghai Han, donghai19@stu.xjtu.edu.cn; Liuyang Zhang, liuyangzhang@xjtu.edu.cn

Cite this article: *Nano Research*, 2026, 19, 94909057 https://doi.org/10.26599/NR.2026.94909057

ABSTRACT: The relentless pursuit of next-generation photonic systems for advanced information processing and terahertz-scale wireless communications necessitates unprecedented precision in arbitrary electromagnetic wave manipulation. By engineering the local phase of individual meta-atoms, metasurfaces enable the precise control of multidimensional optical parameters. Nevertheless, conventional meta-atom design paradigms, heavily reliant on iterative trial-and-error processes, are computationally intensive and resource-demanding. Herein, we propose a task-specific, multifunctional lightweight deep neural network that integrates forward and inverse design modules. This framework enables efficient and accurate exploration within the function space and parameter space. It successfully reconciles the inherent trade-off between computational efficiency and prediction accuracy, thereby enabling a combinatorial inverse design paradigm. This approach, rooted in functional modularity, significantly enhances the efficiency and flexibility of developing highly integrated, multifunctional optical devices. To address the inherent data dependency, we implement a data distillation strategy that reduces the required dataset to just 20% of its original volume while simultaneously enhancing design precision by 12.06%. As a proof of concept, we demonstrate the entire design and validation process of the polarization-multiplexed metalens and vortex-generation metasurfaces. This versatile approach could be readily extended to develop other types of wavefront modulation devices, heralding a new era of customizable and multifunctional meta-devices.

KEYWORDS: deep learning; data distillation; metasurfaces; wavefront manipulation

1 Introduction

Metasurfaces, meticulously engineered artificial structures composed of customized subwavelength units, have garnered significant attention in recent years. This surge of interest stems from its remarkable ability to transcend the diffraction limits inherent to conventional optical systems [1], thereby addressing the longstanding challenge of scale mismatch. These advanced meta-devices enable the precise manipulation of phase [2–4], amplitude [5], and polarization [6], facilitating the arbitrary optical field distributions. Such capabilities empower metasurfaces to perform a plethora of optical functions, encompassing metalens [7,8], diffraction neural networks [9,10], beam steering [6,11], polarization modulation [12], and holography [13,14]. Beyond individual functionalities, the inherently planar, compact, and manufacturing-compatible nature of metasurfaces allows for the integration of multiple optical operations within a single device footprint, thereby significantly enhancing functional density. Therefore, the multifunctional metasurfaces provide emerging possibilities for breaking limitations of traditional bulk optics in degrees of freedom, integrability, and application versatility.

However, the metasurface design methodology remains constrained by repetitive electromagnetic (EM) simulations rooted in empirical structural parameterization [15,16]. Conventional approaches, pioneered through intuitive meta-atom trial-and-error optimization in a forward fashion, rely on classical meta-atom geometries (e.g., elliptical [17], split-ring resonators [18,19], and rectangular patches [20]) governed by physical resonant principles. These methods employ finite-difference time-domain (FDTD) simulations to iteratively map geometric parameters (dimensions, lattice constants, material properties) toward the target spectral or field profiles, thereby achieving desirable EM responses. However, achieving such fine control by precisely tuning geometry and material properties necessitates time-consuming, repetitive EM simulations, which heavily rely on prior expert knowledge and intuition. This trial-and-error approach proves to be inefficient, particularly in high-dimensional parameter spaces [21], where the computational cost becomes prohibitive. While this method is capable of solving "forward problems" (i.e., designing structures to meet predefined performance criteria), it struggles to infer structural parameters given a specific target function or spectral response. Furthermore, the optimization results obtained for one function are often not transferable to others, highlighting the lack of generality inherent in this approach.

The emergence of deep learning has revolutionized metasurface engineering [22], transitioning from intuition-driven structural optimization to sophisticated neural architectures [23–25]. This paradigm shift is achieved by neural architectures that systematically decode light-matter interactions by coupling linear propagation matrices with nonlinear activation operators, thereby establishing a unified computational platform for both forward modeling and inverse design. The existing deep neural networks enable rapid prediction of EM responses directly from geometric parameters [25], achieving a computational acceleration of three orders of magnitude compared to conventional FDTD simulations. Concurrently, inverse design [26–28] has been realized through gradient-based optimization of target wavefront specifications, significantly advancing design efficiency. Mainstream data-driven methods can be broadly categorized into several types. The first category primarily employs discriminative models, such as Multilayer Perceptrons (MLP) [29] or Convolutional Neural Networks [30–32], which learn direct mappings from target responses to parameters. However, the efficacy of these models fundamentally depends on the quality and scale of the training data. To achieve high accuracy, researchers often must devote significant effort to collecting and curating massive datasets, with some models requiring tens of thousands of computationally intensive simulation data points [33]. Preparing such large-scale datasets is extremely time-consuming and constitutes a major computational bottleneck. The second category utilizes generative models, such as Generative Adversarial Networks [26,34–36], Variational Auto-encoders [37,38], and more recently, Diffusion Models [39–41], which have demonstrated strong capabilities in exploring vast design spaces and generating diverse, novel meta-atom designs. Although these models can generate high-quality and diverse structures, they introduce significant additional model complexity, are computationally intensive to train and sample from, and still require high-quality initial datasets to ensure generation quality. A third category incorporates physical equations, such as Physics-Informed Neural Networks [11,24,25,42,43], which integrate physical laws directly into the loss function to reduce reliance on pure data. These approaches, whether reliant on massive data curation or the incorporation of physical priors, face inherent limitations. This highlights the urgent need for streamlined, data-efficient inverse design frameworks that navigate the trade-off between model complexity, data requirements, and computational cost.

In this study, we demonstrate a computationally efficient inverse design framework that shifts the paradigm from a conventional top-down, task-specific optimization to a bottom-up, modular assembly process. As schematically illustrated in Fig. 1, the fundamental optical functions (focusing, beam splitting, etc.) realized by basic meta-atoms are conceptualized as discrete "LEGO blocks". The core of our framework is a bidirectional deep neural network that acts as an "intelligent assembler". It establishes an efficient, inverse mapping between the user-specified target function space (single function, double function, triple function...) and the structure parameter space of the pre-existing module library. It effectively translates complex "function stacking" requirements into tangible "structural realization" scenarios, bypassing the need to redefine a new global optimization problem for each new combination of functions. As a proof-of-principle example, the network successfully predicted a multifunctional metalens exhibiting either orthogonal polarization selectivity or single-polarization operation characteristics. Full-wave EM simulations and experimental validations of customized electric field distributions are systematically conducted throughout the design cycle. The proposed innovative network establishes a new paradigm for intelligent photonic design, paving the way for developing multifunctional and compact meta-devices with unprecedented accuracy.

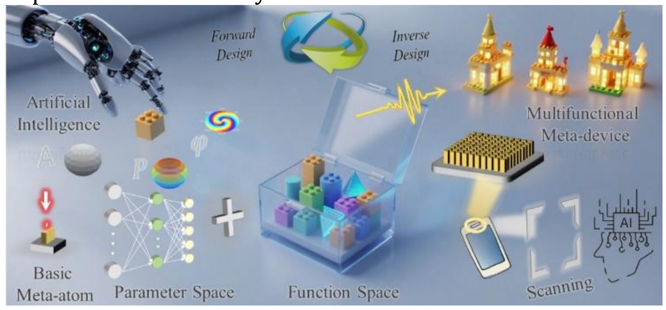


Figure 1 Schematic diagram of multifunctional metasurface design based on a bidirectional neural network, employing a functional modularity and combinatorial design approach. The core concept involves conceptualizing fundamental optical functions—namely, the modulation of amplitude (A), phase (φ), and polarization (P) within individual meta-atoms—as reusable functional modules. This modular analogy is illustrated by symbolic elements: the "AI hand" (upper left) and "scanning brain" (lower right) represent the intelligent DNN agent that selects and verifies the "modules"; the central "Function Space" box represents the user-defined desired functions; and the "Forward Design" and "Inverse Design" arrows on its "lid" symbolize the core bidirectional process connecting function to parameters. The multiple "castles" (multifunctional meta-devices) in the upper right illustrate the framework's combinatorial versatility, showing how one set of "functional modules" (our meta-atom library) can be combined to build many different "models" (multifunctional devices). The bidirectional neural network facilitates efficient global optimization within the complex structural parameter space, enabling the precise identification of the optimal structural design solution, simultaneously exhibiting all prescribed functions. Subsequently, this optimized solution is spatially integrated and inverse-engineered into a single multifunctional metasurface structure through a systematic methodology.

2 Results and Discussion

2.1 Principle and Characteristics of Meta-devices

As illustrated in Fig. 2(a), this investigation employs an optimization-based design approach for silicon dielectric nanopillars characterized by a high refractive index (relative permittivity ≈ 11.9). These geometrically tunable nanopillars with variable dimensions are periodically arranged on a 500-μm-thick silicon substrate with lattice constant Λ = 110 μm to induce Fabry-Pérot resonances capable of achieving phase control for the orthogonal polarization components. EM simulations are conducted using the software CST Microwave Studio, employing periodic boundary conditions in the *xy*-plane and an open boundary condition along the *z*-direction. A transverse electric (TE)-polarized plane wave with fixed incidence angle θ = 0° illuminates the meta-atom, while a probe positioned at the center of the *x*-plane is utilized to monitor the normalized transmission coefficients and phase responses, as quantified in Fig. 2(b). Systematic parameter sweeping across the 5041 (71^2) discrete combinations reveals non-monotonic transmission characteristics: this parametric exploration achieves full 0–2π phase coverage for both TE and transverse magnetic (TM) polarization state. For fixed nanopillar width w, the transmission magnitude initially increases with length l before exhibiting attenuation beyond critical dimensions. Notably, the resultant phase shift distributions demonstrate significant spatial nonuniformity, as evidenced by the irregular projection regions in Fig. 2(c)-(d). This parametric analysis underscores the fundamental challenge in establishing deterministic relationships between geometric parameters (w, l) and target polarization-phase responses, thereby exposing inherent limitations in conventional design methodologies reliant on exhaustive parameter screening (see Supplementary Note 1 for quantitative analysis).

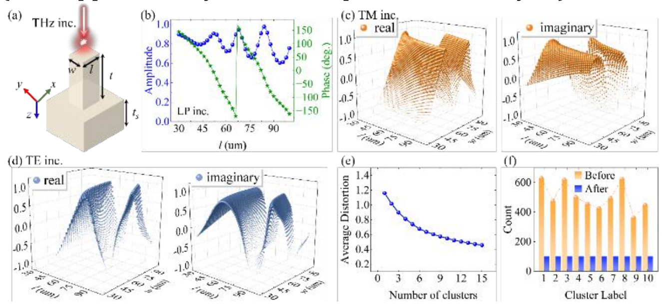


Figure 2 The structural configuration and simulated spectral characteristics of the proposed meta-atom. (a) Schematic illustration of the meta-atom architecture supporting illumination with arbitrary polarization states. (b) Quantitative analysis of TE-polarized transmission amplitude and phase response as functions of the geometric parameter l under normal incidence. (c)-(d) Spatial mapping of real and imaginary components of transmission coefficients derived from systematic parametric sweeps across the (l, w) design space, where l and w denote the longitudinal and transverse dimensions of the nanostructure, respectively. (e) Evaluation of average Euclidean distortion as a function of the candidate number of K-means clusters in the original four-dimensional electromagnetic-response space. (f) Comparative visualization of dataset distribution characteristics before and after implementing the data distillation strategy.

2.2. Hybrid Inverse Design Framework

To enable efficient and scalable inverse design of multifunctional metasurfaces, we develop a hybrid neural network framework grounded in a functional modularity philosophy. Traditional single-function design methods, as shown in Fig. 3(a), typically require a dedicated network for each task and start from a single global target function. If the design requirements change, such as adding a vortex function to a lens, a new and more complex global function must be defined, and the entire optimization process must be re-run from scratch. Our approach starts from a physical perspective to extract reusable functional modules, analogous to building blocks. Through a carefully designed deep neural network model, whether the user's requirement involves a single function or a combination of multiple functions, the optimal components can be identified in the parameter space and assembled into complex structures—namely, multifunctional meta-devices as illustrated in Fig. 3(d). This reusable workflow significantly improves the design efficiency of metastructures.

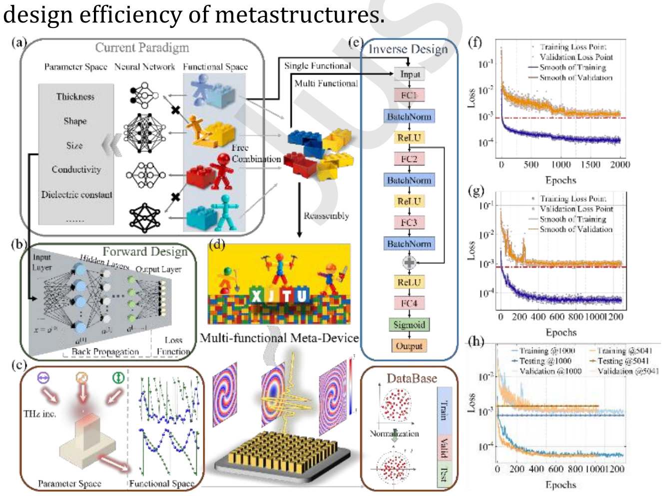


Figure 3 Design strategy and performance of a LEGO-inspired multifunctional metasurface. (a) Conventional one-to-one mapping design based on function and parameters, which offers limited functionality and generalizability. (b) Forward design workflow, which maps structural parameters to spectral responses. (c) Full-wave EM simulations of meta-atoms under non-rotated conditions (θ=0°) across 0.6-0.8 THz, demonstrating polarization-dependent complex amplitude responses. The dataset was normalized, randomly split, and used to train neural networks. (d) Functional modularity concept, enabling the combinatorial inverse design of multifunctional meta-devices from a library of basic functions. (e) By leveraging a pre-trained forward network as a physical validator and optimizing in the response domain, this pipeline ensures that the inverse design network converges to a physically self-consistent set of structural parameters. Training dynamics showing loss evolution across (f) the forward network and (g) the hybrid network, evaluated on training, validation, and test datasets. (h) Comparative analysis of loss functions in the hybrid network before and after data distillation.

To construct a robust dataset for the inverse design of metasurfaces, we develop a comprehensive data processing pipeline, as illustrated in Fig. 3(c). Transmission spectra corresponding to both TE- and TM-polarized incident light are collected across a pre-defined structural parameter space, wherein the resonant unit size varies between 30 μm and 100 μm. Phase values of 0 and 2π are physically equivalent but numerical discontinuous. Therefore, instead of using amplitude and phase, the real and imaginary parts of the two complex transmission coefficients are first assembled into a four-dimensional response vector. K-means clustering and cluster-wise sampling are performed in the original, unnormalized response space. By quantitative analysis of the mean Euclidean distance between each data point and the centroid of its assigned cluster across varying values of the cluster number k, we observe that the average intra-cluster variance initially decreases rapidly with increased k, beyond which the rate of improvement plateaued markedly, exhibiting characteristic elbow-shaped convergence behavior, as represented in Fig. 2(e). Considering the gradual flattening of the distortion curve in Fig. 2(e), the near-plateau of the silhouette score (see Figure S3 in Supplementary Note 3), and the prescribed distilled-data budget, k=10 is selected as a practical compromise between response-space resolution and data efficiency. This stratified sampling approach substantially alleviates class imbalance in the dataset, enhancing model generalization in downstream tasks. After sample selection, the response components and structural parameters are separately normalized using MinMax normalization and subsequently used as inputs to the network.

Building upon the processed dataset, we implemented a hybrid neural network framework comprising a fixed forward model and a trainable inverse model to capture the bidirectional mapping between optical responses and structural parameters. The forward prediction module employs a deep multilayer perceptron (MLP) architecture for spectral response modeling, which has six sequentially connected layers with 128-256-256-128-32-4 neurons (see Supplementary Note 8 for the full ablation study), as outlined in Fig. 3(b). The model maps the two structural parameters to the four components of the complex optical response. An adaptive learning rate scheduler is integrated with early stopping regularization to mitigate overfitting. Model training proceeds in batch mode for up to 2000 epochs, following random partitioning of datasets. The mean squared error (MSE) computation occurs after batch normalization layers, ensuring input magnitudes remain within stable numerical ranges during gradient propagation. The detailed mathematical formulation of the objective function is provided in the Supplementary Note 4. After training, the forward model achieved optimal convergence, with the loss function variation depicted in Fig. 3(f). At this stage, the trained neural network demonstrates sufficient accuracy to functionally replace EM simulation software in predicting the complex amplitude response of meta-atom during inverse design processes.

After training and fixing the forward model, the entire architecture is optimized to minimize the discrepancy between the reconstructed and target optical responses through the complete inverse–forward prediction pipeline. Specifically, for a given target optical response *r*, the inverse model first infers a set of candidate structural parameters $\hat{p} = f_{inv}(r)$, which are subsequently forwarded to the trained forward model to predict the corresponding optical response $\hat{r} = f_{inv}(\hat{p})$. The training process is guided by an MSE loss function. Crucially, this architecture mitigates the practical difficulty associated with the one-to-many inverse mapping by using the pre-trained forward network as a response-domain surrogate validator and computing the loss in the response domain, so that the inverse model is optimized for optical-response consistency rather than a unique geometric label. In this way, the inverse network is guided to converge to a physically self-consistent candidate set of structural parameters for a given target response, rather than being constrained to recover a unique inverse solution. The complete mathematical definition of this loss is available in the Supplementary Note 4.

As illustrated in Fig. 3(e), the inverse subnetwork adopts a deep residual MLP architecture with batch-normalized hidden layers to ensure stable gradient propagation. A Sigmoid activation function is applied to the final output layer to ensure that the predicted structural parameters are confined within a physically admissible range. This architectural refinement is especially critical for the inverse design of metasurfaces, where minute variations in geometric parameters—such as sub-wavelength scale adjustments on the order of 10 μm—can induce abrupt, non-monotonic shifts in the EM response. The trajectory of the loss function throughout the training procedure is presented in Fig. 3(g).

As illustrated in Fig. 3(h), the application of the proposed data distillation framework yields a 12.06% reduction in test error from 8.969×10^{-4} to 7.887×10^{-4}. The complete training pipeline required 1,387.26 s when using the full 5,041-sample dataset and 213.21 s when using the distilled 1,000-sample dataset. The latter corresponds to an 84.6% reduction in the complete training time. After training, the inverse network required 0.578 ms to generate the structural parameters of one meta-atom. Sequential generation of the 3,600 meta-atom parameter sets for the polarization-multiplexed multifunctional device therefore required approximately 2.08 s. A detailed quantitative comparison is provided in Supplementary Note 7. For the same 3,600 meta-atom parameter sets, a straightforward exhaustive lookup-table baseline using the complete 5,041-entry response library would require 3,600×5,041=18,147,600 candidate-response distance evaluations. These results quantify the data, training-time, and inference-efficiency gains achieved within the present parametric design space; the upfront full-wave data-generation cost can be amortized across repeated design tasks, as discussed in Supplementary Table S2.

The 2D-to-4D parametric design space used in this work is relatively low-dimensional and was an intentional choice to balance design flexibility with computational efficiency. As detailed in Table 1, with additional information provided in Supplementary Note 6, those high-dimensional approaches are often forced to use computationally "heavyweight" networks. These complex models inherently require massive datasets to solve single-function design problems. In contrast, our work focuses on the distinct challenge of passive combinatorial multifunctionality. Our streamlined 2D parameterization enables the use of a "lightweight" network with extremely high data efficiency (~1,000 distilled samples) to achieve a flexible combination of functions—a capability most high-dimensional, single-function models do not address.

Table 1 Comparison of data efficiency and problem complexity between this work and representative inverse design methods

Design Method	Complexity	Design Space		Dataset Size	Multifunctional
		Parameter Space	Response Space		
Primary network and an Auxiliary network [44]	Moderate	5D	201D	~30,000	No (Single function: Chiral)
Visual Attention Network[45]	High	5D	600D	~9,000	No (Single function: EIT)
Generative Adversarial Network[46]	High	64 × 64 (pixelated)	32D	~8,000+	No (Single function: Spectrum)
Conditional Generative Adversarial Network[47]	High	80 × 80 (pixelated)	181D	~20000	No (Single function: Multi-resonant)
Transformer Network[47]	High	20D			
Transfer Learning[48]	Moderate	5 × 5/10 × 10/20 × 20 (pixelated)	181D	~10000	(Single function: Far-field)
Unet++ Conditional Generative Adversarial Network[49]	High	16 × 16 (pixelated)	201D	~20000	Yes [Material-Tunable]
Tandem Generative Network[38]	High	32 × 32 (pixelated)	102D	~109600	No (Single function: Spectrum)
Combinatorial DNN [This work]	**Low**	**2D**	**4D**	**~1000**	**Yes (Combinatorial design for passive multifunctionality via functional modularity.)**

2.3. Fabrication and characterization

To verify the mock-up, the experimental validation of a metalens is firstly conducted through systematic prototype development and performance characterization, and its design framework is governed by distinct phase modulation principles derived from diffraction theory. Specifically, the phase profile of a monofocal metalens must satisfy the hyperbolic relationship:

$$\phi(r) = \frac{2\pi}{\lambda}\left(\sqrt{r^2 + f^2} - f\right)$$

* MERGEFORMAT (1)

where *r* denotes the radial coordinate, *λ* represents the operational wavelength, and f corresponds to the focal length. Upon completing computational determination for functional phase distributions, the trained neural network is employed to generate corresponding meta-atom parameters, which are subsequently precisely arranged in predetermined spatial configurations to fabricate the multifunctional meta-devices prototype.

By assembling 3600 discrete meta-atoms in a 60 × 60 square lattice, a square metalens sample with a lateral size of 6.6 mm × 6.6 mm was fabricated. The meta-atoms exhibit parabolic phase retardation to incident radiation. The field distribution of the incident light propagating through the metalens is generated by CST Microwave Studio (see Supplementary Note 2 for quantitative analysis of stochastic structural parameters). Fig. 4(a) depicts the electric field distribution in both the xz-plane and yz-plane, revealing that the electric-field intensity reaches its maximum at 3836.36 μm, demonstrating close agreement with the designed value of 4000 μm. Fig. 4(b) illustrates the electric field distribution in the xy-plane, where a distinct focal spot is observed at the focal point (0, 0). Through the application of Fresnel's approximate diffraction formula:

$$E(x,y,z)=\frac{e^{ikz}}{i\lambda z}\iint E(x',y',0)\,e^{\left\{\left(\frac{k}{2z}\right)\left[(x-x')^2+(y-y')^2\right]\right\}}dx'dy'$$

* MERGEFORMAT (2)

It is observed that wavefront evolution undergoes complex superposition during propagation, characterized by progressive phase modulation and diffraction effects. Consequently, wavefront convergence induces gradual narrowing of the transverse electric field distribution (xy-plane), with phase coalescence and amplitude maximization at the focal point. Fig. 4(c) reveals distinct diffraction patterns and convergence behaviors in pre-focal and post-focal regions, demonstrating that the metalens achieves spatial accumulation of incident radiation through precise phase modulation and diffraction engineering, transforming uniform field distributions into subwavelength focal spots.

The DRIE-related process effect is illustrated in Fig. 4(d), and the fabricated device is shown in Fig. 4(g) (see Supplementary Note 9 for more details on the selection of fabrication techniques). Near-field characterization employed a THz time-domain spectroscopy system integrated with automated 2D translational scanning (see Fig. 4(h)). As shown in Fig. 4(f), the resulting field distributions exhibit exceptional focusing performance, with a coarse scanning area of 8×8 mm² and a refined focal region scan of 1.5×1.5 mm². Based on the intensity profiles presented in Fig. 4(i), the beam widths along the x-axis are determined using the D4σ method derived from the second-order moments. Initially, the total optical power, corresponding to the integral of the intensity distribution, and the centroid coordinates are computed from the measured/simulated intensity data. The centroid coordinates, defined as the first-order moments of the intensity distribution, are obtained as the weighted averages in the x- and y- directions, respectively. Subsequently, the second-order moments of the intensity distribution relative to the centroid are calculated, yielding the squared beam radii along the respective axes. Finally, the beam width for each axis is defined as four times the standard deviation. The simulated beam width along the x-axis is determined to be 0.1934 mm, whereas the experimentally measured value is 0.3241 mm, corresponding to a broadening of 0.1307 mm or approximately 67.6%.

To identify the main origin of this broadening, we further combined stochastic perturbation simulations, DRIE process-window analysis, and SEM-based dimensional metrology, as summarized in Supplementary Notes 2, 9, and 10 and Table S6. The results indicate that lateral critical-dimension shrinkage and corner-shape distortion are the dominant fabrication-related contributors, whereas etching-depth nonuniformity mainly acts as a secondary factor. The measured lateral characteristic dimensions of the fabricated structures are reduced to 18.42, 19.54, and 18.98 μm for the dual-mode, monofocal, and vortex-generating metalenses, respectively, corresponding to dimensional error rates of 38.9%–40.8% relative to the designed minimum dimension. These deviations directly modify the local phase response of the meta-atoms and therefore contribute to beam broadening and background artifacts.

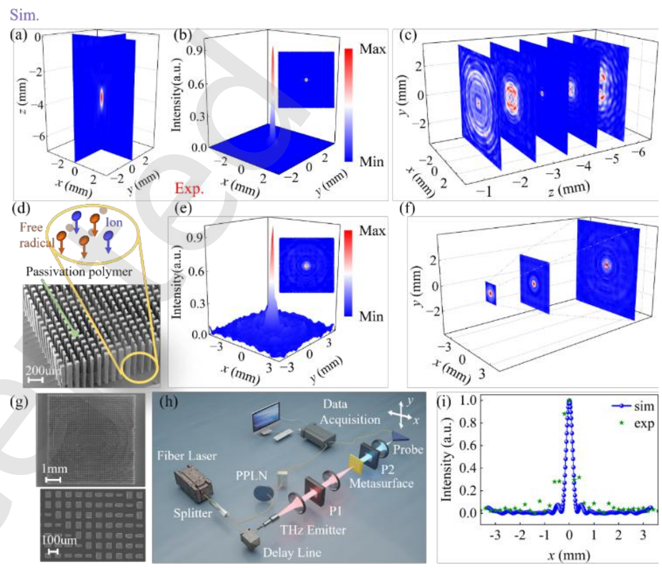


Figure 4 Focusing characteristics of the monofocal metalens consisting of 3600 meta-atoms arranged in a 60 × 60 square lattice with a lateral size of 6.6 mm × 6.6 mm. (a) Cross-sectional electric field distributions in xz- and yz-planes from numerical simulations. (b) Simulated transverse field distribution in the xy-plane. (c) Longitudinal field evolution at varying propagation distances. (d) Cross-section of the trench under the DRIE process showing the different amounts of ions and free radicals that reach the bottom due to the loading effect. (e) Experimentally acquired intensity map in the xy-plane. (f) Scanning-area-dependent field characterization in the xy-plane under three distinct scanning configurations, with detailed scan windows, step sizes, and acquisition rates summarized in Supplementary Note 11 and Table S7. (g) Scanning electron microscopy (SEM) image of the fabricated device through DRIE. (h) Schematic of THz time-domain spectroscopy system, comprising femtosecond laser pump, electro-optic sampling module, and XYZ nanopositioning stage. (i) Comparative intensity profiles at focal position from experimental measurements and computational modeling.

Vortex beams have garnered considerable attention for applications in particle manipulation, optical communications, and high-density data storage. In this study, we demonstrate a metasurface-based vortex beam generator with a topological charge of l = 3, realized through a dual-phase modulation mechanism. The engineered phase distribution is governed by

$$\phi(r)=\frac{2\pi}{\lambda}\left(\sqrt{r^2+f^2}-f\right)+l\theta$$

* MERGEFORMAT (3)

where θ represents the azimuthal angle, l denotes the topological charge, r is the radial coordinate, λ indicates the operational wavelength, and f signifies the focal length. The first term - hyperbolic in nature - compensates for the optical path delay to achieve axial focusing, while the second term introduces an azimuthal phase gradient of 2πl, imparting the beam with orbital angular momentum. By leveraging the Fourier superposition principle implementation, this superimposed phase profile enables constructive interference at the focal plane, generating an annular intensity distribution with phase singularity ($l \neq 0$) as demonstrated in Fig. 5(g). Following phase profile derivation, the trained neural network outputs the optimized meta-atom geometric parameters that are spatially arranged to construct a square vortex-generating metalens with a lateral size of 3.85 mm.

Full-wave EM simulations revealed a toroidal intensity pattern and a helical phase profile at the focal plane of 3450 μm, closely matching the theoretical predictions as depicted in Fig. 5(b-c). For the fabricated sample, DRIE introduced structural non-uniformities primarily due to plasma density gradients and non-uniform ion energy distribution within the reaction chamber. SEM metrology of a representative 500-μm-height nanopillar (Fig. 5(d)) quantified an average height of 487.7 μm in randomly sampled regions. For the vortex-generating metalens, the near-field measurement was performed at the designed focal plane using the same THz near-field scanning system under quasi-collimated THz illumination, with detailed scan parameters summarized in Supplementary Note 11 and Table S7. The measured amplitude and phase profiles on the focal plane exhibit spiral wavefront continuity, and the orbital angular momentum modes and topological charges are consistent with the designed effects (see Fig. 5(e-f)).

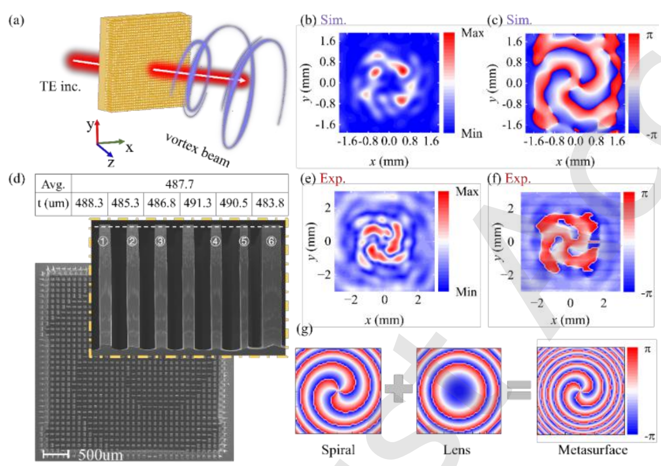


Figure 5 Performance characterization of the vortex-generating metalens. (a) Conceptual schematic of the phase-modulated focusing vortex generator. Numerically simulated (b) amplitude and (c) phase distributions at the focal plane. (d) SEM image of fabricated meta-atoms with statistical height analysis (inset table: measured heights of six representative nanopillars in a randomly sampled region). Experimentally characterized (e) amplitude and (f) phase distributions demonstrating spatial coherence. (g) Fourier-domain phase superposition principle governing vortex formation (topological charge l= 3).

A dual-mode metalens enables rapid mode switching in bio-imaging applications, allowing for high-resolution imaging of cellular responses or drug screening processes[50]. Distinct from conventional metalens that focus incident EM waves into a single point, the dual-focus metalens is designed to simultaneously or selectively focus the wavefront to two spatially distinct foci. The key lies in engineering a spatially varying phase distribution such that the transmitted wavefront undergoes constructive interference at two designated focal points. It should be particularly noted that the focus broadening and energy superposition effects will occur when the inter-focal distance is smaller than the Rayleigh length. Derived from the generalized Snell's law[51], the ideal phase profile $\Phi(x,y)$ at each meta-atom coordinate (x, y) is determined by the argument of complex superposition of two conjugate spherical waves, as defined by $\varphi(x,y)=\arg(e^{i\varphi_1(x,y)}+e^{i\varphi_2(x,y)})$, where λ denotes the operational wavelength and arg[·] represents the phase extraction operation. In this work, two foci are symmetrically positioned with respect to the y-axis at coordinates (0, −1) mm and (0, 1) mm, with a focusing length of 4000 μm. By inverse design using a pre-trained deep neural network, the phase distribution was automatically mapped to generate the geometric parameters of meta-atoms, which were spatially arranged to form a polarization-multiplexed dual-mode metalens composed of 3600 meta-atoms in a 60 × 60 square lattice. The structural parameters of all 3,600 meta-atoms were generated by the trained inverse network in approximately 2.08 s under sequential inference. The assembled full-device model was subsequently evaluated by full-wave simulation and experimental characterization.

Full-wave simulations confirm the expected diffraction-limited focusing at (0, −1) under TE-polarized illumination. When TM-polarized light is applied, the focal point switches to (0, 1). For 45° linear polarization, electric field distributions in xy-plane and xz-plane, shown in Fig. 6(a–c) and Fig. 6(d–f), respectively, exhibit distinct dual-focus characteristics. To experimentally validate the wavefront shaping performance, a dual-mode metalens sample is fabricated and characterized using the same THz near-field scanning imaging system as depicted in Fig. 4(h). Under 45° linearly polarized THz illumination, multiple scan windows were used to capture both the overall dual-focus distribution and the local field features around the focal spots. The scan configurations for different incident polarization states are summarized in Supplementary Note 11 and Table S7. A pronounced focusing effect is observed, as shown in Fig. 6(g). Further measurements are conducted under TE- and TM-polarized THz waves, as shown in Fig. 6(h–i). The scan lengths are set to 8 mm and 6 mm, with a step size 0.2 mm and a scanning speed of 500 pps. As depicted in Fig. 6(j), the discrepancy between experimental and simulated results primarily arises from differences in beam width and intensity. In addition to the discretization of phase modulation by finite-size meta-atoms, the quantitative fabrication-error analysis in Supplementary Notes 2, 9, and 10 indicates that lateral critical-dimension shrinkage, corner-shape distortion, and etching-depth nonuniformity contribute to the experimentally observed beam broadening and intensity redistribution. Among these factors, lateral dimensional deviation is the dominant contributor because it directly modifies the local phase response of each meta-atom. Edge diffraction from the finite device aperture and small incident-angle deviations during measurement may further induce focal tilting and intensity asymmetry. Despite these limitations, the experimentally observed focusing behavior remains consistent with the theoretical design across both large-scale and fine-scale regions, thereby validating the proposed inverse-design strategy and the overall metasurface fabrication workflow within the experimentally verified fabrication-tolerance window.

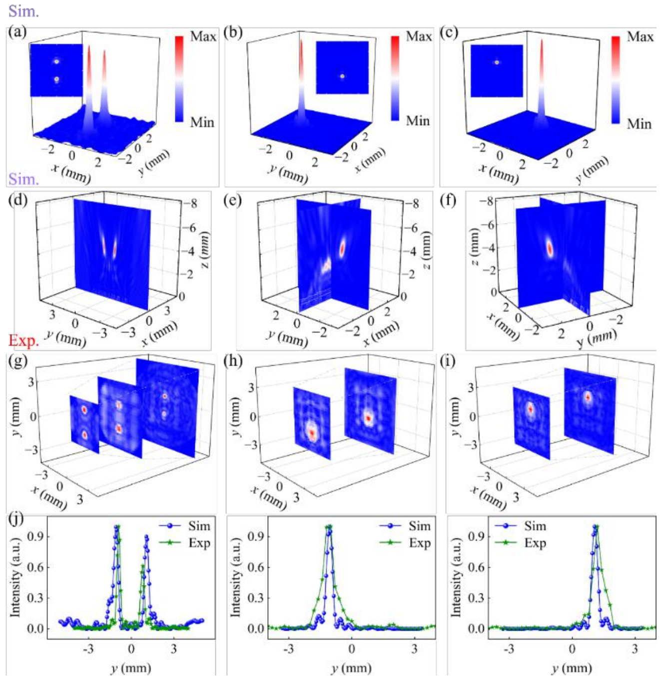


Figure 6 Characterization of polarization-multiplexed dual-mode metalens consisting of 3600 meta-atoms arranged in a 60 × 60 square lattice with a lateral size of 6.6 mm × 6.6 mm. (a-c) Simulated electric field intensity distributions on xy-plane: (a) Dual-focus mode under 45° linear polarization, (b) Single focus at (0, -1) under TE-polarization, (c) Single focus at (0, 1) under TM-polarization. Simulated xz-plane electric field distributions under (d) 45° linear polarization, (e) TE-polarization, and (f) TM-polarization. Experimentally measured xy-plane electric field distributions under (g) 45° linear polarization, (h) TE-polarization, and (i) TM-polarization, obtained using a THz time-domain spectroscopy system. (j) Comparative intensity profiles at focal position from experimental measurements and computational simulation under 45° linear polarization, TE-polarization, and TM-polarization.

3. Conclusion

In summary, inspired by the modular architecture of LEGO systems, we present a paradigm-shifting intelligent metasurface design framework that moves beyond conventional top-down optimization. This innovative methodology enables the unconstrained combinatorial integration of fundamental optical functionalities to construct multifunctional meta-devices, achieved through the synergistic coupling of application-specific deep learning algorithms with data-driven optimization techniques. Enhanced by intelligent training protocols, the prediction accuracy was improved by 12.06% compared to conventional methods (MSE reduced from 8.97×10^{-4} to 7.89×10^{-4}), while requiring only 20% of traditional training data to achieve convergence. Numerical simulations and experimental results validate multiple functional devices, including a metalens generating programmable polarization-dependent dual foci and a vortex generator, and demonstrate precise control of customizable polarization-state distributions in the THz regime. The validated framework provides critical technical support for developing next-generation devices in THz communications and photonic computing, particularly in applications demanding spatiotemporally precise EM wave manipulation. This study not only advances metasurface inverse design methodologies but also establishes an extensible theoretical framework and technical paradigm for artificial intelligence-assisted design in photonic and quantum technologies.

4 Experimental Section

4.1 Fabrication

The fabrication process commenced with 4-inch double-side polished silicon wafers (resistivity >10 kΩ·cm, thickness 1000±15 μm). A sequential cleaning procedure was performed: wafers mounted on PTFE carriers were immersed in acetone bath at ambient temperature for 10 min, followed by isopropanol immersion for 10 min. Subsequent rinsing with ultrapure water was conducted through 10 alternating immersion/drain cycles, with final drying achieved by nitrogen purging. For hard mask formation, wafers were loaded onto a PECVD chuck under vacuum. A stoichiometric SiO_2 layer (4 μm thickness) was deposited using optimized $N_2/N_2O/SiH_4$ gas ratios under stabilized plasma conditions. Post-deposition, thermal treatment was performed on a 90 ℃ hotplate for 120 s. Photolithography involved precise dispensing of 5 mL AZ4620 photoresist (bubble-free) at wafer center, followed by spin-coating at 3000 rpm for 120 s to achieve 20 μm uniform coating. Soft baking at 110 ℃ for 300 s preceded i-line exposure (45 s contact mode) through a chromium photomask. Development in 2.38% TMAH solution with horizontal orbital agitation (300 s) was followed by DI water rinse and nitrogen drying. Optical microscopy (1000× magnification) confirmed pattern fidelity before hard bake (110 ℃, 300 s). Etching proceeded sequentially: ICP etching with SF_6/CHF_3 mixture removed 4 μm SiO_2 (endpoint detection), followed by DRIE using SF_6/C_4F_8 pulsed cycles (Bosch process) for 500 μm silicon etching. Post-etch inspection ensured photoresist integrity prior to O_2 plasma ashing. Final processing included BOE immersion (HF: NH_4F=1:6, 1800 s) for SiO_2 removal, concluding with megasonic cleaning and supercritical CO_2 drying.

4.2 Measurement

A scannable THz near-field microscopy system was employed to characterize the fabricated device, utilizing photoconductive antenna (PCA) technology for both THz wave generation and coherent detection. The system employs photoconductive antenna-based time-domain spectroscopy, operating over 0.2–2.5 THz (spectral resolution <3 GHz) with all-fiber-coupled architecture enabling miniaturized detection module. In the THz generation arm, radiation emitted from the PCA is collimated by a hyperhemispherical silicon lens and subsequently shaped into a quasi-plane wave using an off-axis parabolic mirror (OAP). Through precise synchronization of optical path lengths via a motorized delay stage, time-resolved THz waveforms are acquired through electro-optic sampling. Fourier-transform post-processing yields simultaneous extraction of amplitude spectra and phase profiles across the operational bandwidth. The sample is mounted on a motorized 3D translation stage, enabling precise spatial mapping of both the global and local near-field distributions. For transverse field maps, raster scanning was performed at a fixed z plane. The scan windows and acquisition parameters were selected according to the device and measurement purpose and are summarized in Supplementary Note 11.

Electronic Supplementary Material: Supplementary material is available in the online version of this article at https://doi.org/10.26599/NR.2026.94909057.

Data availability

All data needed to support the conclusions in the paper are presented in the manuscript and the Electronic Supplementary

Material. Additional data related to this paper may be requested from the corresponding author upon request.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (52175115, 92360306), the Fundamental Research Funds for the Central Universities (Grant. 11913224000045), and the National Natural Science Foundation of China (Grant. 62305394).

Declaration of competing interest

All the contributing authors report no conflict of interests in this work.

Author contribution statement

N.Z. conceived the idea. N.Z., D.H.H., and Y.Q.C. performed the experiment. N.Z. and Y.Q.C. discussed and analyzed the measured data. L.Y.Z., Q.R.Y., and X.F.C. supervised the theory and the measurements. N.Z., Z.L.S., D.H.H., and L.Y.Z. prepared the manuscript. All the authors have approved the final manuscript.

References

[1] Zhang, X.; Liu, Z. Superlenses to overcome the diffraction limit. Nat. Mater. 2008, 7, 435–441.

[2] Han, D. H.; Cui, Y. Q.; Hou, Y. S.; Meng, F. Q.; Chen, X. M.; Zhang, L. Y.; Yan, R. Q.; Chen, X. F. 3D-printed all-resin terahertz phase-gradient metamaterials for multiple wavefront manipulations. J. Lightwave Technol. 2025, 43, 3884–3891.

[3] Conrads, L.; Bontke, F.; Mathwieser, A.; Buske, P.; Wuttig, M.; Schmitt, R.; Holly, C.; Taubner, T. Infrared beam-shaping on demand via tailored geometric phase metasurfaces employing the plasmonic phase-change material In3SbTe2. Nat. Commun. 2025, 16, 3698.

[4] Han, D. H.; Li, W. K.; Liu, M.; Chen, X. M.; Zhang, L. Y.; Chen, X. F. Controllable deformation modulated multi-functionality for phase-gradient metamaterials. Laser Photonics Rev. 2024, 18, 2300465.

[5] Zhu, Y.; Zang, X.; Chi, H.; Zhou, Y.; Zhu, Y.; Zhuang, S. Metasurfaces designed by a bidirectional deep neural network and iterative algorithm for generating quantitative field distributions. Light Adv. Manuf. 2023, 4, 104–114.

[6] Ma, Z.; Tian, T.; Liao, Y.; Feng, X.; Li, Y.; Cui, K.; Liu, F.; Sun, H.; Zhang, W.; Huang, Y. Electrically switchable 2^N-channel wave-front control for certain functionalities with N cascaded polarization-dependent metasurfaces. Nat. Commun. 2024, 15, 8370.

[7] Yao, J.; Lai, F.; Fan, Y.; Wang, Y.; Huang, S.-H.; Leng, B.; Liang, Y.; Lin, R.; Chen, S.; Chen, M. K. et al. Nonlocal meta-lens with Huygens' bound states in the continuum. Nat. Commun. 2024, 15, 6543.

[8] Dainese, P.; Marra, L.; Cassara, D.; Portes, A.; Oh, J.; Yang, J.; Palmieri, A.; Rodrigues, J. R.; Dorrah, A. H.; Capasso, F. Shape optimization for high efficiency metasurfaces: theory and implementation. Light Sci. Appl. 2024, 13, 300.

[9] Gu, Z.; Ma, Q.; Gao, X.; You, J. W.; Cui, T. J. Direct electromagnetic information processing with a planar diffractive neural network. Sci. Adv. 2024, 10, eado3937.

[10] Guo, Z.; Tan, Z.; Zang, X.; Zhang, T.; Wang, G.; Li, H.; Wang, Y.; Zhu, Y.; Ding, F.; Zhuang, S. Polarization-selective unidirectional and bidirectional diffractive neural networks for information security and sharing. Nat. Commun. 2025, 16, 4492.

[11] Xu, Y.; Yang, J.-Q.; Fan, K.; Wang, S.; Wu, J.; Zhang, C.; Zhan, D.-C.; Padilla, W. J.; Jin, B.; Chen, J. et al. Physics-informed inverse design of programmable metasurfaces. Adv. Sci. 2024, 11, 2406878.

[12] Bao, Y.; Shi, H.; Wei, R.; Wang, B.; Zhou, Z.; Qiu, C.-W.; Li, B. Efficient gradient-based metasurface optimization toward the limits of wavelength-polarization multiplexing. Nano Lett. 2025, 25, 6340–6347.

[13] Huang, S.-H.; Su, H.-P.; Chen, C.-Y.; Lin, Y.-C.; Yang, Z.; Shi, Y.; Song, Q.; Wu, P. C. Microcavity-assisted multi-resonant metasurfaces enabling versatile wavefront engineering. Nat. Commun. 2024, 15, 9658.

[14] Park, C.; Jeon, Y.; Rho, J. 36-channel spin and wavelength co-multiplexed metasurface holography by phase-gradient inverse design. Adv. Sci. 2025, 12, 2504634.

[15] Qian, C.; Kaminer, I.; Chen, H. A guidance to intelligent metamaterials and metamaterials intelligence. Nat. Commun. 2025, 16, 1154.

[16] Ma, W.; Liu, Z.; Kudyshev, Z. A.; Boltasseva, A.; Cai, W.; Liu, Y. Deep learning for the design of photonic structures. Nat. Photonics 2021, 15, 77–90.

[17] Aigner, A.; Weber, T.; Wester, A.; Maier, S. A.; Tittl, A. Continuous spectral and coupling-strength encoding with dual-gradient metasurfaces. Nat. Nanotechnol. 2024, 19, 1804–1812.

[18] Zhu, R. C.; Wang, J. F.; Fu, X. M.; Liu, X. S.; Liu, T. H.; Chu, Z. T.; Han, Y. J.; Qiu, T. S.; Sui, S.; Qu, S. B. et al. Deep-learning-empowered holographic metasurface with simultaneously customized phase and amplitude. ACS Appl. Mater. Interfaces 2022, 14, 48303–48310.

[19] Dong, B. W.; Zhu, S. Q.; Guo, G. X.; Wu, T.; Lu, X. G.; Huang, W. X.; Ma, H.; Xu, Q.; Han, J. G.; Zhang, S. et al. Switchable Pancharatnam–Berry phases in heterogeneously integrated THz metasurfaces. Adv. Mater. 2025, 37, 2417183.

[20] Yin, Y. Y.; Jiang, Q.; Wang, H. B.; Liu, J. H.; Xie, Y. Y.; Wang, Q. H.; Wang, Y. T.; Huang, L. L. Multi-dimensional multiplexed metasurface holography by inverse design. Adv. Mater. 2024, 36, 2312303.

[21] Zhang, N.; Gao, F.; Wang, R. D.; Shen, Z. L.; Han, D. H.; Cui, Y. Q.; Zhang, L. Y.; Chang, C.; Qiu, C. W.; Chen, X. F. Deep-learning-empowered customized chiral metasurface for calibration-free biosensing. Adv. Mater. 2025, 37, 2411490.

[22] Wu, N.; Sun, Y.; Hu, J.; Yang, C.; Bai, Z.; Wang, F.; Cui, X.; He, S.; Li, Y.; Zhang, C. et al. Intelligent nanophotonics: when machine learning sheds light. eLight 2025, 5, 5.

[23] Wu, Z.; Huang, X.; Yu, N.; Yu, Z. Inverse design of a dielectric metasurface by the spatial coupled mode theory. ACS Photonics 2024, 11, 3019–3025.

[24] Ha, Y. L.; Luo, Y.; Pu, M. B.; Zhang, F.; He, Q.; Jin, J. J.; Xu, M. F.; Guo, Y. H.; Li, X. G.; Li, X. et al. Physics-data-driven intelligent optimization for large-aperture metalenses. Opto-Electron. Adv. 2023, 6, 230133.

[25] Yan, Y. M.; Li, F. J.; Shen, J. Q.; Zhuang, M. Y.; Gao, Y.; Chen, W.; Li, Y. Y.; Wu, Z. L.; Dong, Z. G.; Zhu, J. F. Highly intelligent forward design of metamaterials empowered by circuit-physics-driven deep learning. Laser Photonics Rev. 2025, 19, 2400724.

[26] Ma, W.; Cheng, F.; Xu, Y.; Wen, Q.; Liu, Y. Probabilistic representation and inverse design of metamaterials based on a deep generative model with semi-supervised learning strategy. Adv. Mater. 2019, 31, 1901111.

[27] Han, X.; Fan, Z.; Liu, Z.; Li, C.; Guo, L. J. Inverse design of metasurface optical filters using deep neural network with high degrees of freedom. InfoMat 2021, 3, 432–442.

[28] Karahan, E. A.; Liu, Z.; Gupta, A.; Shao, Z.; Zhou, J.; Khankhoje, U.; Sengupta, K. Deep-learning enabled generalized inverse design of multi-port radio-frequency and sub-terahertz passives and integrated circuits. Nat. Commun. 2024, 15, 10734.

[29] Gao, F.; Zhang, Z.; Xu, Y.; Zhang, L.; Yan, R.; Chen, X. Deep-learning-assisted designing chiral terahertz metamaterials with asymmetric transmission properties. J. Opt. Soc. Am. B 2022, 39, 1511–1519.

[30] An, S.; Zheng, B.; Shalaginov, M. Y.; Tang, H.; Li, H.; Zhou, L.; Ding, J.; Agarwal, A. M.; Rivero-Baleine, C.; Kang, M. et al. Deep convolutional neural networks to predict mutual coupling effects in metasurfaces. Adv. Opt. Mater. 2022, 10, 2102113.

[31] Zhao, Y.; Zhang, N.; Xue, S.; Zhang, L. Terahertz metasurface design based on convolutional neural network. In Proceedings of the 2025 China National Conference on Terahertz Biophysics (CTB 2024); Chang, C.; Qi, F.; Zhang, L.; Hou, L., Eds.; Springer Proceedings in Physics, Vol. 423; Springer: Singapore, 2025; pp 147–152. DOI: 10.1007/978-981-96-4886-3_23.

[32] Zhu, R. C.; Wang, J. F.; Qiu, T. S.; Yang, D. K.; Feng, B.; Chu, Z. T.; Liu, T. H.; Han, Y. J.; Chen, H. Y.; Qu, S. B. Direct field-to-pattern monolithic design of holographic metasurface via residual encoder–decoder convolutional neural network. Opto-Electron. Adv. 2023, 6, 220148.

[33] Yeung, C.; Tsai, R.; Pham, B.; King, B.; Kawagoe, Y.; Ho, D.; Liang, J.; Knight, M. W.; Raman, A. P. Global inverse design across multiple photonic structure classes using generative deep learning. Adv. Opt. Mater. 2021, 9, 2100548.

[34] Chen, H.; Zhang, Z. Inverse design of meta-surfaces based on generative adversarial networks. In Proceedings of the 4th International Conference on Laser, Optics, and Optoelectronic Technology (LOPET 2024), Chongqing, China, August 16, 2024; SPIE: Bellingham, WA, USA, 2024; Vol. 13231, Paper 132312Y. DOI: 10.1117/12.3040163.

[35] Tanriover, I.; Lee, D.; Chen, W.; Aydin, K. Deep generative modeling and inverse design of manufacturable free-form dielectric metasurfaces. ACS Photonics 2023, 10, 875–883.

[36] Zeng, Y.; Cao, H.; Jin, X. Anchor-controlled generative adversarial network for high-fidelity electromagnetic and structurally diverse metasurface design. Nanophotonics 2025, 14, 2923–2938.

[37] Xiong, B.; Xu, Y.; Li, W.; Ma, W.; Chu, T.; Liu, Y. Deep learning design for multiwavelength infrared image sensors based on dielectric freeform metasurface. Adv. Opt. Mater. 2024, 12, 2302200.

[38] Yang, H.; Yang, C.; Li, H. Enhancing high-degree-of-freedom meta-atom design precision and speed with a tandem generative network. ACS Photonics 2025, 12, 1184–1195.

[39] Zhang, Z.; Yang, C.; Qin, Y.; Feng, H.; Feng, J.; Li, H. Diffusion probabilistic model based accurate and high-degree-of-freedom metasurface inverse design. Nanophotonics 2023, 12, 3871–3881.

[40] Zhu, L.; Hua, W.; Lv, C.; Liu, Y. Rapid inverse design of high degree of freedom meta-atoms based on the image-parameter diffusion model. J. Lightwave Technol. 2024, 42, 5269–5278.

[41] Zhang, J.; Qian, C.; You, G. F.; Wang, T.; Saifullah, Y.; Abdi-Ghaleh, R.; Chen, H. S. Harnessing the missing spectral correlation for metasurface inverse design. Adv. Sci. 2024, 11, 2308807.

[42] You, G.; Qian, C.; Tan, S.; Li, E.; Chen, H. Driving deep-learning-based metasurface design with Kramers–Kronig relations. Phys. Rev. Appl. 2024, 22, L041002.

[43] Zhang, J. A.; You, J. W.; Feng, F.; Na, W. C.; Lou, Z. C.; Zhang, Q. J.; Cui, T. J. Physics-driven machine-learning approach incorporating temporal coupled mode theory for intelligent design of metasurfaces. IEEE Trans. Microwave Theory Techn. 2023, 71, 2875–2887.

[44] Ma, W.; Cheng, F.; Liu, Y. Deep-learning-enabled on-demand design of chiral metamaterials. ACS Nano 2018, 12, 6326–6334.

[45] Yin, S.; Zhong, H.; Huang, W.; Zhang, W. Deep-learning-enabled design of terahertz high-Q metamaterials. Opt. Laser Technol. 2025, 181, 111684.

[46] Liu, Z.; Zhu, D.; Rodrigues, S. P.; Lee, K.-T.; Cai, W. Generative model for the inverse design of metasurfaces. Nano Lett. 2018, 18, 6570–6576.

[47] Huang, Y.; Feng, N.; Cai, Y. Artificial-intelligence-generated terahertz multi-resonant metasurfaces via improved transformer and CGAN neural networks. J. Lightwave Technol. 2024, 42, 1518–1525.

[48] Fan, Z. X.; Qian, C.; Jia, Y. T.; Chen, M.; Zhang, J.; Cui, X. S.; Li, E.-P.; Zheng, B.; Cai, T.; Chen, H. S. Transfer-learning-assisted inverse metasurface design for 30% data savings. Phys. Rev. Appl. 2022, 18, 024022.

[49] Xia, H. K.; Chen, S.-L.; Wang, Y. X.; Zhao, Y. J.; Jia, H. P.; Yang, R. C.; Guo, Y. J. Deep-learning-assisted intelligent design of terahertz hybrid-functional metasurfaces with freeform patterns. Opt. Laser Technol. 2025, 181, 112041.

[50] Badloe, T.; Kim, Y.; Kim, J.; Park, H.; Barulin, A.; Diep, Y. N.; Cho, H. S.; Kim, W. S.; Kim, Y. K.; Kim, I. et al. Bright-field and edge-enhanced imaging using an electrically tunable dual-mode metalens. ACS Nano 2023, 17, 14678–14685.

[51] Yu, N.; Genevet, P.; Kats, M. A.; Aieta, F.; Tetienne, J.-P.; Capasso, F.; Gaburro, Z. Light propagation with phase discontinuities: generalized laws of reflection and refraction. Science 2011, 334, 333–337.

Electronic Supplementary Material

Lightweight deep neural network empowers multifunctional metasurface design for arbitrary wavefront manipulations

Nan Zhang, Donghai Han, Zhonglei Shen, Yuqing Cui, Liuyang Zhang, Ruqiang Yan, and Xuefeng Chen

School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China

Address correspondence to Donghai Han, donghai19@stu.xjtu.edu.cn; Liuyang Zhang, liuyangzhang@xjtu.edu.cn

Supporting information to https://doi.org/10.26599/NR.2026.94909057

Supplementary Information

This file includes: Supplementary Notes 1 to 11, Figs S1 to S6, Tables S1 to S7.

Supplementary Note 1: Investigation of Transmission Spectral Modulation Mechanisms through Meta-atom Geometrical Parameter Optimization.

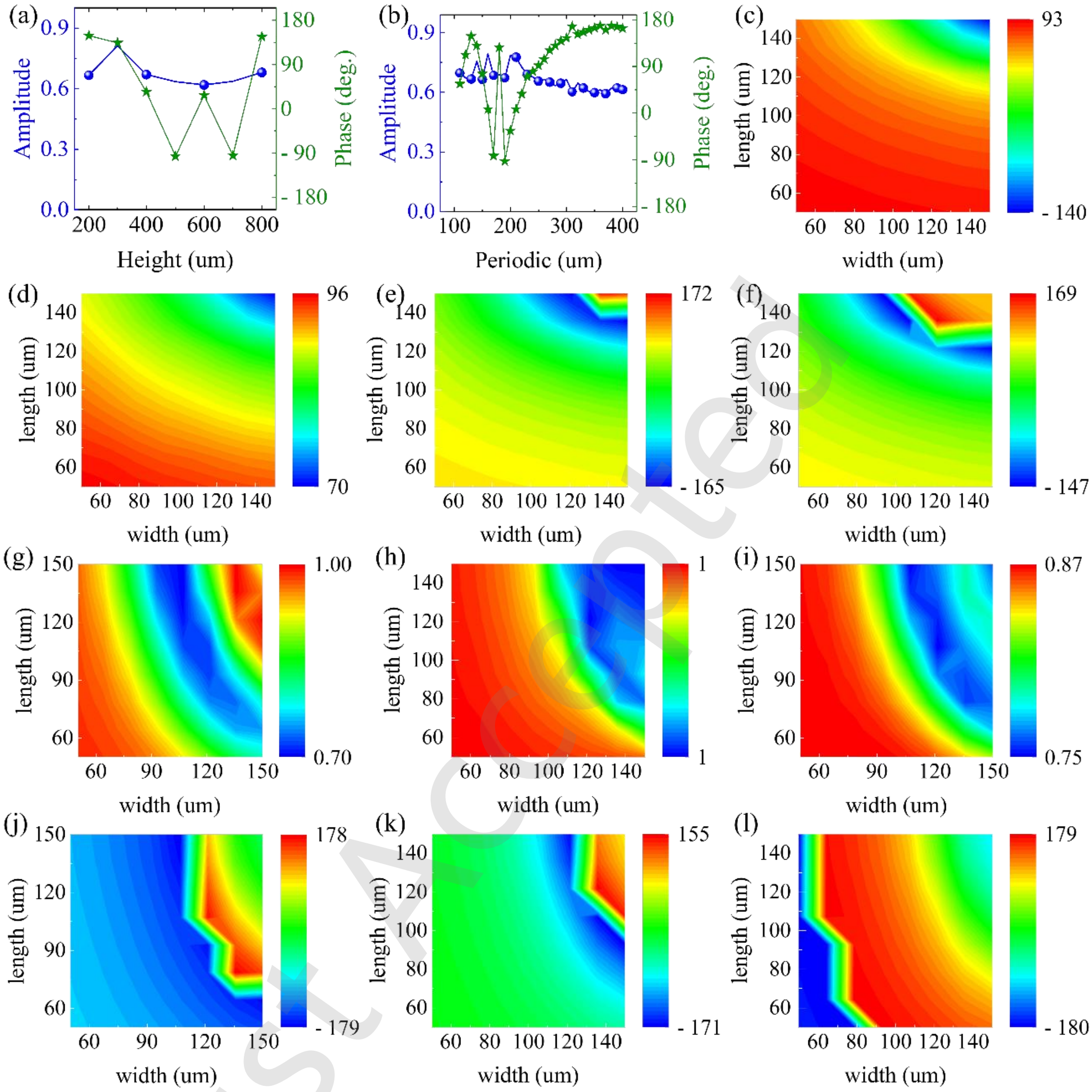


Figure S1 Influence of meta-atom geometry on electromagnetic response. (a) Simulated transmission spectra as a function of nanopillar height, with other parameters fixed. (b) Transmission response under varying structural periodicity. (c–f) Phase coverage achieved at different nanopillar heights: 100 µm, 300 µm, 500 µm, and 700 µm, respectively. (g–i) Transmittance as a function of substrate thickness: 100 µm, 300 µm, and 700 µm. (j–l) Corresponding phase modulation performance for the same substrate thicknesses.

Prior to device design, we systematically investigated how key geometric parameters modulate the electromagnetic responses. With the structural length (l) and width (w) set to 150 µm and the height (h) set to 500 µm, the periodicity parameter p exhibited significant influence on transmission characteristics within the 110-400 µm range. As shown in Fig. S1(b), numerical simulations revealed a decrease in transmission coefficient with increasing periodicity. This behavior indicates that excessively large periodicities compromise transmission efficiency. However, overly small periods are also impractical due to fabrication constraints and the need to maintain adequate spacing between adjacent meta-atoms in the final metasurface assembly. Taking into account both optical performance and manufacturing feasibility, we selected a periodicity of 110 μm as a balanced choice.

The phase modulation capacity showed strong dependence on height when maintaining $l = w = 150$ μm and $p = 300$ μm. As illustrated in Fig. S1(a), the phase coverage range demonstrated correlation with structural height within the 200-800 μm investigation window. Quantitative analysis of phase profiles is presented in Fig. S1(c)–S1(f), corresponding to $h =$ 100μm, 300 μm, 500 μm, and 700 μm, respectively. These results demonstrate that lower heights lead to a narrower phase modulation range, suggesting that insufficient height may limit phase modulation capabilities. However, excessively tall structures pose fabrication challenges due to the aspect ratio limitations. Considering this trade-off, a height of h=500 μm was selected to balance phase tunability with manufacturability.

Finally, we examined the effect of substrate thickness on both amplitude and phase responses. Figure S1(g)-(i) show the amplitude responses for substrates with thicknesses of 100 μm, 300 μm, and 700 μm, respectively, with transmission coefficients consistently maintained at 0.7-1, indicating minimal amplitude sensitivity to substrate thickness. Figure S1(j)–S1(l) further illustrate the corresponding phase responses, revealing complete 0-2 π coverage for substrates ≥ 300 um, while 200 um thickness fails to achieve complete phase modulation. Thus, the impact of substrate thickness on phase behavior is generally limited.

Supplementary Note 2: Investigation of the Impact of Stochastic Structural Parameter Perturbations on Focusing Performance.

To systematically evaluate fabrication tolerances and validate the robustness of the design workflow, we introduced stochastic perturbations into the structural parameters generated by the trained neural network prior to metasurface construction, with subsequent electromagnetic responses quantitatively characterized through full-wave simulations. Figure S2(a) and (b) demonstrate the focusing performance under ±5 μm and ±10 μm deviations in nanopillar length and width, respectively, revealing negligible intensity fluctuations in background regions. In Fig. S2(e) and (f), in addition to ±4 μm perturbations in both length and width, height variations of ±5 μm and ±10 μm were introduced, respectively. The focusing profiles remain essentially unaltered, with deviations again confined to the background region. Similarly, extending lateral deviations to ±6 μm alongside height errors of ±12 μm and ±15 μm as depicted in Fig. S2(g) and (h). Despite the increased fabrication deviations, the focusing functionality is largely preserved, though with slightly more pronounced background artifacts. However, when the magnitude of structural perturbations exceeds these thresholds, the focal spot begins to exhibit noticeable distortions. Figure S2(c) and (d) present the vortex beam generation performance of the designed metasurfaces under structural perturbations. Specifically, Fig. S2(c) includes ±2 μm perturbations in the length and width of the nanostructures, whereas Fig. S2(d) combines ±6 μm planar deviations with an additional ±15 μm variation in height. In both cases, the characteristic spiral phase fronts of vortex beams remain clearly discernible, indicating the robustness of the design even under significant dimensional uncertainties. Finally, Fig. S2(k) and (l) investigate the influence of substrate thickness on the focusing performance, comparing substrates with thicknesses of 175 μm and 400 μm. Notably, performance degradation under all error conditions predominantly manifests as enhanced background noise rather than core optical function failure, confirming the resilience of our inverse design framework to process variations.

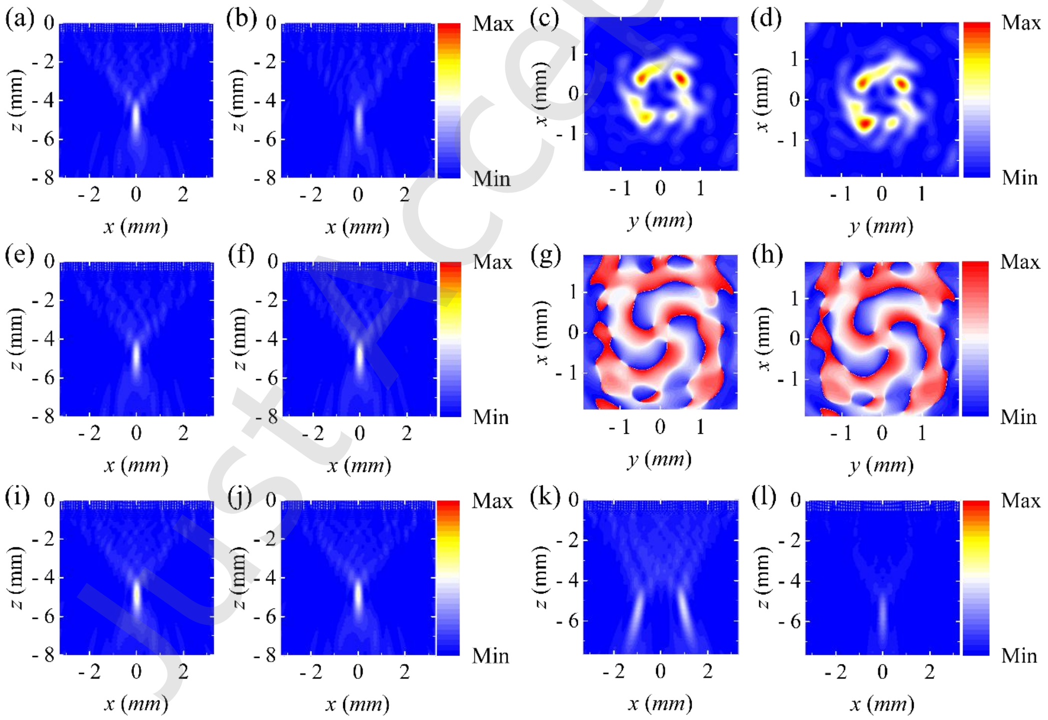


Figure S2 Quantitative analysis of electromagnetic response characteristics with introduced stochastic perturbations in neural network-generated structural parameters.

Supplementary Note 3: Data-Distillation Procedure and Coverage Evaluation

The initial database contained 5,041 full-wave electromagnetic simulations generated from a uniform (71×71) sweep of the nanopillar length (l) and width (w). Both parameters ranged from 30 to 100 μm with an interval of 1 μm. For each parameter combination, the responses of two orthogonal polarization channels were represented using the real and imaginary parts of their complex transmission coefficients. Accordingly, each sample was described by the four-dimensional response vector $r = [\mathrm{Re}(t_x), \mathrm{Im}(t_x), \mathrm{Re}(t_y), \mathrm{Im}(t_y)]$, where t_x and t_y denote the complex transmission coefficients of the two orthogonal polarization channels. The real–imaginary representation was adopted to avoid the numerical discontinuity between physically equivalent phase values close to 0 and 2π.

The data-distillation procedure was performed in the original, unnormalized four-dimensional response space. The candidate number of clusters was evaluated using two complementary metrics. The average Euclidean distortion for *k*=1 to *k*=15 is presented in Fig. 2(e) of the main text, whereas the corresponding silhouette-score analysis for *k*=2 to *k*=15 is shown in Fig. S3. The average distortion decreased from 1.1589 at (*k*=1) to 0.5484 at (*k*=10), corresponding to a reduction of 52.68%. Considering the gradual flattening of the distortion curve, the near-plateau of the silhouette score, and the prescribed 1,000-sample data budget, (*k*=10) was adopted as a practical compromise between response-space resolution and data efficiency, rather than as a mathematically unique optimum. K-means clustering was implemented with $(n_{init} = 10)$ and a fixed random state of 42.

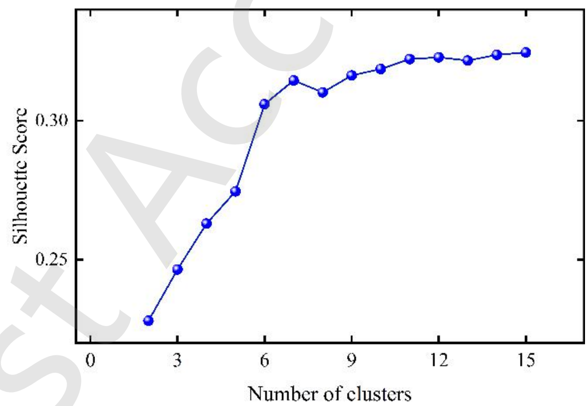


Figure S3 Silhouette scores for candidate cluster numbers from 2 to 15.

The original populations of clusters 0–9 were 607, 419, 479, 623, 511, 474, 461, 613, 509, and 345, respectively. After clustering, samples were selected independently within each response cluster. For a cluster containing more than 100 samples, 100 samples were randomly selected without replacement. If a cluster contained no more than 100 samples, all samples in that cluster were retained. For the present dataset, this procedure retained 100 samples from each of the ten clusters, yielding a distilled dataset of 1,000 samples. As shown in Fig. 2(f), all ten response clusters were retained and their sample populations were equalized.

The same selected indices were simultaneously applied to the response matrix and the structural-parameter matrix, thereby preserving the one-to-one correspondence between each electromagnetic response and its geometry. After cluster-wise sampling, the two pairs of polarization-response components and the structural parameters were separately normalized using MinMax normalization. The distilled dataset was subsequently shuffled and randomly divided into training, validation, and test subsets with proportions of 90%, 5%, and 5%, respectively. All random seeds for K-means initialization, cluster-wise sampling, dataset shuffling, dataset partitioning, and data loading were fixed at 42.

We further performed a quantitative evaluation of parameter-space coverage. The distilled dataset retained the complete ranges of both structural parameters, namely (l, w in [30,100] μm). When the (l-w) domain was divided into a (10×10) grid, all 100 cells occupied by the complete dataset also contained at least one distilled sample, giving a grid-occupancy ratio of 100%. Figure S4(a) compares the distributions of the complete 5,041-sample dataset and the 1,000-sample distilled dataset in the (l-w) parameter plane, showing that the retained samples span both the boundaries and the interior of the admissible design domain. Figure S4(b) presents the empirical cumulative distribution of the nearest-neighbor distances from all 5,041 parameter combinations to the distilled samples in the normalized parameter space. The initial cumulative probability of 19.84% at zero distance corresponds to the 1,000 samples that are themselves included in the distilled dataset. To remove this trivial zero-distance contribution and provide a more stringent assessment of the unsampled region, Fig. S4(c) shows the nearest-neighbor distance distribution exclusively for the 4,041 unselected parameter combinations, with distances reported in μm. For these samples not selected during distillation, the mean, 95th-percentile, and maximum Euclidean distances to the nearest distilled sample in the (l-w) plane were 1.342, 2.236, and 4.123 μm, respectively. These results demonstrate that the distilled samples retain the parameter-space boundaries and provide dense interior coverage without introducing a substantial unsampled region in the present two-dimensional design domain.

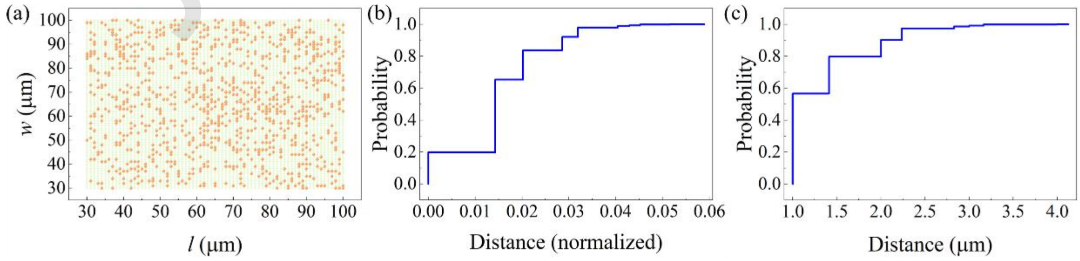


Figure S4 Parameter-space coverage of the distilled dataset. (a) Distribution of the complete 5,041-sample dataset and the 1,000-sample distilled dataset in the (l-w) parameter plane. (b) Empirical cumulative distribution of the nearest-neighbor distances from the complete parameter grid to the distilled samples in the normalized parameter space. (c) Empirical cumulative distribution of the nearest-neighbor distances from the 4,041 unselected parameter combinations to the distilled samples in the l−w plane. Distances are reported in μm.

Response-space coverage was evaluated using cluster retention, component-wise response ranges, and nearest-neighbor distances. All ten response clusters were retained. The distilled-to-complete span ratios for $\mathrm{Re}(t_x), \mathrm{Im}(t_x), \mathrm{Re}(t_y), \mathrm{Im}(t_y)$ were 0.9989, 0.9981, 0.9988, and 0.9925, respectively. For the nearest-neighbor analysis, the four response components were MinMax-normalized using the ranges of the complete 5,041-sample dataset solely to provide dimensionally balanced coverage metrics. For the 4,041 responses not retained during distillation, the mean, median, 95th-percentile, and maximum distances to the nearest distilled response were 0.05971, 0.05664, 0.10218, and 0.24502, respectively. Taken together, the retention of all ten response clusters, the preservation of more than 99.2% of the span of each response component, and the relatively small four-dimensional nearest-neighbor distances indicate that the principal support of the response space is preserved. A consolidated summary of the quantitative coverage metrics in both the structural-parameter space and the electromagnetic-response space is provided in Table S1.

Table S1 Quantitative coverage metrics of the 1,000-sample distilled dataset in the structural-parameter space and the electromagnetic-response space.

Metric	Retained proportion	Occupied 10×10 grid cells	Response clusters retained	
Value	19.84%	100/100	10/10	
Metric	Mean normalized response-space NN distance	Median normalized response-space NN distance	95th-percentile response-space NN distance	Maximum response-space NN distance
Value	0.05971	0.05664	0.10218	0.24502
Metric	Re(t_x) span retained	Im(t_x) span retained	Re(t_y) span retained	Im(t_y) span retained
Value	99.89%	99.81%	99.88%	99.25%
Metric	Mean parameter-space NN distance	Median parameter-space NN distance	95th-percentile parameter-space NN distance	Maximum parameter-space NN distance
Value	1.342 μm	1.000 μm	2.236 μm	4.123 μm

We emphasize that equalized cluster sampling is not intended to reproduce the original empirical probability density. Instead, it preserves all identified response regimes while preventing densely populated response regions from dominating the training process and improving the representation of less frequent but physically relevant responses. The distilled dataset therefore has a more balanced response distribution, while still retaining all ten response clusters identified from the complete dataset.

Supplementary Note 4: Network Details

This section provides the detailed mathematical formulations for the loss functions and network architectures referenced in the main text.

Following cluster-wise sample selection, the four electromagnetic-response components and the two structural parameters were separately normalized using MinMax normalization. The processed dataset was randomly partitioned into training (90%), validation (5%), and test (5%) subsets. The partition was random rather than class-stratified because cluster balancing had already been completed during the preceding distillation step.

Both the forward and inverse networks were trained using the Mean Squared Error (MSE) loss function. The objective function for the forward-prediction module is defined as: $L_{MSE} = \frac{1}{N}\sum_{i=1}^{N} \| \boldsymbol{y}_i - \hat{\boldsymbol{y}}_i \|_2^2$ where N denotes total number of samples in a mini-batch, $y_i \in R^d$ represents ground-truth vector for the i-th sample, $\hat{y}_i \in R^d$ indicates predicted output vector from the neural network.

For the hybrid inverse design framework, the training process is guided by a similar MSE loss function, but computed in the response domain: $L_{inv} = \frac{1}{N}\sum_{i=1}^{N} \left\| r^{(i)} - \hat{r}^{(i)} \right\|^2 = \frac{1}{N}\sum_{i=1}^{N} \left\| r^{(i)} - F_{fwd}(F_{inv}(r))^{(i)} \right\|^2$, where N is the number of training samples, r is the target optical response, and $\hat{r}$ is the response predicted by the fixed forward network (F_{fwd}) after being fed the parameters generated by the inverse network (F_{inv}).

The deep residual MLP architecture used in the inverse subnetwork implements residual modules to ensure stable gradient propagation. This architecture comprises six fully connected layers with progressive dimensional expansion (128→256→512 neurons) followed by feature compression (512→128→32→2 neurons). Each residual module implements the transformation: $h_{l+1} = h_l + \sigma(W_l h_l + b_l)$, where h_l is the hidden representation of layer l, $\sigma(\cdot)$ is the nonlinear activation function, W_l and b_l are the trainable weight matrix and bias vector, respectively.

Supplementary Note 5: Discussion on Applicability Scope and High-Dimensional Challenges.

The data distillation and lightweight network framework proposed in this work is primarily validated within a 2D parameter space (nanopillar length l and width w). In this context, the method is highly effective: based on an initial dataset from 5041 simulations, our strategy reduced the data requirement to 20% while improving prediction accuracy by 12.06%

(MSE decreased from 8.969×10^{-4} to 7.887×10^{-4}). Many practical metasurface devices, such as the metalenses and vortex beam generators demonstrated here, can be effectively described by low-to-medium dimensional parameter spaces (e.g., 2D or 3D). For these scenarios, our method offers an optimal solution for rapid, low-cost design. However, the "full-sweep" sampling strategy used to generate the initial dataset faces scalability challenges. In this work, parameters l and w were sampled from 30-100 μm with a 1 μm step, forming a 71×71 grid (5041 points). The computational cost of this brute-force sampling grows exponentially with dimensionality: extending to 3D (e.g., adding height h) would require $71^3 \approx 357{,}911$ simulations, and to 4D, $71^4 \approx 25{,}411{,}681$ simulations. This exponential growth underscores the infeasibility of full parameter sweeps in high-dimensional spaces. Our data distillation method effectively reduces the required sampling density. Achieving good performance with ~1000 samples (20% of the 2D data) effectively reduces the sampling density from 71 points/dimension to approximately $\sqrt{1000} \approx 32$ points/dimension. If computational resources allow for, say, 10,000 simulations, then in a 3D space, each dimension could have $\sqrt[3]{10000} \approx 21$ sampling points, which might still fall within the effective range of our proposed method. However, for 4D, each dimension would only have $\sqrt[4]{10000} \approx 10$ points, leading to overly sparse data and potentially a sharp performance decline. Therefore, the applicability boundary of our method in its current stage can be defined as "scenarios within low-to-medium dimensional parameter spaces (e.g., 2D-3D) where a high-quality initial dataset can be constructed through a one-time, computationally acceptable investment (e.g., tens of thousands of simulations)."

As dimensionality increases, the brute-force sampling strategy (and subsequent data processing) faces several severe challenges, often referred to as the "Curse of Dimensionality":

- Data Sparsity: Even large numbers of data points become extremely sparse in high-dimensional space, making it difficult to effectively cover and characterize the entire design space.
- Distance Metric Degradation: The effectiveness of distance-based clustering algorithms (like the k-means used here) diminishes in high dimensions, as distances between points tend to become uniform.
- Soaring Model Complexity: Neural networks require more parameters and vastly more data to learn complex high-dimensional mappings, conflicting with the "lightweight" design goal.

Extending the principles of this work to high-dimensional spaces is a clear and critical future direction. The strategy must shift from "brute-force" sampling to intelligent, adaptive data acquisition and generation. Promising pathways identified from the state-of-the-art include: Generative Models[1–3] and Active Learning[4]. Integrating our lightweight network architecture with these advanced data strategies constitutes a promising roadmap for overcoming the dimensionality barrier.

Supplementary Note 6: Comparison with Representative Inverse Design Methods

This comparison is intended to contextualize our work and highlight its specific innovations. We acknowledge that our parametric 2D→4D design space is lower-dimensional than the free-form, pixelated design spaces used in other works. However, this was an intentional choice to support our primary methodological goal: developing a lightweight, data-efficient framework for combinatorial multifunctional design. As this comparison clearly demonstrates, a trade-off exists between design space complexity, network complexity, and data requirements:

- High-Dimensional Designs: Pixelated or free-form approaches offer a larger design space but are forced to use computationally intensive, "heavyweight" network architectures. These models inherently require massive datasets.
- Low-Dimensional Parametric Designs: Even other parametric, low-dimensional designs still require substantial datasets for single-function tasks.
- Our framework intentionally uses a streamlined parametric space. This enables us to use a "lightweight" network (Low Complexity) with extremely high data efficiency. We leverage this efficiency to solve a different and equally important challenge: achieving passive, combinatorial multifunctionality, a capability that most single-task, heavyweight models do not address.

Table 1 clarifies that our work is not aiming to solve the same problem as high-dimensional, data-hungry models. Instead, it offers a distinct, highly efficient paradigm for a different class of design problems, supporting our core claims of a "lightweight" and "efficient" framework for multifunctional devices.

Supplementary Note 7: Quantitative performance comparison of this work with state-of-the-art methods.

To distinguish the one-time offline data-generation cost from the cumulative full-wave evaluations required by a conventional iterative design workflow, a workflow-level comparison is provided in Table S2.

Table S2 Workflow-level and quantitative comparison between conventional iterative full-wave design and the proposed data-distilled DNN framework for the same multifunctional metasurface design task.

Comparison aspect	Conventional iterative full-wave design	Proposed data-distilled DNN framework
Design workflow	Candidate structure → full-wave simulation → response evaluation → parameter modification → repeated simulation until the target or stopping criterion is reached	Target response → inverse network → candidate structural parameters → pretrained forward-network verification → final full-wave validation
Role of the full-wave solver	Inside the design loop.	Offline data-generation and final device validation.
Appropriate measure of simulation burden	Cumulative number of full-wave solver evaluations, (N_{eval}) , rather than dataset size. (N_{eval}) depends on the initial design, target complexity, update strategy, convergence threshold, and designer experience.	Number of offline simulations, effective training-set size, training time, and online inference time can be reported separately and are largely predetermined.
Dataset or simulation size for one task	Target-dependent and not fixed in advance.	5041 offline samples; approximately 1000 samples.
Computational-time model	$(T_{conv} = N_{eval}\overline{t}_{EM} + T_{update})$, where $(\overline{t}_{EM})$ is the mean time of one full-wave simulation and (T_{update}) includes manual or algorithmic parameter updating.	$(T_{DNN} = T_{data} + T_{distill} + T_{train} + T_{infer})$. The first three terms are one-time offline costs, whereas (T_{infer}) is the marginal cost for each new design.
Training time	Not applicable	Approximately 213s.
Online design time	Unknown before convergence.	0.578 ms per meta-atom; approximately 2.08 s for 3600 sequential predictions.
Computational-resource profile	Repeated CPU- and memory-intensive full-wave simulations.	Offline full-wave cost plus lightweight online inference.
First-device computational cost	May be low for simple, well-guided designs.	Includes the upfront cost of generating the dataset and training the model.
Cost for an additional target device	A new target generally requires another iterative search and additional full-wave simulations.	Network inference and final full-device validation.
Reusability of previous computation	Limited reuse of individual prior simulations.	Reusable distilled dataset and trained networks.
Cost scaling with the number of design tasks	The cumulative cost increases approximately with the total number of full-wave evaluations required by all design tasks: $(\sum_{k=1}^{K} N_{eval,k}\overline{t}_{EM})$.	The offline cost is paid once, after which the cumulative cost mainly increases with the much smaller inference cost: $(T_{offline} + \sum_{k=1}^{K} T_{infer,k})$.
Dependence on designer experience	High during geometry selection, parameter tuning, and stopping decisions.	Reduced during online structural generation, still need for dataset, network architecture, and constraint definition.

The meta-atom simulations were performed using CST Microwave Studio 2022. According to the original batch-simulation logs, generation of the complete 5,041-sample database required approximately 70 h on the reported computational platform.

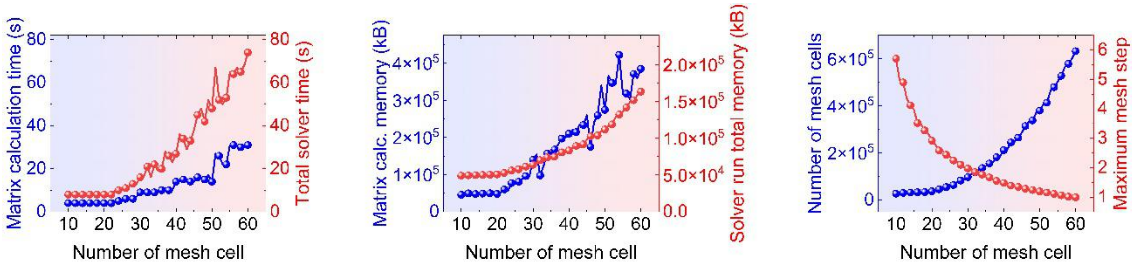


Figure S5 Relationship of Computational Resources (Including Matrix Calculation Time, Total Solver Time, Memory Usage, Maximum Mesh Step, and Number of Mesh Cells) with Simulation Mesh Density.

To further characterize the dependence of the electromagnetic-solver cost on mesh density, an auxiliary resource-scaling benchmark was conducted using the built-in horn-antenna model in CST Microwave Studio 2022 with the time-domain solver. The benchmark was performed on a computer equipped with a 6-core Intel(R) Core(TM) i5-9500 CPU operating at 3.00 GHz. As shown in Fig. S5, the matrix-calculation time, total solver time, memory usage, maximum mesh step, and number of mesh cells exhibit a nonlinear dependence on mesh density. When the mesh-number setting reached 60, the total solver time was 74 s. This auxiliary benchmark is included to characterize the resource-scaling behavior of the full-wave solver and should not be interpreted as a direct timing measurement of the metasurface unit cell.

The computational cost of the proposed workflow was separated into the one-time full-wave data-generation cost, the effective training-set size, the network-training cost, and the online inference cost. According to the original batch-simulation logs, generation of the complete 5,041-sample meta-atom database required approximately 70 h using CST Microwave Studio 2022.

Using the same timing definition, the complete training pipeline required 1,387.26 s when the full 5,041-sample dataset was used and 213.21 s when the distilled 1,000-sample dataset was used. The corresponding reduction in complete training time was $\frac{1387.26-213.21}{1387.26}\times 100\% = 84.6\%$

For the polarization-multiplexed multifunctional metalens containing 3,600 meta-atoms, an exhaustive lookup-table baseline using the complete 5,041-entry response library would require $N_{lookup} = 3,600\times 5,041 = 18,147,600$ candidate-response distance evaluations. This value represents response-matching operations rather than full-wave simulations. The lookup process was not separately timed, and no unmeasured wall-clock value is therefore assigned to it.

After the one-time training stage, the inverse network required 0.578 ms to generate the structural parameters of one meta-atom. Sequential generation of all 3,600 parameter sets therefore required approximately $3,600\times 0.578ms = 2.0808s$

Table S3 provides a direct quantitative comparison of key performance metrics between this work and representative state-of-the-art methods. The comparison includes dataset size, training time, inference time, and the final reported prediction MSE.

Table S3 Quantitative comparison of this work with state-of-the-art data-driven inverse design methods.

Design Method	Tandem Generative Network[5]	Machine Learning[6]	Anchor-Controlled GAN[7]	Generative Deep Learning[8]	This Work
Dataset Size	200,000	8,000	18,768	18,770	1,000
Training Time	16 h	5 h	4.16 h	30min	~3.5 min (213 s)
Inference Time	0.024 s	<30 s	0.42 ms	Not reported in the original publication	0.578 ms
Final MSE	1.81×10^{-3}	1.50×10^{-2}	1.12×10^{-3}	1.27×10^{-2}	7.89×10^{-4}

Supplementary Note 8: Comparison of model performance for different network configurations.

To identify a lightweight architecture that balances prediction accuracy and computational cost, the validation MSE and training time of different network configurations are compared in Table S4.

Table S4 Comparison of model performance, including best validation MSE and total training time, for different network configurations.

Architecture	Val MSE (best)	Train Time (s)	Epochs
64-64-64-64-64-64-64-64	0.001081	213.8697617	1711
128-128-128-128-128	0.001010	203.1319847	2000
128-128-128-128-128-128-128-128	0.001005	254.2848387	2000
128-128-128-128-128-128-128	0.000992	218.7674415	1949
256-256-256-256-256-256-256-256	0.000919	277.5077	2000
512-512-512-512	0.000809	157.8655	1892
128-256-256-128-32	0.000789	189.79	1971

Supplementary Note 9: Analysis of Structural Dimension Compatibility with Fabrication Process.

For terahertz (THz) metalenses operating in the 0.1-10 THz frequency range as subwavelength optical devices, their characteristic dimensions predominantly reside in the micrometer scale, requiring stringent control over fabrication accuracy, sidewall verticality, and surface planarity. Among the most widely adopted etching techniques for high-aspect-ratio microfabrication, the Bosch deep reactive ion etching(DRIE) process currently achieves a minimum industrial linewidth of approximately 1 μm. Further reduction in linewidth may lead to challenges such as polymer residue accumulation, limited gas accessibility, reduced etching rates, structural distortion, or even collapse of the microstructures. The typical achievable aspect ratio (etching depth to feature width) ranges from 10:1 to 20:1. In light of these constraints, a detailed dimensional analysis of the designed structures was conducted, as summarized in Table S5, to ensure feasibility and compatibility with the fabrication process.

Table S5 Characteristic dimensional parameter analysis of designed metalens.

Feature Size (um)	Monofocal metalens	Dual-mode metalens	Vortex-generating metalens
Max length	82.950	99.388	87.979
Min length	30.914	30.154	30.600
Max width	85.848	99.553	83.018
Min width	58.896	30.001	53.505
Max Row pitch	51.104	79.997	56.032
Min Row pitch	24.187	11.346	27.076
Max Column pitch	79.001	79.846	79.354
Min Column pitch	27.294	10.663	22.942

In summary, all critical dimensional parameters (feature sizes: 30.154-99.553 um; pitch ranges: 10.663-79.997 um) satisfy the process window constraints of Bosch etching (minimum linewidth ≥1 um, aspect ratio ≤20:1), confirming the compatibility between designed structures and manufacturing capabilities. This process-window analysis indicates that the designed meta-atoms are manufacturable in principle. However, manufacturability does not guarantee negligible process-induced deviations. The major discrepancy originates from lateral critical-dimension shrinkage and corner-shape distortion during pattern transfer and Bosch DRIE rather than from violation of the nominal process window. Therefore, the most stringent fabrication control should be applied to mask fidelity, lateral etching, sidewall passivation uniformity, and plasma loading during DRIE.

Supplementary Note 10: Three-Dimensional Topography and Critical Dimension Characterization of Microstructures Fabricated via DRIE.

Figure S6(a)-(c) respectively present the critical dimension characterization of microstructures in dual-mode metalens, monofocal metalens, and vortex-generating metalens fabricated through DRIE. Experimental measurements reveal characteristic dimensions of 18.42 μm, 19.54 μm, and 18.98 μm for the respective lens types, demonstrating significant deviations from the designed minimum dimension of 30.154 μm. These systematic processing anomalies, including over-etching, structural shrinkage, and morphological distortion, correspond to dimensional error rates of 38.9%-40.8%. Such discrepancies can be attributed to multiple factors, including photolithographic pattern transfer-induced size reduction; lateral erosion during etching processes; geometric accuracy degradation caused by mask material wear; and plasma loading effects, particularly the etching rate depression in high-density regions due to excessive reactant consumption, contrasted with accelerated etching in low-density zones.

Morphological analysis further identifies substantial geometric distortion at rectangular corners, manifesting as sharpened conical protrusions or irregular "dog-tooth" etching traces, as prominently illustrated in Fig. S6(c). These anomalies are primarily attributed to a combination of physical and chemical factors, including localized enhancement of etching at the corners caused by ion focusing effects during Bosch process cycles, incomplete coverage of the passivation layer at mask corners leading to reduced etch resistance, and the compounded influence of plasma spatial non-uniformity and byproduct accumulation, which further exacerbates etching asymmetry and geometric deformation.

Quantitative depth profiling in Fig. S6(d)-(e) demonstrates microstructure depth variations under uniform design specifications of 500.00 μm. The dual-mode lens exhibits an average etching depth of 493.00 μm with a deviation of 7.00 μm across six randomly sampled microstructures, while the vortex-generating metalens shows greater deviation with 485.82 μm and a deviation of 14.18 μm. This discrepancy results from loading effects and Aspect Ratio Dependent Etching (ARDE), which become more pronounced in densely packed or high aspect ratio structures, leading to locally reduced etching rates. The vortex-generating metalens, characterized by more complex and uneven density distribution, experiences exacerbated etch nonuniformity. Insufficient mask thickness or premature mask degradation may also terminate the etching process early. Furthermore, fluctuations in etching equipment parameters over prolonged durations can adversely affect etch depth consistency.

Figure S6(f) reveals structural integrity defects predominantly occurring at the peripheral regions of the samples. This edge-related phenomenon correlates with uneven distribution of etchant gases and plasma at the wafer edges, resulting in reduced etch efficiency, poor passivation, incomplete structures, or incomplete etch-through. Additionally, edge regions experience lower etching load, which slows the etch rate. Mask materials at edges are more susceptible to thermal damage and plasma erosion, also diminishing their protective capability and aggravating structural loss.

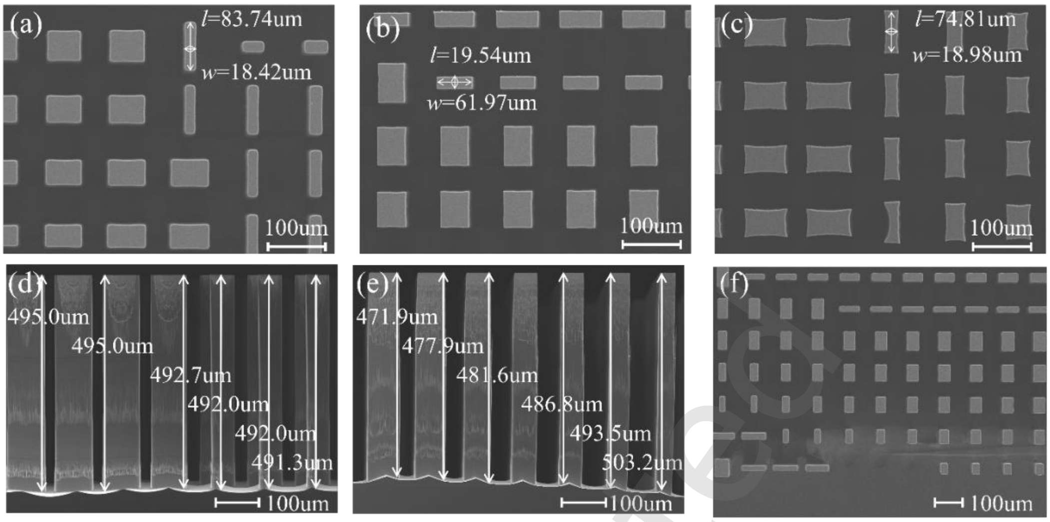


Figure S6 Micro/nano-fabricated sample dimensional characterization via scanning electron microscopy. (a) Geometrical parameter quantification of dual-mode metalens. (b) Geometrical parameter quantification of monofocal metalens; (c) Geometrical parameter quantification of vortex generating metalens; (d) Surface height distribution profile of dual-mode metalens; (e) Surface height distribution profile of vortex generating metalens; (f) Three-dimensional topographical characterization of dual-mode metalens.

Combining the SEM-based dimensional metrology in Fig. S6 with the stochastic perturbation simulations in Fig. S2, we further constructed a quantitative fabrication-error budget, as summarized in Table S6. Table S6 indicates that lateral critical-dimension shrinkage and corner-shape distortion are the dominant fabrication-related error sources in the present process, while etching-depth nonuniformity is a secondary contributor. The substrate-thickness variation and edge-related measurement factors mainly influence the background field and intensity symmetry. Therefore, the pattern-transfer and lateral-etching stages of the Bosch DRIE process require the most stringent control.

Table S6 Quantitative fabrication-error budget and tolerance assessment of the fabricated metasurfaces.

Error source	Process origin	Quantitative deviation or investigated range	Observed optical influence	Relative contribution	Control priority
Lateral CD shrinkage and corner-shape distortion	Photolithographic pattern transfer, mask erosion, lateral etching, incomplete sidewall passivation, ion-induced corner erosion during Bosch DRIE	Measured CD: 18.42, 19.54, and 18.98 μm; error rate: 38.9% to 40.8%;	Beam broadening and increased background artifacts	Dominant	Highest
		simulated lateral perturbation: up to ±10 μm	focal distortion appears beyond the tolerance range		
Etching-depth / height nonuniformity	Loading effect, aspect-ratio-dependent etching, plasma-	Designed height: 500 μm; measured depth: 493.00 and	Mainly increases background artifacts when	Secondary	High

	density gradient, nonuniform ion-energy distribution, mask degradation during long-time DRIE	485.82 μm; deviation: approximately 1.4% and 2.8%;	coupled with lateral errors		
		simulated height perturbation: up to ±15 μm	core focusing and vortex functions remain preserved within the tested range		
Substrate-thickness variation	Wafer-thickness tolerance and thinning or handling variation	Simulated substrate thickness: 175 to 400 μm;	Weaker influence on the core function; mainly contributes to background noise	Minor to moderate	Medium
Edge defects and measurement alignment	Nonuniform etchant-gas and plasma distribution at wafer edges, poor edge passivation, local mask erosion, incomplete etch-through	Localized edge defects; simulated focal position: 3836.36 μm versus designed 4000 μm; measured D4σ beam width: 0.3241 mm versus simulated 0.1934 mm	Intensity asymmetry, focal tilting, and apparent beam broadening	Minor	Medium to low
Overall tolerance window	Combined effect of fabrication and measurement uncertainty	Lateral perturbation up to ±10 μm; combined ±6 μm lateral and ±15 μm height perturbation	Target wavefront functions remain recognizable; degradation mainly appears as background artifacts rather than functional failure	Defines generalizability boundary	Process-dependent recalibration recommended

CD: critical-dimension

Note: The error ranking is based on the combined evidence from SEM metrology, stochastic perturbation simulations, and measured/simulated field-profile comparison. Because the metasurface response is nonlinear and different fabrication errors can be coupled, the listed error sources should not be interpreted as linearly additive contributions. Instead, the table provides a quantitative sensitivity ranking and identifies the fabrication steps requiring the most stringent control.

Supplementary Note 11: Near-Field Scanning Configurations for Different Metasurface Measurements

All near-field measurements reported in the main text were performed using the same THz time-domain spectroscopy near-field microscopy system equipped with a motorized XYZ translation stage. For each transverse xy-plane field map, the z coordinate was fixed at the selected measurement plane, and raster scanning was performed along the x and y directions. Therefore, the different scan windows, spatial steps, and acquisition rates do not indicate the use of different characterization instruments.

The parameter sets correspond partly to different metasurface devices and partly to different measurement configurations applied to the same device. For the monofocal and dual-mode metalenses, multiple nested scan windows were employed to characterize the field distributions at different spatial scales. For the vortex-generating metalens, the scan window was selected to capture both the annular amplitude distribution and the spiral phase profile at the designed focal plane. For the dual-mode metalens, the scanning conditions were further adjusted according to the incident polarization state and the expected focal positions. The complete scanning conditions are summarized in Table S7.

Table S7 Summary of near-field scanning parameters used for different metasurface measurements.

Device	Measurement condition	Scan window	Step size	Scanning speed	Measurement purpose
Monofocal metalens	xy-plane field mapping	8 mm × 8 mm	0.02 mm	500 pps	Large-area overview of the focusing field
		4 mm × 4 mm	0.05 mm	200 pps	Intermediate field-of-view characterization
		1.5 mm × 1.5 mm	0.05 mm	100 pps	Refined focal-region characterization
Vortex-generating metalens	Focal-plane amplitude and phase mapping	6 mm × 6 mm	0.1 mm	500 pps	Characterization of annular amplitude and spiral phase profiles
Dual-mode metalens	45° linearly polarized illumination	10 mm × 10 mm	0.1 mm	500 pps	Large-area dual-focus characterization
		6 mm × 6 mm	0.2 mm	500 pps	Intermediate field-of-view characterization
		4 mm × 4 mm	0.1 mm	500 pps	Local focal-region characterization

pps denotes the number of spatial sampling points acquired per second.

Reference

[1] Zhang Z, Yang C, Qin Y, Feng H, Feng J, Li H. Diffusion probabilistic model based accurate and high-degree-of-freedom metasurface inverse design. Nanophotonics. 2023;12:3871–81.

[2] Zhang J, Qian C, You G, Wang T, Saifullah Y, Abdi‐Ghaleh R, et al. Harnessing the Missing Spectral Correlation for Metasurface Inverse Design. Adv Sci. 2024;11:2308807.

[3] Qian C, Kaminer I, Chen H. A guidance to intelligent metamaterials and metamaterials intelligence. Nat Commun. 2025;16:1154.

[4] Qiqi Dai. High-Asymmetry Metasurface: A New Solution for Terahertz. Adv Sci. 2025:e08610.

[5] Yang H, Yang C, Li H. Enhancing High-Degree-of-Freedom Meta-Atom Design Precision and Speed with a Tandem Generative Network. ACS Photonics. 2025;12:1184–95.

[6] Liu X, Wang P, Xiao C, Fu L, Xu J, Zhang D, et al. Compatible stealth metasurface for laser and infrared with radiative thermal engineering enabled by machine learning. Adv Funct Mater. 2023;33:2212068.

[7] Zeng Y, Cao H, Jin X. Anchor-controlled generative adversarial network for high-fidelity electromagnetic and structurally diverse metasurface design. Nanophotonics. 2025;14:2923–38.
[8] Yeung C, Tsai R, Pham B, King B, Kawagoe Y, Ho D, et al. Global Inverse Design across Multiple Photonic Structure Classes Using Generative Deep Learning. Adv Opt Mater. 2021;9:2100548.