

Article

Open Access

Artificial intelligence-generated photonics: mapping optical properties to subwavelength structures directly via a diffusion model

Shijie Rao, Kaiyu Cui*, Jiawei Yang, Yali Li, Shengjin Wang, Xue Feng, Fang Liu, Wei Zhang and Yidong Huang

Abstract

Subwavelength photonic structures and metamaterials provide revolutionary approaches for controlling light. The inverse design methods proposed for fabricable subwavelength structures are vital for the development of new photonic devices. However, most existing inverse design methods cannot realise direct mapping from optical properties to photonic structures; instead, they rely on forward simulation methods to perform iterative optimization. In this study, we exploit the powerful generative abilities of artificial intelligence and propose a practical inverse design method based on latent diffusion models. Our method directly maps the optical properties to structures without requiring forward simulation and iterative optimization. In this case, the given optical properties can serve as ‘prompts’ and guide the constructed model to ‘draw’ the required photonic structures correctly. Simulations and experiments show that our direct-mapping-based inverse design method can generate fabricable subwavelength photonic structures with high fidelity while following the given optical properties, such as the transmission power, phase, and polarisation responses. This may influence the methods used for optical design and significantly accelerate the research and manufacturing of new photonic devices.

Keywords: Metasurface, Inverse design, Diffusion model

Introduction

Subwavelength structures such as photonic crystals and metamaterials have led to revolutionary approaches to light-field regulation^{1–4}. Because these subwavelength structures cannot be analytically modelled using geometric or wave optics, conventional design approaches are mainly achieved by selecting an optimal match from a predefined library with a limited design space^{5–8} of photonic structures based on forward simulation. The newly developed inverse design approaches⁹ can significantly enlarge the design

space and have shown powerful capabilities in generating less intuitive but more effective photonic structures^{10,11}. However, because existing inverse design strategies usually transform an inverse design problem into an optimization problem and obtain the optimal design using iterative algorithms^{12–18}, the generated photonic structure in each iteration must be forward modelled using numerical simulation methods, such as the finite-difference time domain (FDTD)¹⁹, which is computationally expensive. In addition, these inverse design methods potentially face the common problems encountered by optimization algorithms that involve convergence, efficiency, and global optima.

To overcome the limitations of optimization-based inverse design algorithms, several attempts have been

Correspondence: Kaiyu Cui (kaiyucui@tsinghua.edu.cn)

Department of Electronic Engineering, Beijing National Research Center for Information Science and Technology (BNRist), Tsinghua University, Beijing 100084, China

© The Author(s) 2026



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

made to train deep neural networks (DNNs) and achieve direct mapping from optical properties to photonic structures^{20–34}. Recent research on artificial intelligence (AI)-generated content³⁵ and AI for science has shown great potential for DNNs in various practical fields, such as the generation of realistic images³⁶, chatbots³⁷, medicine³⁸, chemical research³⁹, and mechanical research⁴⁰. To realise practical AI for optics, the gap between optical physics and neural network parameters must be bridged. Although the most advanced AI-based inverse design methods have demonstrated significant improvements, they still encounter substantial challenges in practice, such as the nonuniqueness or existence of solutions, fabrication constraints, limited generalisability, and different distributions of input data between training and deployment. Therefore, they are usually implemented in a predefined and limited design space, in which the inverse problem is a one-to-one mapping^{20–25}. In summary, the lack of effective and accelerated design methods has become a major concern for the further development of new photonic devices^{1,9,10}.

In this study, we propose an inverse design method to achieve direct mapping from optical properties to photonic structures based on a latent diffusion model (LDM)³⁶, named AI-generated photonics (AIGP). We exploit the powerful image-synthesis ability of diffusion models^{41,42} to enlarge the search space and design new photonic structures. In this case, the given optical properties can serve as ‘text prompts’ and guide the model to ‘draw’ the required photonic structures. To achieve such direct mapping, we devised an encoding method for optical properties and designed a prompt encoder network to solve the nonuniqueness problem and provide an interface for designing photonic structures on demand. In addition, a fast forward prediction network is proposed to accelerate the simulation process significantly and realise end-to-end training. We also present a training dataset containing arbitrary shapes that enables as large a design space as possible while meeting the fabrication limits. Compared with existing methods, AIGP addresses key challenges in inverse design, including the nonuniqueness problem, handling unseen inputs, and eliminating the need for iterative optimization. By solving these long-standing issues, a new perspective on photonic design through deep generative models is provided. Moreover, the diffusion network is much easier to train and more powerful than generative adversarial networks⁴³. It can be easily scaled to large sizes and provides strong generative capabilities. In addition, the diffusion network can generate freeform shapes with various topologies. The prompt encoder network addresses the gap between abstract design

demands and realistic optical properties. A more detailed comparison with existing studies is provided in Supplementary Note 1.

Because photonic devices are generally composed of individual subwavelength structures or arrays of building blocks, which are usually referred to as meta-atoms^{5,6,44–48}, we illustrate the powerful direct mapping capabilities of our method with an example of designing meta-atoms from given transmission power, phase, and polarisation properties. Moreover, our method can be easily generalised to other photonic inverse design problems via transfer learning. It can greatly inspire and accelerate the development of various photonic devices and applications such as optical computing, metalenses, hyperspectral imaging chips, structural colours, and beam splitters.

Results and discussion

Direct-mapping inverse design via LDM

Assuming that the given photonic structure is x and the corresponding optical property is y , optimization-based inverse design methods model the forward prediction process $f(\cdot)$ such that $y = f(x)$ and perform the inverse design procedure by solving the optimization problem $\hat{x} = \underset{z}{\text{minimize}} \|y - f(z)\|$. As shown in Fig. 1a, for optimization-based inverse design, the optimization process usually requires a large number of iterative calculations, which is computationally expensive. In contrast, the direct-mapping-based inverse design method attempts to model the inverse function of $f(\cdot)$ such that the inverse design result $\hat{x} = f^{-1}(y)$ can be directly obtained without any iterations. The implementation of this inverse function is shown in Fig. 1b. A subwavelength structure can be described by its geometry, which is usually parameterised by binary-valued pixels that indicate different materials, such as silicon and air. To accelerate the diffusion process and reduce the number of required network parameters, we adopt the latent diffusion³⁶ strategy, which encodes the target geometry into a set of latent parameters to reduce the dimensionality of the design parameters significantly and perform the diffusion process in the latent space. The latent parameters generated by the diffusion network are finally decoded using the image decoder network to obtain the design results. The implementation details of our diffusion network and the image encoder-decoder network are described in the Methods section.

As shown in Fig. 1b, in addition to the latent diffusion network and image encoder-decoder network, a forward prediction network and prompt encoder-decoder network

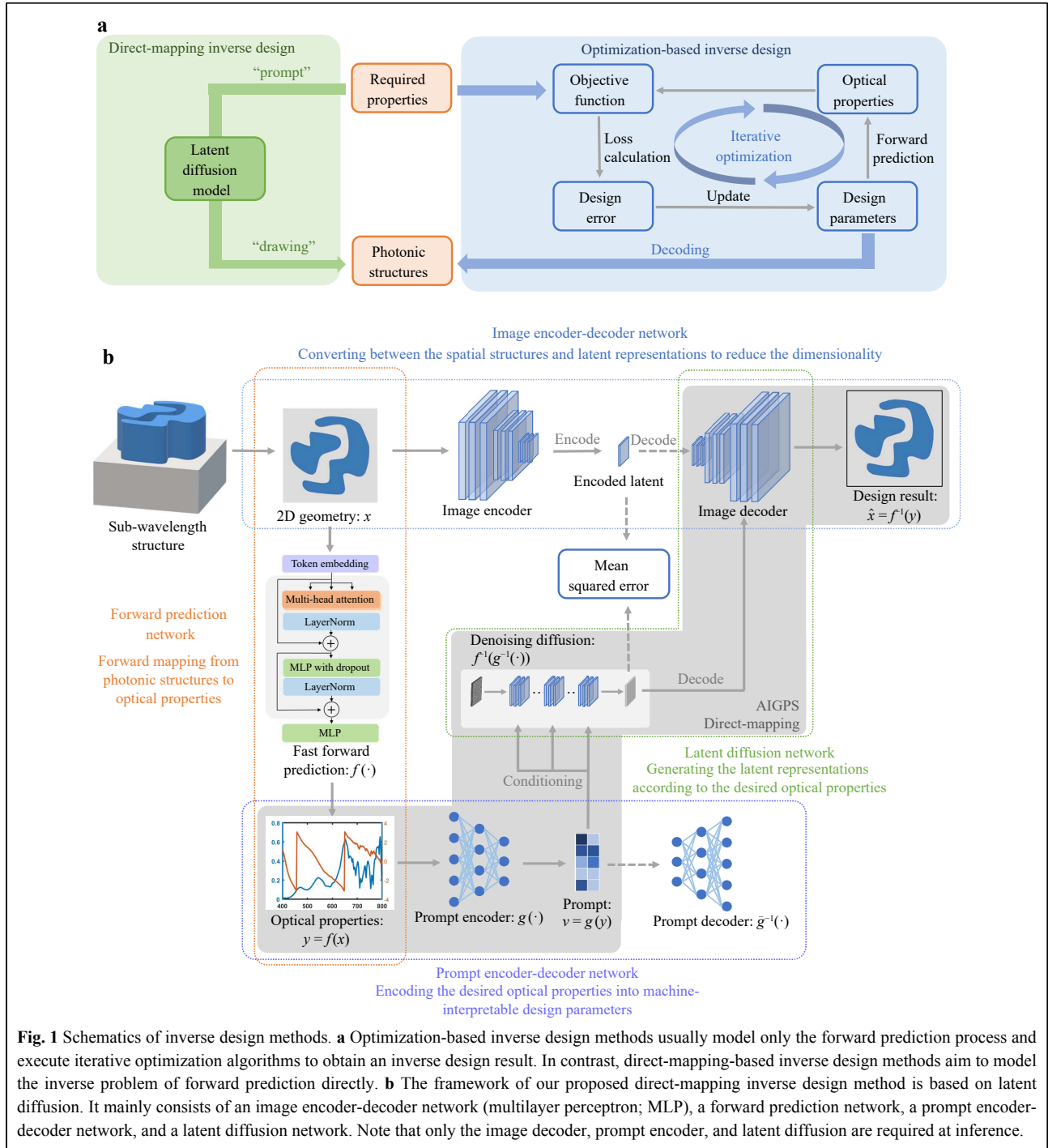


Fig. 1 Schematics of inverse design methods. **a** Optimization-based inverse design methods usually model only the forward prediction process and execute iterative optimization algorithms to obtain an inverse design result. In contrast, direct-mapping-based inverse design methods aim to model the inverse problem of forward prediction directly. **b** The framework of our proposed direct-mapping inverse design method is based on latent diffusion. It mainly consists of an image encoder-decoder network (multilayer perceptron; MLP), a forward prediction network, a prompt encoder-decoder network, and a latent diffusion network. Note that only the image decoder, prompt encoder, and latent diffusion are required at inference.

are included in our inverse design framework. Although forward prediction is not required in the direct-mapping-based inverse design process, it is required to generate input data during the training process. Therefore, to accelerate the forward prediction process, we propose a DNN-based forward prediction network to replace the numerical simulation method and train the diffusion network efficiently.

The prompt encoder network is proposed to perform fuzzy searches because we may not be able to determine a photonic structure that matches the given optical property exactly; however, we attempt to identify compatible solutions in most cases. The prompt encoder network is trained to encode only the required key features of the input optical properties while ignoring other irrelevant details via self-supervised learning. The encoded results

function as conditional controls for the diffusion process and guide the diffusion network to generate subwavelength structures that match the required key features. More specifically, assuming that the prompt encoder network is $g(\cdot)$, its output $v = g(y)$ indicates the reduced properties that contain only key features such as the working band, cutoff frequency, and resonance point of the optical properties. In this manner, the diffusion network is not trained to model $\hat{x} = f^{-1}(y)$, but rather, to model $\hat{x} = f^{-1}(g^{-1}(v))$. Notably, $v = g(y)$ is designed as a many-to-one mapping because the optical properties such as the transmission responses of different meta-atoms may have the same key features. Therefore, our inverse design method can solve the nonuniqueness problem in a free rather than limited design space by learning the one-to-many mapping function $g^{-1}(\cdot)$. Moreover, only the key features, instead of all accurate optical properties, can be sent to the inverse design algorithm. The prompt encoder network can accept abstract inputs, enabling fuzzy searches and providing an interface for on-demand inverse design. Our method is data driven, and the input training dataset is vital for its performance because this dataset determines the search space. Therefore, the training dataset is expected to contain two-dimensional (2D) geometries that are as arbitrary as possible. With reference to the freeform shape generation method⁴⁹, we design an algorithm to build a dataset of arbitrary shapes that satisfies the imposed fabrication limits. These arbitrary shapes provide a much larger design space than shapes generated from predefined parameters, thereby enabling the inverse design of subwavelength structures with complex and superior functions.

To evaluate the capabilities of our direct-mapping inverse design, the principles and experiments of our inverse method are demonstrated via the design of meta-atoms, which are widely adopted in metasurfaces with various functions^{5,15,44,46,50}. Specifically, we use meta-atoms based on a silicon-on-insulator (SOI) with a 220-nm-thick silicon layer as an example and train the network to produce an inverse design subject to the given optical response. We also show that our inverse design method can be easily converted to design meta-atoms based on other materials, such as the commonly used 600-nm-thick TiO_2 layers on SiO_2 substrates. Detailed results of the inverse TiO_2 meta-atom design process are provided in Supplementary Note 2. Because numerical simulations are time consuming, the transmission properties of every meta-atom can be quickly and accurately predicted using the designed forward prediction network using only a small number of simulated meta-atoms as training samples. Although our inverse design method contains several

different networks, which will be further discussed in the following sections, only the image decoder, prompt encoder, and latent diffusion networks are required for inference purposes. The training protocol is described in the Methods section.

Forward prediction network based on Transformer model

Our forward prediction network builds on the Vision Transformer⁵¹ model. As shown in Fig. 2a, we sliced the 2D geometry of the target meta-atom into 16×16 patches; this strategy is similar to the mesh process in an FDTD simulation. The other network details are provided in Supplementary Note 3. The network predicts the complex-valued transmission response of the meta-atom under horizontally polarised incident light. The network outputs the real and imaginary parts of the complex-valued transmission response, which can be further converted into the transmission power and phase spectra.

To study the performance of our forward prediction network, we employed it in a parameter-sweeping task. Parameter sweeps are commonly utilised in meta-atom design tasks to determine the optimal design parameters. We swept the period of the meta-atom shown in Fig. 2a from 300 to 700 nm. The transmission power and phase spectra predicted by our network for different periods are shown in Fig. 2b. We also present the prediction results obtained with a period of $p = 550$ nm in Fig. 2c for comparison with the FDTD simulation results. The network predictions and simulation results were almost identical. Whereas the FDTD method may require over 10^5 iterations to achieve simulation precision, our forward prediction network is a noniterative and end-to-end method, which significantly reduces the computational cost while maintaining high precision. Therefore, the proposed method is approximately 2 000 times faster than the FDTD simulation strategy, and can be even faster when accelerated by a GPU. Additional test results and details are provided in Supplementary Note 4.

After training, the forward prediction network is fixed to train the diffusion network. Its performance indicates its significant potential for quickly obtaining the transmission responses of meta-atoms. Similar DNN-based fast forward prediction strategies have been adopted in previous works^{17,20,23,52}. However, existing approaches usually use predefined parameters, such as the radius and width, as inputs. Therefore, the constructed DNN can only be applied to certain regular geometries. Our prediction network is designed to perform forward predictions for arbitrary shapes. It does not make any assumptions regarding the geometries of the subwavelength structures,

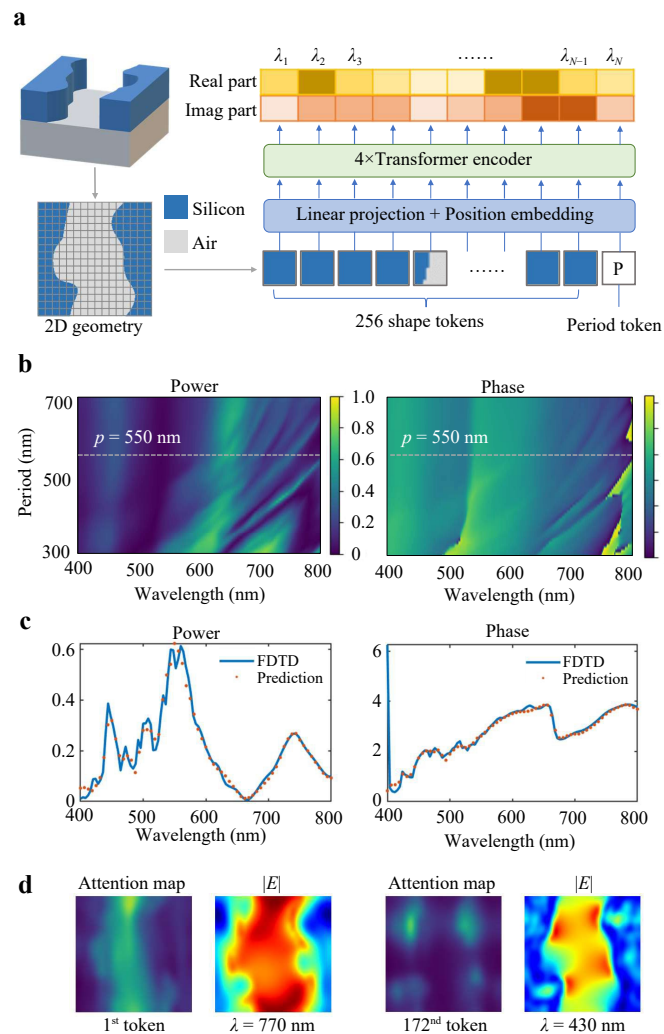


Fig. 2 Forward prediction network and its prediction results. **a** The network is based on the Vision Transformer model. The 2D geometry of the meta-atom was sliced into several patches and sent to the network as input; this strategy is similar to the mesh process in FDTD simulations. The period of the meta-atom was also used as input. **b** The predicted transmission power and phase responses of the meta-atom are shown in **a**. The period parameter was swept from 300 to 700 nm. **c** This subfigure shows the predicted results obtained with a period of $p = 550$ nm, which are indicated by the grey dashed lines in **b**. The prediction fit well with the FDTD simulation results. **d** The attention maps of the Transformer block displayed some similarities to the electric field distribution at several wavelengths.

thereby attaining better generalisability. This method can be an efficient alternative tool for replacing numerical simulations in studies involving metasurfaces and subwavelength structures. Furthermore, the forward prediction network has four Transformer blocks, each of which can output an attention map generated by the self-attention mechanism. Analysing these attention maps provides a possible means of understanding black-box DNN models. For example, by studying the attention maps produced by our network, we found that the first Transformer block mainly extracts the interrelationships between contiguous patches, which are comprehensible because the light field is typically localised in

subwavelength structures. The second block attempts to analyse the interactions between nonadjacent patches, which may facilitate the study of nonlocal light fields. The second attention map exhibits some similarities with the electrical-field distribution at certain wavelengths, as shown in Fig. 2d. The attention maps are further analysed in Supplementary Note 5.

Prompt encoder network trained using self-supervised learning

As mentioned previously, a prompt encoder network is vital for solving nonuniqueness problems. We propose a self-supervised learning strategy inspired by masked

autoencoders⁵³ for training a network that satisfies these requirements. As shown in Fig. 3a, the input of the prompt encoder network is the masking property, which is acquired by applying a certain mask to the input. Thus, some spectral bands of the transmission response are masked (where mask = 0) and unseen by the network. Only the unmasked (where mask = 1) bands are accepted by the network and encoded. By utilising this strategy, the insignificant parts of the optical properties of interest can be masked and the prompt encoder network is guided to output the reduced key features, rather than the entire set of precise features.

To train a high-performance prompt encoder network, another decoder network that enables self-supervised learning is required. The decoder network $\bar{g}^{-1}(\cdot)$ attempts to recover the entire precise property according to the reduced key features v . In this case, $\bar{g}^{-1}(\cdot)$ is designed as a one-to-one mapping rather than a one-to-many mapping to

ensure the uniqueness of its output. Self-supervised learning can be achieved by minimising the mean squared error between the one-to-one pairs of the input property y and recovered property \hat{y} . The detailed training strategy, network architecture, and testing results are presented in Supplementary Note 6. The training dataset includes several realistic transmission responses simulated using FDTD and handcrafted transmission responses, including ideal filter responses, resonance peaks, and random Fourier and Gaussian curves. In this manner, the prompt encoder is trained to accept both realistic and abstract data. After performing self-supervised learning, the decoder network is discarded and the encoder network is fixed to train the diffusion network. Given the design target (input property and mask) shown in Fig. 3a, after encoding is performed through the prompt encoder network, the latent diffusion network can be guided to generate various design results (Fig. 3b). Each result presents a possible solution to the

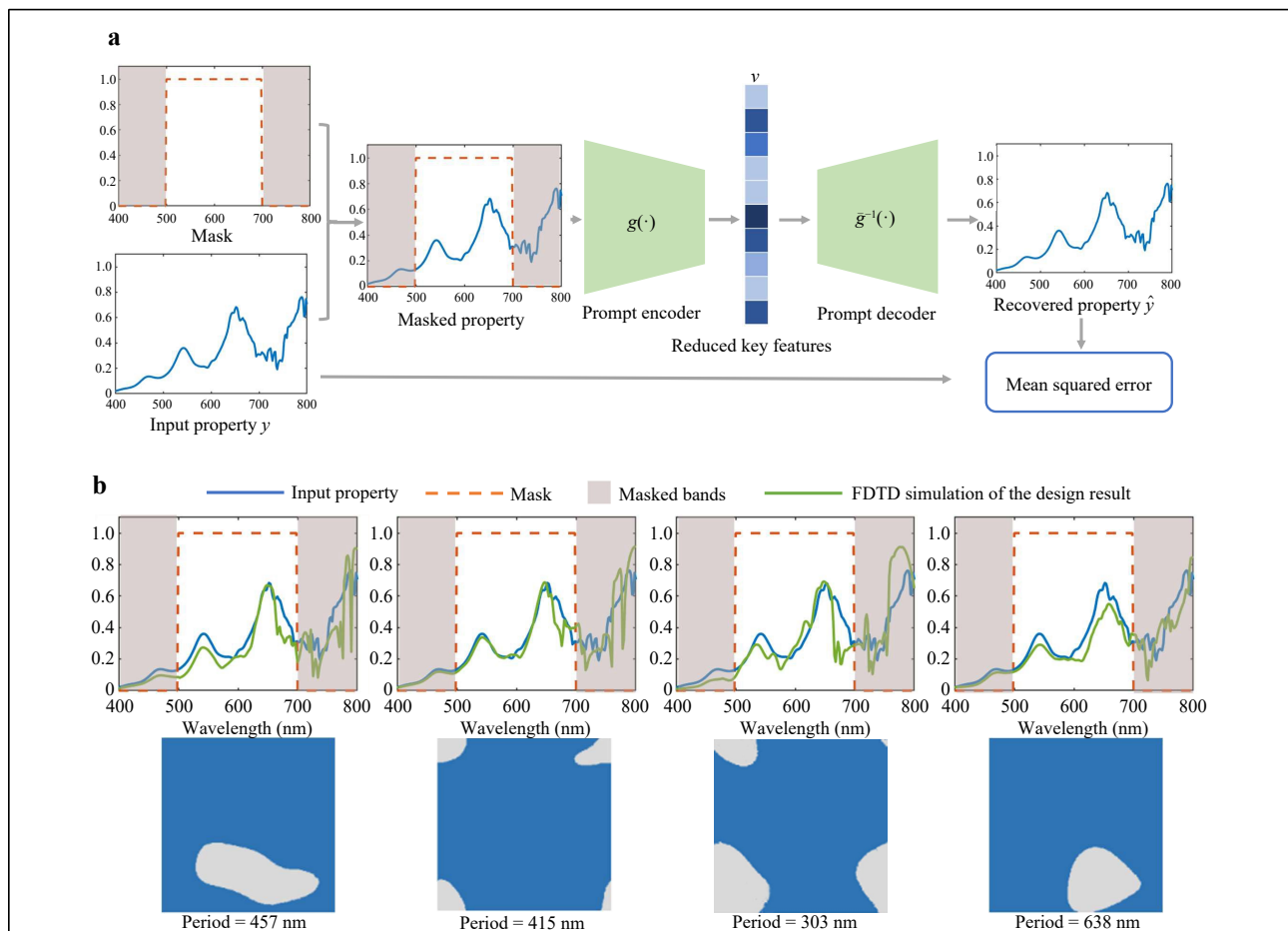


Fig. 3 Schematic of the prompt encoder network. **a** The inputs of the prompt encoder network include the target optical property and a mask. Only the unmasked part of the property is seen by the network. An additional prompt decoder network is adopted to train the prompt encoder network via self-supervised learning. **b** Different inverse design results that are subjected to the same input property and mask are shown in **a**. The proposed prompt encoder enables such a one-to-many mapping to solve the nonuniqueness problem.

inverse design problem, indicating that the proposed prompt encoder enables one-to-many mapping to solve the nonuniqueness problem. Further ablation studies on the prompt encoder network are described in Supplementary Note 12.

Inverse design results produced by the diffusion network

After preparing the forward and prompt encoder networks, the diffusion network can be trained effectively. The workflow of the inverse design process is presented in Supplementary Note 17. To test the direct-mapping-based inverse design ability, we first attempted to map a full-band precise realistic transmission power response to its corresponding meta-atom, as shown in Fig. 4a. The 2D geometry and period of the meta-atom were generated directly by the diffusion network, and forward prediction was not required. We simulated the transmission responses of the generated meta-atoms using FDTD and the results are shown as green curves. The simulated transmission responses fit the design targets well. The results indicate significant inverse design performance. Fig. 4b displays the working process of the latent diffusion network. After several denoising and diffusion steps, the required meta-atom was directly generated to satisfy the given transmission power properties as well as meet the imposed fabrication requirements. Notably, some delicate structures that were too small to be fabricated were automatically eliminated because our training dataset is designed to ensure that the generated shapes satisfy the required fabrication constraints.

C4 symmetry is an important property because meta-atoms with C4 symmetry are polarisation insensitive⁴⁴. Our diffusion network can also be controlled to generate meta-atoms with or without C4 symmetry constraints. Fig. 4c shows the inverse design of a meta-atom with C4 symmetry. The design results also meet the expectations. In addition, several uninteresting bands can be masked and meta-atoms that satisfy the given transmission-power properties in a certain band can be generated. Examples are shown in Fig. 4d, e. These results indicate that the masking strategy and prompt encoder network provide a flexible format for the input design target.

In general, the precise transmission-power response is not fully understood. Instead, we can provide only abstract design parameters. Fig. 4f shows an example of designing a longwave-pass filter. The cutoff wavelength is expected to be approximately 750 nm, and the working band is 650–800 nm. These specifications can easily be converted into the input property (blue curve) and input mask (orange curve), as shown in Fig. 4f, which can be accepted by the

prompt encoder network. We can then obtain the inverse design result, which is also shown in Fig. 4f. The results demonstrate the fuzzy search ability of the proposed inverse design method. Such an ideal filter cannot be achieved using a 220-nm-thick silicon meta-atom. However, our diffusion network attempts to generate solutions that satisfy these requirements as far as possible. The performance of the meta-atom generated by our method was also verified using an FDTD simulation. A further analysis of the physical mechanisms of AIGP is provided in Supplementary Note 13.

To demonstrate the inverse design capabilities of the diffusion network further, we also employed the network to design bandstop and bandpass filters. The results are presented in Supplementary Note 8. Compared with traditional design methods such as parameter sweeping, our inverse design methods can reliably generate the required meta-atoms. This can significantly accelerate and simplify the design of subwavelength structures. For example, to detect a red fluorescent protein in a bioluminescent area, the best approach would be to design a polarisation-independent matched filter. Our diffusion network can directly map the fluorescence spectrum to the meta-atoms within a few seconds (Supplementary Note 9).

In addition to mapping the transmission-power spectrum to the meta-atom, we can achieve direct mapping from the polarisation properties or transmission-phase spectrum to the meta-atom. Fig. 4g shows an example of designing a meta-atom with high-power transmission for horizontally polarised light and low-power transmission for vertically polarised light at a certain wavelength λ_0 . This inverse design ability can be utilised to design a polariser or polarisation beam splitter. Fig. 4h shows another example of designing a meta-atom that produces the same transmission responses at two different wavelengths λ_1 and λ_2 ; this is useful for designing phase modulation devices that work at multiple wavelengths. These results indicate the significant potential of our inverse design method for various applications. A more technical analysis of the latent diffusion network is provided in Supplementary Note 7.

Fabrication and measurement results of direct-mapping inverse-designed structures

Fig. 4b shows that the diffusion network attempts to generate fabricable subwavelength structures, which is essential for the inverse design method. The fabrication limits are learned by the network during the training process. To test the robustness of the designed results and investigate the actual performance of our method, we designed and fabricated various subwavelength photonic

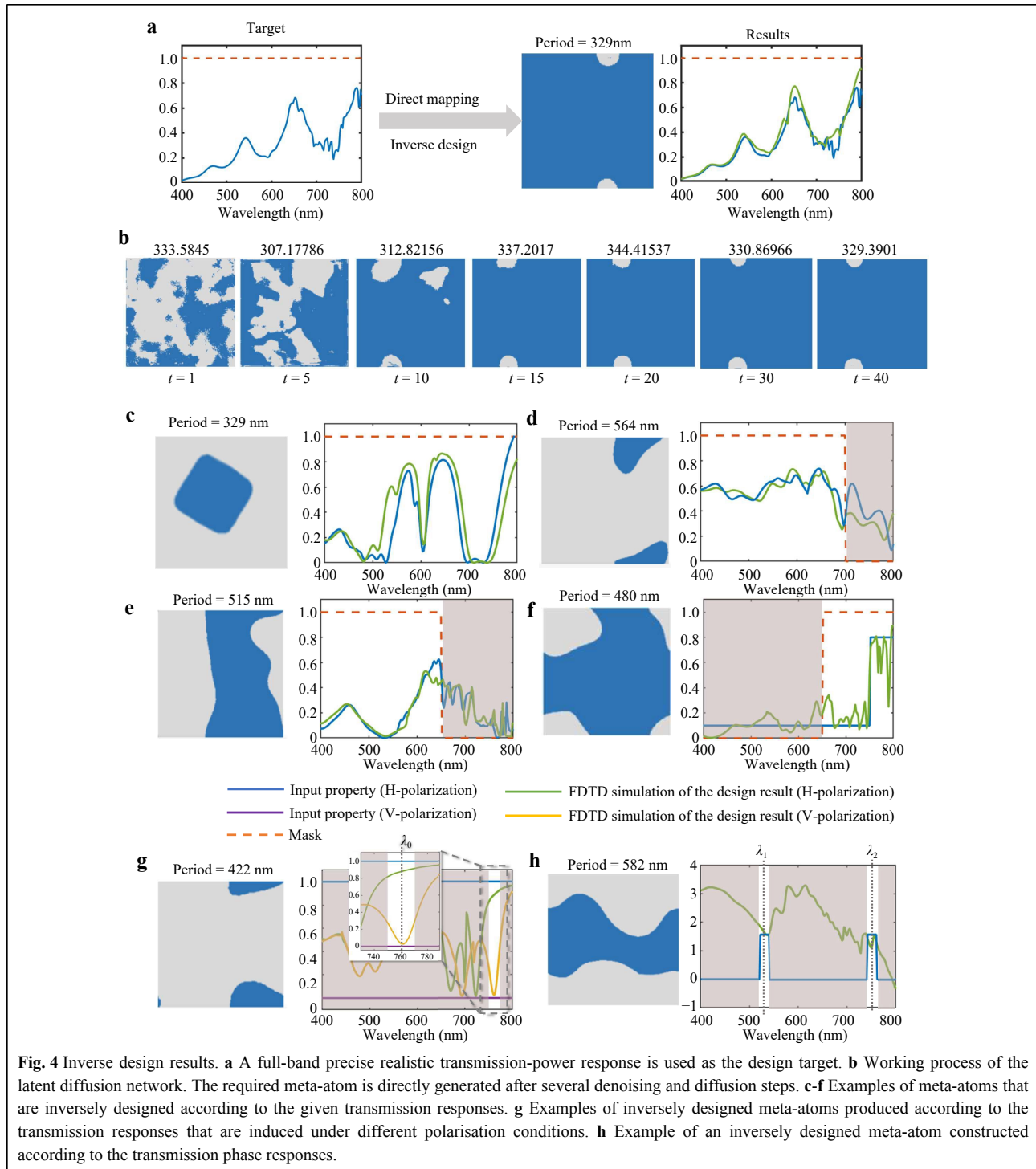
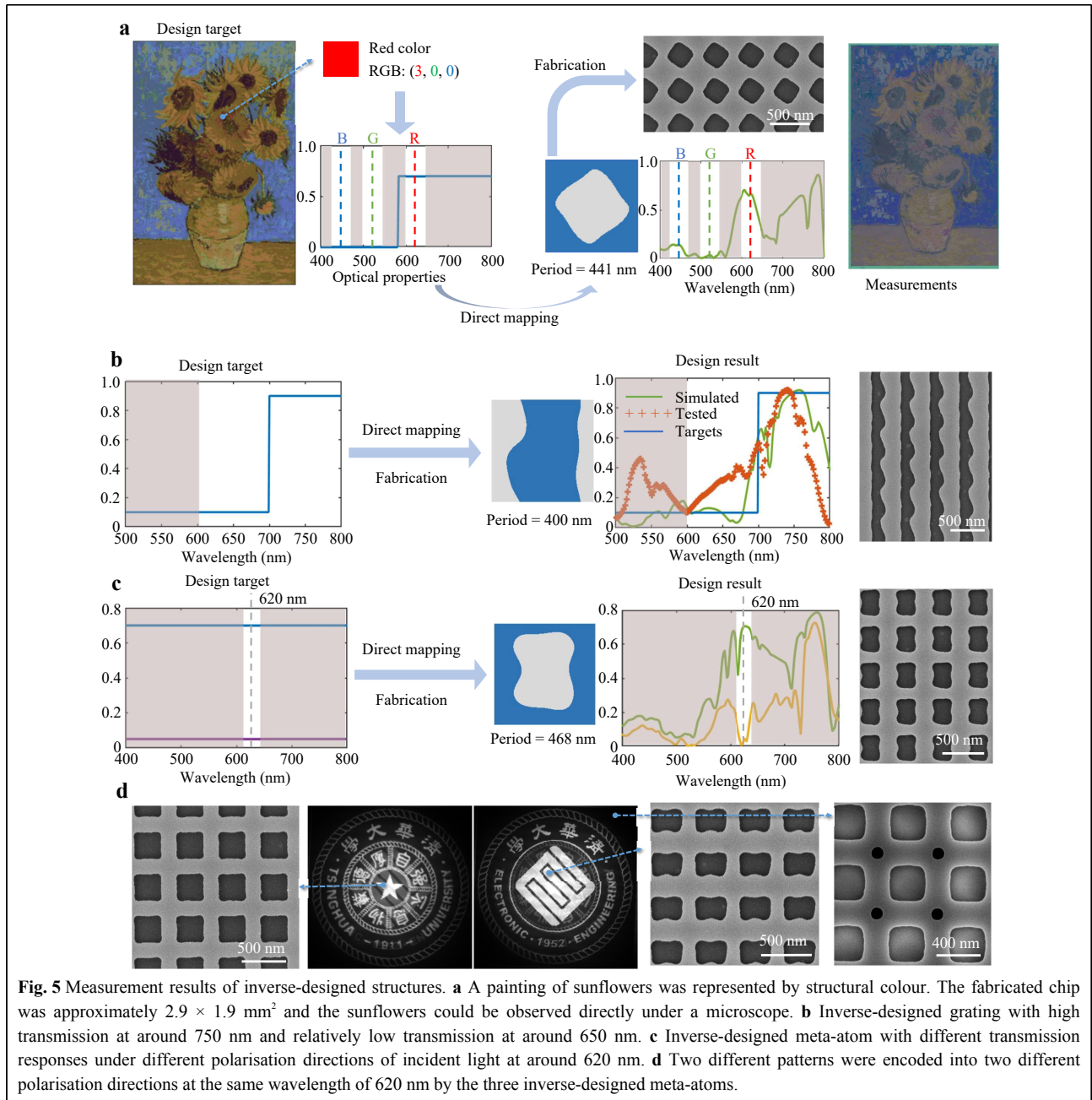


Fig. 4 Inverse design results. **a** A full-band precise realistic transmission-power response is used as the design target. **b** Working process of the latent diffusion network. The required meta-atom is directly generated after several denoising and diffusion steps. **c-f** Examples of meta-atoms that are inversely designed according to the given transmission responses. **g** Examples of inversely designed meta-atoms produced according to the transmission responses that are induced under different polarisation conditions. **h** Example of an inversely designed meta-atom constructed according to the transmission phase responses.

structures, as shown in Fig. 5. First, we mapped various transmission-power responses into meta-atoms directly. We encoded a painting of sunflowers on a chip by designing 64 different structural colours. The inverse design method was fine-tuned using transfer learning to generate meta-atoms based on silicon-on-sapphire (SOS) with a 230-nanometer-thick silicon layer. A C4-symmetry

meta-atom achieved each pixel in the painting. Using our method, we mapped the desired transmission responses at three different wavelengths (623, 528, and 461 nm for red, green, and blue, respectively) to the meta-atoms. The paint was observed directly under a microscope (Fig. 5a). Here, the colour of the sapphire substrate and the fabrication error (approximately 10 nm) resulted in slight chromatism



between the design target and the measurements. A detailed study of fabrication errors is presented in Supplementary Note 16.

We then attempted to design a longwave-pass filter with relatively high transmission in the 700–800 nm bands and low transmission in the 600–700 nm bands. The design target indicated by the green dashed curve in Fig. 5b is an ideal filter that is impossible to implement in practice. However, the proposed method could still provide an approximate solution, as indicated by the blue curves. We fabricated a direct-mapping inverse-designed meta-atom

and measured its transmission spectrum, which is indicated by the red dotted curve.

Finally, we employed the inverse design method to generate meta-atoms according to the given polarisation response (Fig. 5c). We designed three different meta-atoms and encoded two different patterns into two different polarisation directions at the same wavelength (620 nm). The fabricated chip could present different patterns under horizontally and vertically polarised light, as shown in Fig. 5d.

The measurement results demonstrate the practical

capabilities of the proposed inverse design method. This method can generate the desired meta-atoms without iterative optimization. The results also show that our method can be easily generalised to design subwavelength structures of other materials through transfer learning. Moreover, the diffusion model learns the fabrication limits during training. Thus, the generated structures are fabricable. Further details regarding the design and testing of these structures can be found in Supplementary Note 10.

Conclusions

We have proposed a direct-mapping-based inverse design method named AIGP based on a diffusion model. Whereas most existing inverse design methods are based on combinations of optimization algorithms and forward prediction strategies, our inverse design achieves direct mapping from optical properties to photonic structures, thus providing the inverse function of forward prediction. Powered by this direct mapping technique, the proposed method realises a fast inverse design process without requiring forward prediction or iterative optimization. Moreover, our inverse design method is more practical than other approaches because it provides an on-demand interface for accepting abstract design parameters as inputs, rather than precise optical properties. We achieved direct mapping from the transmission power, phase, and polarisation responses to fabricable meta-atoms and verified several inverse design results through fabrication. We also discuss how to perform an inverse design that simultaneously constrains both the amplitude and phase in Supplementary Note 18. Our method can significantly accelerate the development and manufacture of various subwavelength photonic devices.

One of the greatest difficulties encountered when utilising such direct-mapping-based inverse design is the uncertainty of the solution, which may not be unique or may not even exist. To resolve this issue, we introduced a novel framework based on an LDM enhanced by a dedicated prompt encoder network to enable one-to-many mapping and fuzzy search capabilities. If a solution does not exist, our method attempts to design a structure that satisfies the requirements as far as possible via a fuzzy search. However, if the given property is impossible to realise and no close solution can be obtained, such as in the design of a 220-nm-thick silicon meta-atom with high transmission at approximately 400 nm, the diffusion network may generate random outputs because the training dataset does not contain such impossible matching; therefore, the network does not learn to map such an input. Fortunately, our forward prediction network can be used to evaluate the inverse design results rapidly. Therefore, the

confidence of the inverse design results can be determined without further simulations.

In this study, we focused on the inverse design of fabricable meta-atoms. Although meta-atoms have wide applications in various photonic devices, they are relatively small-scale photonic structural units, and are typically studied under periodic boundary conditions. Other photonic devices such as high-performance metalenses require the design of large-scale photonic structures to overcome the limitations of periodic meta-atoms. Our method still has the potential to solve such large-scale inverse design problems because these large-scale structures can also be described by their geometries and compressed by the image encoder network to reduce the number of required design parameters. LDMs are also efficient in generating large-scale images. In addition, our method can be used to design subwavelength structures based on multiple input properties. It is only necessary to modify the prompt encoder network; that is, change the length and dimension of the input feature vector. In this manner, different required properties can be concatenated into a single vector and the diffusion process can also be controlled.

Moreover, our method has significant potential for scaling up to large models. The prompt encoder and forward prediction networks are based on the Transformer, which is the basic building block in large language models such as ChatGPT. The image encoder and diffusion networks are built on the LDM, which has also already been scaled up by stable diffusion. Therefore, our work indicates that Transformer-based and generative deep-learning models offer significant potential for photonics research. Given sufficient training data, it is possible to achieve a large photonics model. This may inspire future AI techniques and research and further empower AI in photonics.

Methods

Implementation of latent diffusion network

The overall architecture of our denoising diffusion model was built on the LDM³⁶. We made some modifications and improvements such that the LDM was suitable for generating meta-atoms with limited training data, and adopted the TensorFlow⁵⁴ framework to implement all DNN models. The core of our solution is the generation mechanism of the diffusion model. In addition to the target optical response y , the model uses a random latent variable ϵ as input. Whereas y alone maps to multiple structures $\{x_1, x_2, \dots, x_n\}$, each pair (y, ϵ_k) is designed to map uniquely to a specific solution x_k . This

transforms an ill-posed one-to-many problem into a well-defined one-to-one mapping, which a neural network can effectively learn. In practice, by sampling different ϵ_k , users can generate a diverse set of structures that satisfy the same target response.

We briefly introduce our denoising diffusion model, and more technical details can be found in the ‘Code availability’ section. First, the 2D geometry of the target meta-atom was described by a binary image with a size of 256×256 . The period of the meta-atom was normalised to the range of (0, 1] and multiplied by the binary image. The image encoder-decoder network is based on a CNN with residual connections and a self-attention mechanism. The image encoder/decoder can encode/decode the image to/from a latent space with a size of $32 \times 32 \times 2$. Denoising diffusion was conducted in the latent space to reduce the incurred computational cost significantly. The diffusion network is based on the U-Net architecture with cross-attention to introduce conditional control. The sampling strategy used in our diffusion process was DDIM⁴², and we adopted a continuous diffusion time, which embeds the noise and signal rates into the diffusion network. In this manner, the number of sampling steps can be changed dynamically at the time of inference. In our experiments, we found that 20 steps were sufficient to generate the required meta-atom effectively.

Freeform meta-atom generation and simulation

The proposed direct mapping algorithm is presented in contrast to existing iterative optimization algorithms. Such a direct mapping cannot be achieved using analytical models. Therefore, we attempted to fit the inverse function of the forward simulation using a data-driven method in which a training dataset is required. A significant computational cost is required to generate such a dataset. However, this computational effort is exerted only once. Once the inverse function is established by training, we can map any transmittance to meta-atoms directly without iterative optimization or forward simulation. The training dataset directly determines the search space of the diffusion network. It is vital to train a high-performance inverse design model and ensure that the subwavelength structures generated by the diffusion network satisfy the imposed fabrication limits. We generated approximately 200 000 shapes as arbitrarily as possible with and without the C4 symmetry constraint. The shapes were then randomly split into training and testing sets in a ratio of 9:1. The periods of the meta-atoms were between 300 and 700 nm. The minimum radius of curvature of the shapes was restricted to 45 nm to satisfy the fabrication limits. If we sampled the period at 10 nm intervals from 300 to 700 nm, we would

obtain 41 different values. We combined these period values with 200 000 different shapes to obtain approximately 10^7 meta-atoms. It would be difficult to simulate all of these meta-atoms via numerical simulations. Therefore, we simulated only approximately 10^5 meta-atoms to train the forward prediction network.

The simulation was performed using the FDTD method under periodic boundary conditions and horizontally polarised incident light. The mesh step was set to 10 nm and the cutoff precision was 10^{-5} . The dt stability factor was set to 0.99. The transmission responses of the other meta-atoms were then quickly predicted.

Training and inference protocols

Each meta-atom was described using a 256×256 image. The values of the pixels were normalised to $[-1, 1]$. The image was encoded in a $32 \times 32 \times 2$ latent space by the image encoder, and the latent space was normalised to a mean of 0 and variance of 1. We then trained the image encoder-decoder network. After training, the image encoder-decoder network achieved a mean absolute error of 0.004 and a root mean square error of 0.028. The forward prediction network uses a binary image representing the 2D geometry and period value separately as inputs, rather than encoding them to form a single image. First, the image encoder-decoder network and forward prediction networks are trained. Then, all transmission responses of the meta-atoms can be predicted quickly via forward prediction, and these responses can be used as the training set to train the prompt encoder network. Finally, the trained image encoder network, forward prediction network, and prompt encoder network are fixed and the diffusion network is trained. At each training step, a batch of shapes is randomly selected from the training set, and for each shape, a period value is randomly generated to form a meta-atom. Subsequently, the transmission responses are predicted and the meta-atoms are encoded in the latent space. The transmission responses are further encoded using the prompt encoder network and used as a conditional control mechanism for the diffusion network. Finally, the diffusion network is trained to denoise the latent space with guidance provided by the input conditional control and signal-to-noise ratio. All networks were trained using the Adam⁵⁵ optimiser with a weight decay rate of 0.000 1, and the loss was calculated as the mean squared error. At the time of inference, only the image decoder, prompt encoder, and diffusion networks are required. However, a forward prediction network can also be used to evaluate the generated results. A quantitative analysis of the computational cost of the training and deployment of AIGP is described in

Supplementary Note 14.

Chip fabrication

The designed subwavelength structures were fabricated by Tianjin H-Chip Technology Group Corporation. These structures were formed on a SOS chip. The intrinsic silicon layer is 230 nm thick and the sapphire layer is 475 μm thick with flatness $<15 \mu\text{m}$. The fabrication process was as follows: Electron-beam lithography was performed at a beam current of 2 nA using a 250-nm-thick ZEP 520A resist layer. Subsequently, the pattern was transferred via inductively coupled plasma etching with a gas mixture of CHF_3 and SF_6 . Finally, the chromium mask was removed using a dry etching process with a combination of CHF_3 and O_2 . The linewidth precision of the fabricated samples was specified as $\pm 5\%$ for features above 200 nm and $\pm 10 \text{ nm}$ for features at or below 200 nm. The minimum line width was 90 nm.

Acknowledgements

The authors would like to thank the Tianjin H-Chip Technology Group Corporation and the Innovation Center of Advanced Optoelectronic Chip and Institute for Electronics and Information Technology in Tianjin, Tsinghua University, for their support during the electron-beam lithography and ICP etching. We thank Yue Zou for language editing. We thank Sheng Xu and Tianhao Liu for their help with device fabrication and testing. This study was supported by the National Key Research and Development Program of China (Grant Nos. 2023YFB2806703 and 2022YFF1501600), National Natural Science Foundation of China (Grant No. U22A6004), Beijing Frontier Science Center for Quantum Information, and Beijing Academy of Quantum Information Sciences.

Author contributions

SR was a major contributor to the proposal and implementation of the direct-mapping inverse design method. KC supervised the study and provided assistance in the meta-atom simulation and device fabrication. JW helped with building the freeform-shaped meta-atom dataset. YL and SW provided advice on the neural network design and training. YH, XF, FL, and WZ provided advice on the experiments and manuscript writing. SR and KC wrote the manuscript with the help of all other authors. All authors have read and approved the final version of the manuscript.

Data availability

The authors declare that the data supporting the findings of this study are available within the paper and its supplementary information files.

Conflict of interest

The authors declare that they have no competing interests.

Supplementary information is available for this paper at <https://doi.org/10.37188/lam.2026.037>.

Received: 20 June 2025 Revised: 07 March 2026 Accepted: 10 March 2026

Accepted article preview online: 11 March 2026

Published online: 20 April 2026

References

- Ma, W. et al. Deep learning for the design of photonic structures. *Nature Photonics* **15**, 77-90 (2021).
- Soukoulis, C. M., Linden, S. & Wegener, M. Negative Refractive Index at Optical Wavelengths. *Science* **315**, 47-49 (2007).
- Cerjan, A. & Fan, S. Complete photonic band gaps in supercell photonic crystals. *Physical Review A* **96**, (2017).
- Chen, W. T., Zhu, A. Y. & Capasso, F. Flat optics with dispersion-engineered metasurfaces. *Nature Reviews Materials* **5**, 604-620 (2020).
- Khorasaninejad, M. et al. Metalenses at visible wavelengths: Diffraction-limited focusing and subwavelength resolution imaging. *Science* **352**, 1190-1194 (2016).
- Miyata, M., Nemoto, N., Shikama, K., Kobayashi, F. & Hashimoto, T. Full-color-sorting metalenses for high-sensitivity image sensors. *Optica* **8**, 1596-1604 (2021).
- Tian, T. et al. Metasurface-Based Free-Space Multi-Port Beam Splitter with Arbitrary Power Ratio. *Advanced Optical Materials* **11**, (2023).
- Chen, X. et al. All-Dielectric Metasurface-Based Beam Splitter with Arbitrary Splitting Ratio. *Nanomaterials* **11**, 1137 (2021).
- Li, Z., Pestourie, R., Lin, Z., Johnson, S. G. & Capasso, F. Empowering Metasurfaces with Inverse Design: Principles and Applications. *ACS Photonics* **9**, 2178-2192 (2022).
- Molesky, S. et al. Inverse design in nanophotonics. *Nature Photonics* **12**, 659-670 (2018).
- Elsawy, M. M. R. et al. Numerical Optimization Methods for Metasurfaces. *Laser & Photonics Reviews* **14**, 1900445 (2020).
- Jensen, J. S. & Sigmund, O. Topology optimization for nano-photonics. *Laser & Photonics Reviews* **5**, 308-321 (2011).
- Fan, J. A. Freeform metasurface design based on topology optimization. *MRS Bulletin* **45**, 196-201 (2020).
- Sell, D. et al. Large-Angle, Multifunctional Metagratings Based on Freeform Multimode Geometries. *Nano Letters* **17**, 3752-3757 (2017).
- Piggott, A. Y. et al. Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer. *Nature Photonics* **9**, 374-377 (2015).
- Jin, Z. et al. Complex Inverse Design of Meta-optics by Segmented Hierarchical Evolutionary Algorithm. *ACS Nano* **13**, 821-829 (2019).
- Inampudi, S. & Mosallaei, H. Neural network based design of metagratings. *Applied Physics Letters* **112**, 241102 (2018).
- Peurifoy, J. et al. Nanophotonic particle simulation and inverse design using artificial neural networks. *Science Advances* **4**, eaar4206 (2018).
- Teixeira, F. L. et al. Finite-difference time-domain methods. *Nature Reviews Methods Primers* **3**, (2023).
- An, S. et al. A Deep Learning Approach for Objective-Driven All-Dielectric Metasurface Design. *ACS Photonics* **6**, 3196-3207 (2019).
- Malkiel, I. et al. Plasmonic nanostructure design and characterization via Deep Learning. *Light: Science & Applications* **7**, 60 (2018).
- Ma, W., Cheng, F. & Liu, Y. Deep-Learning-Enabled On-Demand Design of Chiral Metamaterials. *ACS Nano* **12**, 6326-6334 (2018).
- Liu, D. et al. Training Deep Neural Networks for the Inverse Design of Nanophotonic Structures. *ACS Photonics* **5**, 1365-1369 (2018).
- Kanmaz, T. B. et al. Deep-learning-enabled electromagnetic near-field prediction and inverse design of metasurfaces. *Optica* **10**, 1373 (2023).
- Zhang, T. et al. Efficient spectrum prediction and inverse design for plasmonic waveguide systems based on artificial neural networks. *Photon. Res.* **7**, 368-380 (2019).
- Lee, X. Y. et al. Fast inverse design of microstructures via generative invariance networks. *Nature Computational Science* **1**, 229-238 (2021).
- Jiang, J. & Fan, J. A. Global Optimization of Dielectric Metasurfaces Using a Physics-Driven Neural Network. *Nano Letters* **19**, 5366-5372 (2019).
- Liu, Z. et al. Generative Model for the Inverse Design of Metasurfaces.

- Nano Letters* **18**, 6570-6576(2018).
29. Han, X. et al. Inverse design of metasurface optical filters using deep neural network with high degrees of freedom. *InfoMat* **3**, 432-442 (2021).
 30. Ma, W. et al. Probabilistic representation and inverse design of metamaterials based on a deep generative model with semi-supervised learning strategy. *Advanced Materials* **31**, 1901111 (2019).
 31. Wen, F., Jiang, J. & Fan, J. A. Robust Freeform Metasurface Design Based on Progressively Growing Generative Networks. *ACS Photonics* **7**, 2098-2104 (2020).
 32. Ma, T., Wang, H. & Guo, L. J. OptoGPT: A foundation model for inverse design in optical multilayer thin film structures. *Opto-Electronic Advances* **7**, 240062 (2024).
 33. An, S. et al. Multifunctional metasurface design with a generative adversarial network. *Advanced Optical Materials* **9**, 2001433 (2021).
 34. Ma, T. et al. Benchmarking deep learning-based models on nanophotonic inverse design problems. *Opto-Electronic Science* **1**, 210012 (2022).
 35. Cao, Y. et al. A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt. *arXiv preprint arXiv: 2303.04226* (2023).
 36. Rombach, R. et al. High-resolution image synthesis with latent diffusion models. in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10684-10695.
 37. Achiam, J. et al. GPT-4 Technical Report. *arXiv preprint arXiv: 2303.08774* (2023).
 38. Singhal, K. et al. Large language models encode clinical knowledge. *Nature* **620**, 172-180 (2023).
 39. Boiko, D. A. et al. Autonomous chemical research with large language models. *Nature* **624**, 570-578 (2023).
 40. Bastek, J. H. & Kochmann, D. M. Inverse design of nonlinear mechanical metamaterials via video denoising diffusion models. *Nature Machine Intelligence* **5**, 1466-1475(2023).
 41. Ho, J., Jain, A. & Abbeel, P. Denoising Diffusion Probabilistic Models. *Advances in neural information processing systems* **33**, 6840-6851 (2020).
 42. Song, J. , Meng, C. & Ermon, S. Denoising Diffusion Implicit Models. in *International Conference on Learning Representations*.
 43. Dhariwal, P. & Nichol, A. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems* **34**, 8780-8794 (2021).
 44. Xiong, J. et al. Dynamic brain spectrum acquired by a real-time ultraspectral imaging chip with reconfigurable metasurfaces. *Optica* **9**, 461-468 (2022).
 45. Rao, S. et al. Anti-spoofing face recognition using a metasurface-based snapshot hyperspectral image sensor. *Optica* **9**, 1253-1259 (2022).
 46. Wang, Z. et al. Single-shot on-chip spectral sensors based on photonic crystal slabs. *Nature communications* **10**, 1020 (2019).
 47. Li, T. et al. Revolutionary meta-imaging: from superlens to metalens. *Photonics Insights* **2**, R01 (2023).
 48. Li, Y. et al. Recent progress on structural coloration. *Photonics Insights* **3**, R03 (2024).
 49. Yang, J. et al. Ultraspectral Imaging Based on Metasurfaces with Freeform Shaped Meta-Atoms. *Laser & amp; Photonics Reviews* **16**, 2100663 (2022).
 50. Shi, L. et al. Si-Based Polarizer and 1-Bit Phase-Controlled Non-Polarizing Beam Splitter-Based Integrated Metasurface for Extended Shortwave Infrared. *Nanomaterials* **13**, 2592 (2023).
 51. Dosovitskiy, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. in *International Conference on Learning Representations*.
 52. Gao, L. et al. A Bidirectional Deep Neural Network for Accurate Silicon Color Design. *Advanced Materials* **31**, 1905467 (2019).
 53. He, K. et al. Masked autoencoders are scalable vision learners. in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 16000-16009.
 54. Abadi, M. et al. TensorFlow: a system for Large-Scale machine learning. in *12th USENIX symposium on operating systems design and implementation (OSDI 16)*. 265-283.
 55. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv: 1412.6980* (2014).