Deep Learning-Enabled Rapid Full Electromagnetic Near-Field Prediction and Rapid Inverse Design of Metasurfaces: A Proof-of-Concept Implementation of All-Dielectric Inter-Coupled Nanopillar Fabrics

TEVFIK BULENT KANMAZ,[1] EFE OZTURK,[1,2] HILMI VOLKAN DEMIR,[1,3,*] AND CIGDEM GUNDUZ-DEMIR[4,5]

[1] *UNAM–Institute of Materials Science and Nanotechnology and National Nanotechnology Research Center, Department of Electrical and Electronics Engineering, Department of Physics, Bilkent University, TR-06800 Ankara, Turkey*
[2] *Department of Electrical and Electronics Engineering, METU, TR-06800 Ankara, Turkey*
[3] *LUMINOUS! Centre of Excellence for Semiconductor Lighting and Displays, School of Electrical and Electronic Engineering, School of Physical and Mathematical Sciences, School of Materials Science and Engineering, Nanyang Technological University, 50 Nanyang Avenue, 639798 Singapore*
[4] *Department of Computer Engineering, School of Medicine, KUIS AI Center, Koc University, TR-34450 Istanbul, Turkey*
[5] *School of Medicine, Koc University, TR-34010 Istanbul, Turkey*
[*] *hvdemir@ntu.edu.sg*

Abstract: Metasurfaces generate desired electromagnetic wavefronts using sub-wavelength structures that are much thinner than conventional optical tools. However, their typical design method is based on trial-and-error, which is adversely inefficient in terms of the consumed time and computational power. This paper proposes and demonstrates deep learning-enabled rapid prediction of the full electromagnetic near-field response and inverse prediction of the metasurfaces from desired wavefronts to obtain direct and rapid designs. The proposed encoder-decoder neural network was tested for different metasurface design configurations. This approach overcomes the common issue of predicting only the transmission spectra, a critical limitation of the previous reports of deep learning-based solutions. Our deep learning-empowered near-field model can conveniently be used as a rapid simulation tool for metasurface analyses as well as for their direct rapid design.

1. Introduction

Metasurfaces are artificially designed electromagnetic (EM) structures that allow the realization and engineering of a desired wavefront profile using sub-wavelength metaunits. An array of different orientations of such metaunits, each of which essentially acts as a scatterer, form an optical structure that is ultra-thin compared to their counterpart conventional optical tools [1–5]. The interactions of light with the sub-wavelength units make it possible to control the phase and magnitude of the scattered light [6,7]. Therefore, metasurfaces are competent optical devices that can be utilized for numerous tasks including image formation, focusing, optical vortex generation, spectroscopy, and augmented reality [8–10]. The metaunits composed of all-dielectric materials have been utilized in recent metasurface architectures. These dielectric metaunits are shown to yield minor losses across a wide range of frequencies compared to designs that adopt metal-based units [11–16].

As much as metasurface designs offer benefits in terms of optical performance given their thickness, the design process can be cumbersome and is usually not straightforward. One of the main approaches in their design is to "phase-match" the responses of consecutive units as a designer seeks to achieve the "right interference" of the EM waves [17, 18]. However, this

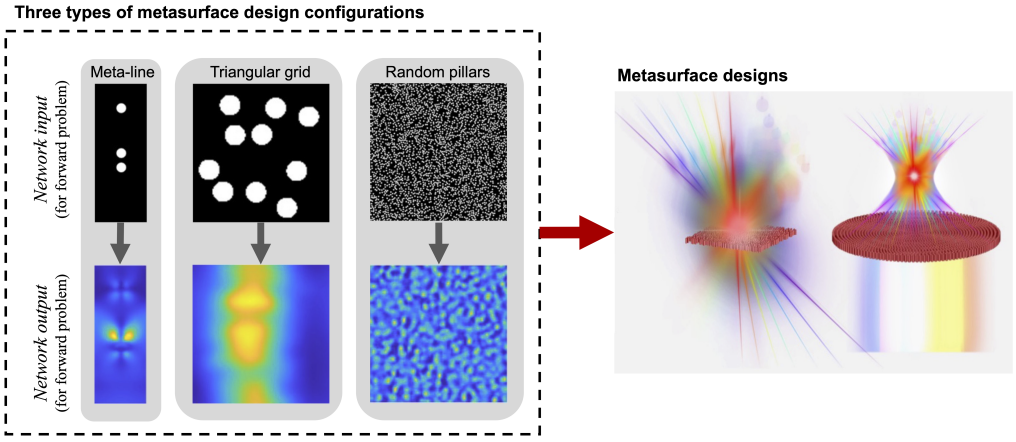
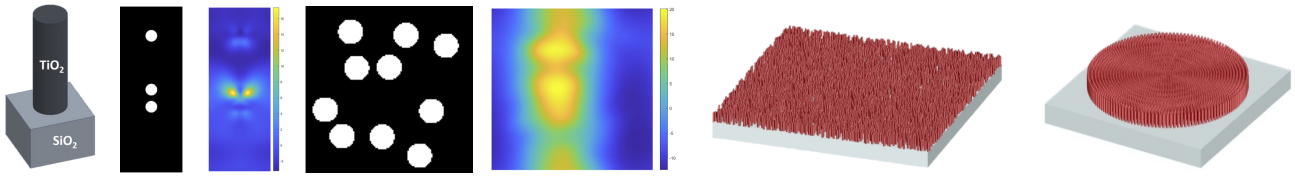


Fig. 1. Input-output representations of three metasurface design configurations selected for this study. For the forward problem, the input is a cross-section nanopillar map representing the metasurface geometry, and the output is a near-field electric field response. The arrow between the input and output indicates the use of our proposed neural network. For the inverse design problem, input-output relations are the opposite.

requires a repetitive process of trail and error. Also, the EM response of a surface may not be a direct superposition of the individual units but rather a collective EM response if there exist inter-coupling effects among the metaunits [19]. Thus, one typically needs to anticipate a design with intuition and optimize it with the metaunit set that is predefined in general or, alternatively, to propose an analytical solution for the model. Nevertheless, finding an analytical solution for an arbitrary design can be practically impossible. Moreover, the EM response of even a single metaunit may become unpredictable pretty quickly with minor changes in the geometry and materials of the units. These circumstances give rise to a strong motivation for the investigation of efficient tools cabaple of finding the EM response. Conventionally, iterative numerical methods and simulation techniques, including FEM (finite element method), FDTD (finite difference time domain), and FIT (finite integration technique) are utilized [20]. These methods provide accurate and reliable solutions even for complex structures. Nonetheless, they are basically brute force methods that require immense computational time and power in order to reach a solution [21, 22]. As the conventional metasurface design process relies on trial and error, the inefficiency of the numerical methods makes it considerably time-consuming for the designer, and thus, requires an experienced metasurface design specialist in most cases.

In this paper, we propose and demonstrate a deep learning-based solution to overcome these disadvantages of conventional methods. Deep neural networks are machine learning models with multiple layers of computational units. These layers have non-linear activation functions, which make it possible for these models to extract highly non-linear relations between the input and output of a system. In the recent decade, many studies have been used to solve many scientific and engineering problems, including the prediction of the EM response of metasurfaces [23–27]. However, these previous studies have their specific limitations. Previously, the analysis of the EM response of metasurfaces was focused only on the metaunits' spectrum response. Additionally, they were limited to a small number of discrete metaunits. This decreases the total number of possible designs but makes it easier to generate a dataset and train a model. Moreover, the inter-coupling effects of the metastructures were generally fully omitted or confined to one axis of interaction only [27]. Overall, these previous studies do not provide full EM information for a complex and high degree-of-freedom (DOF) design that accounts for non-constrained



(a) (b) (c) (d) (e) (f) (g)

Fig. 2. (a) 3D illustration of a single nanopillar used as the metaunit. For the metaline configuration, (b) the refractive index cross-section of an example positioning of the nanopillars and (c) their EM near-field intensity map obtained by the numerical simulation for the forward problem. For the triangular grid configuration, (d) the refractive index cross-section of an example positioning of the nanopillars and (e) their EM near-field intensity map obtained by the numerical simulation. (f) Illustration of randomly-positioned nanopillars. (g) Illustration of circularly positioned nanopillars, which will be used for the achromatic metalens design.

inter-coupling effects. Thus, they cannot offer a fully capable alternative for numerical solvers such as FDTD, for example, for the implementation of all-dielectric inter-coupled nanopillar fabrics.

In this work, different than the previous literature, we introduce a ***multi-task deep neural network design*** to predict the complete phase map information for a high-DOF inter-coupled metasurface geometry (***forward problem***). In addition, we propose to use a ***single-task deep neural network*** with a similar architecture for the ***inverse design problem***, in which the aim is to find the metasurface geometry from a given EM phase profile [28]. To the best of our knowledge, there exist no conventional tools or deep learning-based models for the solution of this inverse design problem, which makes our design the first to achieve such metasurface geometry prediction. Working with three types of metasurface design configurations, each of which has a different DOF (Fig. 1), our experiments revealed that the proposed neural networks showed high performance both for the forward and the inverse design problems.

2. Method

In this work, two main problems are studied: 1) the forward problem of EM near-field response prediction, in which the input is a cross-section map of a metasurface, and the output is an EM near-field intensity map and 2) the inverse design problem of metasurface geometry prediction, in which the input is an EM phase profile and the output is a metasurface geometry. Both of these prediction problems are addressed for three types of metasurface design configurations, as illustrated in Fig. 1, by designing deep neural networks.

2.1. Metasurface Design Configurations

The selected three configurations, with different DOFs, use different positioning of the fixed-size pillars of the same material. In our experiments, we fixed the size and the material type since one can achieve almost any desired phase profile by changing the positions of the pillars, and thus, their inter-coupling.

In general, one may prefer metaunits that lead to devices and metasurface designs as efficient and broadband as possible. Although the metaunit designs that use the Panchartam-Berry phase approach produced promising results in recent years [29], the resulting devices suffer from low device efficiency and large lattice sizes in general. The lattice structures are forced to be large as it is mandatory to reduce inter-coupling effects between the metaunits. On the contrary, metalens structures that utilize the inter-coupling effects between the metaunits do not suffer

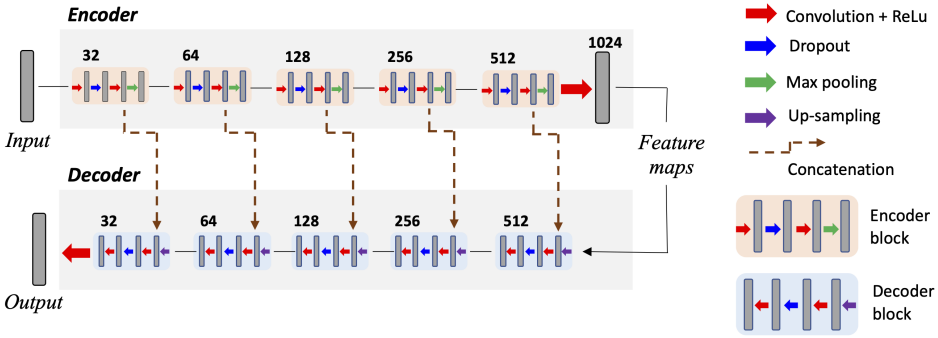


Fig. 3. Schematic overview of the encoder and decoder architectures used in our network designs. Each arrow corresponds to a network operation(s) distinguishable by its color. The numbers on the top of each encoder/decoder block are the number of feature maps used by the block. The height and width of the block's input change with the input image size selected for a particular configuration. Note that similar to this illustration, our single-task design uses one decoder that uses the feature maps, which are the outputs of the encoder (see Fig. 4). In our multi-task design, there are multiple decoders that use the same feature maps learned by the shared encoder (see Fig. 5).

from significant lattice constraints and show great efficiency [27, 30]. Despite the performance level of these designs, near the optimum device operating frequency, they tend to fail at providing the same efficiency over a broad spectrum [31]. Considering these undesired limitations, in our designs, we use a dielectric metaunit that has good inter-coupling and broadband efficiency over the 400-700 nm visible range. The TiO_2 nanopillar optimized in Ref. [32] is adopted as our fundamental design block. These nanopillars have a radius of 45 nm and a height of 600 nm and are placed on top of a SiO_2 substrate (Fig. 2a).

The ***metaline configuration*** is used as a proof of concept, in which three pillars are aligned vertically. In other words, the simulated structure is a "metaline" that is periodic and infinite in one direction. In particular, the position of one pillar is fixed at the center of the cross-sectional area, and the other two meta pillars' vertical distances to the center change in each input sample. The simulation data were collected at a wavelength of 550 nm. The resolution of a two-dimensional (2D) cross-sectional input image is set to 128×48 pixels. The pillars are represented on this input image as circles with a fixed radius. With the selected image resolution and the radius, this geometry of pillars ensures that the interactions between the pillars are confined to the vertical axis only. This reduces the problem's complexity and DOF. In this work, we use this configuration to show that the inter-coupling effects between metapillars can be predicted with a deep neural network. An example positioning of the pillars and its EM near-field intensity map obtained by simulation are illustrated in Figs. 2b and 2c, respectively.

In the ***triangular grid configuration***, the pillars are randomly placed on a triangular grid, which limits the number of places that a pillar can be located (see Fig. S1). The lattice spacing is set to 5 nm, which is small enough to induce the coupling of the pillars from every direction. The coupling interactions between the pillars are, therefore, not restricted to one axis but effective in all directions. Likewise, the simulation data were collected at a wavelength of 550 nm. In this configuration, the input image size is fixed as 128×128. As a result of this fixed size, the number of randomly-located metapillars changes from one sample to another (in our experiments, this number varies from 10 to 27 due to the packing limit). Note that this configuration has the limitation of having a fixed total area, which will be relaxed in the next configuration. An example pillar positioning and its simulated map are given in Figs. 2d and 2e, respectively.

The ***random pillar configuration*** is used for the case where there are no constraints on the

number or position of the pillars and on the total simulation area. The random positioning of the pillars is illustrated in Fig. 2f. This configuration simulates the case where an arbitrarily large simulation area is chosen to calculate its EM near-field response. In our experiments, metapillars are randomly located across a simulation area of 2048×2048 pixels. This results in an average of 2950.7 ± 13.8 pillars in each simulation. Here it is worth noting that, although we fixed this simulation area in the experiments, our neural network design provides a generic and computationally feasible solution that can be applied to an arbitrarily selected area (please see Sec. 3.2 for the details). Furthermore, to show the applicability of this solution on different and multiple frequencies, for this configuration, we collected the simulation data at seven spectral points in the range of 400 and 700 nm; these points are equally spaced in the frequency domain.

To solve this problem numerically with FDTD, the computational time may take prohibitively long for large simulation areas [33], making the simulation infeasible. On the other hand, the proposed neural network solution works much more efficiently, even for large areas. To demonstrate large-area simulation in a real-world application, the neural network trained for the random pillar configuration is transferred and fine-tuned to design an achromatic metalens [32], which is shown in Fig. 2g. With the selected pillar radius, this achromatic metalens design covers a simulation area of 1410×1410 pixels, different than the simulation area previously selected for the random pillar configuration. Nevertheless, the proposed approach is applicable.

2.2. Deep Neural Network Designs

This work uses deep neural networks with the U-Net architecture, which is a very well-known model in the field of computer vision [34]. This architecture is an encoder-decoder network, where the encoder accepts an image (a 2D map) as its input and the decoder generates another image that has the same resolution as the input image. In the forward problem, the input image is a binary map specifying the metasurface geometry. To acquire the input, refractive index cross-section maps just above the substrate are converted to binary maps to represent the metasurface. In these binary maps, the space (pixel positions) occupied by the pillars is represented with 1 (white), and the empty space (filled with air) is represented with 0 (black). The output of the forward problem is the EM near-field response, which is considered as a six-channel image containing the real and imaginary parts of the three Cartesian components of the vectorial electric field; i.e., Re(E_x), Im(E_x), Re(E_y), Im(E_y), Re(E_z), and Im(E_z). For the inverse design problem, the inputs and the outputs are the opposite of the forward problem.

In order for the encoder-decoder network to generate (predict) an output from a given input, the encoder path extracts features directly on the input using a series of convolutional and pooling (downsampling) layers. The decoder path then constructs the output map from the extracted features by upsampling and convoluting them in consecutive layers. In the U-Net architecture, long-skip connections (concatenation operations) are added between the corresponding intermediate layers of the encoder and decoder paths to preserve details in this output construction. The schematic overview of the encoder and decoder architecture used in our network designs is sketched in Fig. 3.

In our designs, the encoder path includes five consecutive blocks of two convolutions and one max pooling, whereas the decoder path consists of five consecutive blocks of upsampling, concatenation, and two convolutions. All convolutions use 3×3 filters and are followed by the rectified linear unit (ReLU) activation function. The dropout layer, with a factor of 0.2, is added after the first convolution to prevent overfitting. The pooling/upsampling layers use 2×2 filters. The number of feature maps used by the first encoder block is 32. This number is doubled after each pooling layer and halved after each upsampling layer. The block number is selected as five since it yields a sufficient field of view to extract coupling information between the pillars with the selected filter sizes and the image resolution. Note that this is one of the typical architectures used by the U-Net based models [34].

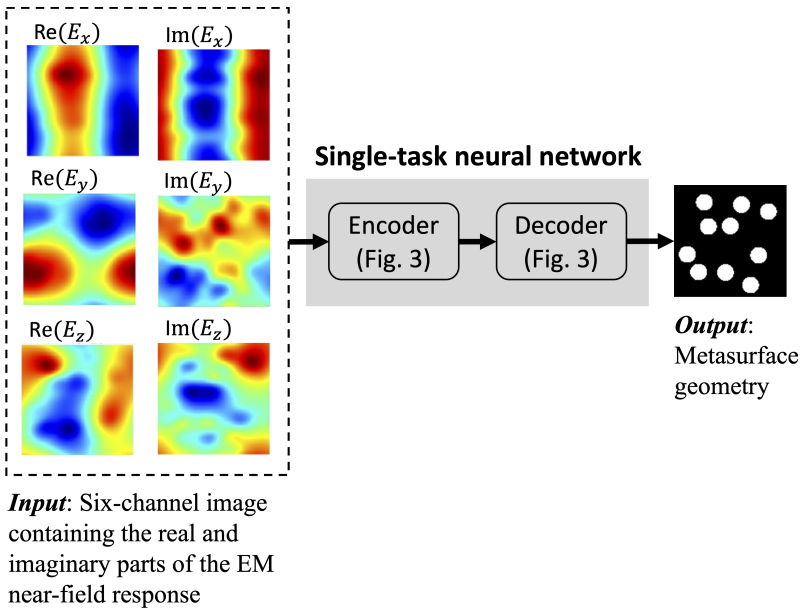


Fig. 4. For the triangular grid configuration, the single-task neural network is designed for the inverse design problem. Similar single-task networks are used for the metaline and random pillar configurations.

In this work, we propose two neural network designs. The first one is a single-task network for the inverse design problem. This network has one encoder and one decoder since one-channel output map, representing the metasurface geometry, is estimated from multi-channel input maps, representing the Cartesian components in an EM phase profile (Fig. 4). The second one is a multi-task network designed for the forward problem to predict the real and imaginary parts of the Cartesian components of the EM near-field intensity map from the metasurface geometry. This network includes one shared encoder path and six decoder paths, one for predicting each real or imaginary part (Fig. 5). Here we use a multi-task network since the multi-task learning paradigm is known as an effective means to predict different but related tasks. Learning multiple tasks from a single shared encoder decreases the likelihood of overfitting since this requires learning a shared representation that works adequately well for all the tasks [35].

The networks were designed and trained in Python using the Tensorflow framework. The network weights were optimized by backpropagation that used the categorical cross-entropy and mean squared error loss functions for the single-task and multi-task networks, respectively. The AdaDelta optimizer was used to adaptively adjust the learning rate. The training was performed on a GPU (GeForce RTX 2080). For the aforementioned three configurations, the training, validation, and test datasets were generated using the Lumerical FDTD solver. Refractive index monitors and frequency domain field monitors were used in the simulations to obtain the input-output pairs (see Table S1 and Fig. S2 for more training and data collection details).

3. Results

This section provides configuration-specific details of the network designs, separately for the forward and the inverse design problem. Besides, it presents quantitative and visual results obtained for each configuration.

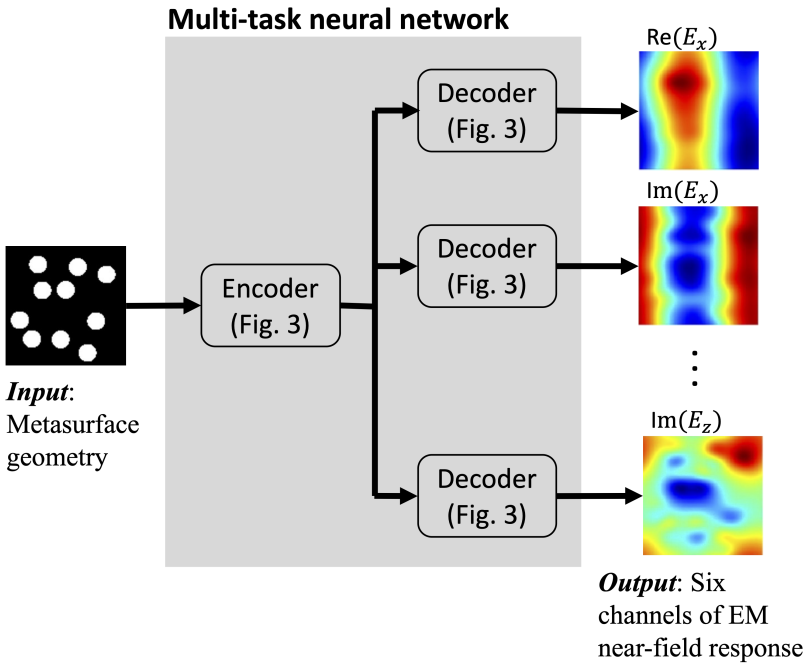


Fig. 5. For the triangular grid configuration, the multi-task neural network is designed for the forward problem. Similar multi-task networks are used for the metaline and random pillar configurations.

3.1. Forward Problem for Metaline and Triangular Grid Configurations

The multi-task network design illustrated in Fig 5 was used to predict six-channel EM near-field response from metasurface geometry input maps. Subsequently, these six channels were used to calculate the corresponding EM near-field intensity map. Visual results obtained by the proposed network on three exemplary test set samples are illustrated in Figs. 6a-6c and Figs. 6d-6f for the metaline and triangular grid configurations, respectively. Additionally, for each sample, the mean squared error (MSE) was calculated between the EM near-field intensity (or irradiance) maps obtained by simulation and predicted by the proposed multi-task network. As the unit of irradiance is W/m^2, the MSE values are given in the unit of W^2/m^4. For the metaline and triangular grid configurations, the MSEs averaged over test set samples were 1.29e-03±0.12e-03 and 5.76e-07±0.02e-07, respectively. These low MSEs indicated that the maps were predicted quite accurately, as also supported by the visual results.

3.2. Forward Problem for Random Pillar Configuration

The random pillar configuration allows using an arbitrarily selected area and collecting simulated data at an arbitrary number of spectral points. Thus, we make two modifications to the network design and its use.

First, the previous configurations were simulated for a single spectral point at a particular wavelength. Thus, we designed a multi-task network with six decoders, each of which predicted a real or an imaginary part of the Cartesian components of the EM near-field response. On the other hand, the random pillar configuration allows to simulate data at N different spectral points, which necessitates training 6×N decoders at the same time. In our design, this number would be 42 for the selected seven spectral points. This would correspond to simultaneously optimizing the weights of a larger network, which required more training data and demanded more powerful

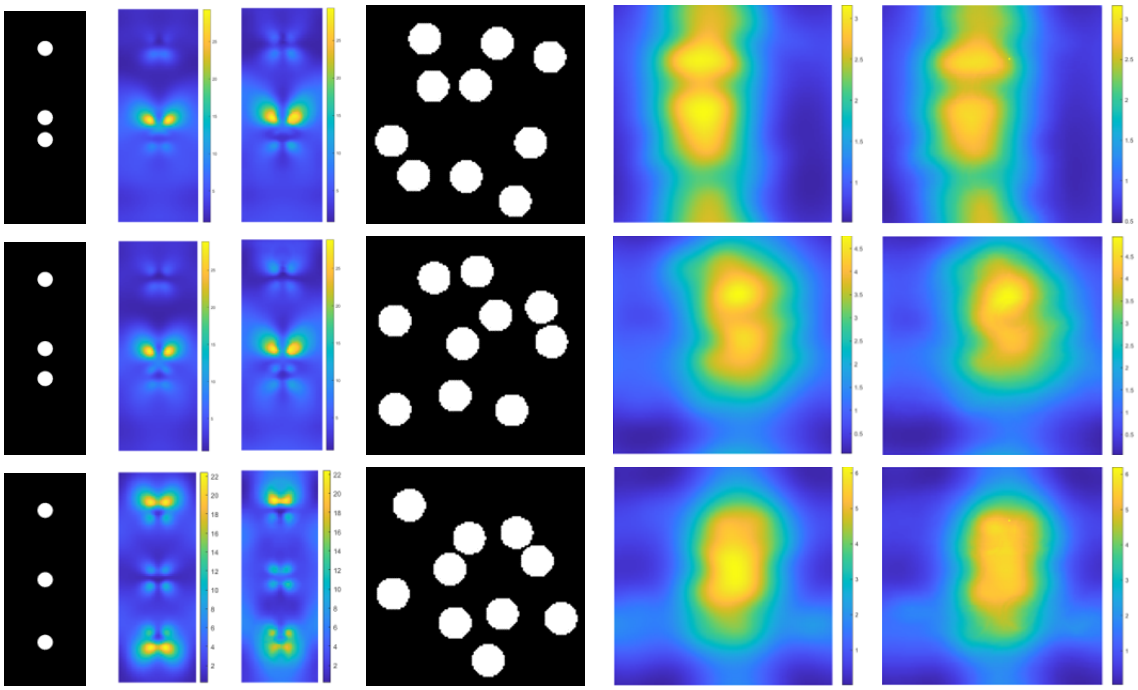


Fig. 6. For the metaline and triangular grid configurations, visual results obtained on exemplary test set samples. For the metaline configuration, (a) metasurface geometry maps and EM near-field intensity maps (b) obtained by simulation and (c) predicted by the proposed multi-task network for the forward problem. The mean squared errors for these three samples were 1.3e-03, 2.1e-03, and 2.4e-03, respectively. For the triangular configuration, (a) metasurface geometry maps and EM near-field intensity maps (b) obtained by simulation and (c) predicted by the proposed multi-task network for the forward problem. The mean squared errors for these three samples were 5.8e-07, 2.5e-06, and 1.6e-06, respectively.

processors and larger memory resources. As a result, at some point, simultaneous training of all these decoders would become infeasible. To alleviate this problem, we used six independently trained multi-task networks, one for a real or an imaginary part of one Cartesian component of the EM near-field response. Each network has one encoder and multiple decoders, as given in Fig. 5. But this time, each decoder predicts a response obtained for one of the seven spectral points (see Fig. S3). The encoder and decoder architectures of these networks are the same as those given in Fig. 3.

The second modification was to handle arbitrarily selected large simulation areas. To this end, we designed all networks to take 256×256 input tiles that were cropped out of a metasurface geometry map. Networks were trained on the tiles randomly cropped out of the training samples. Then, to predict the EM near-field intensity map of a test sample, we estimated the maps for overlapping tiles and averaged all predictions estimated for the same pixel. The overlapping tiles were obtained by sliding a window over the map with an increment of 64 pixels (Fig. 7) In our experiments, we only considered the predictions in the middle 128×128 section of each window. We made this choice as well as selected the tile size as 256 considering the inter-coupling effects between the pillars (see Sec. 3.3 for details). The increment size of 64 was small enough to

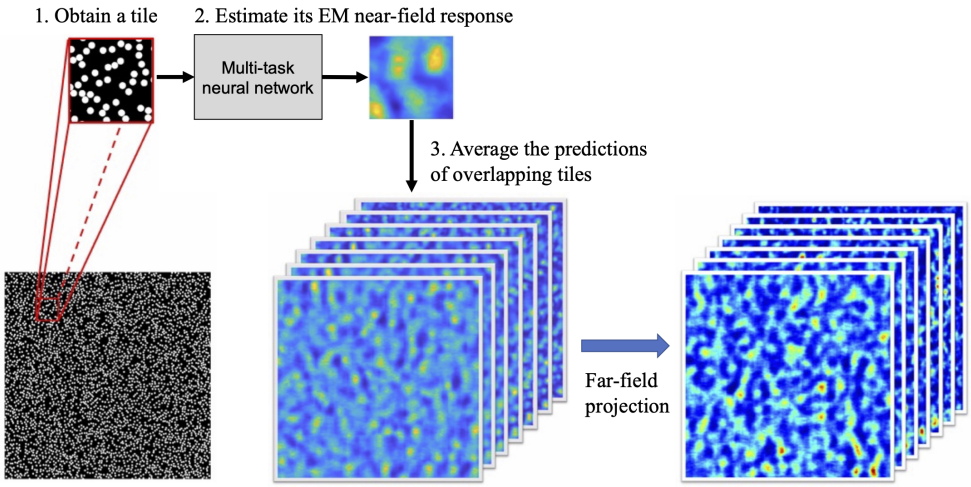


Fig. 7. An illustration of estimating EM near-field and far-field responses in the random pillar configuration by the sliding window approach.

make multiple predictions for the same pixel, but at the same time, large enough to cause only negligible speed-down in the computational time.

In addition to analyzing the performance of our network on predicting EM near-field responses, we examined its effects on EM far-field response projections. To do so, we separately calculated the far-field response of the simulated and the predicted near-field response, using the Fresnel approximation, and compared these far-field responses visually and quantitatively. Note that one can calculate the EM far-field response by taking the convolution of the EM near-field response with the free space transfer function [36]. However, the analytical convolution integral makes calculations almost impossible for many cases except for the most simple diffraction geometries, and hence, one needs approximations. Also note that, in our experiments, we used the Fresnel approximation (see Fig. S4 and the explanation below it for more details) as it is one of the approximations that can be used in most cases but also contains fewer assumptions compared to the others [37].

As aforementioned, for the random pillar configuration, data were simulated at seven spectral points, which were again equally spaced in the frequency domain, corresponding to the spectral range of 400-700 nm. For each spectral point, the mean squared error (MSE) between the maps obtained by simulation and predicted by the proposed multi-task neural network was calculated. Table 1 reports the average MSE calculated on the test set samples for each spectral point separately. As seen in this table, the errors were acceptably low, which also led to accurate visual results. For an example spectral point (with 535.3 THz frequency corresponding to 560 nm wavelength), the visual results obtained on an exemplary test sample are given in Fig. 8. The visual results (EM near-field and far-field intensity maps obtained by simulation and prediction) on the same test sample for the other spectral points can be found in Fig. S5.

For EM near-field intensity maps, it is quite important to predict the regions with localized electric field responses. In order to understand the effectiveness of our networks in predicting such peaks, we conducted an additional experiment: First, we compared the responses of the simulated and predicted EM near-field maps against a threshold and identified peak regions in each map as the pixels with responses greater than this threshold. Then, the overlapping peak regions in the simulated and predicted maps were considered as true positives and the precision, recall, and f-score metrics were calculated for quantitative evaluation. The visual and quantitative results of this experiment are shown in Fig. S6 and Table S2. These results indicated that our

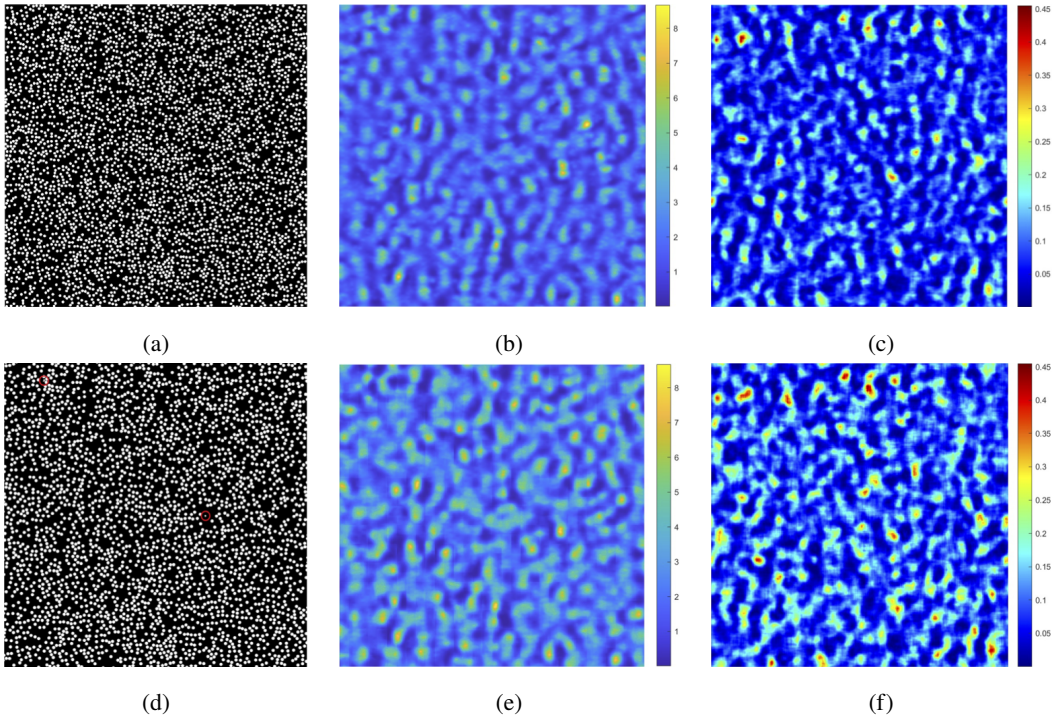


Fig. 8. (a) An example of the refractive index cross-section of randomly-positioned nanopillars. (b) Its EM near-field intensity map obtained by simulation. (c) The EM far-field intensity map calculated from the simulated EM near-field intensity map using the Fresnel approximation. (d) The refractive index cross-section estimated from the EM near-field response by the proposed single-task network for the inverse design problem. Here two red circles indicate the position of the missed pillar positions in this estimation. (e) The EM near-field intensity map predicted by the proposed multi-task network. (f) The EM far-field intensity map calculated from the predicted EM near-field intensity map using the Fresnel approximation. Please note that near-field intensity maps are shown using the parula colormap whereas far-field intensity maps are shown using the jet colormap for better visualization.

Table 1. For the random pillar configuration, the mean squared error (MSE) for the prediction of EM near-field and far-field intensity maps at different frequencies. These results were the average MSEs obtained on the test set samples.

Frequency (THz)	Wavelength (nm)	Near-field (MSE)	Far-field (MSE)
428.3	700	5.0e-08	4.6e-06
481.8	622	8.0e-08	3.9e-06
535.3	560	1.8e-07	1.5e-07
588.9	509	2.0e-07	6.5e-07
642.4	467	4.3e-07	4.2e-10
695.9	430	4.7e-07	3.2e-07
749.5	400	5.6e-07	5.2e-08

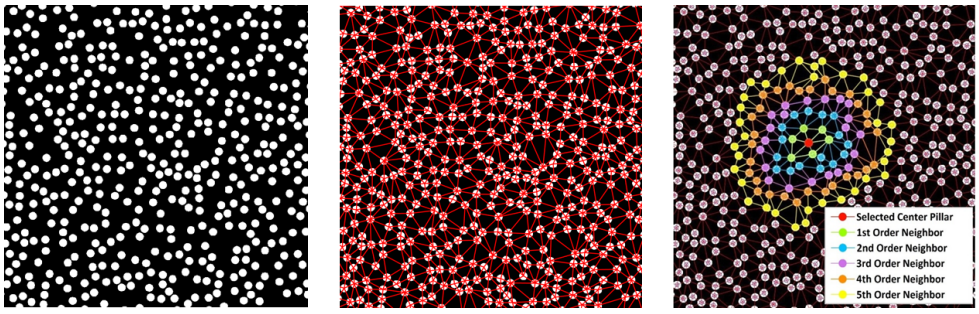
networks were able to detect the peak regions with high accuracy, leading to f-scores greater than 77.00 percent for the first six frequencies. For the seventh one, the f-score was 73.78 percent. We repeated this experiment for the EM far-field intensity prediction and its results are also reported in SI.

3.3. Analysis of Tile Size Selection

The electric field above a pillar is not determined by that pillar alone but also by pillars in its specified proximity (neighborhood) due to the intercoupling effect. In our experiments, we selected the tile size, considering this intercoupling effect, as follows:

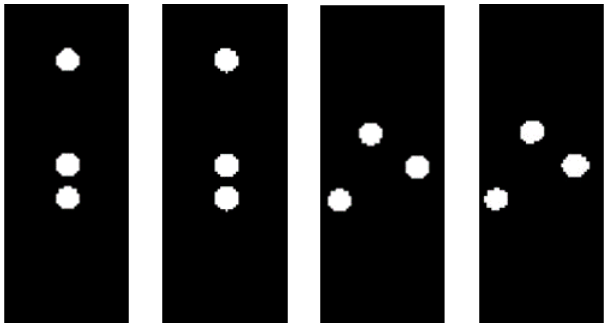
First, for each pillar p, the first k-th neighborhood was determined by constructing a graph on all pillars and visiting the other pillars starting from pillar p by the breadth-first traversal. This graph was constructed considering all pillars as vertices and assigning graph edges between the pillars by Delaunay triangulation [38], which is a common way of constructing a planar graph. For an example refractive index cross-section map, red lines in Fig. 9b correspond to the edges of a graph constructed on the pillars shown in Fig. 9a by Delaunay triangulation. Two pillars (vertices) are called as immediate neighbors if they are the end points of the same triangle edge. The k-th order neighbors of pillar p are defined as the pillars that are visited in the k-th hop of the breadth-first traversal starting from p (in other words, the pillars that are k edges far away from pillar p). The k-th order neighbors of a selected pillar are illustrated in Fig. 9c.

Then, the tile size was determined using an iterative method. At the beginning, a pillar was selected as the center, all other pillars were removed, and an FDTD simulation was run. The method would stop if the error between this simulation and the full simulation, containing thousands of pillars, was small enough. Otherwise, the pillars in the next neighborhood were also included, another FDTD simulation was run, and the error margin was calculated. In our experiments, we observed that the deviation from the full simulation was less than 1 percent after including the fourth-order neighbors (Fig. S7). Thus, we selected the tile size as 256×256 that was adequately covers pillars in the fourth- to fifth-order neighborhood. Additionally, for pillars located towards the edges of a tile, this neighborhood information is not available in every direction. In order not to be negatively affected from such incomplete information, we decided



(a) (b) (c)

Fig. 9. Illustration of determining the k-th neighborhood of a pillar. (a) Refractive index cross-section map of randomly-placed pillars. (b) Graph edges (red lines) assigned by Delaunay triangulation on the given pillars. (c) For the pillar indicated with red, the k-th order neighbors determined by the breadth-first traversal on the graph.



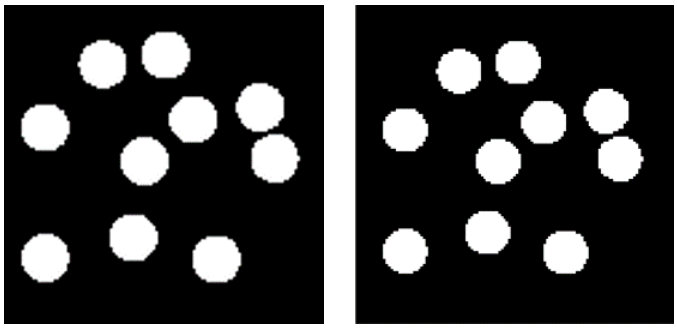
(a) (b) (c) (d)

Fig. 10. For the metaline configuration, visual results obtained for the inverse design problem. For an exemplary test set sample, (a) the original simulation design and (b) the inverse design prediction. For another sample that was used to understand the robustness of the network, (c) the original simulation design and (d) the inverse design prediction.

to use only the predictions in the middle 128×128 section of each tile. Note that the selection of only the middle section would not cause incomplete predictions as the tiles were cropped by sliding windows with an increment of 64 pixels.

3.4. Inverse Design Problem for Metaline and Triangular Grid Configurations

The single-task network design illustrated in Fig. 4 was used to predict the metasurface geometry from the EM near-field intensity maps. The experiments revealed that the predictions were quite successful for both configurations. For the metaline and triangular grid configuration, the visual result predicted by the proposed network on an exemplary test set sample is depicted in Figs. 10 and 11, respectively, together with the original simulation design that was used to create the EM near-field monitor data, which were the network inputs for this prediction. As also seen in these figures, a near-perfect prediction performance was achieved. Additionally, we designed a specific test for the metaline configuration to understand the robustness of its network on the positioning of the three pillars. This test aimed to investigate the performance of the geometry prediction when the three pillars were arbitrarily located. Fig. 10c shows such a test sample, which is much



(a) (b)

Fig. 11. For the triangular grid, (a) the original simulation design of an exemplary test set sample and (b) its inverse design prediction.

Table 2. For the inverse design prediction in the random pillar configuration, the precision, recall, and f-score percentages obtained on the test set samples.

	Precision	Recall	F-score
Sample 1 (Fig. 8)	98.90	99.08	98.99
Sample 2 (Fig. S8)	99.01	99.14	99.07
Sample 3 (Fig. S8)	98.87	99.01	98.94
Sample 4 (Fig. S8)	99.11	99.04	99.08
Sample 5 (Fig. S8)	99.29	99.42	99.35

more complex that the training set samples that allow only one-dimensional pillar positioning. As seen in Fig. 10d, the network successfully generalized learned interactions between pillars on even such arbitrarily located pillars.

3.5. Inverse Design Problem for Random Pillar Configuration

The single-task network was tested on five test set samples. The original simulation design and the prediction are given in Fig. 8 for the first sample, and in Fig. S8 for the others. For quantitative assessment, first, the number of true positive (TP) pillars was calculated comparing the pillars P that were predicted by the network with the pillars S in the original simulation design. A predicted pillar $p_i \in P$ is true positive if its centroid was found inside $s_j \in S$ and if there exist no other predicted pillars whose centroids were found inside the same pillar s_j. Then, the precision=TP/$|P|$, recall=TP/$|S|$, and f-score metrics were calculated. The f-score metric is the harmonic mean of precision and recall. Table 2 reports these metrics for each test set sample separately. This table revealed that the inverse design prediction gave near-perfect performance scores.

3.6. Verification on Achromatic Lens Design

To assess the applicability of the proposed network designs on a real-world application, we conducted experiments on a specific achromatic metalens design given in [32], which consists of

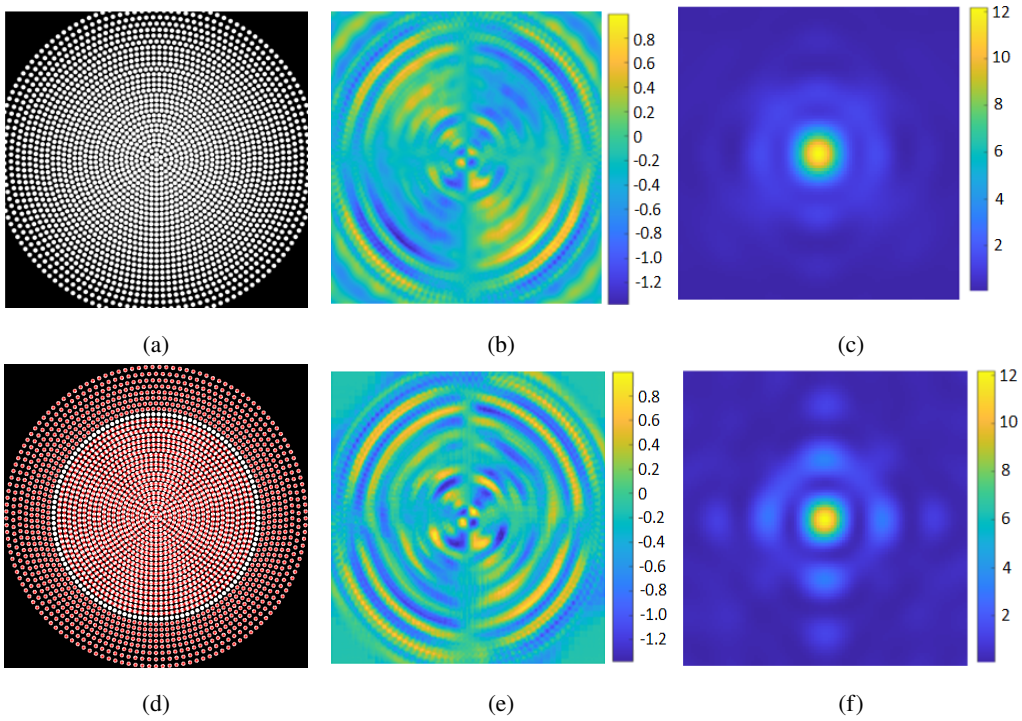


Fig. 12. (a) Refractive index cross-section of a circular achromatic metalens. (b) Its EM near-field intensity map obtained by simulation. (c) EM far-field intensity map calculated from the simulated EM near-field intensity map. (d) Refractive index cross-section estimated from the EM near-field response by the proposed single-task network for the inverse design problem. Here predictions are indicated as red circles and the boundaries of pillars in the original design, shown in (a), are indicated as white for comparison. (e) EM near-field intensity map predicted by the proposed multi-task network. (f) EM far-field intensity map calculated from the predicted EM near-field intensity map.

a periodically structured metasurface configuration (Fig. 12). To better predict the EM near-field intensity map from this specific geometry, we transferred the weights of the neural network pretrained for the random pillar configuration and fine-tuned these weights on new samples that also contained periodic and dense nanopillar structures. Next, the newly trained network was used to predict the EM near-field intensity map for the achromatic metalens design. This transfer learning is illustrated in Fig. 13. Here it is important to note that the new training samples were not cropped from any part of the achromatic metalens design but from synthetically generated structures containing pillars on a regular grid (one example is shown in Fig. 13). As a result, there existed no bias in the training samples towards overfitting this achromatic metalens design.

For the achromatic metalens design displayed in Fig. 12a, the full EM near-field intensity maps obtained by simulation and predicted by the proposed neural network are illustrated in Figs. 12b and 12e, respectively. The MSE between these maps was 4.23e-07. The main assumptions of the Fresnel approximation were not satisfied for this particular application, and thus, another approach was necessary to calculate the far-field projections. To this end, we used the "farfieldexact3d" script method, available in the Lumerical FDTD solver [39], as follows. The predictions were fed to the FDTD environment again and labeled as near-field DFT monitor information. Then, the "farfieldexact3d" script method was called to calculate the EM far-field response at the desired

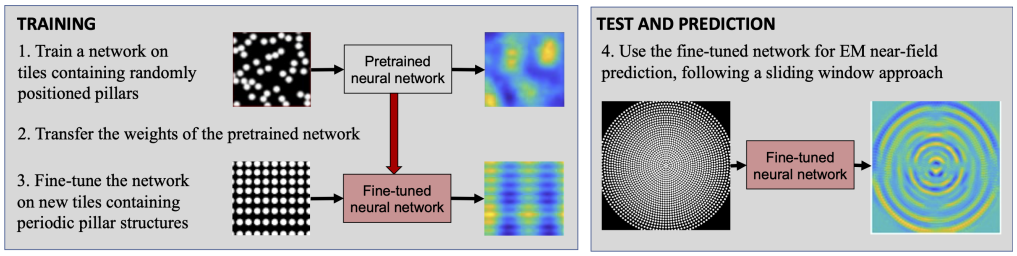


Fig. 13. Illustration of transfer learning for achromatic metalens design prediction.

points on the 3D space. The same script method was utilized for the simulation data as well, and the EM far-field intensity maps found by the same procedure are compared. These maps are illustrated in Figs. 12c and 12f, for the simulation and prediction, respectively. The MSE between these far-field maps was 1.14e-07.

Finally, we tested our inverse design prediction network on this achromatic metalens design. One predicted metasurface geometry map is shown in Fig. 12d. However, since inverse designs are not unique for the metasurface design process (i.e., since one can achieve the same EM response with different designs theoretically), we trained and tested multiple models. In Figs. 14a-14c, three different inverse designs (corresponding to three separate models) predicted for the same achromatic lens were illustrated. In all these three visuals, the first eight concentric circles were predicted at the same location as the pillars in the original design. However, the ninth layer was not predicted in any of these three runs; the inside of this missing layer turned out to be different for different runs.

The EM far-field response at the focal plane of a circular lens is more important compared to the EM near-field response. Thus, all these three inverse models were simulated with the Lumerical FDTD solver to determine their EM near-field and far-field responses and the calculated EM far-field responses are presented in Figs. 14e-14g. The EM far-field intensity map of the original design is also given in Fig. 14d for comparison. As seen in these visuals, intensity distributions of the EM far-field responses were almost identical while focusing efficiency in the predictions was slightly lower than the original design. This might be attributed to the following: In a neural network design, it is very common to normalize the input data to enhance network learning. Thus, we followed the same approach in our experiments and trained the network on the normalized EM near-field responses. This normalization, however, did not take into consideration the impact of free-space propagation in a circular design on the EM near-field intensity predictions. Despite this limitation, the EM far-field responses across all three inverse designs displayed similar characteristics, regardless of their focusing efficiency differences.

4. Discussion and Conclusion

Our experiments revealed that encoder-decoder networks, with the U-Net architecture, proved to predict EM near-field responses and design metasurface configurations. Additionally, they showed to possess several advantages over iterative numerical methods as well as the other neural networks designed for similar tasks. First, the proposed network designs reduced the required computational time compared to the iterative numerical methods, which had to perform a complicated, computationally demanding, and time-consuming simulation for each EM near-field response. On the other hand, predicting these responses by a trained neural network was faster. Note that even though the network training could take more time, this training should be undertaken only once and it could be used for many predictions without further training.

Second, compared to the previous neural networks implemented for metasurface design, our proposal of using an encoder-decoder network enabled to predict the full EM near-field intensity

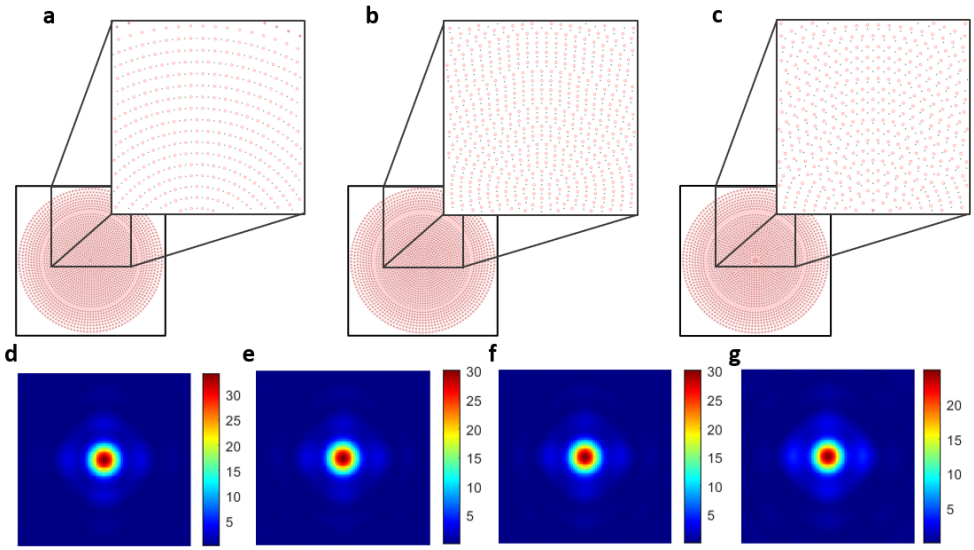


Fig. 14. (a)-(c) Three inverse designs predicted by three separately trained networks for the same achromatic lens. (d) EM far-field intensity map calculated for the original design. (e)-(g) EM far-field responses obtained by simulating the predicted inverse designs, shown in (a)-(c), with the Lumerical FDTD solver.

map rather than just analyzing the transmission spectra, which only focus on one parameter. Thus, our proposal provided a full EM simulation result that could be analyzed and processed further. Third, our proposed encoder-decoder network design facilitated a systematic prediction of a metasurface pillar geometry, namely the inverse design problem. The traditional simulation tools and methods did not provide a direct inverse design prediction and relied heavily on the trial-and-error approach. Consequently, our innovative design strategy of directly mapping a corresponding EM near-field response to the metasurface design holds immense potential to revolutionize the design approach. Additionally, the inverse solution did not set any constraints on the total number and location of metaunits as well as the total simulation area. On the other hand, such constraints existed in previous studies for the inverse design of metasurfaces.

In summary, this work introduced a novel deep learning approach for predicting the full EM near-field response of metasurfaces. It provided a complete solution for EM near-field responses, making it the first account of deep-learning approach in the literature to achieve this task. Moreover, this approach offered a solution for the inverse design problem, which aims to find the metasurface geometry from the corresponding EM near-field response, for the first time. The proposed approaches were faster and computationally less demanding than the traditional methods. Moreover, the type of the device, the operating frequency range, and the type of design geometry do not primarily affect the performance of our approach. Therefore, it is possible to apply the proposed approach in analyzing and designing different electromagnetic and photonic devices. One can explore these possibilities as future work.

5. Funding

This research is supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK), under its 1004 - Center of Excellence Support Program grant no:20AG001. H.V.D. also acknowledges the support from TUBA."

6. Acknowledgements

We would like to extend our heartfelt gratitude to two individuals who played crucial roles in the completion of this manuscript.

First, we would like to express our sincere appreciation to Rumon Miah for his valuable assistance in the initial simulations. Rumon's expertise, insightful suggestions, and technical guidance significantly influenced the direction and methodology of our research.

Second, we would like to acknowledge the contributions of Selahattin Cansız in the deep learning aspect of this manuscript. Selahattin's profound understanding of deep learning algorithms, architectures, and meticulous reviews greatly enhanced the quality and rigor of our research.

Finally, we would like to express our appreciation to all those who supported us during the course of this work, whether through providing resources, offering valuable insights, or offering moral support. Your contributions have been invaluable in shaping the outcome of this manuscript.

7. Disclosures

The authors declare no conflicts of interest.

References

1. P. D. Peter, "light scattering by small particles. h. c. van de hulst. wiley, new york; chapman & hall, london, 1957. xiii+470 pp. illus. 12." Science **127**, 477–478 (1958).
2. J. D. Jackson, *Classical electrodynamics* (Wiley, New York, 1975), 2nd ed.
3. F. Walter, G. Li, C. Meier, S. Zhang, and T. Zentgraf, "Ultrathin nonlinear metasurface for optical image encoding," Nano Letters **17**, 3171–3175 (2017).
4. C. D. Robert, K. Mohammadreza, C. Wei Ting, O. Jaewon, and C. Federico, "Broadband high-efficiency dielectric metasurfaces for the visible spectrum," Proceedings of the National Academy of Sciences **113**, 10473–10478 (2016).
5. M. Decker, I. Staude, M. Falkner, J. Dominguez, D. N. Neshev, I. Brener, T. Pertsch, and Y. S. Kivshar, "High-efficiency dielectric huygens' surfaces," Advanced Optical Materials **3**, 813–820 (2015).
6. J. Qiang, H. Leyong, G. Guangzhou, L. Junjie, W. Yongtian, and H. Lingling, "Arbitrary amplitude and phase control in visible by dielectric metasurface," Opt. Express **30**, 13530–13539 (2022).
7. A. C. Overvig, S. Shrestha, S. C. Malek, M. Lu, A. Stein, C. Zheng, and N. Yu, "Dielectric metasurfaces for complete and independent control of the optical amplitude and phase," Light Sci Appl **8**, 92 (2019).
8. E. Arbabi, A. Arbabi, S. M. Kamali, Y. Horie, M. Faraji-Dana, and A. Faraon, "Mems-tunable dielectric metasurface lens," Nat Commun **9**, 812 (2018).
9. D. Neshev and I. Aharonovich, "Optical metasurfaces: new generation building blocks for multi-functional optics," Light Sci Appl **7**, 58 (2018).
10. A. Hammad, R. Arbab Abdur, M. Husnul, A. Muhammad Mahmood, M. Nasir, and N. Sadia, "Phase engineering with all-dielectric metasurfaces for focused-optical-vortex (fov) beams with high cross-polarization efficiency," Opt. Mater. Express **10**, 434–448 (2020).
11. Y. Wenhong, S. Xiao, Q. Song, Y. Liu, Y. Wu, S. Wang, J. Yu, J. Han, and D. P. Tsai, "All-dielectric metasurface for high-performance structural color," Nature Communications **11** (2020).
12. Y. Q. Hu, X. D. Wang, X. H. Luo, X. N. Ou, L. Li, Y. Q. Chen, P. Yang, S. Wang, and H. G. Duan, "All-dielectric metasurfaces for polarization manipulation: principles and emerging applications," Nanophotonics **9**, 3755–3780 (2020).
13. P. Genevet, F. Capasso, F. Aieta, M. Khorasaninejad, and R. Devlin, "Recent advances in planar optics: from plasmonic to dielectric metasurfaces," Optica **4**, 139–152 (2017).
14. C. D. Robert, K. Mohammadreza, C. Wei Ting, O. Jaewon, and C. Federico, "Broadband high-efficiency dielectric metasurfaces for the visible spectrum," Proceedings of the National Academy of Sciences **113**, 10473–10478 (2016).
15. N. F. Yu and F. Capasso, "Flat optics with designer metasurfaces," Nature Materials **13**, 139–150 (2014).
16. M. Khorasaninejad, Z. Shi, A. Y. Zhu, W. T. Chen, V. Sanjeev, A. Zaidi, and F. Capasso, "Achromatic metalens over 60 nm bandwidth in the visible and metalens with reverse chromatic dispersion," Nano Letters **17**, 1819–1824 (2017).
17. X. Ding, F. Monticone, K. Zhang, L. Zhang, D. Gao, S. N. Burokur, A. de Lustrac, Q. Wu, C.-W. Qiu, and A. Alù, "Ultrathin pancharatnam–berry metasurface with maximal cross-polarization efficiency," Advanced Materials **27**, 1195–1200 (2015).
18. Z. Junxiao, Q. Haoliang, C. Ching-Fu, Z. Junxiang, L. Guangru, W. Qianyi, L. Hailu, W. Shuangchun, and L. Zhaowei, "Optical edge detection based on high-efficiency dielectric metasurface," Proceedings of the National Academy of Sciences **116**, 11137–11140 (2019).
19. L. Q. Cong, Y. K. Srivastava, and R. Singh, "Inter and intra-metamolecular interaction enabled broadband high-efficiency polarization control in metasurfaces," Applied Physics Letters **108** (2016).
20. Y. Vahabzadeh, N. Chamanara, K. Achouri, and C. Caloz, "Computational analysis of metasurfaces," IEEE Journal on Multiscale and Multiphysics Computational Techniques **PP** (2017).
21. X. Lin, Y. Rivenson, N. T. Yardimei, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, "All-optical machine learning using diffractive deep neural networks," Science **361**, 1004–+ (2018).
22. J. Peurifoy, Y. C. Shen, L. Jing, Y. Yang, F. Cano-Renteria, B. G. DeLacy, J. D. Joannopoulos, M. Tegmark, and M. Soljacic, "Nanophotonic particle simulation and inverse design using artificial neural networks," Science Advances **4** (2018).
23. S. An, C. Fowler, B. Zheng, M. Y. Shalaginov, H. Tang, H. Li, L. Zhou, J. Ding, A. M. Agarwal, C. Rivero-Baleine, K. A. Richardson, T. Gu, J. Hu, and H. Zhang, "A deep learning approach for objective-driven all-dielectric metasurface design," ACS Photonics **6**, 3196–3207 (2019).
24. C. C. Nadell, B. Huang, J. M. Malof, and W. J. Padilla, "Deep learning for accelerated all-dielectric metasurface design," Opt Express **27**, 27523–27535 (2019).
25. F. Ghorbani, S. Beyraghi, J. Shabanpour, H. Oraizi, H. Soleimani, and M. Soleimani, "Deep neural network-based automatic metasurface design with a wide frequency range," Sci Rep **11**, 7102 (2021).
26. T. S. Qiu, X. Shi, J. F. Wang, Y. F. Li, S. B. Qu, Q. Cheng, T. J. Cui, and S. Sui, "Deep learning: A rapid and efficient route to automatic metasurface design," Advanced Science **6** (2019).
27. S. S. An, B. W. Zheng, M. Y. Shalaginov, H. Tang, H. Li, L. Zhou, Y. X. Dong, M. Haerinia, A. M. Agarwal, C. Rivero-Baleine, M. Kang, K. A. Richardson, T. Gu, J. J. Hu, C. Fowler, and H. L. Zhang, "Deep convolutional neural networks to predict mutual coupling effects in metasurfaces," Advanced Optical Materials **10** (2022).
28. Z. Li, R. Pestourie, Z. Lin, S. G. Johnson, and F. Capasso, "Empowering metasurfaces with inverse design: Principles and applications," ACS Photonics **9**, 2178–2192 (2022).
29. M. V. Berry, "Quantal phase-factors accompanying adiabatic changes," Proceedings of the Royal Society of London

Series a-Mathematical and Physical Sciences **392**, 45–57 (1984).
30. O. Akin and H. V. Demir, "High-efficiency low-crosstalk dielectric metasurfaces of mid-wave infrared focal plane arrays," Applied Physics Letters **110** (2017).
31. A. Nagarajan, K. van Erve, and G. Gerini, "Ultra-narrowband polarization insensitive transmission filter using a coupled dielectric-metal metasurface," Optics Express **28**, 773–787 (2020).
32. H. B. Yağcı and H. V. Demir, "'meta-atomless' architecture based on an irregular continuous fabric of coupling-tuned identical nanopillars enables highly efficient and achromatic metasurfaces," Applied Physics Letters (2020).
33. R. Osgood jr. and X. Meng, *Numerical Methods* (Springer International Publishing, 2021).
34. O. Ronneberger, P. Fischer, and T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation* (2015), book section Chapter 28, pp. 234–241, Lecture Notes in Computer Science.
35. R. Caruana, "Multitask learning," Machine Learning **28**, 41–75 (1997).
36. M. J. Barth, R. R. McLeod, and R. W. Ziolkowski, "A near and far-field projection algorithm for finite-difference time-domain codes," Journal of Electromagnetic Waves and Applications **6**, 5–18.
37. P. Pellatfinet, "Fresnel diffraction and the fractional-order fourier-transform," Optics Letters **19**, 1388–1390 (1994).
38. L. Chen and J. C. Xu, "Optimal delaunay triangulations," Journal of Computational Mathematics **22**, 299–308 (2004).
39. "farfieldexact3d - script command – ansys optics," .