Designing Multiplexed Supercell Metasurfaces with Tandem Neural Networks

Christopher Yeung1,2, Ju-Ming Tsai1, Brian King1, Benjamin Pham1, Julia Liang1, David Ho1, Mark W. Knight2, and Aaswath P. Raman1,*

1Department of Materials Science and Engineering, University of California, Los Angeles, CA 90024, USA
2Northrop Grumman Corporation, Redondo Beach, CA 90278, USA

*Corresponding Author: aaswath@ucla.edu

Abstract: Complex nanophotonic structures hold the potential to deliver exquisitely tailored optical responses for a range of device applications. Metal-insulator-metal (MIM) metasurfaces arranged in supercells, for instance, can be tailored by geometry and material choice to exhibit a range of absorption properties and resonant wavelengths. With this flexibility however comes a vast space of design possibilities that classical design paradigms struggle to efficiently navigate. To overcome this challenge, here we demonstrate that a deep tandem neural network approach can efficiently generate multiplexed supercells through inverse design. By using a training dataset with several thousand full-wave electromagnetic simulations in a design space of over three trillion possible designs, the deep-learning model can accurately generate a wide range of complex supercell designs given a spectral target. Beyond inverse design, we show that our approach can also be used to explore the limits of broadband absorption and emission in such supercell configurations. Our demonstration of high-dimensional supercell inverse design with deep neural networks is generally applicable to any complex nanophotonic structure composed of multiple subunit elements.

Keywords: nanophotonics; supercells; plasmonic metasurfaces; deep learning, tandem neural network
1. Introduction

Nanophotonic structures, including metasurfaces and metamaterials, have greatly expanded our ability to tailor light-matter interaction and deliver new functionalities for information processing and sensing applications [1], [2], [3], [4]. As demand for advanced capabilities and high-performance nanophotonic devices grow, multimodal implementations with interconnected ensembles of optical sub-components, including in supercells, have shown great promise in delivering tailored responses with respect to many optical characteristics [5], [6], [7], [8]. For example, complex spatial arrangements within photonic crystal circuits have yielded high-efficiency spatial mode conversion [9]. Similarly, by employing metasurfaces that contain periodic arrays of meta-atoms with different geometric parameters, a range of useful behaviors including out-of-plane beam deflection and mirroring can be demonstrated [10]. Although the incorporation of numerous distinct subunit elements within a photonic structure is desirable, it is accompanied by an exponential increase in design costs as a result of the increased dimensionality of the associated design space [11].

Conventional design processes for complex metasurfaces and photonic crystals rely on electromagnetic (EM) simulations that are iteratively optimized by tuning key design parameters until the desired optical properties are obtained. Techniques that have been employed include evolutionary algorithms [12], topology optimization [13], [14], [15], and adjoint-based methods [16], [17]. In the context of supercells and complex/non-periodic arrangements, methods such as Schur complement domain decomposition and overlapping-domain approximation have yielded compelling results [18], [19]. As the unit cell of a metasurface increases in size and complexity, however, computation times from iterative optimization can rapidly escalate from hours to potentially days or weeks. Additionally, optimizations must be repeated and reconfigured for every new target, thus requiring a substantial amount of computational resources and often some prior intuition on the capability of a particular class of nanophotonic structures. These computational costs are further compounded by the fact that only the final optimized results are preserved; any prior data generated in an optimization cycle is not typically reused in the future [41]. As a result, iterative design methods become increasingly inefficient over time [20].

In response to the need for more efficient design strategies, recent advancements in deep neural networks (DNNs) have found application in nanophotonic design [21]. DNNs are now well established in many applications including natural language processing, drug discovery, materials design, and medical diagnosis [22], [23], [24]. In the photonics context, DNNs have shown promise in generating results that are less noisy and more fabricable than other machine learning-based methods (such as generative adversarial networks [25], [26]) by directly predicting key geometric parameters (e.g., resonator widths, lengths, radii, etc.). Previously explored DNN-based photonics design include the forward and inverse modeling of multi-shell nanoparticles, multi-layer thin-films, and various classes of metasurfaces [27], [28], [29]. In a forward-modeling DNN, the network takes structural parameters as inputs and predicts optical properties such as the absorption spectra. By contrast, an inverse-modeling DNN accepts target optical properties as inputs and generates matching structural parameters. By leveraging a one-time investment of EM simulation training data (where all of the generated data is used for training), an accurate inverse-modeling DNN has the potential to design devices that outperform those produced by inverse optimization, while generating the same designs with orders-of-magnitude faster speeds [30]. While promising, prior studies of DNNs for nanophotonic design have primarily focused on individual scatterers or periodic structures with single-unit cell elements and relatively narrowband operation [31], [32], [33]. The possibility of employing a machine learning-based approach to designing complex supercells has not been
previously investigated, but holds the potential to overcome key challenges in meeting challenging target
capabilities with complex nanophotonic architectures.

A particular category of periodic metasurface structure that has shown promise in supercell
configurations is the metal-insulator-metal (MIM) metasurface absorber. Periodic MIM absorbers yield
strong resonances that are narrowband in nature, where the wavelength of the resonance peak can be shifted
by changing the shape of the resonator [34], [35], [36], [37]. By adopting simple supercell configurations,
which contain more than one resonator geometry, multi-resonant and broadband absorption behavior has
previously been realized [38], [39], [40]. The design and optimization of more complex supercells however
remains an open challenge, but holds the potential of yielding a broader range of spectral selectivity than
has previously been achieved.

In this article, we investigate the inverse design of large multiplexed supercell metasurfaces with
over 100 subunit elements that can achieve a diverse set of broadband spectral responses. Specifically, we
focus on engineering arbitrary bandwidth absorbers operating in the mid- and long-wave infrared regime
(4-12 µm) by designing supercell metal-insulator-metal (MIM) metamaterial absorbers through a deep
learning approach. To navigate the large design space that comes with the increased dimensionality of
supercells, we employed a tandem neural network (shown conceptually in Figure 1A). We demonstrate that
with a training dataset of several thousand simulations, in a high-dimensional design space with over three
trillion possible design combinations, the network can inverse-design narrowband, multi-resonance, and
broadband-absorption supercell metasurfaces with high degrees of accuracy. Furthermore, we show that
the network itself can be harnessed to explore the physical limits of this class of metasurfaces.

2. Results and discussion

2.1 Data preparation for deep learning

Figure 1B presents the detailed implementation of our tandem neural network approach for the
inverse design of supercell metasurfaces. First, we defined an array of MIM resonators, as seen in the
bottom-right (blue section) of the schematic. We used gold cross-shaped resonators of 100 nm thickness
with a 100 nm gold backing and 200 nm Al2O3 spacer. This class of metasurfaces was derived from existing
literature on selective thermal emitters and exhibits narrowband resonances in the MIR range [42]. The
cross-shaped resonators had fixed widths (500 nm) and variable lengths (1.4-3 µm in 0.1 µm steps). Each
resonator array represents a quadrant of the supercell and resembles a hexagonal close-packed (HCP) lattice
with a twin boundary, where the individual resonators are mirrored along the diagonal plane. The quadrant
is then mirrored along the x- and y-axes to create a four-fold symmetric supercell. The HCP configuration
is designed to maximize the area density (and therefore the resonance efficiency) of the supercell, while the
four-fold symmetry ensures the structure is s- and p-polarization independent under normal incidence. At
the designated positions of each cross-shaped resonator, we used the corresponding resonator lengths to
create a vector for each supercell design. A structure is represented by $D_A = [l_1, l_2, \ldots, l_m]$, with $l_m$ being the
length of the $m$-th resonator (where $D_A$ and $D_B$ are vectors with distinct $l$-values). These vectors were then
used as the supercell design parameters for deep learning. We then converted the supercell design
parameters into three-dimensional MIM structure models and used the models to perform full-wave EM
simulations (Lumerical FDTD) over the spectral range of 4-12 µm at normal incidence, obtaining a 800-
point absorption spectrum for each structure. With this approach, we simulated the absorption spectrum ($A$)
for randomly generated design parameters ($D$) to create training data pairs ($D, A$) for the neural network.
Figure 1. Inverse design of supercell metasurface designs with a range of underlying symmetries using a tandem neural network approach. (A) A target absorption spectrum is defined, and the matching design parameters for a multiplexed array of plasmonic resonators are generated. (B) Data preparation schematic for deep learning. Supercell design parameters representing resonator lengths and positions are converted into 3D models. Full-wave electromagnetic simulations are performed on the models. The design parameters and corresponding absorption spectra are used to train the tandem neural network.

2.2 Network characterization and evaluation

The performance of a tandem network hinges on the accuracy of the forward-modeling network as well as the breadth and size of the training dataset. Thus, we sought to optimize the architecture of the forward-modeling network and to ensure that the size of our training dataset maximizes the network’s implementation efficiency and predictive capabilities. The forward-modeling network is trained to predict an absorption spectrum ($A'$), given a set of design parameters ($D$) as inputs. Our optimized architecture consists of a 25-neuron input layer (matching the vector size of the supercell design parameters $D$, with values normalized from 0 to 1), and four fully-connected hidden layers with 50, 100, 200, and 400 neurons,
respectively. The output layer contains 800 neurons. Figure S1 presents the hyperparameter tuning results, where — in addition to the number of layers and neurons — the Adam optimizer, batch size of 10, and sigmoid activation functions yielded the lowest validation loss. The training and validation losses of the forward-modeling network are found in Figure 2A (on the left plot), where 10% of the total training dataset (3,600 instances of \(D\) and \(A\)) was used for validation. To prevent overfitting, an early stopping procedure was added to the training process (after 20 epochs of \(\Delta \text{Validation MSE} = 0\)), and the convergence of both losses can be observed. In addition, example test results of the forward network from the validation dataset, on the middle and right plots of Figure 2A, indicate that its predictions (given an input \(D\) from the training dataset) match well with the targets (of the corresponding \(A\)).

Using the same hyperparameters as the forward network with an inverted sequence of hidden layers and neurons (400, 200, 100, and 50), we trained an inverse-modeling network for the prediction of \(D\) given an input \(A\). Model validation was performed by simulating the predicted \(D\), then comparing the target (\(A\)) and simulated spectra. As shown in Figure 2B, the validation loss in the inverse model is higher than the training loss, indicating that the model is overfitting. This behavior is reflected in the example validation test results (to the right of the loss plot), where the network fares poorly when predicting parameters for absorption spectra outside of the training set.

Poor convergence of the inverse-modeling network is a direct result of the nonuniqueness problem [27], where the multiple mappings between an EM response and its available structural parameters may confound the network’s learning process. To account for the nonuniqueness problem, we implemented the tandem architecture by coupling the inverse-modeling network with a pretrained forward-modeling network. The inverse-modeling network was trained by minimizing the loss function between the input absorption spectrum (\(A\)) and the spectrum predicted by the forward-modeling network (\(A'\)), where \(A'\) is generated by the same \(D\) predicted by the inverse-modeling network. As in Ref. [26], we define loss as the mean squared error between \(A\) and \(A'\):

\[
\text{MSE} = \frac{1}{n} \sum (A'_i - A_i)^2.
\]

Thus, rather than requiring the network to generate designs that are similar to the designs from the training dataset (which may vary significantly due to the issue of nonuniqueness), the loss function converges if the target and predicted spectra (\(A\) and \(A'\)) are similar. Accordingly, training the tandem network resulted in strong convergence of both training and validation losses, and the example tests reveal that the accuracy of the tandem implementation surpasses that of the inverse-modeling architecture alone (Figure 2C).

The number of supercell design parameters in \(D\) (\(m = 25\)) and the range of values therein (9 different cross lengths) mean there are a total of \(3.81 \times 10^{12}\) possible supercell designs in the investigated design space. Thus, to expedite our deep learning efforts, we sought to minimize simulation time and to ensure that the quantity of our training data is amenable to the network’s intended functions. We trained the optimized tandem network under various dataset size increments, as shown in Figure S2. We then validated each model with a fixed dataset across all increments (using 300 reserved data instances) and with datasets created by randomly selecting 10% of the data instances within each increment. We observe that in both cases, the validation loss converges to \(2.5 \times 10^{-3}\) at approximately 3,000 data instances, demonstrating that the final model resulted in the optimal performance while minimizing the amount of data required for deep learning by sampling less than \(1 \times 10^{-7}\%\) of the total design space.
2.3 Inverse design of multi-resonance and broadband metasurfaces

We utilized the optimized tandem network to generate new supercell metasurface designs with a broad range of spectral properties. Figure 3 presents a series of test cases comparing the target network inputs to the simulated results of the corresponding output designs. The inset images show the spatial geometries of each supercell designed by the network. For example, as shown in Figure 3A, after specifying a narrowband target with a full width half maximum (FWHM) of 0.5 µm, the network generated a periodic layout that aligns with the target spectra with over 90% accuracy, and prior literature [42]. Similarly, in Figure 3B and 3C, dual-narrowband and triple-narrowband designs were created (with sharp resonances at two and three discrete wavelengths) that closely match their respective targets. In these multi-resonance structures, the supercells include additional cross dimensions that are associated with distinct resonances.

The ability to construct an array of resonator geometries suggests that different resonant modes can be superimposed to achieve responses of arbitrary bandwidth [43]. Accordingly, we tasked the neural network with designing metasurfaces with various broadband characteristics (FWHM > 1 µm). In Figure
3D, a broadband structure with a FWHM of 1.5 µm is shown, and in Figure 3E, we increased the complexity of the target to design a structure with dual-broadband absorption peaks. Lastly, in Figure 3F, we demonstrate the design of a broadband graybody structure that encompasses the entire MIR range of resonance wavelengths captured by the training dataset (5-9 µm).

In the design of the aforementioned multi-resonance and broadband structures, the network not only defined the resonator dimensions required to achieve resonances at the target wavelengths, but also determined their appropriate placements within the lattice in order to reach the target absorption amplitudes. For example, as shown in Figure 4A, the triple-narrowband structure possesses three primary cross lengths that are responsible for resonances at 5.2, 7.2, and 8.6 µm. The high absorption amplitudes are also attributed to the periodic and short-range ordered arrangements (repeating patterns spanning 1-2 subunit cell distances) of the resonators, which result in the strong dipole resonances seen in the electric field enhancement plots. When short-range order is disrupted and converted to long-range order (patterns spanning beyond 2 subunit cell distances), as shown in Figure 4B, the net absorption spectra is drastically altered. Thus, by systematically predicting the subunit resonator dimensions as well as their spatial positions, the trained network can accurately and rapidly design multiplexed metasurfaces with a range of underlying symmetries.

**Figure 3.** Inverse design of new supercell metasurfaces with the tandem neural network. The structures exhibit (A) narrowband, (B) dual-narrowband, (C) triple-narrowband, (D) broadband, (E) dual-broadband, and (F) graybody behaviors. Blue lines indicate the target spectra used as inputs to the network, and orange lines represent the simulated results of the output design parameters. Inset images show the physical layouts of the network-generated supercells.

### 2.4 Exploring the physical limits of metasurface designs

In practice, it is difficult to combine multiple distinct resonant modes in a single metasurface while minimizing hybridization between modes and maintaining high absorption per unit area [43], [44]. Thus, multiplexed resonator structures impose an inherent tradeoff between broadband response and maximum absorption. We here highlight that the instantaneous calculation speed of the developed neural
network can be harnessed to explore the physical limits and the structure-property relationships of classes of metasurfaces. To enable this, we use the pre-trained forward-modeling network as a validation mechanism for the design parameters predicted by the inverse-modeling tandem network. As one example, we specified design targets using Lorentzian functions of increasing bandwidth (FWHM of 0.2-4 µm centered at 7 µm), illustrated in Figure 5A. The tandem network outputs were then fed into the forward-modeling network, and the resulting design predictions were compared to the initial targets. In this approach, the forward network effectively serves as a high-speed surrogate EM solver, replacing the FDTD software that was used to generate the training data.

Figure 4. Relationships between supercell absorption properties and their subunit resonator spatial distributions. The absorption spectra and corresponding electric field profiles for the (A) triple-narrowband and (B) graybody structures reveal the dependence of absorption amplitude on short-range and long-range ordered resonators, respectively.

The design predictions reveal that when an unobtainable target was specified, the network designs a structure with the closest possible solution in the context of this class of supercell metasurfaces. As a result, we observe that as the target bandwidth increases, the discrepancy between the target response and the closest design (measured by the MSE) increases as well (Figure 5B). This in turn allows us to numerically infer the limits of achievable broadband resonances for this class of resonant metasurfaces. Building upon these observations, we can estimate the upper bounds of the absorption at various bandwidths, and derive the following relationship for this metasurface class’ absorption design limits ($R^2 = 0.98$):

$$A_{\text{max}} = 0.0004f^2 - 0.0305f + 1.0214,$$

where $A_{\text{max}}$ is the maximum absorption and $f$ is the FWHM in THz. As an additional example of discovering application-specific design insights through the neural network, we can calculate the average normal-incidence emissivity of the optimized supercell metasurfaces within defined target bandwidths:
\[ \overline{\varepsilon} = \frac{\int_{v_1}^{v_2} I_{BB}(T,v) \varepsilon(v) dv}{\int_{v_1}^{v_2} I_{BB}(T,v) dv}. \] (2)

Here, \( I_{BB}(T, v) = \frac{2h \nu^3}{c^2} \frac{\nu}{e^{h\nu/k_BT} - 1} \) is the spectral radiance of a blackbody at temperature \( T \), where \( h \) is Planck’s constant, \( k_B \) is the Boltzmann constant, \( c \) is the speed of light, and \( \nu \) is frequency. The lower and upper bounds of the integral \((v_1 \text{ and } v_2)\) are derived from the evaluated spectral range (4-12 µm). \( \varepsilon(v) \) is the metasurface’s spectral emittance, which is equal to \( A(v) \) by Kirchoff’s law. In this case, by querying the neural network in a cyclic manner to solve for emittance (at various temperatures) as a function of the target bandwidth, we can find the relationship between the two parameters (Figure 5B) in a remarkably short time frame (less than one minute). Overall, we observe that as the bandwidth sought (FWHM) increases, the MSE between a target with maximum absorption across the entire bandwidth increases and the maximum absorption point decreases. Furthermore, the integrated normal incidence emittance increases as the additional bandwidth compensates for the decreases in the peak absorption/emittance value. However as can be seen in Figure 5B the precise relationship is complex and depends both on the bandwidth being specified and the temperature of the metasurface because the blackbody spectral radiance changes with temperature. By training a neural network that is tasked with the inverse design of complex supercell metasurfaces, we thus demonstrate that the same framework can be strategically leveraged to rapidly identify design trends and dependencies associated with application-specific properties.

**Figure 5.** Probing metasurface design limits using the tandem network. (A) Tandem network inputs (targets) for various Lorentzian functions (FWHM of 0.6, 1, 2, and 3 µm centered at 7 µm) and the corresponding forward network outputs (design predictions). The network is unable to identify designs that exceed the bandwidth / maximum absorption limits of this class of supercell metasurfaces (B) Network-determined design trends and metrics, including the MSE between target and design responses, thermal emittance of the metasurface, and max absorption as functions of FWHM (THz).
3. Conclusions

In this article, we demonstrated a machine learning approach to the inverse-design of multiplexed supercell metasurfaces with over 100 subunit elements. By forming a cascaded architecture with an inverse-modeling and forward-modeling network, the tandem network can overcome the nonuniqueness problem and successfully learn a high-dimensional design space of over three trillion possible designs using only 3,600 data instances. Through the superposition and coupling of multiple resonant modes in a compact region, the network can further efficiently design supercell structures with a range of symmetries that yield narrowband, broadband, and multi-resonant responses. The network not only predicts the geometric parameters for an array of resonators (e.g., resonator widths, lengths, radii, etc.), but also selects their optimum spatial arrangement towards satisfying a specified target, thereby enabling additional degrees of complexity in metasurface design. Though we sought to maximize implementation efficiency by minimizing the required training data, we expect that the performance of our tandem network can be improved with more training data and a larger network architecture. Furthermore, we demonstrate that the network itself can be utilized to study the physical limits of the investigated class of metasurfaces. By using the forward-modeling network as a full-wave EM simulator, high-speed parameter sweeps can be performed to capture property-specific design trends such as maximum absorption and thermal emittance as a function of bandwidth. Importantly, our results show that DNN-based approaches can efficiently design and characterize large-scale supercell metasurfaces with numerous discrete resonators. We believe our results can expedite the development of supercell-class nanophotonic structures, which may in turn yield new tailored capabilities not achievable through conventional periodic nanostructures.

4. Supplementary Material

Included in the supplementary material are details regarding hyperparameter optimization and training data size.

Acknowledgements: This work was supported by the Sloan Research Fellowship from the Alfred P. Sloan Foundation.

References

[1] Olthaus J, Schrinner P, Reiter D. Optimal Photonic Crystal Cavities for Coupling Nanoemitters to Photonic Integrated Circuits. Adv Quantum Technol 2020, 3:1900084.
[2] Yoshimi H, Yamaguchi T, Ota Y, Arakawa Y, Iwamoto S. Slow light waveguides in topological valley photonic crystals. Opt Lett 2020, 45:2648-2651.
[3] Bin Tarik F, Famili A, Lao Y, Ryckman J. D. Robust optical physical unclonable function using disordered photonic integrated circuits. Nanophotonics 2020, 20200049.
[4] Mittapalli V, Khan H. Excitation Schemes of Plasmonic Angular Ring Resonator-Based Band-Pass Filters Using a MIM Waveguide. Photonics 2019, 6(2):41.
[5] Ding F, Wang Z, He S, Shalaev VM, Kildishev AV. Broadband high-efficiency half-wave plate: a supercell-based plasmonic metasurface approach. ACS Nano 2015, 9(4):4111-9.
[6] Aoni RA, Rahmani M, Xu L, et al. High-efficiency visible light manipulation using dielectric Metasurfaces. Sci Rep 2019, 9(1):1-9.
[7] Wu PC, Tsai WY, Chen WT, et al. Versatile polarization generation with an aluminum plasmonic metasurface. Nano Lett 2017, 17(1):445-52.
[8] Ma Q, Chen L, Jing HB, et al. Controllable and programmable nonreciprocity based on detachable digital coding metasurface. Adv Opt Mater 2019, 7(24):1901285.
[9] Liu V, Miller DA, Fan S. Ultra-compact photonic crystal waveguide spatial mode converter and its connection to the optical diode effect. Opt Express 2012, 20(27):28388-97.
[10] Guo X, Ding Y, Chen X, Duan Y, Ni X. Molding Free-Space Light with Guided-Wave-Driven Metasurfaces. 2020, arXiv preprint, arXiv:2001.03001.
[11] Hegde RS. Deep learning: a new tool for photonic nanostructure design. Nanoscale Adv 2020, 2:1007–1023.
[12] Gondarenko A, Lipson M. Low Modal Volume Dipole-like Dielectric Slab Resonator. Opt Express 2008, 16:17689-17694.
[13] Kao CY, Osher S, Yablonovitch E. Maximizing Band Gaps in Two-Dimensional Photonic Crystals by Using Level Set Methods. Appl Phy. B: Lasers Opt 2 2005, 81:235-244.
[14] Piggott AY, Lu J, Lagoudakis KG, et al. Inverse Design and Demonstration of a Compact and Broadband on-Chip Wavelength Demultiplexer. Nat Photonics 2015, 9(6):374-377.
[15] Shen B, Wang P, Polson R, Menon R. An Integrated Nanophotonics Polarization Beamsplitter with 2.4x2.4 μm2 Footprint. Nat Photonics 2015, 9:378-382.
[16] Oskooi A, Mutapcic A, Noda S, et al. Robust Optimization of Adiabatic Tapers for Coupling to Slow-Light Photonic-Crystal Waveguides. Opt Express 2012, 20:21558-21575.
[17] Seliger P, Mahvash M, Wang C, Levi A. Optimization of Aperiodic Dielectric Structures. J Appl Phys 2006, 100:034310.
[18] Verweij S, Liu V, Fan S. Accelerating simulation of ensembles of locally differing optical structures via a Schur complement domain decomposition. Opt Lett 2014, 39(22):6458-61.
[19] Lin Z, Johnson SG. Overlapping domains for topology optimization of large-area metasurfaces. Opt Express 2019, 27(22):32445-53.
[20] Elesin Y, Lazarov BS, Jensen JS, Sigmund O. Time domain topology optimization of 3D nanophotonic devices. Photonic Nanostruct 2014, 12(1):23-33.
[21] Yeung C, Tsai JM, King B, Kawagoe Y, Ho D, Knight M, Raman AP. Elucidating the Behavior of Nanophotonic Structures Through Explainable Machine Learning Algorithms. ACS Photonics. 2020.
[22] Abiodun OI, Jantana A, Omolara AE, Dada KV, Mohamed NA, Arshad H. State-of-the-art in artificial neural network applications: A survey. Heliyon 2018, 4(11):e00938.
[23] Muhammad W, Hart GR, Nartowt B, Farrell JJ, Johung K, Liang Y, Deng J. Pancreatic cancer prediction through an artificial neural network. Frontiers in Artificial Intelligence 2019, 2:2.
[24] Conduit B, Jones N, Stone H, Conduit G. Design of a nickel-base superalloy using a neural network. Mat and Des 2017, 131:358-365.
[25] So S, Rho J. Designing nanophotonic structures using conditional deep convolutional generative adversarial networks. Nanophotonics 2019, 8:1255–1261.
[26] Liu Z, Zhu D, Rodrigues SP, Lee KT, Cai W. Generative Model for the Inverse Design of Metasurfaces. Nano Lett 2018, 18:6570–6576.
[27] Liu D, Tan Y, Khoram E, Yu Z. Training deep neural networks for the inverse design of nanophotonic structures. ACS Photonics 2018, 5(4):1365-9.
[28] Peurifoy J, Shen Y, Jing L, et al. Nanophotonic particle simulation and inverse design using artificial neural networks. Sci Adv 2018, 4(6):eaar4206.

[29] An S, Fowler C, Zheng B, et al. A deep learning approach for objective-driven all-dielectric metasurface design. ACS Photonics 2019, 6(12):3196-207.

[30] Jiang J, Fan JA. Global optimization of dielectric metasurfaces using a physics-driven neural network. Nano Lett 2019, 19(8):5366-72.

[31] Ma W, Cheng F, Liu Y. Deep-learning-enabled on-demand design of chiral metamaterials. ACS Nano 2018, 12(6):6326-34.

[32] Inampudi S, Mosallaei H. Neural network based design of metagratings. Appl Phys Lett 2018, 112(24):241102.

[33] Harper ES, Coyle EJ, Vernon JP, Mills MS. Inverse design of broadband highly reflective metasurfaces using neural networks. Phys Rev B 2020, 101(19):195104.

[34] Ogawa S, Kimata M. Metal-insulator-metal-based plasmonic metamaterial absorbers at visible and infrared wavelengths: a review. Materials 2018, 11(3):458.

[35] Vorobyev AY, Topkov AN, Gurin OV, Svich VA, Guo C. Enhanced absorption of metals over ultrabroad electromagnetic spectrum. Appl Phys Lett 2009, 95(12):121106.

[36] Ye YQ, Jin Y, He S. Omnidirectional, polarization-insensitive and broadband thin absorber in the terahertz regime. JOSA B 2010, 27(3):498-504.

[37] Chen HH, Su YC, Huang WL, Kuo CY, Tian WC, Chen MJ, Lee SC. A plasmonic infrared photodetector with narrow bandwidth absorption. Appl Phys Lett 2014, 105(2):023109.

[38] Ma Y, Chen Q, Grant J, Saha SC, Khalid A, Cumming DR. A terahertz polarization insensitive dual band metamaterial absorber. Opt Lett 2011, 36(6):945-7.

[39] Shen X, Cui TJ, Zhao J, Ma HF, Jiang WX, Li H. Polarization-independent wide-angle triple-band metamaterial absorber. Opt Express 2011, 19(10):9401-7.

[40] Luo H, Cheng YZ, Gong RZ. Numerical study of metamaterial absorber and extending absorbance bandwidth based on multi-square patches. Eur Phys J B 2011, 81(4):387-92.

[41] Gao L, Li X, Liu D, Wang L, Yu Z. A bidirectional deep neural network for accurate silicon color design. Adv Mater 2019, 31(51):1905467.

[42] Liu X, Tyler T, Starr T, Starr AF, Jokerst NM, Padilla WJ. Taming the blackbody with infrared metamaterials as selective thermal emitters. Phys Rev Lett 2011, 107(4):045901.

[43] Fan RH, Xiong B, Peng RW, Wang M. Constructing metastructures with broadband electromagnetic functionality. Adv Mater 2019, 1904646.

[44] Ma W, Wen Y, Yu X. Broadband metamaterial absorber at mid-infrared using multiplexed cross resonators. Opt Express 2013, 21(25):30724-30.
Supplementary Material

Forward-Modeling Network Optimization

To train the tandem network for inverse design, we first optimized the architecture of the forward-modeling network through extensive hyperparameter tuning. Figure S1 shows the tuning results, where we compare the validation loss of a starting controlled architecture to the losses of networks trained after changing a single dependent variable. Implemented through the TensorFlow framework, the controlled architecture consists of 2 hidden layers, each with 100 neurons, sigmoid activation functions, a batch size of 10 data instances, and the Adam optimizer. The learning rate is 0.001 with an exponential decay of $10^{-5}$. The tested dependent variables include: number of hidden layers, number of neurons within each layer, activation function, batch size, and optimizer.

Figure S1A shows a comparison of different batch sizes (10, 100, and 1000), where we observe noticeable increases in loss as the batch size was increased. In Figure S1B, we tested three commonly used optimization algorithms: Adam, stochastic gradient descent (SGD), and RMSprop. SGD yielded higher losses than the other algorithms while Adam and RMSprop resulted in similar losses. However, RMSprop plateaued much sooner than Adam, indicating that the network was able to improve further with Adam. In Figure S1C, we compared the following activation functions: Sigmoid, TanH (hyperbolic tangent), ReLU (rectified linear unit), Leaky ReLU, and Parametric ReLU. Here, we observe that the Sigmoid and TanH functions resulted in the lowest losses, while the loss progression was more stable and ultimately lower with Sigmoid. We then tested various numbers of neurons and layers (Figure S1D and S1E), and found that increasing from 100 to 200 neurons garnered small improvements, while increasing the number of layers alone yielded insignificant changes. Figure S1F shows the full integration of the individually-optimized hyperparameters and the use of more elaborate combinations of neurons and layers, which resulted in considerable overall performance improvements. From these tests, we found that the 50-100-200-400 neuron architecture has the best performance. Adding more neurons to the optimized architecture did not improve the performance any further and unnecessarily increased training time.

Figure S1. Forward network hyperparameter tuning. Validation loss of a controlled architecture in comparison to various dependent variables, including: (A) batch size, (B) optimizer, (C) activation function, (D) number of neurons, (E) number of layers, (F) combined hyperparameter settings.
(D) number of neurons per layer, and (E) number of hidden layers. (G) Final optimized architecture comparison.

**Training Dataset Size Analysis**

To expedite our deep learning efforts, we sought to minimize the total simulation time and therefore the amount of training data. However, we must also ensure that the dataset size is large enough to maximize the network’s ability to learn supercell design. In that regard, as shown in Figure S2A, we trained the optimized tandem network architecture with increasing increments of training data and evaluated the corresponding validation losses. First, we trained the network with a training set size of 300, and recorded the network’s final validation loss using two different validation datasets. The first validation dataset was a fixed group of 300 data instances. This dataset was intended to monitor the network’s growth as it trained toward a predetermined set of goals. The second validation dataset comes from randomly splitting 10% of the data instances within the available training set. The validation loss derived from the second dataset informs how well the network is able to generalize from the amount of data it learned from. We repeated this validation process after increasing the training set size in 300-increment steps, and found that both validation losses began to converge to $2.5 \times 10^{-3}$ at approximately 3,000 data instances. The reported losses represent the averaged results of 5 training cycles for each training set size. Error bars were omitted due to negligible differences in the range of losses. Thus, we demonstrate that the final model resulted in the optimal performance while minimizing the amount of data required for deep learning.

![Figure S2](image.png)

**Figure S2.** Training set size analysis. Validation loss vs. training set size with fixed and randomly split validation datasets.