Probabilistic Representation and Inverse Design of Metamaterials Based on a Deep Generative Model with Semi-Supervised Learning Strategy

Wei Ma, Feng Cheng, Yihao Xu, Qinlong Wen, and Yongmin Liu*

The research of metamaterials has achieved enormous success in the manipulation of light in a prescribed manner using delicately designed subwavelength structures, so-called meta-atoms. Even though modern numerical methods allow for the accurate calculation of the optical response of complex structures, the inverse design of metamaterials, which aims to retrieve the optimal structure according to given requirements, is still a challenging task owing to the nonintuitive and nonunique relationship between physical structures and optical responses. To better unveil this implicit relationship and thus facilitate metamaterial designs, it is proposed to represent metamaterials and model the inverse design problem in a probabilistically generative manner, enabling to elegantly investigate the complex structure–performance relationship in an interpretable way, and solve the one-to-many mapping issue that is intractable in a deterministic model. Moreover, to alleviate the burden of numerical calculations when collecting data, a semisupervised learning strategy is developed that allows the model to utilize unlabeled data in addition to labeled data in an end-to-end training. On a data-driven basis, the proposed deep generative model can serve as a comprehensive and efficient tool that accelerates the design, characterization, and even new discovery in the research domain of metamaterials, and photonics in general.

Artificially constructed metamaterials make it possible to design macroscopic optical responses from microscopic subwavelength structures that mimic atoms in ordinary materials. Although the degrees of freedom in the meta-atom design provides tremendous flexibility in engineering the optical properties of metamaterials, this bottom-up design strategy can hardly be generalized into concrete and practical guidelines. At present, the design process of metamaterials mostly relies on physics-inspired methods, resorting to human knowledge such as physical insights revealed by simplified analytical modeling, similar experience transferred from previous practice, and intuition obtained by scientific reasoning. For example, many meta-atoms inherited traditional antenna designs with geometries like rectangle, cross, bowtie, V-shape, H-shape, and so on, whose first-order response is approximated by electrical dipole resonance with relevant scaling effect. Some other designs guided by physical intuition include ring-like structures that exhibit strong magnetic resonances induced by the incident magnetic field, dielectric building blocks that can induce both electric and magnetic resonances leading to better control of the phase of the scattered light, or the spectra line-shape tailoring by introducing coupling among different resonant modes.

Despite the exciting results obtained by these physics-inspired designs, this methodology basically relies on a trial-and-error process, usually involving numerical methods like finite-difference-time-domain (FDTD) or finite element method (FEM) to iteratively solve Maxwell’s equations. The low efficiency and thus limited exploration of the design varieties tend to easily omit the optimal solution. The inverse design approaches start from the opposite end, and try to optimize certain objective functions describing the desired performance. Common approaches for inverse problems include genetic algorithm, level set methods, and topology optimization, which, however, are still stochastic searching algorithms that are time-consuming and deteriorate rapidly as the design space grows. Different from numerical calculations, data-driven methods based on machine learning (ML) solve the optimization problem from statistical perspectives, so that the solution to optimize a target can be approximately generalized from numerous design examples. With the rapidly accumulated data and thus booming of deep learning (DL), the state-of-the-art in many research domains, such as speech recognition, computer vision, natural language processing, and decision making, has been pushed far beyond conventional methods. Deep neural networks simulate biological signal processing that allow computational models to learn multiple
levels of abstract representations of data layer by layer,[29] with superior advantages to discover intricate structures in large data sets by using the backpropagation algorithm in training. More recently as a powerful tool, DL has been applied to many other research fields, such as material science,[30,31] chemistry,[32–34] laser physics,[35] particle physics,[36] quantum mechanics,[17,38] and microscopy,[39] showing great potential to circumvent the drawbacks of traditional methods in these areas.

Neural networks have been used to solve the design, optimization, and prediction problems of electromagnetism in some early works,[40–42] but the model capability and performance were limited, largely due to the simple model structure and the lack of data. More recent works sought to deal with the inverse design by DL under various scenarios, including plasmonic waveguides,[43] optical power splitters,[44] plasmonic metamaterials,[45] chiral metamaterials,[46] and nanophotonic particles.[47] Despite different design targets and network architectures, the common idea behind these works is to model the relationship between the design parameters and optical response as a bidirectional mapping, which is only able to deal with a few design parameters in a small range of applications. More fatally, treating metamaterial design as a one-to-one mapping as regression problem is inconsistent with the physical intuitions, since drastically different metamaterial structures could produce very similar optical responses. To remedy this issue, tandem training strategy has been used to avoid instable training loss and force the inverse network to converge to one possible solution,[48] but this operation sacrifices the design variety and leads to limited generalization ability.

To mitigate all these challenges, here we propose a probabilistic graphic model as a comprehensive solution to the metamaterial design problem. The metamaterial patterns are represented as 2D images with the ability to describe any random geometries beyond fixed parameterization.[49,50] Our model aims to follow the heuristic design procedure from human experience in the sense that similar geometries are clustered together as possible candidates, and similar optical responses are obtained by varying the design within each geometry group. To achieve this goal, we introduce latent variables as a probabilistic representation of metamaterial design by incorporating a variational auto-encoder (VAE) structure in our model,[51] which encodes the designed pattern together with the corresponding optical response into a latent space. By forcing a given prior distribution on the latent space, new designs can be reconstructed from the latent variables by sampling in the latent space, allowing many design varieties satisfying the same requirements. Unlike a very recently proposed generative model using generative adversarial network (GAN)[50] that needs a pretrained simulator to guarantee the inverse design process, our model solves both the forward and inverse problems at the same time, and can be trained in an end-to-end manner. More notably, by introducing latent variable to encode metamaterial design, the proposed deep generative model offers interpretability and can utilize unlabeled data in a semi-supervised learning strategy to improve the model performance,[52] which can fully exploit human experience on possible metamaterial geometries while alleviating the intensive burden of numerical calculations to collect data. This represents another unique aspect of our work. With the probabilistic representation capability, our model can produce much more diverse candidates for the inverse design, which, in conjunction with human knowledge, would open a new paradigm for discovering sophisticated metamaterials and photonic structures with prescribed, exotic optical properties and functionalities.

Figure 1a schematically shows the proposed framework to model metamaterials, which can simultaneously realize forward prediction from given designs and inverse retrieval of the possible designs from required optical responses. As shown in the left part of the figure, the meta-atom under investigation is a sandwiched structure working in the reflective configuration, which is composed of a continuous metallic ground plane, a dielectric spacer, and a top metallic resonator. The thickness of the resonator and spacer is fixed at 50 and 100 nm, respectively, while the period of the unit cell is 2 µm in both directions. We treat metamaterial design as a 2D binary image of resolution 64 × 64, where 1 stands for metal and 0 stands for air. In this way, arbitrary design geometries can be properly represented, avoiding the limit of parameterization. Considering two different linear polarization conditions and optical reciprocity, we will focus on three unique reflection spectra, that is, x polarization input with x polarization output (R_{xx}), y polarization input with y polarization output (R_{yy}), and x polarization input with y polarization output (R_{yx}), which are identical to another cross-polarized reflection (R_{xy}). The reflection spectra of interest are set in the mid-infrared region from 40 to 100 THz, and discretized into 31 data points each with 2 THz interval. More details about data preparation can be found in the Experimental Section.

From physical insights, the forward prediction is a many-to-one mapping, since any meta-atom structure supporting the same resonant mode can produce the same optical response within a certain range of tolerance, which can be modeled deterministically as a regression problem. While the inverse design has to deal with the opposite, one-to-many mapping or structured output problem, allowing diverse predictions.[53] To solve this problem, we introduce a latent variable which, subject to a given prior distribution, encodes the meta-atom pattern together with its optical response. Therefore, the inverse design can be modeled as a generative process, by first sampling a latent variable from the latent space, and then reproducing the design from the sampled variable. This process introduces diversity due to the stochastic sampling step. The overall comprehensive problem including forward prediction, inverse generation, and latent space construction are solved jointly by the proposed deep generative model. We further illustrate the one-to-many mapping issue and the advantage of our model in Section 1 of the Supporting Information.

As indicated in Figure 1a, there are three types of variables in our deep generative model: input variable x (geometric pattern of metamaterial structure), output variable y (three distinct reflection spectra), and latent variable z (compressed code of the design). The overall model builds up the probabilistic relationship among these variables, which accounts for different functionalities when applying the model to the metamaterial design. Shown as solid boundaries in Figure 1b, our deep generative model can be decomposed into three submodels, namely, recognition model for the encoding process of metamaterial structures into latent space, prediction model for
the forward prediction of reflection spectra of metamaterial with given geometric pattern, and generation model for the inverse design process of metamaterial given required spectra. The detailed mathematical formulation of the deep generative model and loss objective for training can be found in Sections 2 and 3 of the Supporting Information.

To realize the deep generative model that provides a comprehensive solution to the metamaterial design and characterization, we implement the three probabilistic submodels using four deep neural networks, each with deliberately designed structure for its specific function. Specifically, they are feature extraction network, prediction network, recognition network, and reconstruction network as shown in Figure 1b. In order to fully exploit human experience on possible metamaterial geometries while avoiding intensive numerical calculations to collect data, the entire deep generative model is trained in an end-to-end manner with both labeled and unlabeled data employing a semi-supervised learning strategy. We show that, with the aid of unlabeled data, the model performance is obviously improved (Table S2, Supporting Information). This means the proposed model can efficiently learn from similar metamaterial patterns without corresponding optical response obtained by numerical simulations, which alleviate the burden in data acquisition compared with other supervised learning counterpart. More detailed implementation of the model with semi-supervised training strategy and network architecture can be found in Sections 4–6 of the Supporting Information.

To make a clear demonstration of the encoder–decoder configuration of our model, we first train the model for 300 000 steps on three basic geometries, namely, cross, split ring, and h-shape. In Figure 2a, the evolution of the reproduced images is illustrated at certain training steps. Given the input images on the leftmost column, our model gradually produces more and more accurate reconstruction of the inputs as the training proceeds. Finally, at the last step when training ends, the test input images are faithfully reconstructed with high fidelity, indicating an accurate distribution of meta-atom patterns described by the proposed model. Before evaluating the performance of the proposed deep generative model, we first check the structure of the latent space where the metamaterial design is encoded. Since the dimension of the latent space is 20, we use t-distributed stochastic neighbor embedding (t-SNE) method to reduce the dimension to 2 for visualization purpose.

In Figure 2b we plot the 2D distribution of the encoded test data in three basic geometry groups. The three geometry groups are clearly separated into three clusters of cross, split ring, and h-shape, respectively. Without providing the corresponding labels for the shapes, our model automatically learns to distinguish different shapes through the encoding–decoding training iterations on both labeled and unlabeled data. As
described by the KL-divergence term in Equation S1, we have applied a standard Gaussian prior distribution of the latent variables independent of the geometric patterns and reflection spectra, \( p_\phi(z) \sim \mathcal{N}(0, I) \). The result in Figure 2b indicates that the latent distribution learned by our model approximates the spherical Gaussian prior in the 20-dimensional latent space, while keeping the fine structure within the prior according to different design patterns.

On the other hand, by feeding both image features and reflection spectra into the recognition network (Figure 1b), the VAE structure actually models the distribution of design patterns conditioned on its optical response. Naturally, besides the shape information, the latent variables are expected to encode the information on optical response at the same time. Therefore, we proceed to check, within each geometry clusters, how their respective optical responses are distributed. Since the reflection spectra are continuous and thus cannot be divided into categories like design geometries, we choose to manually classify the optical response by certain criteria to see whether the latent space can distinguish different spectra line shape to some extent. Out of all 3000 test samples, we first pick the designs that have strong resonant responses, with the dips of spectra \( R_{xx} \) and \( R_{yy} \) smaller than 0.7 and the peak of spectrum \( R_{xy} \) smaller than 0.3. Then the designs with the strongest resonance of \( R_{xx} \) at a frequency smaller than 60 THz are defined as type 1 optical response, while those with the strongest resonance of \( R_{xx} \) at a frequency larger than 60 THz are defined as type 2 optical response. Figure 2c illustrates the distribution of the selected test data according to their different optical responses (different geometry groups are roughly indicated by the dashed ellipses). Not surprisingly, the two types of optical response are separated into different clusters within each geometry group, especially for split ring and h-shape. In the split ring group, type 1 responses are clustered at two ends while type 2 responses are centered in the middle. In the h-shape group, the two types of responses are also separated with a clear gap. Therefore, the latent variables learned by our model, with the dimension of only 20, not only encode the design patterns with the size of \( 64 \times 64 \), but also reflect the associated optical response with the dimension of 93 in a compact but informative way.

The proposed deep generative model is a comprehensive system for metamaterial design, representation, and characterization. As shown in Figure 3, we demonstrate the performance of our model by evaluating it on three samples randomly drawn from the test data set. The ground-truth geometry and reflection spectra obtained by numerical simulations are plotted as solid lines and insets in Figure 3a,d,g. The forward prediction performance is first evaluated by feeding the ground-truth geometry to the prediction model, and the output spectra, each discretized into 31 data points, are plotted as scattered hollow circles in the same figures. The excellent agreement between the predicted spectra by our model and the numerically simulated spectra clearly confirms that our model can function as an effective simulator for fast metamaterial characterization.

For the inverse retrieval of the designs, we pick the 31 equidistant data points from the solid lines of each numerically simulated spectra in Figure 3a,d,g, and the concatenated 93-dimensional vectors act as one input to the generation model. Another input is the 20-dimensional latent variable, which is sampled from the prior standard Gaussian distribution of the latent space. The generation model accepts these two inputs, and produces the binary image representing meta-atom design. Instead of using only the required spectra as the input, the stochastically sampled latent variable allows diverse outputs.

Figure 2. Reproduced geometry during training process after certain steps a). Visualization of the latent space by reducing the dimension from 20 to 2 using t-SNE. The distribution of the meta-atom design highlighted by its shape b) and by the two spectra types c).
from the generation model, making it possible to solve the one-to-many mapping problem.

Since the true latent distribution is different from the approximated Gaussian prior, as described by the KL-divergence in Equation S1 in the Supporting Information, the sampled latent variable is not able to produce highly accurate retrieval from the given spectra in all cases. However, thanks to the prediction capability, we can feed the generated image back to the prediction model and compare the predicted spectra with the required spectra. Through this self-checking procedure, those highly precise retrieval results can be obtained by simply applying a threshold on the difference between forward prediction and requirements. Under a sum-square-error threshold of 0.2, we give two retrieval results demonstrated in Figure 3b,c,e,f,h,i. The insets are the retrieved design patterns, and the corresponding reflection spectra are obtained by numerically simulating the retrieved design. All the generated designs reproduce the corresponding input reflection spectra with high fidelity, indicating that our model can effectively link the meta-atom design and optical response through the probabilistic representation by latent variables.

Apart from the three representative geometry groups, our model can be easily extended by training on more diverse data sets. To illustrate the generality of the proposed model, we have collected some other training data on various geometries like rectangle, ellipse, arc, bowtie, L-shape, etc. Then a transfer learning method is applied to fine tune the trained model on the new data, which achieves good results on both forward prediction and inverse retrieval. These additional results are presented in Section 7 of the Supporting Information.

Figure 3. Evaluation of the deep generative model. a,d,g) are the samples from test data set where the solid lines and insets are the ground-truth reflection spectra and design pattern, respectively. The hollow circles are the output from prediction model, i.e., the predicted spectra discretized into 31 data points each. b,c,e,f,h,i) are the retrieved designs and their corresponding optical responses.
Although our model performance is quite satisfactory on the test data set in both forward prediction and inverse retrieval, the real-world application usually requires on-demand inverse design of metamaterial with artificially defined optical response according to specific tasks. Therefore, instead of strictly feasible spectra as in Figure 3, a robust inverse design model should also be able to retrieve possible designs from a rough sketch of the predefined spectra. Based on the nature of electromagnetic resonance, we choose to describe each reflection spectrum as the sum of multiple Lorentz line shapes. Therefore, for each on-demand requirement on optical response, we only need to specify the resonant frequency $\omega_0$, bandwidth $\Delta$, and amplitude $A$ for one single resonant feature approximated by Lorentz line shape. The complete reflection spectrum for one set of input-output polarization configuration is given by summing up all the component resonances, as described below

$$ R_{xx,yy} = \sum \left( 1 - \frac{A_i}{1 + \left( \omega - \omega_0 \right) / \left( \Delta / 2 \right) ^2} \right) $$

The on-demand retrieved results are shown in Figure 4 for three cases with two generated meta-atom designs in each case. Similar to the previous model evaluation, the required spectra are fed to the generation model together with the sampled latent space from Gaussian prior, and the raw retrievals are filtered by a threshold of the difference between required spectra and the output from prediction model in order to guarantee accuracy. In Figure 4a, we assume a single resonance for both $R_{xx}$ and $R_{yy}$ with the same bandwidth of 4THz and amplitude $A = 0.7$, at the frequency of 60 and 80 THz respectively. $R_{xy}$ is set as a constant of 0.1 across the frequency range. We demonstrate two out of the many retrieved meta-atom designs in the insets of Figure 4a, and the corresponding reflection spectra obtained by numerical simulations are plotted together as solid lines. Despite the very different meta-atom geometries, the optical responses of these retrieved designs agree fairly well with the required reflections. Similar results are also acquired in two other cases. In Figure 4b, we require that only $R_{xx}$ has a single strong resonance at 70 with bandwidth of 4 THz and resonant dip of 0.1, while other two reflection spectra are set as constant of 0.1 or 1. In Figure 4c, only $R_{yy}$ is required to have a dual band resonance with $A_1 = 0.7$, $\omega_{0,1} = 60$ THz, $\Delta_1 = 4$ THz, $A_2 = 0.7$, $\omega_{0,2} = 80$ THz, $\Delta_2 = 4$ THz, leaving the rest two spectra as constant of 0.1 or 1. In all cases, as long as the prescribed requirements are realizable for the sandwiched structure of the meta-atom, the proposed deep generative model can stably...
make diverse retrievals that best approximate the requirements. More results of on-demand retrieved designs are presented in Section 8 of the Supporting Information (see Figure S5, Supporting Information).

Moreover, by sampling the latent variables from specific parts of the latent space, we can have prior geometric restrictions on the retrieved patterns, and even perform arithmetic operations in latent space to control the output geometry, which offer great flexibility in the on-demand inverse design of metamaterials. These additional results are presented in Section 8 of the Supporting Information (see Figure S6, Supporting Information).

Even though our model is demonstrated to perform well on planar metamaterial under linear polarization, it should be noted that the proposed deep learning architecture, which models a deterministic process in one direction and a generative process in the other direction, can be readily extended to other inverse design problems of metamaterials and nanophotonics as well. Here a typical example is the design of chiral metamirrors, which enables selective reflection of designated circularly polarized light without reversing its handedness, yet high absorption of the other polarization state at certain wavelengths is maintained. Unlike linearly polarized incidence, which primarily interacts with the meta-atom along a certain dimension, circularly polarized light interacts with meta-atom in a more complex way, especially for a 3D structure. Due to the highly nonlinear relationship between geometric chirality and chiroptical responses, the design of such chiral metamaterial structures is not straightforward but still primarily depend on iterative trial-and-error simulations on certain fixed geometries.

To apply our model in the chiral metamirror design, we consider a double layer structure with two meta-atoms separated by two dielectric spacers backed by a reflective ground plane. In this case, we fix both dielectric spacer thicknesses at 300 nm, and the unit cell period and working frequency range are the same as previous planar structures. To account for the top and bottom geometries, the input of the feature extraction network and output of the reconstruction network are altered to two-channel binary images. The chiroptical response is characterized by circular dichroism (CD), which is defined as the absorption difference between left-circularly polarized (LCP) incidence and right-circularly polarized (RCP) incidence. Similarly, the CD spectra are discretized into 31 points before being fed to the model.

Clearly, retrieving a double-layer metamaterial design from given CD spectra that describe its chiroptical response becomes a one-to-many mapping inverse problem, which is readily tractable by the proposed deep generative model. Except the input and output format, the rest of our model is kept unchanged when adapted to the chiral metamirror design. As shown in Figure 5a,b, if we require a CD spectrum with a resonant frequency at 60 THz, bandwidth of 4 THz, and amplitude $A = 0.6$, our model can produce multiple designs that satisfy the requirement. The geometries of both the top and bottom patterns retrieved by the model are plotted as insets of the simulated CD spectra. In the third column, we plot the three reflection spectra of the retrieved design under circularly polarized incidence, where the subscript L and R denote LCP and RCP polarization, respectively. Due to the reciprocity, $R_{LR} = R_{RL}$. Despite the very different full reflections, the CD spectra of both designs satisfy the requirement in the first column quite well. Similar results are also obtained if we require a negative CD value at 80 THz, with the retrieved CD agreeing well with the requirement (Figure 5c,d). The results on chiral metamaterial design manifest that our model is capable of unveiling highly nonintuitive and nonlinear relationship between a subwavelength structure and its optical response, which will substantially accelerate the design of complex photonic designs with novel optical properties.

To conclude, we propose and develop a deep generative model that elegantly solves the metamaterial design problem. Instead of building up a one-to-one mapping between meta-atom geometry and optical response that contradicts physical intuition, our model contains an encoder–decoder structure that first encodes the metamaterial design into a latent space, from which the latent variables are sampled as probabilistic representation of the metamaterial. On top of the latent variable, we can easily realize diverse retrieval of the meta-atom geometry given the required spectra, solving the one-to-many mapping problem in the inverse design. The forward prediction of optical responses is achieved by a deterministic prediction model, which shares the extracted features with the encoder. Moreover, the encoder–decoder configuration in our model allows to utilize both labeled and unlabeled data in a semi-supervised learning scheme, which helps to exploit human experience to maximum extent while alleviating the burden of numerical simulations. The high efficiency and versatility of the proposed deep generative model make it a promising candidate in the research field of metamaterials and photonics, where the design is mainly based on physics-inspired methods in conjunction with numerical simulations from trial and error. The latent variable encoding and the subsequent generative process in our model render the inverse design more stable, diverse, and versatile, and can be readily extended to other research domains of photonic and material science, enabling on-demand design, characterization, and even new discoveries in various applications.

Experimental Section

In order to demonstrate the performance of the proposed deep generative model, metamaterial patterns were collected from several geometry groups such as cross shape, split ring, h-shape, L-shape, rectangle, ellipse, bowtie, arc, etc. These geometry groups were selected because they have been adopted by previous works as efficient and fabrication-feasible candidates of meta-atoms and thus represent typical human experience in metamaterial design. To increase diversity, all possible design parameters such as length, width, relative offset, and rotational angle of different parts of the geometries were randomly sampled, and then random global distortion was applied to the pattern before discretized to $64 \times 64$ binary image. Therefore, the generated metamaterial pattern covered a vast variety in design which can hardly be described by a few fixed parameters. Considering the achievable feature size in practical fabrication processes, an image processing method was also proposed to post-process the model output, which can help to modify or eliminate designs that does not meet a prescribed fabrication tolerance. More details can be found in Section 9 of the Supporting Information.

10 000 patterns were collected with numerically calculated spectra (by CST Microwave Studio) for cross, split ring, and h-shape, which
serve as the basic designs to test this model performance under the semi-supervised learning strategy. Another 10 000 labeled samples of other various design patterns were collected for transfer learning, in order to test the generality of our model on highly diverse design patterns. 1000 data were also prepared for each geometry groups as test set. For the stacked chiral metamaterials design, 3000 possible shape pairs were randomly sampled from all the geometry groups for both top and bottom resonators, and the metamaterial response was simulated under circularly polarized incidence.

In the numerical simulation, CST Microwave Studio was employed to generate the reflection spectra data. The spacer was modeled as a lossless dielectric with refractive index of 2, and gold was treated by Drude model. The proposed model was constructed under the open-source machine learning framework of TensorFlow.

![Graphs showing on-demand inverse design of double layer chiral meta-mirrors given specific CD spectra. Positive CD resonance at 60 THz (a,b) and negative CD resonance at 80 THz (c,d). Retrieved top and bottom meta-atom design (insets in the second columns) and their corresponding full reflection spectra (third columns) from numerical simulation under circularly polarized incidence.](image-url)
Supporting Information
Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements
The authors acknowledge the financial support from the Office of Naval Research (N00014-16-1-2409) and National Science Foundation (EEC-1916839).

Conflict of Interest
The authors declare no conflict of interest.

Keywords
deep learning, metamaterials, photonics