Article

A Deep Convolutional Mixture Density Network for Inverse Design of Layered Photonic Structures

Rohit Unni, Kan Yao, and Yuebing Zheng

ACS Photonics, **Just Accepted Manuscript** • DOI: 10.1021/acsphotonics.0c00630 • Publication Date (Web): 07 Sep 2020

A Deep Convolutional Mixture Density Network for Inverse Design of Layered Photonic Structures

Rohit Unni,[1,2,†] Kan Yao,[1,2,†] and Yuebing Zheng[1,2,*]

[1]*Walker Department of Mechanical Engineering, The University of Texas at Austin, Austin, Texas 78712, USA*

[2]*Texas Materials Institute, The University of Texas at Austin, Austin, Texas 78712, USA*

[*]Corresponding Author: zheng@austin.utexas.edu

ABSTRACT: Machine learning (ML) techniques, such as neural networks, have emerged as powerful tools for the inverse design of nanophotonic structures. However, this innovative approach suffers some limitations. A primary one is the non-uniqueness problem, which can prevent ML algorithms from properly converging because vastly different designs produce nearly identical spectra. Here, we introduce a mixture density network (MDN) approach, which models the design parameters as multimodal probability distributions instead of discrete values, allowing the algorithms to converge in cases of non-uniqueness without sacrificing degenerate solutions. We apply our MDN technique to inversely design two types of multilayer photonic structures consisting of thin films of oxides, which present a significant challenge for conventional ML algorithms due to a high degree of non-uniqueness in their optical properties. In the 10-layer case, the MDN can handle transmission spectra with high complexity and under varying illumination conditions. The 4-layer case tends to show a stronger multimodal character, with secondary modes indicating alternative solutions for a target spectrum. The shape of the distributions gives valuable information for post-processing and about the uncertainty in the predictions, which is not available with deterministic networks. Our approach provides an effective solution to the inverse design of photonic structures and yields more optimal searches for the structures with high degeneracy and spectral complexity.

KEYWORDS: *deep learning, artificial neural networks, multilayer structures, nanophotonics, inverse design, non-uniqueness*

Modern nanophotonic structures, including metamaterials and plasmonic structures, feature a wide range of optical responses for various applications.[1-3] The optical properties of nanophotonic devices, unlike those of their bulk optical counterparts being largely determined by the material properties,[4] have strong dependence on not only the constituent materials but also the geometry of individual building blocks that are subwavelength in size, their arrangement, and the illumination conditions such as the polarization and angle of incidence.[5, 6] These variables form a hyperspace of possible designs, where each set of parameters uniquely defines a design of a nanophotonic structure with a certain optical property. Inverse design of nanophotonic devices is thus a task of searching this space for an optimal set of parameters that can produce the desired optical response.[7, 8] However, the parameter space can be enormous, and the relationship between the designs and optical properties is complex and usually implicit. With limited physical intuitions to guide the search, the traditional trial-and-error framework is inefficient in improving from the initial guess. Computational techniques such as genetic algorithms[9, 10] and topology optimization[11-14] have been utilized for inverse design as well. These techniques drive the design improvement by optimizing certain objective functions and enable discovery of solutions not available to intuition-based methods. However, despite versatile applicability and scalability, computational techniques require recurring computational efforts for every design request. All these limitations motivate the need for innovative design methods.[11, 15, 16]

A promising approach for inverse design is the use of machine learning (ML) algorithms to calculate the required parameters.[8, 17, 18] ML algorithms operate by leveraging large labeled datasets to learn complex relationships and to optimize an objective function mapping the inputs to the outputs. In the case of photonic structures, the inputs are the optical properties, e.g., spectra, and the output labels are the design parameters. A variety of nanophotonic structures such as multilayer nanoparticles,[19, 20] metasurfaces,[21-24] metagratings,[25, 26] split ring resonators,[27] compound metamolecules,[28] and color generation[29, 30] have been inversely designed using ML algorithms. ML has also been utilized to predict optical properties and electromagnetic field distributions of different structures[31, 32] and to decode optical images, videos and spectra.[33] While ML has shown to be extraordinarily powerful in learning complex and unintuitive relationships within datasets to make accurate predictions, it still suffers from several issues.

Photonic structures are particularly vulnerable to the non-uniqueness, or the many-to-one problem.[34] It is not uncommon for structures with wildly divergent designs to produce nearly

identical optical properties (Fig. 1a). The non-uniqueness causes problems for convergence because ML algorithms typically aim to optimize a single mapping from inputs to outputs by assuming a one-to-one correspondence between them. In this one-to-one paradigm, there is a single "correct" answer for each output, and the algorithm's goal is to update its parameters until the correct answer can be obtained as often as possible. However, when there are multiple correct answers for a given input, the algorithm could be conflicted on how to adjust, and convergence is thus not guaranteed. For photonic structures that are particularly sensitive to their designs, such as multilayer thin films (Fig. 1b), there is a high degree of non-uniqueness in the data, presenting a significant challenge for modeling. Moreover, the optical properties of many photonic structures are intriguing. Various physical mechanisms, including sophisticated resonances from different excitation conditions, coupling between neighboring units, and interference, could all come into play, resulting in a dense population of distinct spectral features across the wavelengths of interest. In the inverse design, finding a solution that can accurately reproduce the desired optical response of a photonic structure with such high complexity is another fundamental challenge.

Thus far, nearly all ML-based inverse design works have utilized artificial neural networks (NNs) as the underlying model. NNs work by a large series of interconnected processing nodes mapping the inputs to the outputs, where the internal parameters are tuned by repeatedly feeding training data to refine the accuracy of the outputs (Fig. 1c).[35] Standard NNs make deterministic predictions, embodied by discrete values, at each output neuron corresponding to a design variable. Nevertheless, this paradigm struggles with convergence issues if there exist multiple solutions that are comparable or the spectra contain very complex features. Here, we introduce a mixture density network (MDN) as a novel approach to these outstanding limitations. MDNs operate by modeling the final output as a probability distribution of possible values rather than single discrete values as with standard NNs (Fig. 1c).[36] Being successfully implemented for applications such as speech inversion,[37] volatility prediction,[38] and material modeling,[39] MDNs have not been applied for inverse design. In theory, MDNs can handle arbitrary amounts of non-uniqueness in the datasets and capture any number of degenerate solutions. The shape of the distributions also gives information about the confidence of the model's predictions, allowing for a wide range of sampling techniques to optimize the design and find alternative solutions that avoid the issues with the standard deterministic NNs. To demonstrate this concept, we develop a deep convolutional mixture density network model for the inverse design of multilayer photonic structures. We show

the efficacy of our approach by modeling the transmittance of transverse magnetic (TM) polarized light through a 10-layer structure of stacked alternating oxides with arbitrary angles of incidence. Despite the simplicity of the structure, it produces spectra with sharp and closely spaced peaks and dips because of interference. Our MDN model retrieves these complex features well. Moreover, enabled by the probability distribution of the design variables, post-processing sampling is developed to further improve the accuracy. We also apply our model to a simpler 4-layer structure to highlight its ability to produce multimodal distributions for cases of non-uniqueness. Our approach offers a more widely applicable solution to the many-to-one problem while being able to handle much more complicated optical data inputs.

Results and Discussion

Network Architectures. We begin by taking a closer look into the current limitations of NN-assisted inverse design. First and most importantly, there is still no reliable method for fully dealing with the degeneracy problem that arises from the non-unique response-to-design mapping. With a standard NN (Fig. 2a, left panel), such mapping may pull the weights in different or even opposite directions in the hyperspace, resulting in difficult convergence or converging in between degenerate ground truth solutions (Fig. 2a, right panel). Early work has attempted to divide the data into groups that eliminate duplicates;[40] but this is not feasible for photonic applications where the non-uniqueness is more severe and would require training of many different models separately. Dimensionality reduction has also shown some promise in relieving non-uniqueness, but it suffers from limited applicability as the reduced spaces are not guaranteed to be one-to-one mapped.[41] Another approach that has been proposed is the tandem network architecture (Fig. 2b, left panel).[27, 34] By attaching a pretrained forward modeling network to the end of an inverse design network, tandem networks relax the requirements of converging, which alleviates the non-uniqueness issue but does not completely solve it. When there are multiple candidate solutions to a certain design request, the weights in the inverse design network will still see conflicting gradients that hinder effective converging. Furthermore, if convergence does occur, the network returns a single solution, which is not guaranteed to be the ground truth design, for each input and ignores other viable outputs (Fig. 2b, right panel). A second issue with the current NNs is that they provide limited knowledge about the outputs, which is informative for optimizing the design and understanding the underlying physics. Given that inverse design is a regression task, a given

prediction is expected to have some deviation from the correct answer. However, it is difficult to estimate the magnitude of this deviation without knowing the ground truth values, as the user has only the average deviation across the entire dataset without more granular information than that. In tandem architectures, where the error can propagate between multiple networks, this problem is strongly amplified. The only way to know the viability of a predicted design is to simulate or experimentally measure it. If it proves to be wrong, the model has no recourse because it is deterministic and will always output the same wrong answer. Thus, it would be extremely valuable for real-world applications to have a model to output a prediction as well as further information that can be used to optimize the predicted design, interpret the reliability of that prediction, and suggest alternative solutions.

To address the above challenges, especially the non-uniqueness, we adopt the concept of MDN that operates differently in making predictions. Standard NNs have the output neurons correspond directly to the discrete values of each output for design variables. In contrast, our MDNs model the outputs as a mixture of several Gaussian probability distributions, which are sampled for individual design variable predictions. As illustrated in Fig. 2(c), the output neurons correspond to the parameters of these distributions, with each parametrized by a mean μ and a variance σ. In the following, the terms "distributions" and "modes" are sometimes interchanged when referring to individual peaks in the mixed probability curve determined by the MDN output. For a given design variable, these distributions are summed with a weight parameter π into the final probability distribution. The final output defines a probability density function over the continuous space of each design variable, with the function output being the relative probability of a given design value. Therefore, rather than a standard NN, which outputs a single estimate of what it believes the correct value is, an MDN outputs an estimate of what it believes the chances are for every possible design value to be correct. A trained network will always produce the same probability density function. However, this distribution can be sampled multiple times in post-processing for different design variable values, making the whole framework non-deterministic.

We demonstrate our technique on the inverse design of layered photonic structures as shown in Fig. 1(b). The structure consists of 10 layers of alternating SiO_2 and TiO_2 illuminated from the top by TM-polarized light at a variety of angles of incidence. The optical properties, including transmittance, are simulated by solving the Fresnel equations in MATLAB[42] for wavelengths between 300 and 1000 nm. The design variables include the thickness of each layer

and the angle of incidence, forming a vector in 11 dimensions. The thickness of each oxide layer is between 10 and 300 nm and the angle of incidence is between 0° and 40°. Larger-angle incidence and transverse-electric (TE)-polarized incidence are studied as separate cases (see Supporting Information, Section 1). For each sample of the dataset, we randomly choose each design variable uniformly within the designated ranges and discretize the spectrum at 301 evenly spaced points. At the listed conditions, the structures feature unusually high spectral complexity and strong sensitivity to the design variables as compared to most metasurfaces and photonic devices, which is challenging for NNs to make accurate predictions. We generate a dataset of 144,000 samples and split it into 70% training and 30% test. The former dataset is used to train the model, and the latter is used to evaluate the model's accuracy to ensure that it does not overfit to the training data and can generalize to new samples it has not seen. The design variables are rescaled such that the minimum possible value of each variable (30 nm for thicknesses, 0° for angle) is adjusted to 0, the maximum value (300 nm for thicknesses, 40° for angle) is adjusted to 1, and intermediate values are linearly mapped between 0 and 1. The rescaling is necessary because the values of the angle of incidence are orders-of-magnitude greater than the layer thicknesses, which can hurt the early portions of training. We emphasize that incorporation of variables of different natures (e.g., layer thicknesses and illumination conditions) in the design vector of NNs has been rarely explored. This attempt, nevertheless, could extend the applicability of ML to inverse design tasks and enable complete searches in the design hyperspace.

Inverse Design with MDNs. We implement the MDN by utilizing three sets of convolutional and pooling layers at the input, followed by three fully connected layers and finally the adapted layer with 16 mixtures at the output. The convolutional layers pass a series of filters over successive values in one layer and perform a convolution with the filter weights. For data such as the transmission spectra used here, these filters typically learn to identify important spectral features such as the location of peaks and valleys, and the pooling layers group nearby data, both of which allow the training to run more efficiently than the fully connected layers that have far more weights to optimize. For the cost function, we utilize the negative log-likelihood metric. The full details on the architecture and hyperparameters of the network are provided in the Supporting Information (see Sections 2 and 3). The network is trained for 600 epochs (Fig. 3a), reaching an average negative log-likelihood of −5 with the training data and −4.5 with the test data. The range of possible values of log-likelihood is highly dependent on the nature and distribution of the data

being modeled and has little intrinsic interpretability by itself. We present more detailed qualitative and quantitative measures of the model's accuracy in the following section. The learning rate is dynamically adjusted several times during the training, dropping by a factor of 1.4 whenever the cost for the test data does not decrease for 10 successive epochs, resulting in the stepwise sharp drops in the training curve. The final models were trained on the Stampede2 supercomputer at the Texas Advanced Computing Center,[43] with training times ranging between 30 and 40 seconds per epoch and a total computation time of 7 hours.

The final model produces distributions consistently centered near the ground truth values with associated uncertainty quantified by the width of the distribution modes. We demonstrate the capability of trained model to inversely design structures. The optical properties of the predicted designs are computed using the aforementioned MATLAB code for verification. The 10-layer structure shows extremely high spectral complexity and high sensitivity to small deviations in the design variables, making the dataset extraordinarily challenging to model through conventional means. Despite this, the sophistication and depth of the initial layers of our architecture and the flexibility afforded by the MDN approach allow us to achieve levels of accuracy consistent with prior NNs trained on far simpler spectra.[34] We also train our model on a simplified 4-layer case limited to normal incidence to demonstrate the higher capabilities of our model when handling simpler structures. For this model, we train 50,000 samples split into training and test datasets at the same ratio as the 10-layer case, using the same architecture, with training stopping at 100 epochs. The number of samples in our study roughly scales linearly with the number of design variables, while the data requirement can be relaxed for structures with lower spectral complexity and at the cost of longer training.

We discuss the results of the 10-layer structure first. Due to the degree of variance in most predicted distributions and the high sensitivity of the 10-layer case to small changes in the input thicknesses, one may expect that sampling the model only once will typically yield poor designs. Since each variable is sampled independently, it is likely that at least one or two values will diverge from the ground truth enough to cause the real spectrum to diverge, making it necessary to take multiple samples. As an initial run, we take the center of the most prominent mode of the distribution, since the 10-layer structure displays quasi-unimodal character for many cases and the mode is centered at or close to the ground truth value. Figure 3(b) presents the comparison between a ground truth spectrum selected from the test dataset and the spectrum of the predicted design

when plugged back into simulation. Even with a naive sampling strategy, the predicted design shows a moderate degree of correspondence in the location and magnitude of the peaks and valleys for wavelengths below 370 nm and above 450 nm, while obvious deviations occur for the most irregular spectral features between 370 and 450 nm.

Next, a more sophisticated sampling strategy is devised to enhance the accuracy of the model. As schematically shown in Fig. 4(a), we implement a simple and quick post-processing method to further refine the designs. Briefly, we sweep through the design vector and test new guesses based on the probability distributions until a best design is found. Detailed procedures are described in Section 4 of the Supporting Information. The improvement by post-processing is significant, which can be visualized when we revisit the design request in Fig. 3(b). Figure 4(b) and 4(c) compare the design vectors and spectra between the ground truth and two designs without and with post-processing refinement. Individual design values after post-processing generally match the ground truths better, which helps to resemble all the complex spectral features with high accuracy. For a randomly selected sample of 50 spectra from the test dataset, our post-processing method improves the root mean squared error (RMSE) between ground truth and the spectra produced by the MDN by an average of 42%. A more detailed accounting of the simulated designs is given in the Supporting Information (Section 5). We note that the computational cost of the post-processing, measured by the number of iterations of sampling, results from the unusually high spectral complexity and non-uniqueness of the present structure, and for normal design tasks, it can be effectively reduced to the level comparable to those of the conventional optimization methods. Even for the multilayer structures, high degrees of improvement can still be achieved with lower numbers of iterations, and the full data on these results is given in the Supporting Information (Figure S3, Figure S4). It is also expected that smarter sampling strategies (e.g., optimization methods that can process all variables at a time or an independent neural network) can make MDNs more efficient and accurate.

The accuracy of our model across the whole dataset is first shown by comparisons of model output distributions for single design variables with the ground truth values for both sets of data, with one example of each randomly selected (Fig. 5a). As stated before, the output of the model for a single design variable comprises a probability density function. The *x*-axis corresponds to the possible values for the design variable, rescaled from 0 to 1 as in the training data. The units and scale of the *y*- axis have no physical interpretation; rather, it corresponds to the relative likelihood

that a given discrete value for the design variable in the continuous space of [0,1] will be chosen when sampling the distribution. We find that the local maxima in the distributions, i.e., the peaks in probability, tend to match closely with the ground truth values. However, the simpler 4-layer structure often tends to feature multimodal distributions, i.e., multiple peaks (see Supporting Information, Section 6 for statistics). We sample 4400 design variables from each of the test datasets of our four MDN models and measure the error of the predictions, as plotted in a histogram in Fig. 5b. We define the error as the absolute value of the difference between the center of the most prominent peak in the probability density function for a given design variable and the ground truth value in the normalized scale. The histogram plots the counts of different ranges of error, with most values falling close to zero error.

We further demonstrate the advantages of our method as compared to an alternative approach to the many-to-one problem. We train a deep tandem network on the same 10-layer dataset. The forward network is trained with fully connected layers and connected to the inverse network using a similar architecture of convolutional and pooling layers used in the first section of the MDN model. The network is trained with RMSE as the cost function. The forward network converges to an error of 0.06 while the full tandem network converges to 0.08. The high complexity of the 10-layer transmission spectrum causes some error in the forward network, which propagates to the inverse portion. This uncertainty can cause the model to produce designs that result in inaccurate spectra when plugged into simulation despite reproducing the ground truth spectrum through the neural network (Fig. 5c). In order to make a fair comparison with the results of our MDN with post-processing, we apply an analogous post-processing method to the prediction of the tandem network (see Supporting Information, Section 4 for details). Nonetheless, the result of the tandem network is not improved by an appreciable amount. In sharp contrast, much higher correspondence can be achieved by the MDN approach. This illustrates that the probabilistic nature of MDNs offers unique optimization opportunities that are impossible with standard NNs.

The MDN trained for the 4-layer structure produces narrower peaks. It can also show multimodal distributions in cases where degenerate solutions are viable in producing the desired spectrum. The 4-layer structure showcases MDN's capability to uniquely address the many-to-one problem without sacrificing viable solutions. We use a selected example shown in Fig. 6 to illustrate this. A desired spectrum from the test dataset is fed into the model. The predicted distribution mixture parameters are used to construct the full probability distribution for each of

the four design variables, i.e., thicknesses of the four layers in the structure (Fig. 6a). Different designs are sampled from these distributions and fed back into simulation to see how well the simulated spectra match the desired spectrum. Some layers display multimodal character to their outputs, and designs can be obtained by taking samples from different modes to find alternative solutions away from the ground truth design from the dataset. Selected results are summarized in Fig. 6b for comparison. To illustrate how well these secondary modes can capture degenerate solutions, we do not apply any post-processing for this example. Each design variable for the three different designs is taken from the center of one of the distribution modes. Although the closest agreement with the desired spectrum can be achieved for sample 1 where all the four variables are close to the ground truths, sample 3 containing one thickness sampled from a secondary mode can still display reasonably good agreement, particularly at shorter wavelengths. The distributions also display a range of shapes, with some narrow peaks and some wide or merged ones. The diversity in the shape of the distributions illustrates another unique advantage of our model. Standard NNs are deterministic and will always produce the same prediction for a given request once trained. The only measure of the model's uncertainty is the average deviation across the entire dataset, which is merely the mean of a wide distribution of possible deviations, and one has no way of knowing where a given prediction may fall on that distribution outside of trial-and-error. The only way to assess the validity is to test the predicted design for its optical response by simulations or experiments. If the response differs significantly from the requested one, there is no additional information as to how far the design is from a correct one, and the model has no way of outputting any alternates. For complex design tasks, the model will give incorrect answers for an appreciable number of inputs and be useless. With our MDN model, one can extract additional information from the shape of the distributions to help address these challenges.

We sample both MDN models for the 4-layer and 10-layer cases (5000 times each) with different requested spectra and collect information about the widths of all the contained modes and the errors with respect to the ground truth. Full details on the calculation of the mode widths are provided in the Supporting Information (see Section 6). The error is defined as the difference between the center of the mode containing the ground truth and the ground truth itself, and we compare this error to the width of the selected mode. In other words, we wish to know that when a value is chosen near the center of a mode, how likely our choice is to be close to the correct value. In the most ideal case, the center of at least one of the modes would always correspond

exactly to a correct value for the design variable and the error would be zero. In practice, however, there are cases for which the true value is not exactly at the center of a mode, similar to how standard NNs can predict incorrectly. When comparing across the whole dataset, we find a moderate positive correlation between width and error, with a correlation coefficient of 0.6. The full scatter plot of this sampled data is given in the Supporting Information (Figure S10). This means, in general, that the narrower a peak in the distribution is, the more likely its center is close to an exact solution, which allows the model to provide information about its confidence in the predictions. A widely spanned mode indicates high uncertainty, where the optimal value is more likely to be some distance away from the peak, while a very narrow peak indicates high confidence with that specific variable. An example can be seen in Fig. 6a, where the distribution for the thickness of layer 1 has a secondary mode with the largest width of any in the four distributions, and the design that samples from this mode (i.e., design 3), has more noticeable deviation from the ground truth than design 2, which samples from the narrower secondary modes in layers 2 and 4.

The information from the distribution shape also offers an explanation to the improvement from our post-processing method. In a probability density function, narrow peaks will be more likely to be sampled close to their center, and broad peaks, which spread the probability density over a larger range, are more likely to explore further areas of the design space. This allows a much greater chance of finding the ground truth or another degenerate value when the initial output of the MDN is not centered correctly. When the ground truth is not known, as with real-world applications, an optimal design could still be reached by resampling the marginal distributions with the post-processing or other conventional optimization techniques,[44] rather than searching the entire design space. Even the most accurate NNs trained on simple tasks will give wrong answers in some cases. However, in cases with incorrect predictions, MDN can still aid in restricting the search space and provide information to find the true design, showing superior applicability.

Lastly, we show that the proposed method can be used to search for designs that produce idealized fictitious spectra. In Figure 7, aiming to achieve functional devices with high transparency over a certain range of wavelengths, we generate three fictitious spectra as desired responses (see Supporting Information, Section 7 for details). The post-processing method is also modified to limit the optimization to the desired wavelength ranges. Interestingly, despite the fictitious spectra not resembling any of the ones in the training or test datasets, the MDN for the 4-layer structure have successfully found solutions that match the transparent windows for all the

requests. These results confirm that, once trained, NN-based models can be extended to solve other design tasks without additional computational costs.

Our MDN model demonstrates the flexibility to effectively handle both structures with complex optical responses and the simpler structures with multiple degenerate solutions, making it an ideal candidate to extend the applicability of inverse design to a wider range of nanophotonic devices. The ability to handle sharp spectral features is extremely valuable in modeling many thin film structures such as one-dimensional photonic crystals.[45] Even our simplified 4-layer structure displays the sharper spectral features than those shown in previous works. It is expected that our model can attain more accurate predictions for structures like metagratings[25] and layered nanoparticles[19] while capturing degenerate solutions. It can also prove fruitful for the inverse design of anti-reflective coatings[46] and metasurfaces for wavefront shaping.[47] The model's ability to estimate its confidence in the predictions increases its viability in all real-world applications.

Conclusion

In conclusion, the proposed MDNs exhibit unique advantages in addressing many of the limitations, especially the non-uniqueness issue, shown so far with NNs for the inverse design of photonic structures. By modeling the designs as probability distributions, the MDNs can alleviate the difficulty of convergence when multiple structures result in identical optical responses, and further discover these viable degenerate solutions with estimable confidence instead of blindly converging to one of them in a deterministic manner. Despite a high degree of non-uniqueness in the input dataset, our implemented deep convolutional MDN can make accurate predictions for multilayer photonic structures with high spectral complexity and produce multimodal distributions representing multiple viable designs for the simpler structures. The distributions' uncertainty adds stochasticity to the predictions, allowing for finding of the ground truth values even when the output is not centered correctly, and provides further information on the accuracy of the model, making this method more viable for real-world applications.

Transfer learning can be leveraged for our MDNs to take advantage of the information gleaned from one structure to more efficiently learn an inverse design approach for a similar structure without having to train a brand new model from scratch, and it has shown some success on modeling multilayer structures featuring simple optical spectra.[48] Another approach that has been proposed for a different purpose and can be potentially useful for solving the non-uniqueness

problem is the use of a variational autoencoder (VAE) to compress input and output data into a latent space, which is sampled by a generative model to create candidate design geometries.[49] The VAE approach offers another probabilistic approach to modeling photonic structures with degenerate solutions; however, it still suffers from a limited range of applications. The generative model employed is primarily suited to geometric designs that can be represented with an image and does not translate well to designs encompassing discrete variables with a wide range of possible values such as the dimensions and compositions of materials. Furthermore, in practice, the VAE encoding of the latent space shows limited ability to suggest alternative designs that are topologically different.[21, 49] Such limitations can be overcome by combining NNs with other techniques such as topology optimization,[24, 25] or, employing new design frameworks,[50] to search the design space more thoroughly, efficiently, and practically. Even the most well designed NNs will have a distribution of possible errors, where any given single prediction may be on the high error side, failing to produce a viable design suggestion. For structures with a highly complex relation between design and response, being able to approximate multiple unintuitive alternative solutions as well as the ground truths can significantly improve the model's ability to suggest accurate designs, as it can spread the expected error over multiple possibilities providing that only one design needs to exactly match the requested spectrum.

Finally, we envision our MDN method can be expanded to the more sophisticated photonic devices. Many aspects of the model such as the loss function, architecture, and hyperparameters could benefit from further improvement and grid-searching. A more sophisticated sampling strategy can be devised to take care of the information in the distributions to make more accurate predictions. The model can be adjusted so that the distributions of each design variable are not trained separately, allowing for a full covariance matrix to be learned, which could yield better predictions and new physical intuitions about the structure-property relations.

†These authors contributed equally.

Acknowledgements

The authors would like to acknowledge the financial support from the National Aeronautics and Space Administration (NASA) Early Career Faculty Award (80NSSC17K0520) and the National Institute of General Medical Sciences of the National Institutes of Health (DP2GM128446). They

also thank the Texas Advanced Computing Center (TACC) at the University of Texas at Austin for providing high-performance computing resources.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at doi:

Additional models for large angles of incidence and TE polarization; Architectures of NN models; Hyperparameters; Post-processing procedure; Model performance deviations and spectral complexity; Correlation between mode uncertainty and width, and mode statistics; Generation of idealized spectra.

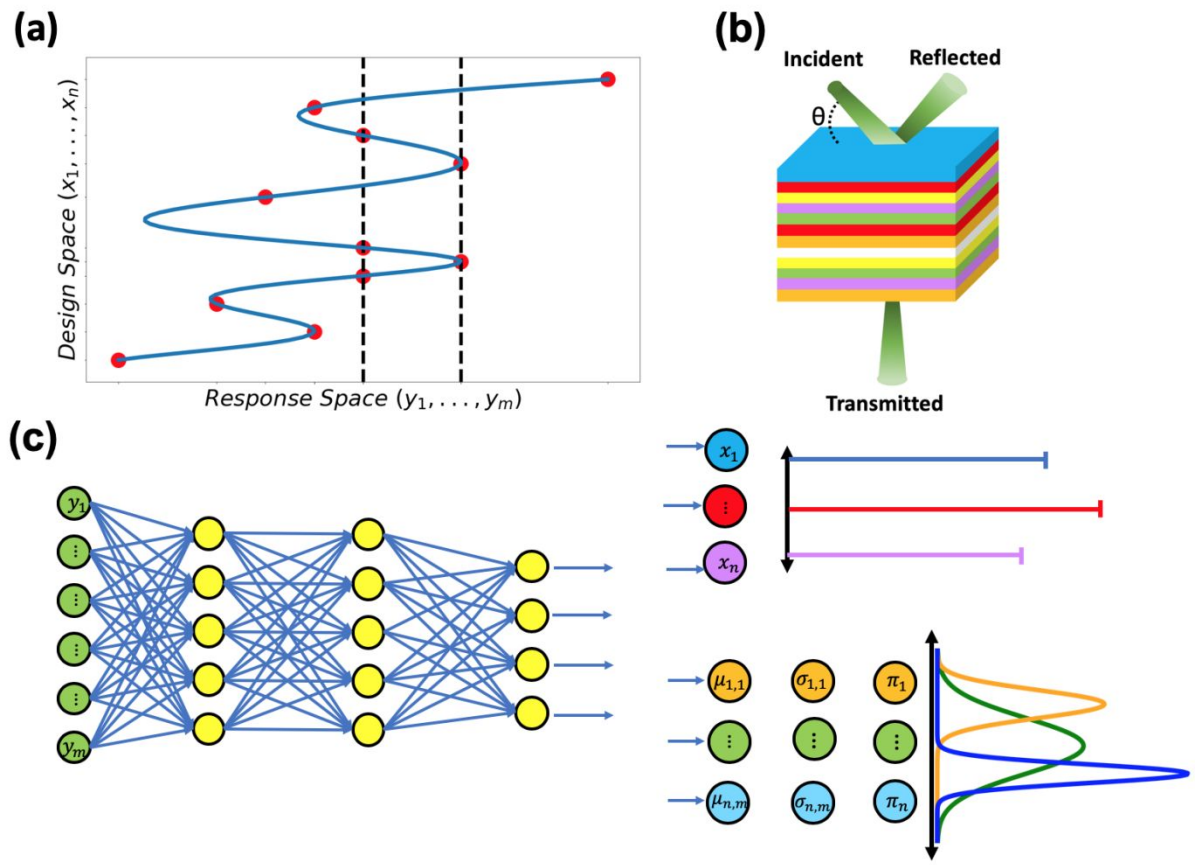


Figure 1: (a) Illustration of many-to-one problem in the inverse design of photonic structures with ML. The goal is to learn a function that maps the optical properties (y_i) in the response space to the design variables (x_i) in the design space. However, when multiple designs correspond to identical optical responses (as indicated by dashed vertical lines), there is no deterministic function that can model all solutions. (b) Multilayer dielectric thin films represent a class of structures that feature high degeneracy and complexity in their optical properties, such as transmission and reflection. (c) Illustration of neural network (NN) architectures with differences between standard network and the mixture density network (MDN). NN is constructed from layers of neurons connected to one another. In the case of fully connected layers (left), each neuron is connected to every neuron in the next layer. At the output of a standard network (top right), the output neurons correspond directly to the desired discrete values for individual design variables. In contrast, for an MDN (bottom right), the output neurons are parameters of probability distributions that are combined to model all the candidate designs.

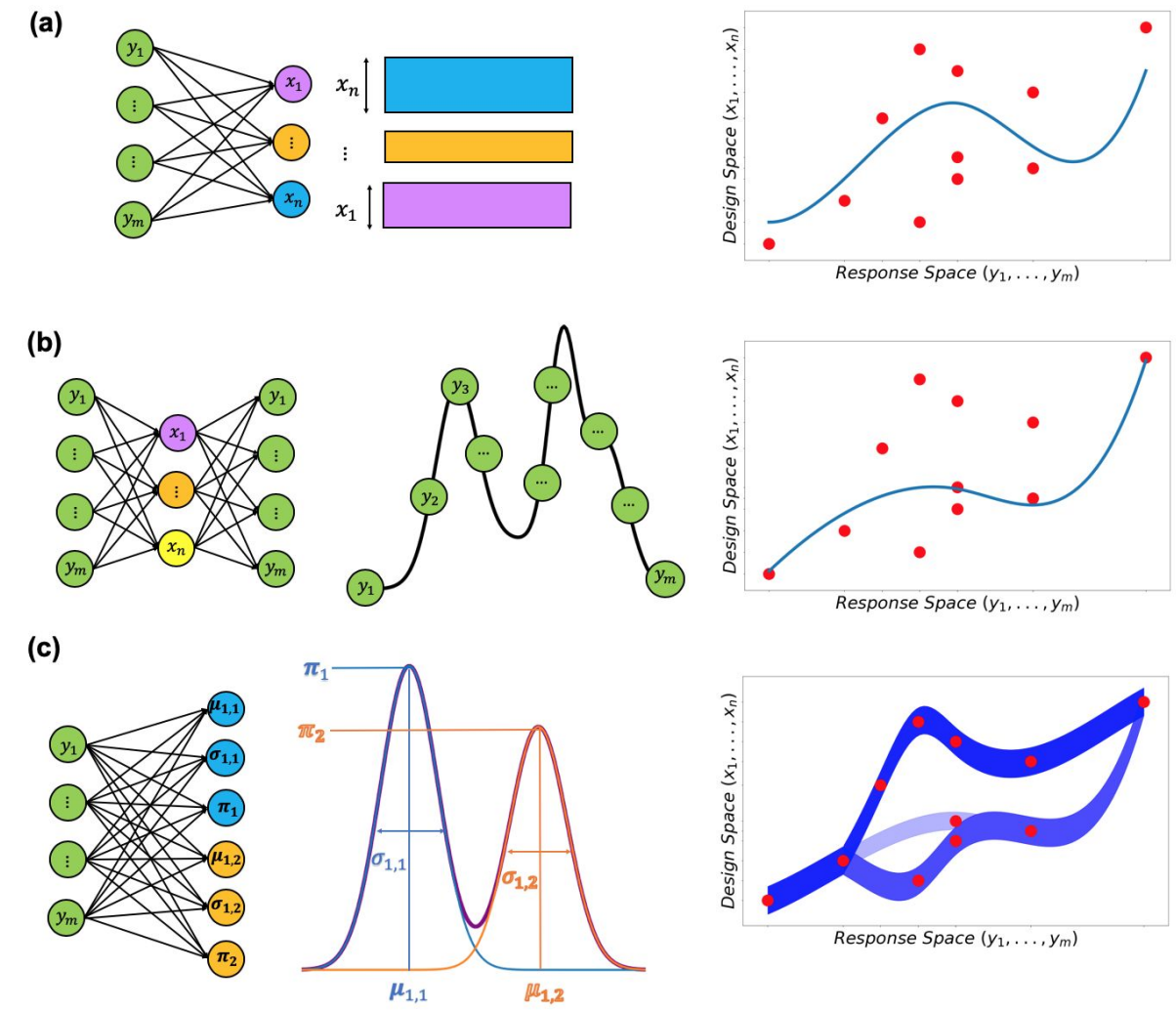


Figure 2: Different types of NN models for solving the many-to-one problem. (a) A standard NN maps straight from the response (y_1, …, y_m) to the design (x_1, …, x_n). Attempting to learn a deterministic mapping can cause the model to try to predict in between degenerate solutions (red dots), and even prevent it from converging with unique solutions. (b) For the tandem network approach, the response is mapped to the design, which is connected to a pretrained and frozen modeling network that maps the design to the response. This can still have difficulty in converging, and an optimal solution may ignore other viable options. (c) An MDN produces a mixture of multiple Gaussian distributions for each design variable. Each distribution in the mixture is parametrized by a mean μ, a variance σ, and weight π. Probability distributions are represented by the shaded strips rather than a single line for deterministic mapping (right panels). The MDN can capture all degenerate solutions through multimodal distributions, with the relative strength of the modes visualized by the opacity (related to π) of each strip.

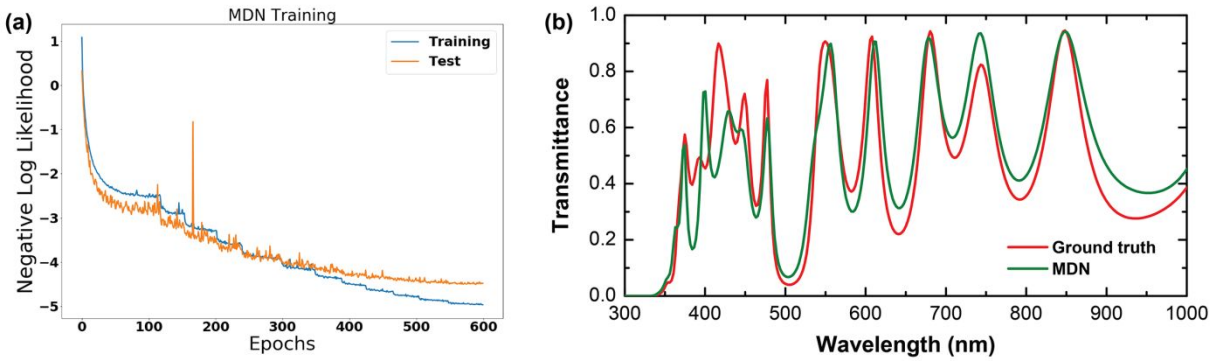


Figure 3: (a) Learning curve of the MDN trained for the 10-layer photonic structure for both the training and test data over 600 epochs, stopped early to prevent overfitting. The loss function minimized is the average negative log-likelihood. The test error is approximated from samples of the test data, which can occasionally cause large spikes due to sampling error. (b) Comparison of a requested spectrum (red curve) with the spectrum produced by the design suggested by the MDN model (green curve).

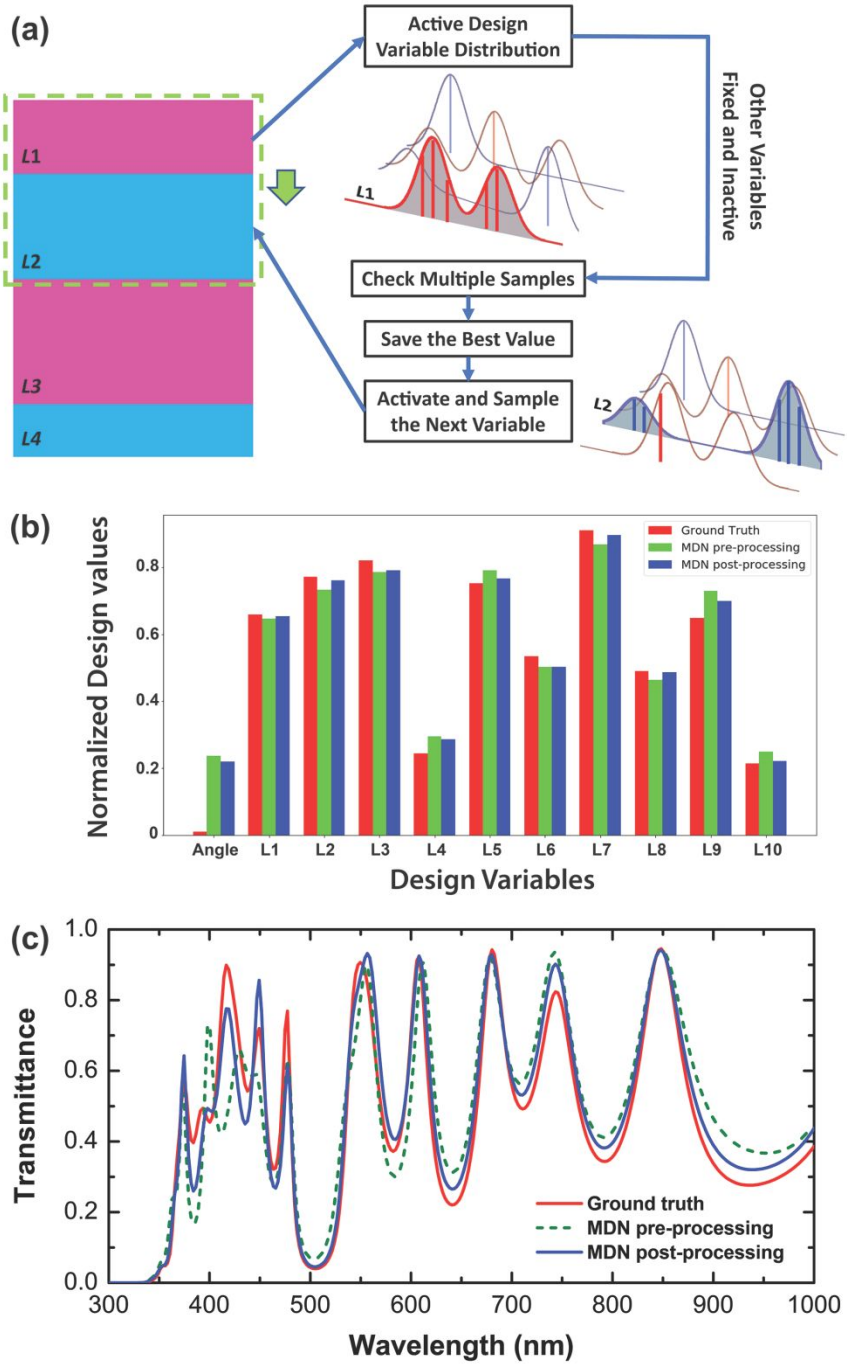


Figure 4: (a) Post-processing procedure. Following the output of the MDN, each individual design variable is sampled around the peaks of its probability distribution for a better design while other variables are fixed. *L*1, *L*2, *L3, and L4*denote the distributions for the thicknesses of layer 1, layer 2, layer 3, and layer 4. (b) Comparison of the discrete design values of the ground truth structure corresponding to the requested spectrum (red) in Fig. 3b, the design produced by the MDN without post-processing refinement (green), and the design after post-processing (blue). "Angle"

corresponds to the angle of incidence, while L1, …, L10 refer to the thicknesses of layers 1 through 10. (c) Comparison of the transmittance spectra of the three designs in (b). Obvious improvements can be seen in the post-processing result.

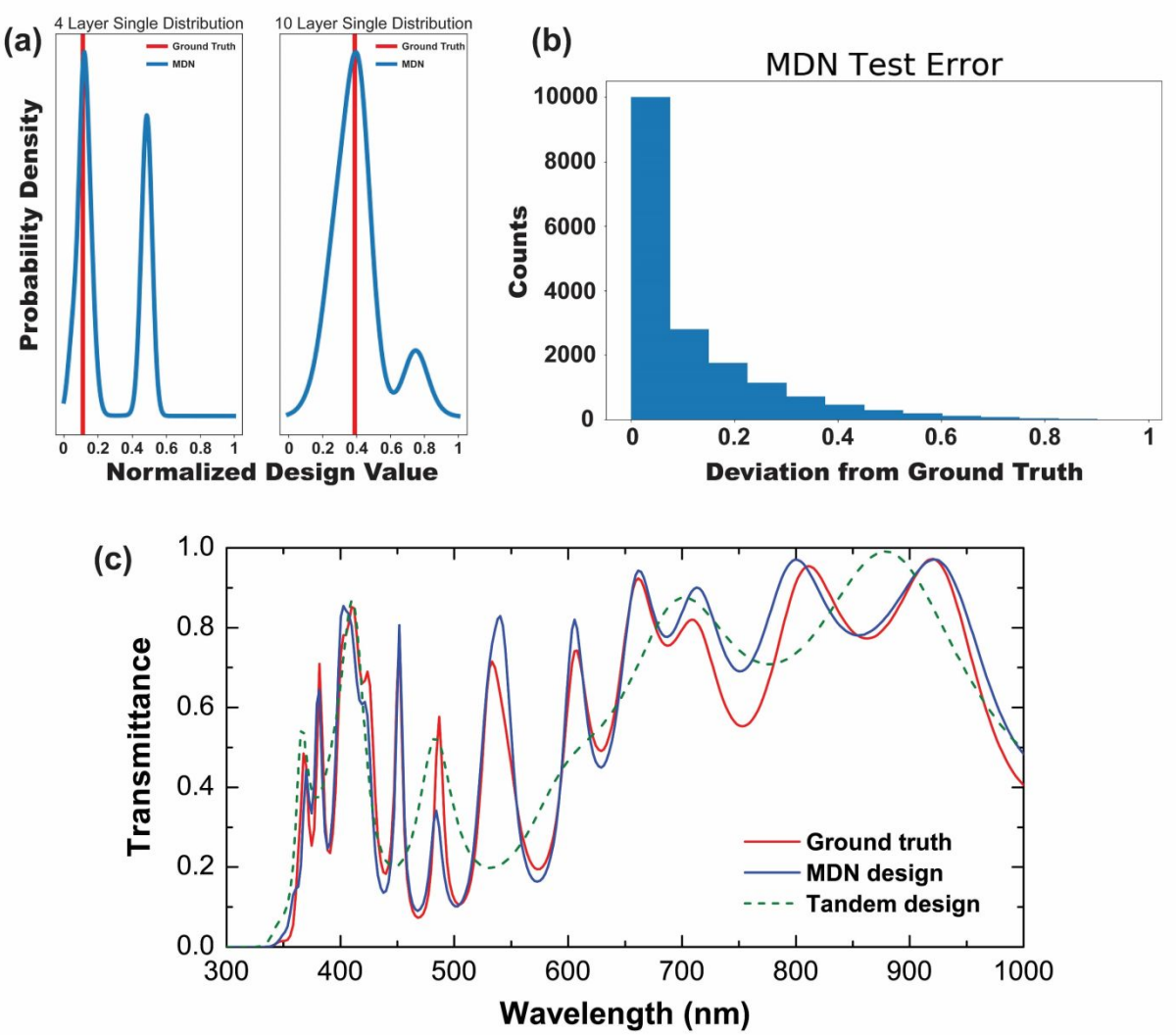


Figure 5: (a) Visualization of model distributions produced, with respect to the rescaled design values, with a randomly selected sample for the 4-layer (left) and 10-layer (right) datasets, respectively. Ground truths are denoted by the vertical red lines. The 4-layer case tends to have stronger multimodal characteristics by showing a second peak comparable in amplitude to the optimal solution near the ground truth. (b) Histogram of the deviation of the design variables from the ground truth for 4400 design variables from the test dataset for each of the four MDN models. All design variable comparisons are flattened into single one-dimensional dataset. (c) Comparison of a requested spectrum for the 10-layer structure (red curve) with the spectra produced by designs suggested by the MDN model (blue curve) and the tandem model (green dashed curve) trained on the same data. Both models are followed by a post-processing module.

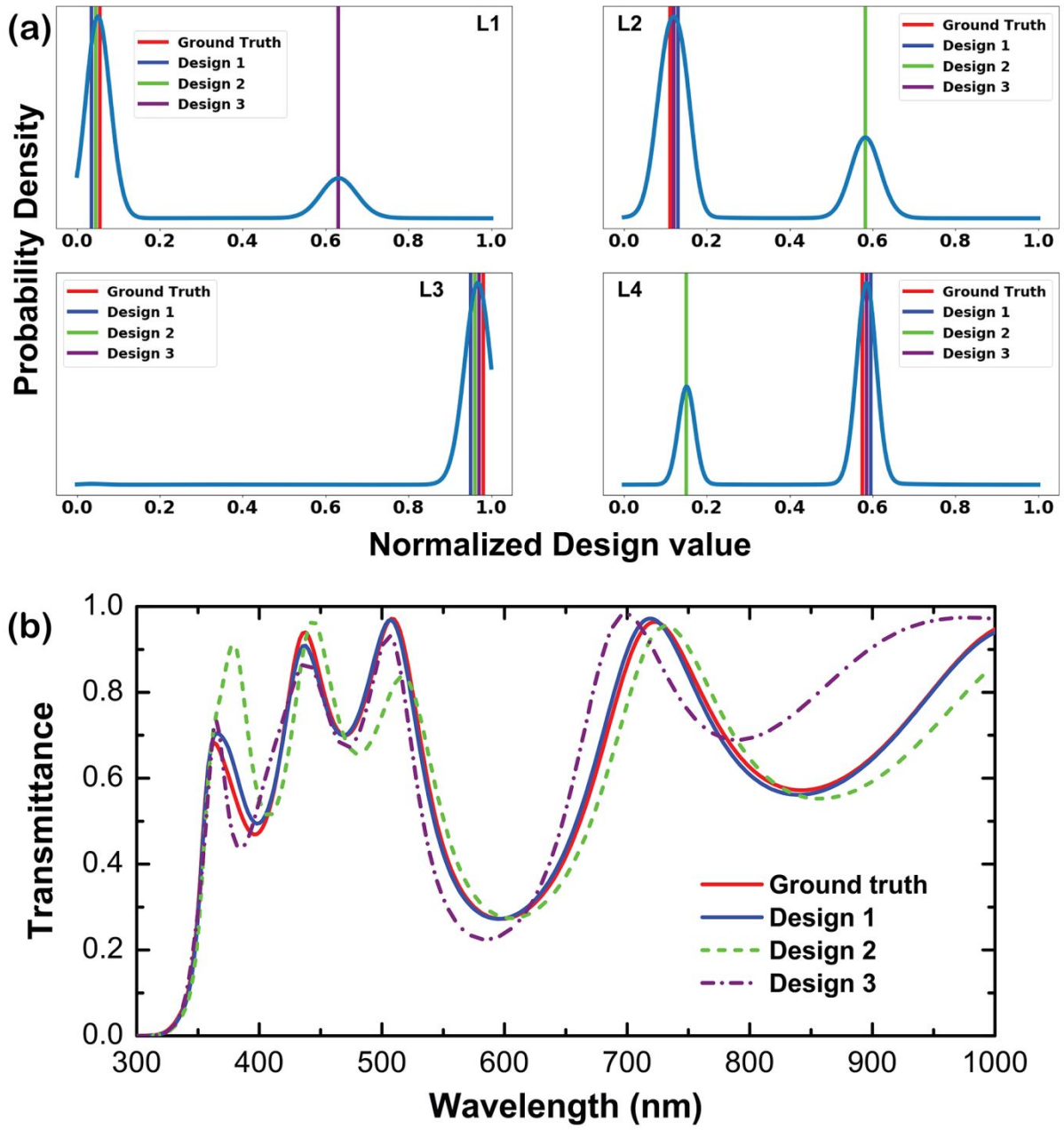


Figure 6: (a) MDN-produced distributions for a 4-layer design. Each panel represents one of the four design variables, i.e., thicknesses of the four layers in the structure. Vertical lines denote the samples taken for simulation, with the original ground truth design denoted by the red line. (b) Spectra produced by each of the sampled designs in (a) as compared to the requested spectrum (red curve).

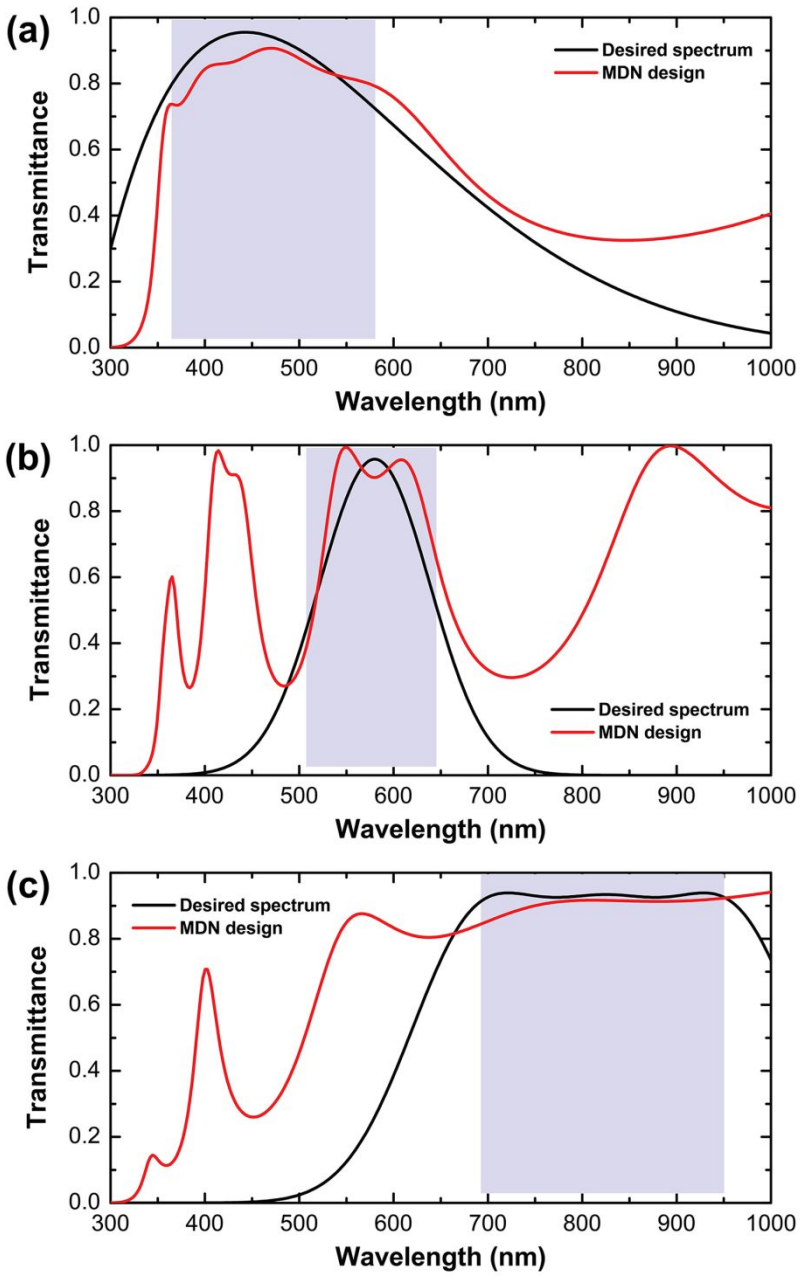


Figure 7: Designs of 4-layer structures for producing fictitious optical properties. Idealized spectra with high transmittance at different wavelengths and over different bandwidths, denoted by the shaded areas, are input as the desired spectra. The MDN and a slightly adapted post-processing module output realistic designs that have the best fit with the desired spectra across the wavelengths of interest.

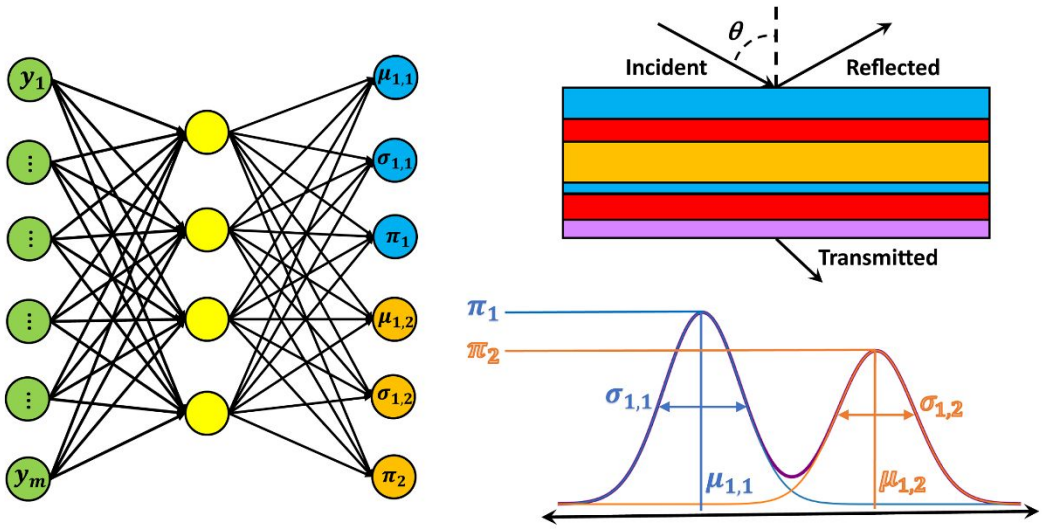
References

(1) Yu, N.; Capasso, F. Flat optics with designer metasurfaces. *Nature Materials* **2014,** *13* (2), 139-150, DOI: 10.1038/nmat3839.
(2) Kildishev, A. V.; Boltasseva, A.; Shalaev, V. M. Planar Photonics with Metasurfaces. *Science* **2013,** *339* (6125), 1232009, DOI: 10.1126/science.1232009.
(3) Yao, K.; Liu, Y. Plasmonic metamaterials. **2014,** *3* (2), 177, DOI: https://doi.org/10.1515/ntrev-2012-0071.
(4) Naik, G. V.; Shalaev, V. M.; Boltasseva, A. Alternative Plasmonic Materials: Beyond Gold and Silver. *Advanced Materials* **2013,** *25* (24), 3264-3294, DOI: 10.1002/adma.201205076.
(5) Koenderink, A. F.; Alù, A.; Polman, A. Nanophotonics: Shrinking light-based technology. *Science* **2015,** *348* (6234), 516-521, DOI: 10.1126/science.1261243.
(6) Baranov, D. G.; Zuev, D. A.; Lepeshov, S. I.; Kotov, O. V.; Krasnok, A. E.; Evlyukhin, A. B.; Chichkov, B. N. All-dielectric nanophotonics: the quest for better materials and fabrication techniques. *Optica* **2017,** *4* (7), 814-825, DOI: 10.1364/OPTICA.4.000814.
(7) Molesky, S.; Lin, Z.; Piggott, A. Y.; Jin, W.; Vucković, J.; Rodriguez, A. W. Inverse design in nanophotonics. *Nature Photonics* **2018,** *12* (11), 659-670, DOI: 10.1038/s41566-018-0246-9.
(8) Yao, K.; Unni, R.; Zheng, Y. Intelligent nanophotonics: merging photonics and artificial intelligence at the nanoscale. *Nanophotonics* **2019,** *8* (3), 339-366, DOI: 10.1515/nanoph-2018-0183.
(9) Goldberg, D. E.; Holland, J. H. Genetic Algorithms and Machine Learning. *Machine Learning* **1988,** *3* (2), 95-99, DOI: 10.1023/A:1022602019183.
(10) Lu, C.; Liu, Z.; Wu, Y.; Xiao, Z.; Yu, D.; Zhang, H.; Wang, C.; Hu, X.; Liu, Y.-C.; Liu, X.; Zhang, X. Nanophotonic Polarization Routers Based on an Intelligent Algorithm. *Advanced Optical Materials* **2020,** *n/a* (n/a), 1902018, DOI: 10.1002/adom.201902018.
(11) Jensen, J. S.; Sigmund, O. Topology Optimization for Nano-Photonics. *Laser Photonics Rev.* **2011,** *5*, 308.
(12) Yang, J.; Fan, J. A. Topology-Optimized Metasurfaces: Impact of Initial Geometric Layout. *Opt. Lett.* **2017,** *42*, 3161.
(13) Fan, J. A. Freeform metasurface design based on topology optimization. *MRS Bulletin* **2020,** *45* (3), 196-201, DOI: 10.1557/mrs.2020.62.
(14) Lin, Z.; Groever, B.; Capasso, F.; Rodriguez, A. W.; Lončar, M. Topology-Optimized Multilayered Metaoptics. *Physical Review Applied* **2018,** *9* (4), 044030, DOI: 10.1103/PhysRevApplied.9.044030.
(15) Sigmund, O. On the usefulness of non-gradient approaches in topology optimization. *Structural and Multidisciplinary Optimization* **2011,** *43* (5), 589-596, DOI: 10.1007/s00158-011-0638-7.
(16) Geremia, J. M.; Williams, J.; Mabuchi, H. Inverse-problem approach to designing photonic crystals for cavity QED experiments. *Physical Review E* **2002,** *66* (6), 066606, DOI: 10.1103/PhysRevE.66.066606.
(17) Malkiel, I.; Mrejen, I.; Nagler, A.; Arieli, U.; Wolf, L.; Suchowski, H. Plasmonic nanostructure design and characterization via Deep Learning. *Light: Science & Applications* **2018,** *7* (1), 1-8, DOI: doi:10.1038/s41377-018-0060-7.
(18) Hegde, R. S. Deep learning: a new tool for photonic nanostructure design. *Nanoscale Advances* **2020,** *2* (3), 1007-1023, DOI: 10.1039/C9NA00656G.
(19) Peurifoy, J.; Shen, Y.; Jing, L.; Yang, Y.; Cano-Renteria, F.; DeLacy, B. G.; Joannopoulos, J. D.; Tegmark, M.; Soljačić, M. Nanophotonic particle simulation and inverse design using artificial neural networks. *Science Advances* **2018,** *4* (6), DOI: 10.1126/sciadv.aar4206.
(20) He, J.; He, C.; Zheng, C.; Wang, Q.; Ye, J. Plasmonic nanoparticle simulations and inverse design using machine learning. *Nanoscale* **2019,** *11* (37), 17444-17459, DOI: 10.1039/C9NR03450A.
(21) Liu, Z.; Zhu, D.; Rodrigues, S. P.; Lee, K.-T.; Cai, W. Generative Model for the Inverse Design of Metasurfaces. *Nano Letters* **2018,** *18* (10), 6560-6576, DOI: 10.1021/acs.nanolett.8b03171.
(22) Tittl, A.; John-Herpin, A.; Leitis, A.; Arvelo, E. R.; Altug, H. Metasurface-Based Molecular Biosensing Aided by Artificial Intelligence. *Angewandte Chemie International Edition* **2019,** *58* (42), 14810-14822, DOI: 10.1002/anie.201901443.
(23) Nadell, C. C.; Huang, B.; Malof, J. M.; Padilla, W. J. Deep learning for accelerated all-dielectric metasurface design. *Optics Express* **2019,** *27* (20), 27523-27535, DOI: 10.1364/OE.27.027523.
(24) Kudyshev, Z. A.; Kildishev, A. V.; Shalaev, V. M.; Boltasseva, A. Machine-learning-assisted metasurface design for high-efficiency thermal emitter optimization. *Applied Physics Reviews* **2020,** *7* (2), 021407, DOI: 10.1063/1.5134792.
(25) Jiang, J.; Sell, D.; Hoyer, S.; Hickey, J.; Yang, J.; Fan, J. A. Free-Form Diffractive Metagrating Design Based on Generative Adversarial Networks. *ACS Nano* **2019,** *13* (8), 8872-8878, DOI: 10.1021/acsnano.9b02371.

(26) Inampudi, S.; Mosallaei, H. Neural Network Based Design of Metagratings. *Appl. Phys. Lett.* **2018,** *112*, 241102.
(27) Ma, W.; Cheng, F.; Liu, Y. Deep-Learning-Enabled On-Demand Design of Chiral Metamaterials. *ACS Nano* **2018,** *12* (6), 6326-6334, DOI: 10.1021/acsnano.8b03569.
(28) Liu, Z.; Zhu, D.; Lee, K.-T.; Kim, A. S.; Raju, L.; Cai, W. Compounding Meta-Atoms into Metamolecules with Hybrid Artificial Intelligence Techniques. *Advanced Materials* **2020,** *32* (6), 1904790, DOI: 10.1002/adma.201904790.
(29) Sajedian, I.; Badloe, T.; Rho, J. Optimisation of colour generation from dielectric nanostructures using reinforcement learning. *Optics Express* **2019,** *27* (4), 5874-5883, DOI: doi:10.1364/OE.27.005874.
(30) Gao, L.; Li, X.; Liu, D.; Wang, L.; Yu, Z. A Bidirectional Deep Neural Network for Accurate Silicon Color Design. *Advanced Materials* **2019,** *31* (51), 1905467, DOI: 10.1002/adma.201905467.
(31) Li, Y.; Xu, Y.; Jiang, M.; Li, B.; Han, T.; Chi, C.; Lin, F.; Shen, B.; Zhu, X.; Lai, L.; Fang, Z. Self-Learning Perfect Optical Chirality via a Deep Neural Network. *Physical Review Letters* **2019,** *123* (21), 213902, DOI: 10.1103/PhysRevLett.123.213902.
(32) Wiecha, P. R.; Muskens, O. L. Deep Learning Meets Nanophotonics: A Generalized Accurate Predictor for Near Fields and Far Fields of Arbitrary 3D Nanostructures. *Nano Letters* **2020,** *20* (1), 329-338, DOI: 10.1021/acs.nanolett.9b03971.
(33) Wang, H.; Rivenson, Y.; Jin, Y.; Wei, Z.; Gao, R.; Günaydın, H.; Bentolila, L. A.; Kural, C.; Ozcan, A. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nature Methods* **2019,** *16* (1), 103-110, DOI: 10.1038/s41592-018-0239-0.
(34) Liu, D.; Tan, Y.; Khoram, E.; Yu, Z. Training Deep Neural Networks for the Inverse Design of Nanophotonic Structures. *ACS Photonics* **2018,** *5* (4), 1365-1369, DOI: 10.1021/acsphotonics.7b01377.
(35) LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015,** *521* (7553), 436-444, DOI: 10.1038/nature14539.
(36) Bishop, C. M. *Mixture Density Networks*; Aston University: Neural Computing Research Group, 1994.
(37) Richmond, K. In *Trajectory Mixture Density Networks with Multiple Mixtures for Acoustic-Articulatory Inversion*, Advances in Nonlinear Speech Processing, Berlin, Heidelberg, 2007//; Chetouani, M.; Hussain, A.; Gas, B.; Milgram, M.; Zarader, J.-L., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 2007; pp 263-272.
(38) Schittenkopf, C.; Dorffner, G.; Dockner, E. J. In *Volatility Prediction with Mixture Density Networks*, ICANN 98, London, 1998//; Niklasson, L.; Bodén, M.; Ziemke, T., Eds. Springer London: London, 1998; pp 929-934.
(39) Morand, L.; Helm, D. A mixture of experts approach to handle ambiguities in parameter identification problems in material modeling. *Computational Materials Science* **2019,** *167*, 85-91, DOI: https://doi.org/10.1016/j.commatsci.2019.04.003.
(40) Kabir, H.; Wang, Y.; Yu, M.; Zhang, Q. J. Neural Network Inverse Modeling and Applications to Microwave Filter Design. *IEEE Trans. Microwave Theory Tech.* **2008,** *56* (4), 867.
(41) Kiarashinejad, Y.; Abdollahramezani, S.; Adibi, A. Deep learning approach based on dimensionality reduction for designing electromagnetic nanostructures. *npj Computational Materials* **2020,** *6* (1), 12, DOI: 10.1038/s41524-020-0276-y.
(42) Born, M.; Wolf, E. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 7 ed.; Cambridge University Press: Cambridge, 1999.
(43) Texas Advanced Computing Center (TACC). The University of Texas at Austin.
(44) Su, L.; Vercruysse, D.; Skarda, J.; Sapra, N. V.; Petykiewicz, J. A.; Vučković, J. Nanophotonic inverse design with SPINS: Software architecture and practical considerations. *Applied Physics Reviews* **2020,** *7* (1), 011407, DOI: 10.1063/1.5131263.
(45) Joannopoulos, J. D.; Johnson, S. G.; Winn, J. N.; Meade, R. D. *Photonic Crystals: Molding the Flow of Light* Princeton University Press: 2008.
(46) Keshavarz Hedayati, M.; Elbahri, M. Antireflective Coatings: Conventional Stacking Layers and Ultrathin Plasmonic Metasurfaces, A Mini-Review. *Materials (Basel)* **2016,** *9* (6), 497, DOI: 10.3390/ma9060497.
(47) Wang, Z.; Li, T.; Soman, A.; Mao, D.; Kananen, T.; Gu, T. On-chip wavefront shaping with dielectric metasurface. *Nature Communications* **2019,** *10* (1), 3547, DOI: 10.1038/s41467-019-11578-y.
(48) Qu, Y.; Jing, L.; Shen, Y.; Qiu, M.; Soljačić, M. Migrating Knowledge between Physical Scenarios Based on Artificial Neural Networks. *ACS Photonics* **2019,** *6* (5), 1168-1174, DOI: 10.1021/acsphotonics.8b01526.
(49) Ma, W.; Cheng, F.; Xu, Y.; Wen, Q.; Liu, Y. Probabilistic Representation and Inverse Design of Metamaterials Based on a Deep Generative Model with Semi-Supervised Learning Strategy. *Advanced Materials* **2019,** *31* (35), 1901111, DOI: 10.1002/adma.201901111.

(50) Angeris, G.; Vučković, J.; Boyd, S. P. Computational Bounds for Photonic Design. *ACS Photonics* **2019,** *6* (5), 1232-1239, DOI: 10.1021/acsphotonics.9b00154.

For Table of Contents Use Only



Title: A Deep Convolutional Mixture Density Network for the Inverse Design of Layered Photonic Structures

Authors: Rohit Unni, Kan Yao, and Yuebing Zheng

Synopsis: In the inverse design of photonic structures, a mixture density network (MDN) models the design parameters as multimodal probability distributions instead of discrete values, improving convergence and discovery of degenerate solutions when the response-to-design mapping is non-unique.