CHINESE OPTICS LETTERS

Vol. 24, No. 1 | January 2026

Neural network inspired backpropagation algorithm for nanophotonic optimization

Lingjie Fan (范灵杰)[1,2], Ang Jiang (姜 昂)[1,2], Jiajun Wang (王佳俊)[1], Wenzhe Liu (刘文哲)[1,3*], Minjia Zheng (郑敏嘉)[1**], Lei Shi (石 磊)[1,2,3,4***], and Jian Zi (资 剑)[1,2,3,4]

[1] State Key Laboratory of Surface Physics, Key Laboratory of Micro- and Nano-Photonic Structures (Ministry of Education) and Department of Physics, Fudan University, Shanghai 200433, China

[2] Shanghai Engineering Research Center of Optical Metrology for Nano-fabrication, Shanghai 200433, China

[3] Institute for Nanoelectronic Devices and Quantum Computing, Fudan University, Shanghai 200438, China

[4] Collaborative Innovation Center of Advanced Microsstructures, Nanjing University, Nanjing 210093, China

*Corresponding author: wzliu@fudan.edu.cn

**Corresponding author: mjzheng@fudan.edu.cn

***Corresponding author: lshi@fudan.edu.cn

Received May 30, 2025 | Accepted August 11, 2025 | Posted Online November 18, 2025

In recent years, optimizing the design of photonic structures has become a key area of nanophotonic research. Traditional methods like shape optimization and topology optimization each have their advantages, but also limitations. Shape optimization is fast and easy to manufacture but struggles with achieving the best possible performance, especially for complex designs. Topology optimization, on the other hand, can create more innovative designs but often results in structures that are hard to fabricate. In this work, we propose, to our knowledge, a new method that combines the best features of both shape and topology optimization. We apply this method to design photonic devices like gratings, which are used in many optical applications. To make the process more efficient, we construct backpropagation in nanophotonic structures inspired by neural networks. This allows us to quickly calculate the gradients of the parameters, speeding up the optimization process. We design a highly efficient blazed grating and a polarization beam splitter (PBS) using our method. Better results are achieved than the previous, including finding the structure with high first-order performance over a broader spectrum and improving the performance of the PBS from 96% to 98.8%. By reducing optimization time from over 11,000 h to just 38 h, our method opens up new possibilities for faster and more efficient optical designs.

Keywords: inverse design; backpropagation; nanophotonic structure; optimization.
DOI: 10.3788/COL202624.013601

1. Introduction

Driven by the growing demand for advanced photonic devices, optical inverse design has established itself as a critical engineering paradigm that bridges computational optimization with nano-fabrication capabilities[1,2]. The goal of optical inverse design is to use optimization algorithms to generate structures with desired optical responses[3,4]. Different types of optical inverse design methods and optimization algorithms, with their own advantages and restrictions, are employed for feasible and reliable designs according to the specific requirements of different applications[5–9].

Optical inverse design can be categorized into two types: shape optimization and topology optimization approaches[10–12]. Shape optimization employs geometric primitives (e.g., triangles, cylinders, and rectangles) as building blocks, achieving target optical responses through adjustments of critical parameters (width, height, position, and diameter). The parametric nature of this approach naturally lends itself to gradient-based optimization, where analytic differentiation directly calculates parameter sensitivities while inherently preserving geometric manufacturability constraints. This results in computationally efficient solutions within low-dimensional design spaces, accompanied by human-interpretable geometries that facilitate fabrication[13–15]. However, shape optimization can be cumbersome when designing for large bandwidth applications, discovering structures that approach theoretical performance limits, or finding novel optical responses[16]. In contrast, topology optimization adopts a material distribution paradigm, discretizing the design domain into pixels where each grid point continuously varies material

composition between two candidate dielectrics[17–19]. To navigate the exponentially growing design space (often exceeding 10^4 variables) in topology optimization, an adjoint-based method has emerged as the essential framework, leveraging dual electromagnetic field solutions to compute all material gradients in a single computation cycle. This representation enables unprecedented design freedom, generating intricate, non-intuitive patterns that may achieve theoretical performance limits. However, topology optimization still faces challenges: since the optimized structure is represented as a pixelated image, it may contain internal holes or isolated islands. Some of these features can be difficult to process.

To leverage the strengths of both approaches, combining topology optimization with shape optimization could be a promising solution. Taking a grating design as an example, we can divide the grating structure into hundreds of layers, with each layer composed of a simple shape, such as a rectangle. The overall model then becomes a topology optimization problem, where new structures are generated by continuously adjusting the parameters of each layer. The hundreds of parameters provide the flexibility needed to fine-tune the performance. For each layer, shape optimization can be applied by adjusting only the width, height, and position of the rectangle, which helps avoid the issue of internal holes and isolated islands within the structure. For the optimization algorithm, inspired by the success of neural network backpropagation, we construct the backpropagation in nanophotonic structures, and draw a parallel between multilayer nanophotonic structures and multilayer neural networks[19–21]. Incident light fields, as inputs to the nanophotonic structures, propagate through each layer according to the transfer matrix method. Transfer matrices describe the light propagation between two neighboring layers, analogous to the weights and activation functions in neural networks. This analogy allows us to optimize the parameters of the simple shapes in each layer of the nanophotonic structure through a backpropagation process, similar to how neural networks are trained.

In this work, we propose a topology optimization algorithm combined with shape optimization, which takes advantage of the strengths of both approaches. Additionally, based on Maxwell's equations, we construct a backpropagation process for nanophotonic structures to accelerate the optimization. As a practical example, we apply our proposed algorithm to grating design. First, we design a large-bandwidth blazed grating with high efficiency from the ultraviolet to near-infrared spectral regions. Next, we optimize a grating used as a polarization beam splitter (PBS), achieving a polarizing separation efficiency of 98.8%, which is higher than the previously reported value of 96%. Finally, we use shape optimization to further improve performance and determine the critical parameters of each structure.

2. Results

2.1. Backpropagation of nanophotonic structures

Figure 1(a) shows the analogy between a nanophotonic structure and the corresponding neural network. For the nanophotonic structure, the incident light is decomposed into several

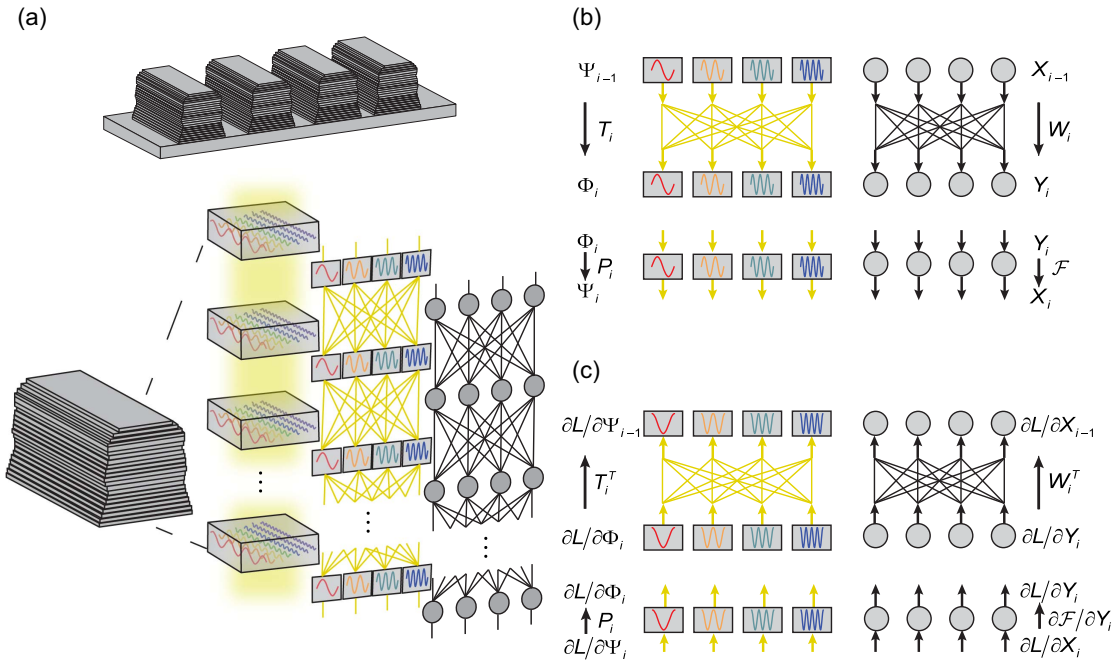

Fig. 1. Analogy between nanophotonic structures and neural networks. (a) In nanophotonic structures, eigenmodes propagate independently within the bulk layer, while they are coupled at the interfaces between neighboring layers. In neural networks, data flow independently through each neuron, with neighboring neurons connected by weighted connections that couple the data. (b) The forward computation process in nanophotonic structures and neural networks. (c) The backpropagation process in nanophotonic structures and neural networks.

eigenmodes in each layer. Each eigenmode propagates independently within the bulk layer, with its phase changing as it travels. These eigenmodes are coupled together at the interface between adjacent layers[22]. Similarly, in neural networks, the data of each channel propagate independently through the neurons, and the outputs of neighboring neurons are coupled through weight connections. In this analogy, a "bulk layer" refers to a homogeneous material region aligned along the light propagation direction (typically the z-axis), characterized by distinct optical properties (e.g., refractive index and thickness) and physical dimensions comparable to the wavelength of light. To reduce the computational time required for optimization, the backpropagation process is constructed based on the analogue between the nanophotonic structure and the neural network. A side-by-side comparison of the two is shown in Figs. 1(b) and 1(c). The left panels in Figs. 1(b) and 1(c) display the interfaces and eigenmodes of each layer in the nanophotonic structures, while the right panels show the corresponding weight connections, neurons, input data, and output data in the neural network. For the nanophotonic structure, $\mathbf{\Psi}_{i-1} = [\psi_{i-1}^1, \psi_{i-1}^2, \cdots, \psi_{i-1}^m]^T$ represents the eigenmodes of the upper layer at the ith interface, and $\mathbf{\Phi}_i = [\phi_i^1, \phi_i^2, \cdots, \phi_i^m]^T$ represents the eigenmodes of the lower layer at the ith interface. For the neural network, $X_{i-1} = [x_{i-1}^1, x_{i-1}^2, \cdots, x_{i-1}^m]^T$ represents the input to the ith weight connection, and $Y_i = [y_i^1, y_i^2, \cdots, y_i^m]^T$ represents the output of the ith weight connection.

For the propagation of light at the interface of nanophotonic structures, the eigenmodes of neighboring layers couple and redistribute. The eigenmode $\mathbf{\Psi}_{i-1}$ of the upper layer is related to the eigenmode $\mathbf{\Phi}_i$ of the lower layer by the transfer matrix $\mathbf{T}_i$:

$$\begin{bmatrix} \phi_i^1 \\ \phi_i^2 \\ \vdots \\ \phi_i^m \end{bmatrix} = \mathbf{T}_i \begin{bmatrix} \psi_{i-1}^1 \\ \psi_{i-1}^2 \\ \vdots \\ \psi_{i-1}^m \end{bmatrix}. \tag{1}$$

The transfer matrix $\mathbf{T}$ between neighboring layers is determined by the boundary conditions of the electromagnetic field[23,24]. Similarly, in the propagation of data through the weight connections in neural networks, the data from different neurons in neighboring layers are coupled and redistributed. The output data Y_i of the weight connection is related to its input data X_{i-1} by the weight matrix $\mathbf{W}_i$:

$$\begin{bmatrix} y_i^1 \\ y_i^2 \\ \vdots \\ y_i^m \end{bmatrix} = \mathbf{W}_i \begin{bmatrix} x_{i-1}^1 \\ x_{i-1}^2 \\ \vdots \\ x_{i-1}^m \end{bmatrix}. \tag{2}$$

For the propagation of light through the bulk layer of nanophotonic structures, each eigenmode $\mathbf{\Phi}_i$ propagates independently. A phase shift is applied to each eigenmode within the layer. The eigenmode $\mathbf{\Psi}_i$ on the downward side is related to the eigenmode $\mathbf{\Phi}_i$ on the upward side by the propagation matrix $\mathbf{P}_i$, which is diagonal and manipulates the phase of each eigenmode independently:

$$\begin{bmatrix} \psi_i^1 \\ \psi_i^2 \\ \vdots \\ \psi_i^m \end{bmatrix} = \mathbf{P}_i \begin{bmatrix} \phi_i^1 \\ \phi_i^2 \\ \vdots \\ \phi_i^m \end{bmatrix}. \tag{3}$$

The propagation matrix $\mathbf{P}_i$ is diagonal, with each element representing a phase factor:

$$\mathbf{P}_i = \begin{bmatrix} e^{-j\phi_1} & & & 0 \\ & e^{-j\phi_2} & & \\ & & \ddots & \\ 0 & & & e^{-j\phi_m} \end{bmatrix}. \tag{4}$$

Similarly, in neural networks, the activation function $\mathcal{F}(Y_i)$ is applied independently to each neuron[25]. The output data X_i of the neurons are related to the input data Y_i by

$$\begin{bmatrix} x_i^1 \\ x_i^2 \\ \vdots \\ x_i^m \end{bmatrix} = \begin{bmatrix} \mathcal{F} & & & 0 \\ & \mathcal{F} & & \\ & & \ddots & \\ 0 & & & \mathcal{F} \end{bmatrix} \begin{bmatrix} y_i^1 \\ y_i^2 \\ \vdots \\ y_i^m \end{bmatrix}, \tag{5}$$

where the activation function $\mathcal{F}(Y_i)$ is applied element-wise, resulting in a diagonal matrix.

The comparison above shows that the propagation behavior of light in nanophotonic structures is analogous to the propagation of data in neural networks. By analogy with the backpropagation process in neural networks, a similar backpropagation process can be established for nanophotonic structures. If the mean squared error is used as the loss function L to evaluate the difference between the output and the target, the backpropagation process begins by calculating the gradient of the loss function. Using the chain rule, the gradients of the parameters in each layer can then be calculated in turn.

For the backpropagation of data in neural networks, the gradient of Y_i is related to the gradient of X_i by the derivative of the activation function $\frac{\partial L}{\partial Y_i}$:

$$\begin{bmatrix} \frac{\partial L}{\partial y_i^1} \\ \frac{\partial L}{\partial y_i^2} \\ \vdots \\ \frac{\partial L}{\partial y_i^m} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathcal{F}}{\partial y_i^1} & & & 0 \\ & \frac{\partial \mathcal{F}}{\partial y_i^2} & & \\ & & \ddots & \\ 0 & & & \frac{\partial \mathcal{F}}{\partial y_i^m} \end{bmatrix} \begin{bmatrix} \frac{\partial L}{\partial x_i^1} \\ \frac{\partial L}{\partial x_i^2} \\ \vdots \\ \frac{\partial L}{\partial x_i^m} \end{bmatrix}. \tag{6}$$

Similarly, for the backpropagation of light in the nanophotonic structure, the gradient of the eigenmode $\mathbf{\Phi}_i$ is related to the gradient of $\mathbf{\Psi}_i$ by the propagation matrix $\mathbf{P}_i$:

$$\begin{bmatrix} \frac{\partial L}{\partial \phi_i^1} \\ \frac{\partial L}{\partial \phi_i^2} \\ \vdots \\ \frac{\partial L}{\partial \phi_i^m} \end{bmatrix} = \mathbf{P}_i^T \begin{bmatrix} \frac{\partial L}{\partial \psi_i^1} \\ \frac{\partial L}{\partial \psi_i^2} \\ \vdots \\ \frac{\partial L}{\partial \psi_i^m} \end{bmatrix}. \tag{7}$$

For the backpropagation of data in the weight connections of neural networks, the gradient of Y_i is related to the gradient of X_{i-1} by the transposed weight matrix $\mathbf{W}_i^T$:

$$\begin{bmatrix} \frac{\partial L}{\partial x_{i-1}^1} \\ \frac{\partial L}{\partial x_{i-1}^2} \\ \vdots \\ \frac{\partial L}{\partial x_{i-1}^m} \end{bmatrix} = \mathbf{W}_i^T \begin{bmatrix} \frac{\partial L}{\partial y_i^1} \\ \frac{\partial L}{\partial y_i^2} \\ \vdots \\ \frac{\partial L}{\partial y_i^m} \end{bmatrix}. \tag{8}$$

In the same way, for the backpropagation of light in the nanophotonic structure, the gradient of the eigenmode Ψ_{i-1} in the upper layer is related to the gradient of Φ_i in the lower layer by the transposed transfer matrix $\mathbf{T}_i^T$:

$$\begin{bmatrix} \frac{\partial L}{\partial \psi_{i-1}^1} \\ \frac{\partial L}{\partial \psi_{i-1}^2} \\ \vdots \\ \frac{\partial L}{\partial \psi_{i-1}^m} \end{bmatrix} = \mathbf{T}_i^T \begin{bmatrix} \frac{\partial L}{\partial \phi_i^1} \\ \frac{\partial L}{\partial \phi_i^2} \\ \vdots \\ \frac{\partial L}{\partial \phi_i^m} \end{bmatrix}. \tag{9}$$

Then, the gradient is propagating through the nanophotonic structures. The gradients of the shape parameters in each layer could be obtained for the optimization process.

2.2. Inverse design of blazed gratings

Figure 2 shows the optimization process of a blazed grating. The goal of optimization is to find a new structure with high first-order diffraction efficiency over the range from 200 to 900 nm, while minimizing the diffraction efficiency of other orders. This enables the blazed grating to be used over a broad spectrum and reduces the impact of stray light. To achieve this, the loss function L is given by

$$L = \sum_{i=200}^{900} (f_i^{-2} - 0)^2 + \sum_{i=200}^{900} (f_i^{-1} - 1)^2 + \sum_{i=200}^{900} (f_i^0 - 0)^2 + \sum_{i=200}^{900} (f_i^1 - 0)^2 + \sum_{i=200}^{900} (f_i^2 - 0)^2, \tag{10}$$

where f_i^{-2}, f_i^{-1}, f_i^0, f_i^1, and f_i^2 are the −2nd-order, −1st-order, 0th-order, 1st-order, and 2nd-order diffraction efficiencies of grating at wavelength i. A triangular grating is chosen as the initial structure, which is divided into 100 layers as shown in Fig. 2 (a). Each layer consists of a rectangular shape with optimizable parameters such as width, height, and position. The initial structure exhibits high 1st-order diffraction efficiency at the blazed wavelength of 420 nm, but the 1st-order diffraction efficiency quickly decreases in the ultraviolet and near-infrared ranges. At wavelengths of 200 and 900 nm, the 1st-order diffraction

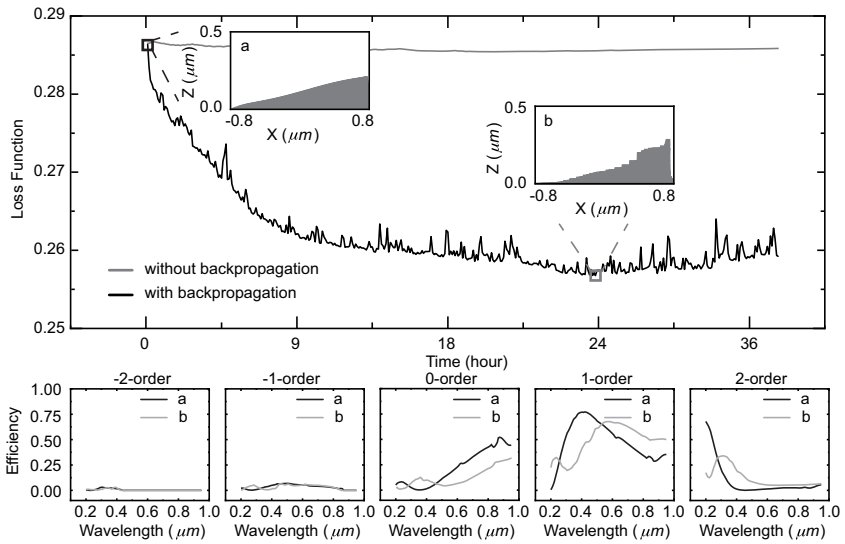

Fig. 2. Optimization process of the blazed grating. The loss functions L of the optimization with backpropagation (black line) and without backpropagation (gray line) are shown in the upper panel. The convergence speed of optimization processes without backpropagation is significantly lower than that with backpropagation. The corresponding optimized structures and spectra at different stages are shown. (a) The initial structure for blazed grating optimization is shown in the inset, and the black lines represent the diffraction efficiency of the initial structures. (b) The final optimized structure for blazed optimization is shown in the inset, and the gray lines represent the diffraction efficiency of the optimized structures.

efficiency of the initial structure is only 0.006 and 0.354, respectively.

During the optimization process, the loss function L continuously decreases, as shown in Fig. 2. The optimization is completed in 38 h, aided by backpropagation for nanophotonic structures. (For comparison, the optimization process without backpropagation would have taken 11,400 h.) The initial and optimized structures and their optical responses are shown in Fig. 2. In terms of structure, there are no internal holes or isolated islands due to the optimization of each layer's shape. A novel structure, with improved performance compared to the triangular grating, is discovered. This structure is recognized as a dual-blazed grating, which has been previously reported to achieve broadband diffraction efficiency. Our optimization algorithm rediscovers this structure through the optimization process. In terms of the optical response, the 1st-order diffraction efficiency is improved in the ultraviolet and near-infrared ranges, while the 0th-order and 2nd-order diffraction efficiencies decrease during the optimization. The first-order diffraction efficiency of the final optimized structure at 200 and 900 nm increases to 0.25 and 0.50, respectively, compared to the initial structure's parameters of 0.006 and 0.354. The 0th-order diffraction efficiency at 900 nm decreases from 0.5 to 0.25, and the 2nd-order diffraction efficiency at 200 nm decreases from 0.7 to 0.2. This shows that the final optimized structure provides high performance over a broader spectrum.

2.3. Inverse design of PBSs

The PBS is used to divide incident light into two orthogonally polarized beams. The most common type of PBS is based on birefringent materials, multilayer thin films, and grating structures. Recently, a transmission PBS was designed and fabricated using a deep-etched rectangular- and triangular-groove fused-silica grating with high −1st-order diffraction efficiency for TE polarization and high 0th-order diffraction efficiency for TM polarization[26,27]. Different polarized light travelling through the grating is assigned to a different diffraction channel. This grating design demonstrates high efficiency (over 96%) across the C+L band range for both TE and TM polarizations. In this work, by applying the algorithm proposed, we aim to discover new grating structures that exhibit higher diffraction efficiency and a greater extinction ratio than previously reported rectangular and triangular gratings. Figure 3 shows the optimization process of the PBS. The loss function L is defined as

$$L = \sum_{i=1480}^{1620} (f_i^{\mathrm{TM},0} - 1)^2 + \sum_{i=1480}^{1620} (f_i^{\mathrm{TM},-1} - 0)^2 + \sum_{i=1480}^{1620} (f_i^{\mathrm{TE},0} - 0)^2 + \sum_{i=1480}^{1620} (f_i^{\mathrm{TE},-1} - 1)^2 \quad (11)$$

where $f_i^{\mathrm{TM},0}$ and $f_i^{\mathrm{TM},-1}$ are the 0th-order and −1st-order diffraction efficiencies for TM polarization, and $f_i^{\mathrm{TE},0}$ and $f_i^{\mathrm{TE},-1}$ are the 0th-order and −1st-order diffraction efficiencies for TE polarization.

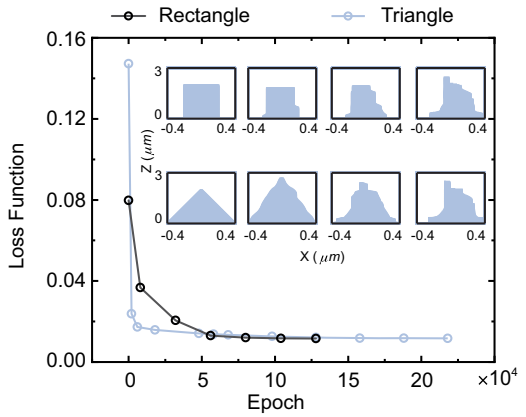

Fig. 3. Optimization process of the PBS. Two different initial structures, rectangular and triangular, are selected for the optimization, and similar optimized results are obtained.

After optimizing the grating structures, the loss function decreases from 0.147 to 0.012. This indicates that the final optimized grating achieves high efficiency (with diffraction efficiency exceeding 98.8%) over the entire C+L band range. To confirm the robustness of the results and their independence from the initial structure, two different initial grating structures—rectangular and triangular—were selected for optimization. Four different stages of the optimization process are shown in Fig. 3. Despite starting with different initial structures and optical responses, the final optimized gratings are nearly identical, and their loss function L converges to the same value of 0.012 after 100,000 optimization epochs.

2.4. Critical parameters

The optimized structures obtained using our proposed optimization algorithm exhibit two key features: shape transitions (main features) and rough edges. To demonstrate that the observed spectral changes are due to shape transitions rather than the rough edges, we remodel the optimized structures by preserving the main features and eliminating the roughness, as shown in Fig. 4(a). The critical parameters—period a, heights h_1, h_2, h_3, and widths w_1, w_2, w_3 —are defined for the remodeled structures. We then apply shape optimization to adjust these critical parameters. The corresponding loss functions are shown in Fig. 4(b). The loss function converges to the same values as those obtained with topology optimization: 0.26 for the blazed grating and 0.013 for the PBS. This demonstrates that similar performance can be achieved through shape transitions alone, while retaining the key features. For the blazed grating, the final optimized critical parameters are as follows: period $a = 1.6069\,\mu\mathrm{m}$, heights $h_1 = 1.3258\,\mu\mathrm{m}$, $h_2 = 0.3968\,\mu\mathrm{m}$, $h_3 = 0.2624\,\mu\mathrm{m}$, and widths $w_1 = 0.2704\,\mu\mathrm{m}$, $w_2 = 0.2180\,\mu\mathrm{m}$, $w_3 = 0.0868\,\mu\mathrm{m}$. For the PBS, the final optimized critical parameters are: period $a = 0.9554\,\mu\mathrm{m}$, heights $h_1 = 0.3380\,\mu\mathrm{m}$, $h_2 = 1.1271\,\mu\mathrm{m}$, $h_3 = 2.0518\,\mu\mathrm{m}$, and widths $w_1 = 0.7591\,\mu\mathrm{m}$, $w_2 = 0.3849\,\mu\mathrm{m}$, $w_3 = 0.1557\,\mu\mathrm{m}$.

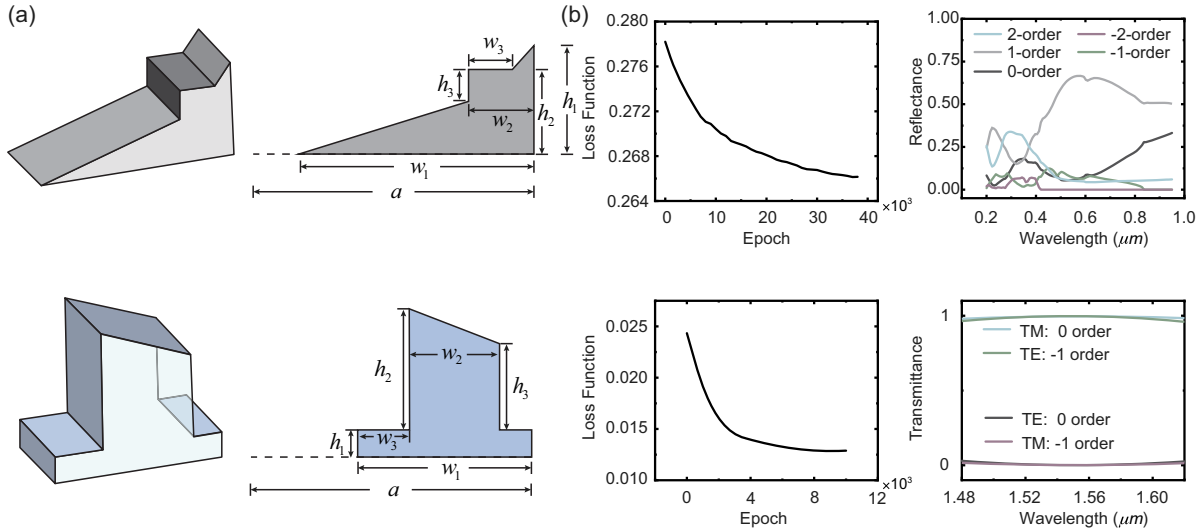

Fig. 4. Remodeling of the optimized structures. (a) The critical parameters, including period a, heights h_1, h_2, h_3, and widths w_1, w_2, w_3, are defined and optimized through shape optimization. (b) The loss function for the shape optimization and the spectrum of the remodeled structure.

3. Methods

3.1. Implementation of the algorithm

A custom program is written in C language to implement the algorithm, in favor of C language's speed. The entire framework can be divided into three parts: linear algebra, simulation, and optimization. The linear algebra module provides basic mathematical functions based on BLAS and LAPACK[28–30]. The simulation module is designed to construct the calculations and backpropagation process for nanophotonic structures. The optimization module uses the Adam algorithm to optimize these nanophotonic structures[31].

3.2. Platforms for testing

To provide a fair performance comparison, we tested our homemade program on a desktop computer equipped with an Intel (R) Core(TM) *i*7–10700F CPU @ 2.90 GHz with 16 cores. The task at hand was optimizing a blazed grating structure, and the results showed significant performance improvements when using backpropagation for nanophotonic structures. With backpropagation of nanophotonic structures, the optimization process was completed in 38 h. Without backpropagation, the same optimization task required approximately 11,400 h—a dramatic difference in computational time.

3.3. Stochastic gradient descent

The optimization algorithm used in this work is the Adam algorithm, belonging to a class of stochastic gradient descent algorithms. Stochastic gradient descent computes the gradient using only one random training example per iteration, and takes noisy, high-variance steps based on the subset of data. It is faster per-iteration computation, suitable for large-scale data, but may exhibit oscillations or temporary increases in loss, as shown in Fig. 2. The advantage of this optimization algorithm is that it is not confined to local minima and can escape local minima during the optimization process, continuously exploring the parameter space. At the same time, this also means that its loss does not decrease monotonically.

3.4. Backpropagation of the output layer

In transmission mode, the diffraction efficiencies f_i of each order i are defined as the ratio of the power flux S_i of each order to the incident power flux S_{inc}:

$$f_i = \frac{S_i}{S_{\text{inc}}}. \tag{12}$$

According to the power flux calculation formula[37], we can derive the relationship between the power flux S_i and the transmission coefficient t_i:

$$S_i = \text{Re}\left(t_i^{\dagger} kp_n \frac{t_i}{q_n \omega}\right), \tag{13}$$

where $t_i^{\dagger}$ is the conjugate transpose of t_i, kp_n is the matrix of the eigen equation for calculating the eigenmode Φ_n in output layer n, q_n is the eigenvalue of eigenmode Φ_n in the output layer (the z-component of the wave vector of the eigenmode in the output layer), and ω is the angular frequency of the light.

For backpropagation, first, we can propagate the partial derivatives from $\frac{\partial L}{\partial f_i}$ into the power flux $\frac{\partial L}{\partial S_i}$:

$$\frac{\partial L}{\partial S_i} = \frac{\partial L}{\partial f_i}\frac{1}{S_{\text{inc}}}. \tag{14}$$

Then, the derivatives in the power flux S_i are extended to the transmission coefficient t_i, eigenvalues q_n, and eigenmatrix kp_n:

$$\frac{\partial L}{\partial t_i} = \left(\left(kp_n \frac{t_i}{q_n \omega}\right)^{\dagger} + \frac{kp_n^T t_i^{\dagger}}{q_n \omega}\right) \frac{\partial L}{\partial S_i},$$
$$\frac{\partial L}{\partial q_n} = -\left(kp_n^T t_i^{\dagger} \frac{t_i}{q_n^2 \omega}\right) \frac{\partial L}{\partial S_i},$$
$$\frac{\partial L}{\partial kp_n} = t_i^{\dagger} \left(\frac{t}{q_n \omega}\right)^T \frac{\partial L}{\partial S_i}. \tag{15}$$

The transmission coefficients t_i are computed via the scattering matrix or transfer matrix. Subsequently, through backpropagation, the partial derivatives $\frac{\partial L}{\partial t_i}$ can be propagated to the scattering matrix or transfer matrix and further propagated to structural parameters.

4. Discussion

We propose an optimization method that combines topology and shape optimization. Specifically, we apply this approach to grating design, where we divide the structure into multiple layers, each composed of simple shapes (e.g., rectangles). To make the process more efficient, we construct backpropagation in nanophotonic structures inspired by neural networks.

We demonstrate the effectiveness of our method through two case studies: the design of a large bandwidth blazed grating and a PBS. For the blazed grating, our approach achieves high diffraction efficiency across the ultraviolet to near-infrared spectrum. For the PBS, we optimize a grating to achieve a polarizing separation efficiency of 98.8%, surpassing previous reports of 96%. The optimization process is significantly accelerated by our backpropagation method, reducing the computation time from 11,400 h to just 38 h for the blazed grating design. Additionally, we use shape optimization to fine-tune the designs and extract critical parameters for each structure.

Our results demonstrate that combining topology and shape optimization with backpropagation can produce highly efficient, novel photonic structures in a fraction of the time traditionally required. This method holds significant potential for advancing optical design in a variety of applications, with potential for scalability and further optimization in complex nanophotonic systems[32,33]. For optical metrology and 3D printing, our method can be applied to the reconstruction of the multilayer photonic crystal and layered patterned structures[34]. The design approach could also be extended to the antenna design[35,36]. Newly shaped nano-antennas without pixelated internal holes or isolated islands could be found through our method.

4.1. Selection criteria for the number of layers

The selection criteria for the number of layers consist of two steps. Start by optimizing a simple shape, adjusting its structural parameters such as the length and width of a rectangle, until the loss can no longer be reduced. Then, without altering the overall shape, split it into two smaller shapes, for example, dividing a rectangle into two smaller rectangles stacked vertically. Repeat the two steps until the loss can no longer be decreased or the loss meets the design requirements. A more practical example is shown in Fig. 5, where we increase the number of points without altering the polygon's shape.

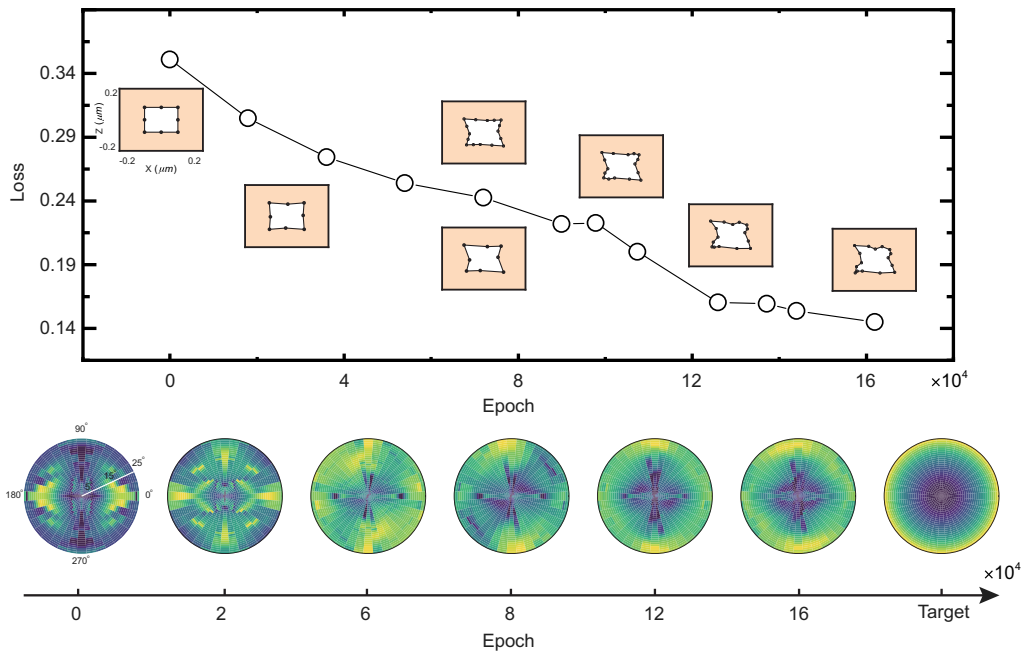

Fig. 5. Results of the optimized photonic crystal slab. The loss and structures in a unit cell during the optimization are shown in the upper panel. The optimization target and the optimized optical response, the real part of Jones matrix elements r_{ss} within an incident angle range of 0°–25°, and a full azimuthal range of 0°–360° are shown in the lower panel.

4.2. Full three-dimensional optimization

We optimized an arbitrary polygonal structure within a unit cell for a photonic crystal slab. Some scattered points in the unit cell are connected counterclockwise to form the polygon, and our backpropagation algorithm ultimately propagates the derivatives to the coordinates of these points. This allows us to update the positions of these points in each iteration, thereby achieving the evolution of the polygon. During the optimization process, we can increase the number of points without altering the polygon's shape, thereby adding more degrees of freedom for optimization and enabling continuous improvement, leading to a steady decrease in loss, as shown in Fig. 5. Therefore, our method is applicable not only to layered patterns along the z-axis but also to non-layered patterns within the x–y plane, enabling full three-dimensional optimization.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Nos. 12234007, 12321161645, 12221004, T2394480, T2394481, and 12404427), the National Key R&D Program of China (Nos. 2023YFA1406900 and 2022YFA1404800), and the Science and Technology Commission of Shanghai Municipality (Nos. 22142200400, 21DZ1101500, 2019SHZDZX01, 23DZ2260100, 24YF2702400, and 24142200100).

References

1. S. Molesky, Z. Lin, A. Y. Piggott, *et al.*, "Inverse design in nanophotonics," Nat. Photonics **12**, 659 (2018).
2. J. Kim, J. Kim, J. Kim, *et al.*, "Inverse design of nanophotonic devices enabled by optimization algorithms and deep learning: recent achievements and future prospects," Nanophotonics **14**, 121 (2025).
3. N. Wang, W. Yan, Y. Qu, *et al.*, "Intelligent designs in nanophotonics: from optimization towards inverse creation," PhotoniX **2**, 1 (2021).
4. S. Colburn and A. Majumdar, "Inverse design and flexible parameterization of meta-optics using algorithmic differentiation," Commun. Phys. **4**, 1 (2021).
5. L. Zhu, Y. Li, Z. Yang, *et al.*, "An on-demand inverse design method for nanophotonic devices based on generative model and hybrid optimization algorithm," Plasmonics **19**, 1279 (2024).
6. Z. Wang, W. Ji, T. Zhao, *et al.*, "Efficient inverse design method of AWG based on BPNN-PSO algorithm," Opt. Commun. **552**, 130080 (2024).
7. A. Sun, S. Xing, X. Deng, *et al.*, "Edge-guided inverse design of digital metamaterial-based mode multiplexers for high-capacity multi-dimensional optical interconnect," Nat. Commun. **16**, 2372 (2025).
8. D. Liu, Y. Tan, E. Khoram, *et al.*, "Training deep neural networks for the inverse design of nanophotonic structures," ACS Photonics **5**, 1365 (2018).
9. J. Peurifoy, Y. Shen, L. Jing, *et al.*, "Nanophotonic particle simulation and inverse design using artificial neural networks," Sci. Adv. **4**, eaar4206 (2018).
10. P. Dainese, L. Marra, D. Cassara, *et al.*, "Shape optimization for high efficiency metasurfaces: theory and implementation," Light Sci. Appl. **13**, 300 (2024).
11. R. E. Christiansen and O. Sigmund, "Inverse design in photonics by topology optimization: tutorial," J. Opt. Soc. Am. B **38**, 496 (2020).
12. Y. Gao, Q. Chen, S. Pian, *et al.*, "Inverse design in flat optics," Photonics Nanostruct. Fundam. Appl. **52**, 101074 (2022).
13. N. K. Brown, B. Young, B. Clark, *et al.*, "Optimized design of interlocking metasurfaces," Mater. Des. **233**, 112272 (2023).
14. H. Zhong, B. Li, Y. Lin, *et al.*, "Topology optimization of broadband polarization conversion metasurfaces in microwave region," Results Eng. **25**, 104575 (2025).
15. E. Zhu, Z. Zong, E. Li, *et al.*, "Frequency transfer and inverse design for metasurface under multi-physics coupling by Euler latent dynamic and data-analytical regularizations," Nat. Commun. **16**, 2251 (2025).
16. M. Choi, J. Park, J. Shin, *et al.*, "Realization of high-performance optical metasurfaces over a large area: a review from a design perspective," npj Nanophotonics **1**, 31 (2024).
17. Y. Liu, Y. Oda, and K. Sasahara, "Shape and topology optimization method with generalized topological derivatives," Int. J. Mech. Sci. **284**, 109735 (2024).
18. T. Yamada, K. Izui, S. Nishiwaki, *et al.*, "A topology optimization method based on the level set method incorporating a fictitious interface energy," Comput. Methods Appl. Mech. Eng. **199**, 2876 (2010).
19. S. Zhu, J. Hu, J. Qi, *et al.*, "Sketch-guided topology optimization with enhanced diversity for innovative structural design," Appl. Sci. **15**, 2753 (2025).
20. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," Nature **323**, 533 (1986).
21. S. So, T. Badloe, J. Noh, *et al.*, "Deep learning enabled inverse design in nanophotonics," Nanophotonics **9**, 1041 (2020).
22. Z. Qian, L. Shan, X. Zhang, *et al.*, "Spontaneous emission in micro- or nanophotonic structures," PhotoniX **2**, 21 (2021).
23. R. B. Balili, "Transfer matrix method in nanophotonics," Int. J. Mod. Phys. Conf. Ser. **17**, 159 (2012).
24. D. Beutel, I. Fernandez-Corbaton, and C. Rockstuhl, "Treams – a T-matrix-based scattering code for nanophotonics," Comput. Phys. Commun. **297**, 109076 (2024).
25. P. Bohra, J. Campos, H. Gupta, *et al.*, "Learning activation functions in deep (spline) neural networks," IEEE Open J. Signal Process. **1**, 295 (2020).
26. B. Wang, C. Zhou, S. Wang, *et al.*, "Polarizing beam splitter of a deep-etched fused-silica grating," Opt. Lett. **32**, 1299 (2007).
27. J. Zheng, C. Zhou, J. Feng, *et al.*, "Polarizing beam splitter of deep-etched triangular-groove fused-silica gratings," Opt. Lett. **33**, 1554 (2008).
28. J. Dongarra and J. Waśniewski, "High performance linear algebra package - LAPACK90," in *Advances in Randomized Parallel Computing* (1999), p. 241.
29. J. Choi, J. J. Dongarra, R. Pozo, *et al.*, "ScaLAPACK: a scalable linear algebra library for distributed memory concurrent computers," in *Proceedings of the Fourth Symposium on the Frontiers of Massively Parallel Computation* (1992), p. 120.
30. P. E. Bjørstad and T. Sørevik, "Data-parallel BLAS as a basis for LAPACK on massively parallel computers," in *Linear Algebra for Large Scale and Real-Time Applications* (1993), p. 13.
31. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," arXiv:1412.6980 [cs.LG] (2017).
32. Y. Chen, F. Meng, G. Li, *et al.*, "Topology optimization of photonic crystals with exotic properties resulting from Dirac-like cones," Acta Mater. **164**, 377 (2019).
33. F. Meng, S. Li, H. Lin, *et al.*, "Topology optimization of photonic structures for all-angle negative refraction," Finite Elem. Anal. Des. **117**, 46 (2016).
34. L. Wang, S. Zhang, Q. Wang, *et al.*, "Fabrication of three-dimensional (3D) woodpile structure photonic crystal with layer by layer e-beam lithography," Appl. Phys. A **95**, 329 (2009).
35. J. Sun, E. Timurdogan, A. Yaacobi, *et al.*, "Large-scale nanophotonic phased array," Nature **493**, 195 (2013).
36. I. Ahmad, S. Ullah, J. Ud Din, *et al.*, "Maple-leaf shaped broadband optical nano-antenna with hybrid plasmonic feed for nano-photonic applications," Appl. Sci. **11**, 8893 (2021).
37. V. Liu and S. Fan, "S4: a free electromagnetic solver for layered periodic structures," Comput. Phys. Commun. **183**, 2233 (2012).