ACS Photonics

pubs.acs.org/journal/apchd5

Article

Enhancing Adjoint Optimization-Based Photonic Inverse Design with Explainable Machine Learning

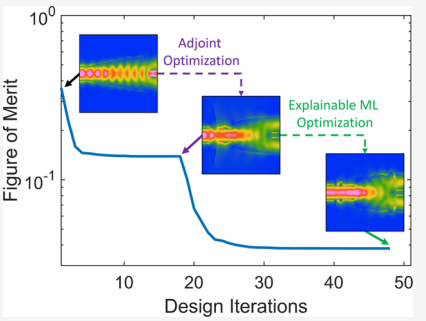
Christopher Yeung, David Ho, Benjamin Pham, Katherine T. Fountaine, Zihan Zhang, Kara Levy, and Aaswath P. Raman*

Cite This: *ACS Photonics* 2022, 9, 1577−1585

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: A fundamental challenge in the design of photonic devices, and electromagnetic structures more generally, is the optimization of their overall architecture to achieve a desired response. To this end, topology or shape optimizers based on the adjoint variable method have been widely adopted due to their high computational efficiency and ability to create complex freeform geometries. However, the functional understanding of such freeform structures remains a black box. Moreover, unless a design space of high-performance devices is known in advance, such gradient-based optimizers can get trapped in local minima valleys or saddle points, which limits performance achievable through this inverse design process. To elucidate the relationships between device performance and nanoscale structuring while mitigating the effects of local minima trapping, we present an inverse design framework that combines adjoint optimization, automated machine learning, and explainable artificial intelligence. Integrated with a numerical electromagnetic simulation method, our framework reveals structural contributions toward a figure-of-merit (FOM) of interest. Through an explanation-based reoptimization process, this information is then leveraged to minimize the FOM further than that obtained through adjoint optimization alone, thus overcoming the optimization's local minima. We demonstrate our framework in the context of waveguide splitter design and achieve between 39 and 74% increases in device performance relative to state-of-the-art adjoint optimization-based inverse design across a range of telecom wavelengths. Our results highlight machine learning strategies that can substantially extend and enhance the capabilities of a conventional, optimization-based inverse design algorithm while revealing deeper insights into the algorithm's designs.



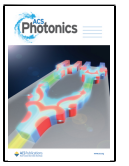
KEYWORDS: *nanophotonics, deep learning, explainability, adjoint optimization, automated machine learning*

INTRODUCTION

Effectively optimizing nanophotonic structures is key to their use in a broad range of optical applications. For example, photonic integrated circuits, metasurfaces, and guided-wave systems can be geometrically manipulated at subwavelength scales to deliver a wide range of functionalities.[1−5] However, a large design space must be rapidly explored in order to optimize the geometry for a particular application. To effectively navigate such a design space, gradient-based optimization algorithms such as the adjoint variable method have been widely adopted to design nonintuitive or irregularly shaped electromagnetic structures that are highly efficient at accomplishing a particular goal. By calculating the shape derivatives at all points in space using only two electromagnetic simulations per iteration,[6] adjoint optimizations are orders of magnitude more computationally efficient than alternative optimization methods and capable of achieving state-of-the-art performance.[6−9]

Although adjoint optimization-based methods have been successfully applied to a variety of photonic systems,[10−13] the method's reliance on gradient-based information means that the method is local in nature and therefore bounded by the corresponding limitations. Specifically, since the design space for electromagnetic structures is predominantly nonconvex, adjoint optimizations (or gradient-based optimization algorithms in general) are susceptible to getting stuck in local minima valleys or saddle points (hereon collectively referred to as local minima).[14,34] Thus, unless a region of high-performance devices is known in advance, multiple optimization runs are needed (typically by using random starting points) to arrive at a single optimization target.[15] To overcome these limitations, recent efforts have combined machine learning (ML) with adjoint optimization. For example, population-based inverse design was demonstrated using global topology-optimization networks, or GLOnets,[16] which integrate the adjoint method directly into the training process. Alternative strategies also include the integration of ML and

Received: October 25, 2021
Published: April 21, 2022



https://doi.org/10.1021/acsphotonics.1c01636
ACS Photonics 2022, 9, 1577−1585

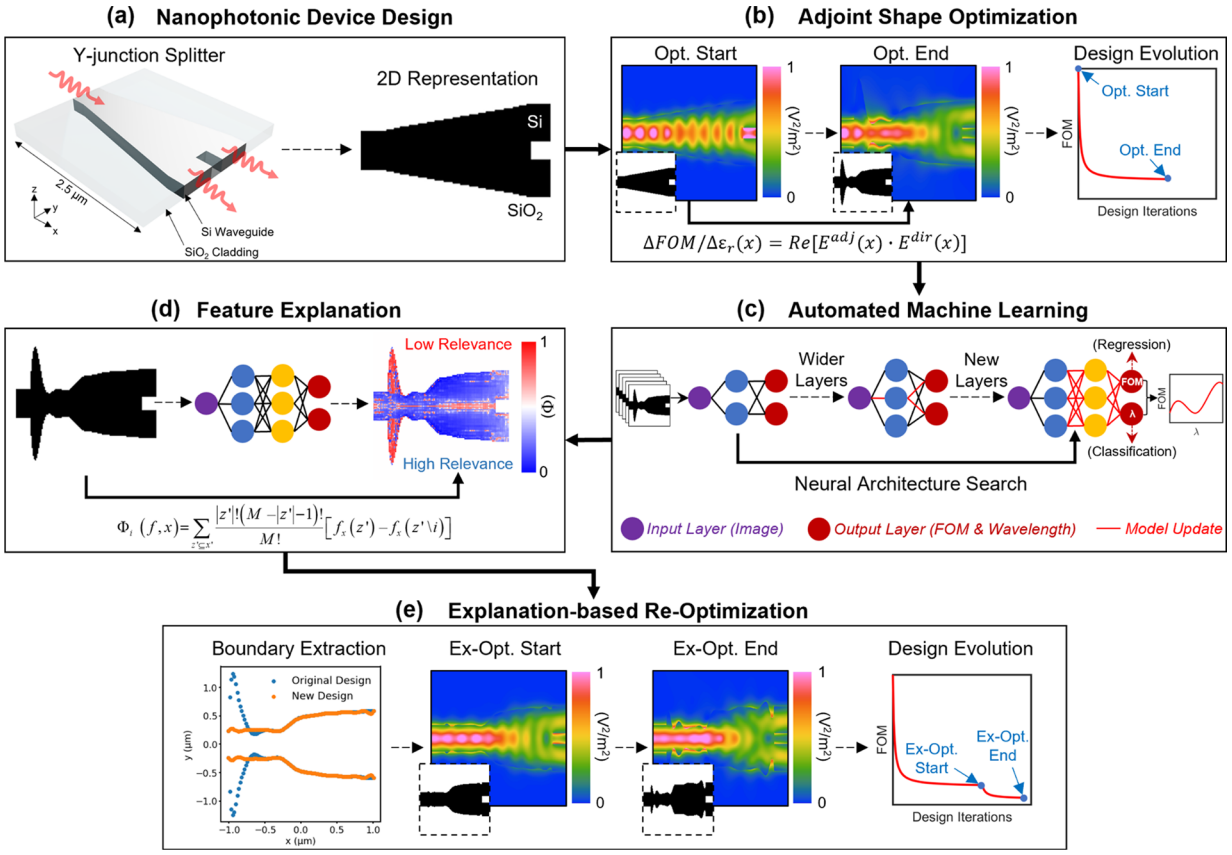


Figure 1. (a) Nanophotonic device optimization: silicon-on-insulator Y-junction splitter for telecom applications. (b) Multiple adjoint optimization runs are applied to the Y-splitter design at various target wavelengths. (c) Results of the optimizations are used as training data in automated machine learning (AutoML) to train a neural network, where the inputs are images and the outputs are device figure-of-merits (FOM) and target wavelengths. (d) Explainable AI algorithms are used on the neural network to capture feature explanations, (e) which are used to optimize device performance further by allowing the algorithm to escape its local minima.

adjoint optimization as a two-step process, where the ML component performs an initial global-search approximation, and then the optimization improves design performance further.[16−18,40] Although both approaches can improve upon the algorithm's performance, the underlying issue of local minima trapping remains unaddressed, since the integration and use of a gradient-based optimizer inherently indicates that the issue is still present. In this regard, metaheuristic techniques such as simulated annealing have been proposed to escape local minima in the search process.[19,20] This method may directly address the issue of local minima trapping through neighbor-based exploration, but its application to photonic shape and topology optimization has been severely limited due to relatively low computational efficiency (on the order of 1000 iterations).[21,22]

To comprehensively address the issue of local minima trapping in adjoint optimization-based inverse design, we seek to first identify the root of the problem and find what caused the algorithm to get trapped in the first place. In the context of the optical structures being optimized, arriving at certain geometric elements and their resulting electromagnetic response must contribute to guiding the optimization toward suboptimal results. To discover the geometric features responsible for local minima trapping and to then overcome them, we employ an explainable artificial intelligence (XAI)-based approach where a neural network is trained using data output from the adjoint optimization method. XAI serves as a promising candidate for addressing local minima trapping due to its well-known ability to reveal a model's decision making process as well as the contributing factors thereof (i.e., addressing the black box problem).[23−25] For example, XAI can reveal the spatial regions of a nanophotonic structure that contribute to the presence or lack of an absorption peak.[26] Thus, to explain the causes of local minima trapping in gradient-based adjoint optimization and subsequently use this information to prevent the optimization from converging onto suboptimal states, we present an XAI-based framework that utilizes the relationships between device efficiency and nanoscale structuring to increase overall inverse design optimization performance.

METHODS

We demonstrate our optimization framework in the context of Y-splitter waveguide design (Figure 1a), where the objective is to optimize the shape of the silicon−oxide interface to maximize power transmission efficiency from an input port to two output ports of the same width. Here, we represent the objective function as a decreasing figure-of-merit (FOM) which ranges from 1 to 0, where 0 represents ideal performance. With this definition in place, adjoint shape optimizations are performed on an initial Y-splitter design to minimize the FOM (Figure 1b) across a range of target wavelengths in the telecommunication window (1.3−1.8 μm). The design and FOM information from the optimizations are used (as training data) in conjunction with neural architecture

https://doi.org/10.1021/acsphotonics.1c01636
ACS Photonics 2022, 9, 1577−1585

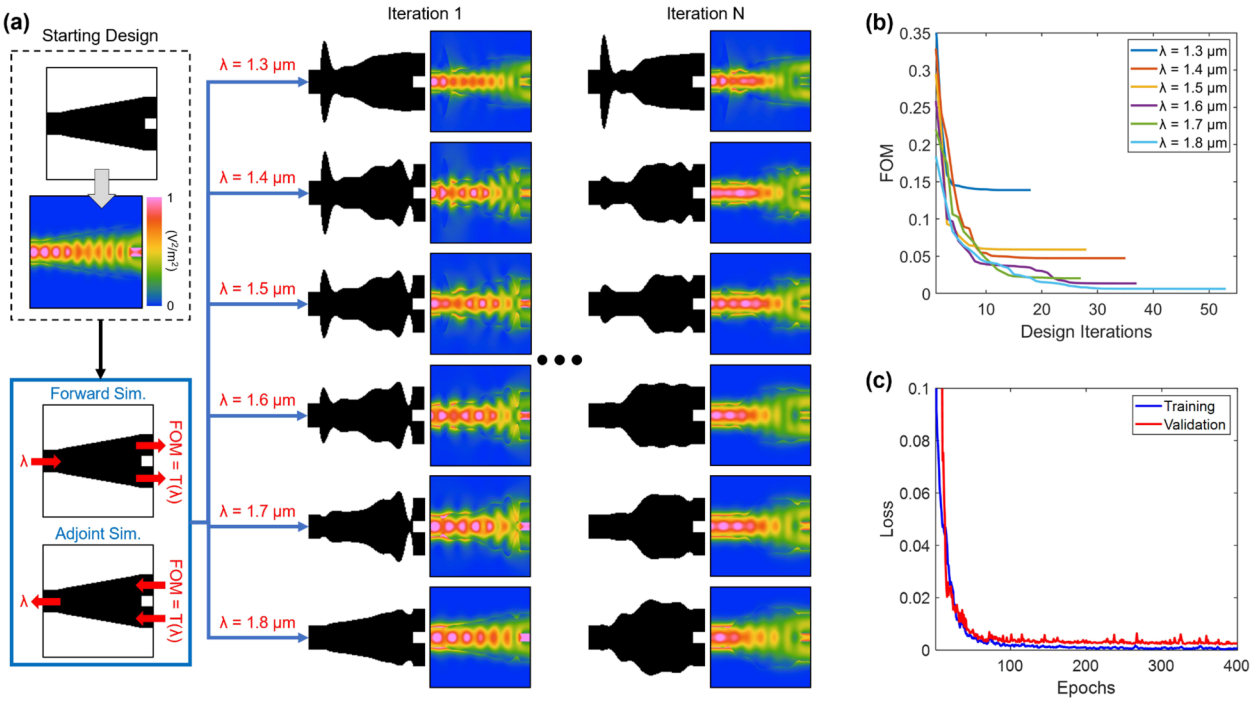


Figure 2. (a) Training data generation. Adjoint optimization runs are performed on random starting designs (35% fill fraction starting design shown) across target wavelengths ranging from 1.3 to 1.8 μm to produce high-performance devices in the telecom window. (b) FOM vs design iterations across each optimization run. (c) Training and validation losses for the AutoML-optimized neural network shows high training accuracy and model convergence.

search to automate the training of an ML model (AutoML). After training, the model learns the relationships between the device structure and performance by accurately predicting the FOM and target wavelength of an input design (Figure 1c). We then use a suite of XAI algorithms, SHapley Additive exPlanations or SHAP (a post hoc explanation technique based on game theory[27]), on the model to extract the structure−performance relationships as "feature explanation" heatmaps (Figure 1d). By interrogating our trained ML model, the explanations here inform the structural features that contribute to the FOM of interest. Using this information, we devise a boundary extraction algorithm that takes the explanations and makes guided design changes that enable the optimization to escape its local minima (Figure 1e). These design changes then provide a new starting point for the local adjoint optimization method, which allows the method to reach lower FOMs, or higher device performances,than previously achievable (at multiple target wavelengths).

RESULTS AND DISCUSSION

Adjoint Optimization and Convolutional Neural Network Training. We first developed our training dataset by performing multiple adjoint shape optimization runs on a starting Y-splitter design (Figure 2a). We applied a widely adopted implementation of the adjoint method that is integrated with a commercial finite-difference time-domain solver.[28] The 2D cross-sections of the Y-splitter designs are represented as black and white images, where the black and white pixels represent the permittivity of silicon and SiO_2, respectively. In our configuration of the adjoint method, the optimizable region is the area between the input and output ports, while the port sizes remain fixed. The optimizable geometry within this region is defined using the level set method and cubic spline interpolations.[28,29] Each optimization run was performed on randomized starting designs (waveguide structures with 25, 35, 40, and 50% fill fractions; collectively shown in Figure S2) using different operating wavelengths as optimization objectives (1.3−1.8 μm in 0.1 μm steps) to produce a collection of device designs with gradual performance improvements. Performance improvement is indicated by a decreasing FOM (as design iteration increases), until a plateau is reached. As shown in Figure 2a (here, the 35% fill fraction starting design), each design iteration consists of a forward and adjoint (i.e., time-reversed) simulation, which calculates the shape derivative over the entire optimizable region and modifies the geometry (per iteration) in proportion to the FOM gradient.[6] At the final iteration "N" (which may vary across each optimization run), device geometry is tailored to achieve maximum attainable performance with respect to the sought target. In our application of the adjoint method, the FOM represents the power coupling of guided modes and is defined as follows:

$$\mathrm{FOM} = \frac{1}{P(\lambda)} \int |T_0(\lambda) - T(\lambda)| \mathrm{d}\lambda \tag{1}$$

where λ is the evaluated wavelength, T is the actual power transmission through the output ports, T_0 is the ideal power transmission, and P is the source power (in Watts). Thus, the FOM is the difference between the input and output transmission normalized by the power injected by the source, which results in the maximum performance at 0. Figure 2b shows the results of each optimization run, where the collective FOM information and corresponding designs (at each iteration; not including the starting design) are used as training data for deep learning.

From the optimized structures shown in Figure 2, unique geometries are obtained for each target wavelength, which in

https://doi.org/10.1021/acsphotonics.1c01636
ACS Photonics 2022, 9, 1577−1585

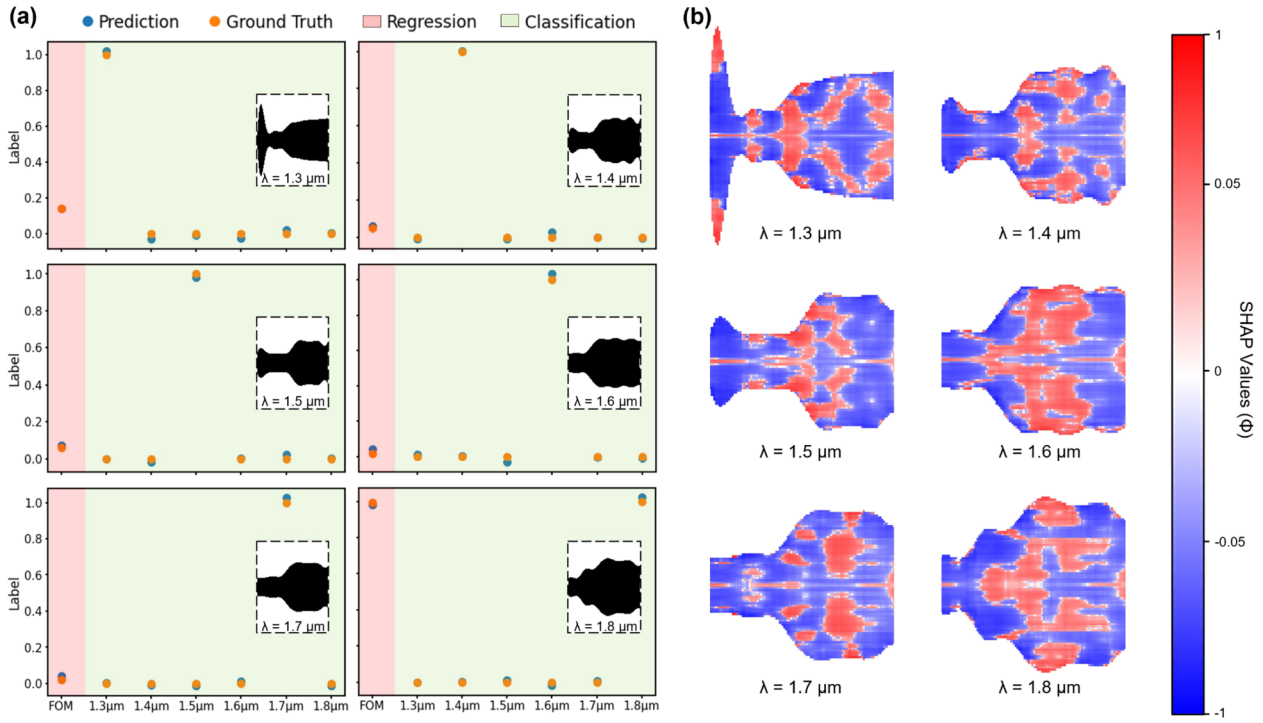


Figure 3. (a) Comparisons between model predictions and ground truths, for FOM (regression) and target wavelength (classification) values, show that the model accurately learned the relationship between the device structure and performance. Inset images show the model inputs, which are adjoint-optimized devices. (b) SHAP explanation heatmaps of the optimized devices reveal the structural features that contribute positively (blue) or negatively (red) toward optimal device performance. Note that this is the reverse of conventional SHAP definitions due to our desired FOM being minimized.

turn yield a range of FOM values. Therefore, the FOM and target wavelength are coupled with one another, and both are dependent on the waveguide structure. Thus, to ensure that our model simultaneously learns both of these properties, which in turn captures more information regarding the structure than models trained on the properties individually, we designed a single neural network that takes the Y-splitter geometries as inputs (here, 128 × 64 pixel images, or 2.5 × 1.25 μm^2 domains) and outputs both FOM and target wavelength. In the particular design space we explored, over 600 input and output pairs were generated for the neural network. For ease of training, target wavelengths were converted into categorical labels, where a position-specific output node value of 1 represents the wavelength of a specific design, while the other positional nodes equal 0. For example, a target wavelength (T_λ) of 1.3 μm is represented as $T_{1.3}$ = [1,0,0,0,0,0], 1.4 μm is $T_{1.4}$ = [0,1,0,0,0,0], and this pattern is repeated up to 1.8 μm. Alternatively, argmax(T_λ) or the index of the maximum value along the vector represents the target wavelength. Combined with a floating point value (ranging from 0 to 1) to serve as the FOM, we devise a training data structure that is amenable to both classification and regression-based tasks. With this input−output relationship defined, as well as a 90:10 training−validation data split, we use neural architecture search (AutoKeras[30]) with image blocks to automate the deep learning process by testing different model variants across multiple trials. We observe that the optimal architecture was identified after 12 trials, which had a validation loss of 9.1×10^{-5}. The final training and validation losses of each trial are presented in Figure S1a, and the evolution of the convolutional neural network (CNN) architecture from the first trial to the last is shown in Figure S1b. The optimized CNN possesses five convolutional blocks (with 512, 256, 128, 64, and 32 filters, respectively) followed by a dense layer. Each block contains Leaky ReLu, batch normalization, and max pooling layers. Training progression of the optimized architecture is shown in Figure 2c, where a strong convergence between the training and validation losses can be observed. We further verified our model's performance through cross validation and overfitting analyses (found in Tables S1 and S2 and Figures S12 and S13 of the Supporting Information).

Structure Explanation and Reoptimization. After training our machine learning model, we next sought to explain the relationship between the overall shape and FOM such that this information can be leveraged to potentially further optimize the devices, and overcome any local minima the adjoint method may have arrived at. To verify that the model properly learned the structure−FOM relationship, we passed the final design iterations (of each target wavelength) into the trained model and compared the ground truth outputs to the model's predictions. The comparison is shown in Figure 3a, where we observe a strong match (over 90% accuracy) between the predictions (blue points) and ground truths (orange points). The inset images in Figure 3a are model inputs. From this result, we can infer that the model accurately learned the key features on the optimized structures, which contribute to the target wavelength-specific FOM values. Therefore, we can utilize XAI to reveal the structure−performance relationships of each device. Specifically, we employed an explanation strategy for photonic design—using SHAP—to highlight the device feature contributions to their respective FOM.[26] These feature contribution heatmaps (represented as SHAP values, $\phi(x,y)$, ranging from −1 to 1)

https://doi.org/10.1021/acsphotonics.1c01636
ACS Photonics 2022, 9, 1577−1585

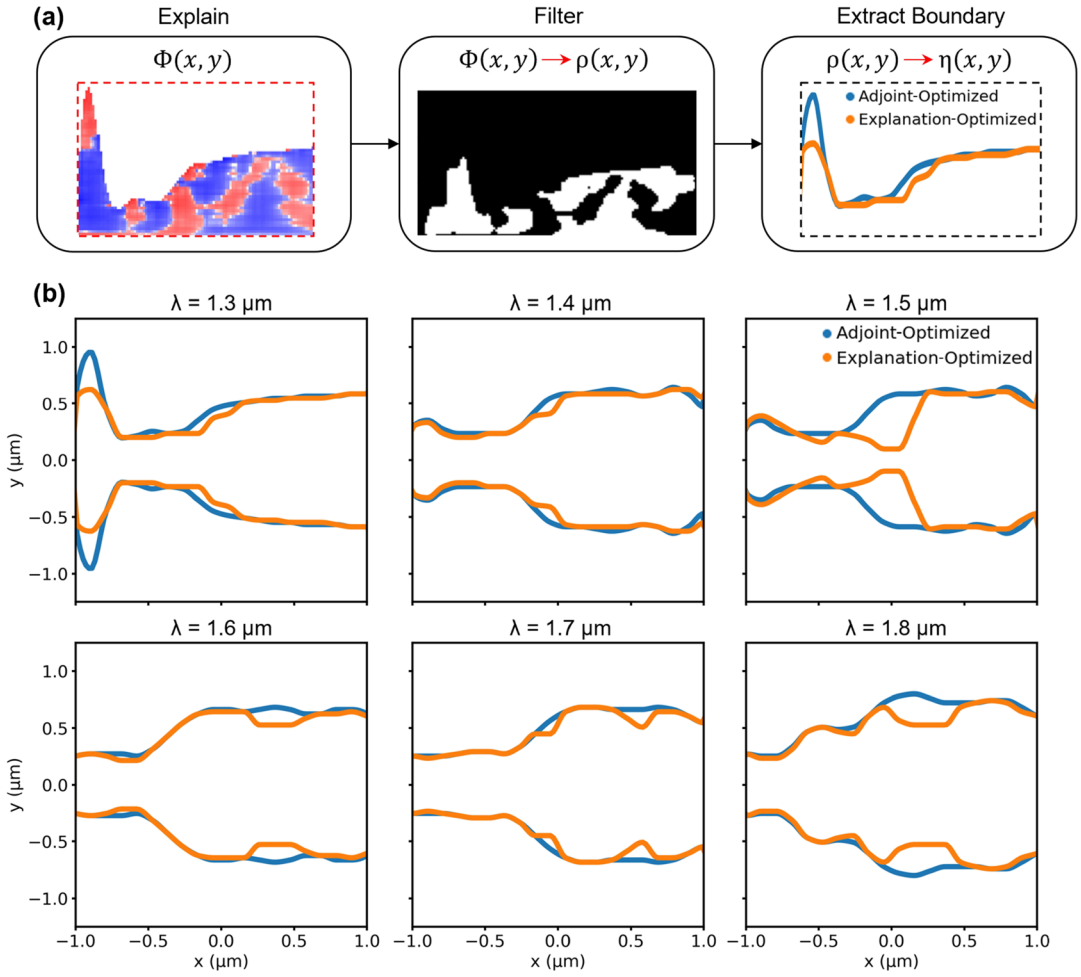


Figure 4. (a) Schematic representation of our explanation-optimization algorithm and workflow, consisting of explanation, filtering, and boundary extraction steps. (b) Comparison between the adjoint-optimized and explanation-optimized geometries.

are illustrated in Figure 3b, where the blue and red pixels indicate positive and negative contributions toward the FOM, respectively. We note that this is the reverse of conventional SHAP definitions due to our desired FOM being minimized. We then leverage the information captured by the XAI algorithm, and manipulate the structure accordingly, to assess its effect on device performance.

To determine how to practically use the SHAP values (represented as red/blue heatmap pixels), we first note that high concentrations of blue pixels are located throughout a majority of each structure, while the center of the structures and select portions of the outer boundaries contain large regions of red pixels. In this regard, since the training data solely consist of geometries with varying degrees of shape changes at the SOI (silicon-on-insulator) boundary, and no geometry change is introduced within the structure (i.e., no material subtractions or white pixels are inside the main island of black pixels), we focus our analysis on the SHAP values located at the structure boundary rather than the center. Following the aforementioned principles of positive and negative contributions, we define the red regions along the structure boundaries as negative contributions toward device performance that should be removed from the design. Accordingly, we devised a boundary extraction algorithm to systematically adjust the shape of the adjoint-optimized devices using the explanation heatmaps. The algorithm, conceptually illustrated in Figure 4a, consists of an initial filtering procedure, which identifies the red-to-blue transition points along the structure boundary. In this procedure, a binarization function is applied to the SHAP values that convert the structure into existing and nonexisting elements (shown in the center of Figure 4a as white and black pixels, respectively). Thresholds for binarization ($\rho(x,y)$) are given by the following step function:

$$\rho(x, y) = \begin{cases} 1 \text{ for } \Phi(x, y) \leq 0 \\ 1 \text{ for } \Phi(x, y) > 0 \end{cases} \quad (2)$$

where $\rho(x,y) = 1$ and $\rho(x,y) = 0$ indicate existing and nonexisting elements, respectively. A median filter is applied to the binarization to reduce noise. Next, we "draw" a new boundary around the existing elements (Figure 4a, right) by capturing an array of points $\eta(x,y) = [X_i;Y_i]$ in which $X_i = [x_1,x_2,...,x_i]$ and $Y_i = [y_1,y_2,...,y_i]$ are vectors of length i. X_i is an evenly spaced set of x-coordinate values from the left to the right of the image. Since i ultimately determines the resolution of the shape, we set its value to 20 points (matching the interval used in the initial optimization runs) to ensure that the optimizable geometry is within a feasible fabrication range. This interval equates to 100 nm spacing along the x axis, which is well within CMOS lithography resolutions. Each point on the spline can range from 0 to 1.25 μm in 20 nm steps Thus,

https://doi.org/10.1021/acsphotonics.1c01636
ACS Photonics 2022, 9, 1577−1585

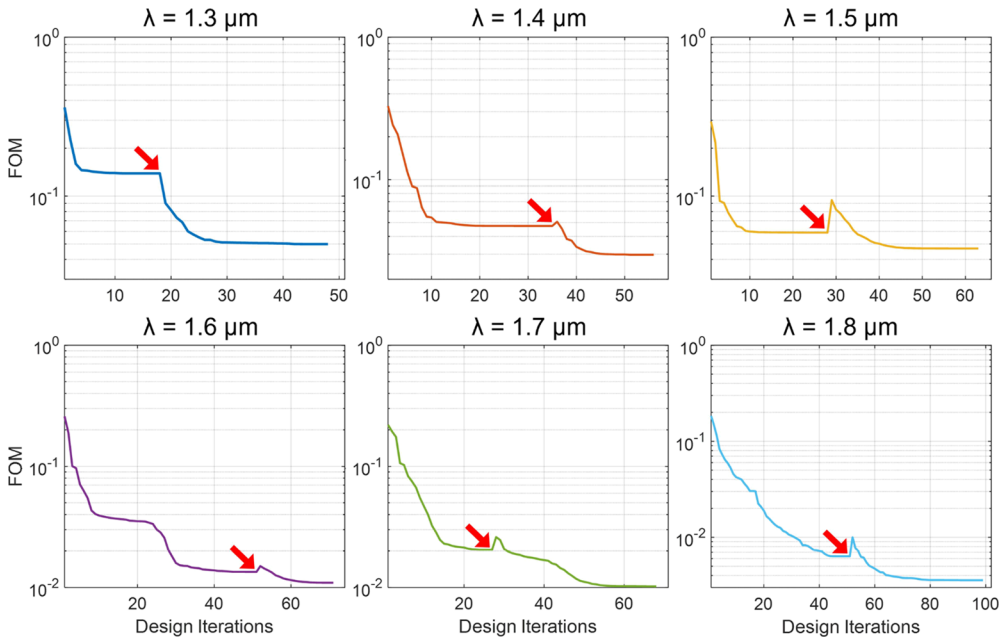


Figure 5. Two-stage optimization of SOI waveguide designs, across target wavelengths ranging from 1.3 to 1.8 μm, using the 35% fill fraction starting design. Red arrows indicate the end of the first adjoint optimization stage and the beginning of the second explanation-based reoptimization stage. Final FOM values are improved by 39%, on average, across all target wavelengths.

the number of parameter permutations describing the design are in the order of 1×10.[30] To find the y-coordinate values in Y_i, we apply Algorithm 1 in the Supporting Information.

Using Algorithm 1, we raster the image (from top to bottom) across all values in X_i to find the points where existing elements are found (indicated by $P(x,y) = 1$), and then mark these points for Y_i. For quality purposes, we add the α hyperparameter to enhance robustness by reducing sharp changes in the structure as a result of filtering or noise from the explanations. We apply this workflow to each wavelength-specific adjoint-optimized structure from the previous step and present the new "explanation-optimized" results in Figure 4b. As an example of our method's application, for the 1.3 μm target design, we note that the explanation algorithm deemed the large vertical spike near the input port as a negative (red) contribution. After applying our explanation-based boundary extraction process, the height of the spike was reduced.

To assess whether the explanation (or SHAP value)-based modifications to the optimized structures (e.g., the spike reduction) yielded meaningful or effective contributions, we simulated the explanation-optimized designs and used them as new starting points for a second stage of adjoint optimization runs. In Figure 5, we show the FOM evolutions over the entire optimization cycle of the 35% fill fraction starting design. The explanations and optimization cycles of the remaining starting designs can be found in Figures S2−S5. The red arrows indicate the end of the first optimization stage and the beginning of the second explanation-based reoptimization stage. Further observation revealed that in the second stage of the 1.3 μm target design, reducing the vertical spike immediately reduced the FOM from 0.139 to 0.090 (a 35% improvement) at the first iteration, while the end of the optimization resulted in a final FOM of 0.050 (a 64% improvement compared to the end of the first stage). This result indicates that the explanation-based modifications overcame a saddle point in the original adjoint optimization process. In some of the other examples (e.g., 1.4−1.8 μm), the first step of the second stage did not always result in an immediate FOM reduction, particularly when the FOM value was already exceedingly low (<0.075). We validate in Figure S6 that this increase in the FOM is due to the optimization getting stuck in a local minima valley instead of a saddle point, since the FOM must first increase before the algorithm can identify a lower global minima, particularly when modifying the design from where the optimization algorithm ended. However, across all the optimization targets, the end of every second-stage optimization consistently resulted in a lower final FOM than the first-stage FOM (a 39% decrease on average). Moreover, an increase in FOM followed by a further global reduction is indicative of an objective function that was previously stuck in a local minimum.[14] Thus, we demonstrate that our explanation-based reoptimization technique is capable of enhancing the performance of the adjoint optimization algorithm by allowing the FOM to escape its local minima for various optimization targets and performance ranges. We note that this entire workflow only used two optimization runs per target: one for feature contribution learning and the other for local minima escape or global FOM reduction. As previously mentioned, alternative methods at identifying lower minima typically involve repeated optimizations at random starting points or metaheuristic approaches, which can scale well-beyond two optimization runs per target (in the order of 1000 iterations in the case of simulated annealing).[21,22]

In the proceeding sections, we further assess the performance of the presented optimization scheme by conducting a number of additional tests, including (1) an evaluation of the ability for SHAP to immediately improve the FOM (if the adjoint optimization is stopped prematurely) to determine the link between structure modifications and FOM improvements (i.e., an "early stop" analysis), (2) the applicability of our approach on a smaller dataset and (3) different material systems, and (4) comparisons of our approach against arbitrary perturbations to the adjoint-optimized structure (i.e., a

https://doi.org/10.1021/acsphotonics.1c01636
ACS Photonics 2022, 9, 1577−1585

"random change" analysis). First, to verify that the model is actually learning how to modify the structure, in our "early stop" analysis, we removed the portion of the training data where the adjoint optimization reached the local minima, retrained our model, and repeated our explanation-based modification. We observe in Figure S7 that across all target wavelengths, the final FOM obtained through the SHAP explanations is lower than the best available design, thereby confirming that the model is learning how to modify the structure to improve the FOM (based on the information it was given to learn).

Since there is substantial precedence in the existing literature on using explainable artificial intelligence with small training datasets (in the order of several dozens to hundreds of data points), particularly to reduce the burden of data collection or computation costs,[35−39] we have repeated our study on a reduced dataset using only the optimization results from a single starting design (35% fill fraction). Results of this analysis are shown in Figures S8 and S9, where we also observe performance improvements (43% on average) for all target wavelengths, which suggests that the proposed approach is applicable (to a degree) to smaller datasets.

Next, we evaluated the generalizability of the proposed framework by applying it to other contemporary nanophotonic design challenges. We note that prior studies have successfully applied XAI to alternative nanophotonic structures, such as metasurfaces, and demonstrated performance enhancements in the form of spectral property tuning[26] (though no optimization algorithm integration was employed). Thus, we focused this generalizability analysis on material alternatives. In this regard, over the past few years, Si_3N_4 has emerged as a promising alternative to silicon in photonic systems. Compared to silicon, Si_3N_4 has lower propagation losses and does not exhibit two-photon absorption in the telecommunication range[3,31−33] (among other pros and cons). Accordingly, we performed the same two-stage optimization study on an Si_3N_4 waveguide (for the same Y-splitter starting geometry) and found that the second-round optimizations were also able to surpass the results of the first at every target wavelength (Figure S10). Across all the test cases, an average FOM improvement of 74% was achieved.

Finally, to show that the achieved performance enhancements were not simply obtained through arbitrary perturbations to the optimized structure, we performed an additional "random change" analysis where we randomly modified the first-stage structures, repeated the second stage of optimizations, and compared the results. This comparison is presented in Figure S11, where five random modifications (defined in the Supporting Information) were made to each structure. Across the 30 tests performed on the six optimized designs, all of the randomly modified structures possess higher FOM values (i.e., lower performance) than those of the explainability-optimized devices, while only two "random change" results fall within 25% of the explainability-optimized device performance. Additionally, 28 tests produced higher final FOM values than the initial optimized designs. Therefore, not only are the random changes ineffective in terms of escaping the local minima, but they can also inadvertently push the optimization into a worse state compared to where the optimization started at. As such, we demonstrate that our XAI-based approach is not stochastic in nature, but can deterministically tune a structure in order to maximize performance. Through the preceding series of tests, we show that the presented approach is generally applicable to numerous applications of adjoint optimization for electromagnetic design, including those with different constituent materials, structures, and optimization targets.

CONCLUSIONS

In summary, we presented an inverse design framework that extends the capabilities of gradient-based shape or topology optimization algorithms for photonic inverse design while elucidating the relationships between device performance and nanoscale structuring. Our framework combines adjoint optimization, AutoML, and XAI to enhance device performance beyond that which is obtainable through the optimization algorithm alone. We applied our method to SOI waveguide design and showed that the optimization algorithm initially reaches a performance plateau (i.e., local minima). After utilizing XAI to reveal the device's structural contributions toward a designated FOM (where 0 represents ideal performance), we leveraged this information (in conjunction with a boundary extraction algorithm) to push the optimization out of its local minima and reduce the FOM further. Across a range of performance-plateaued devices optimized for various wavelengths within the 1.3 to 1.8 μm telecom window, our method was able to improve device performance by an average of 39%. The entire procedure only requires two optimization runs per optimization target, which is potentially more computationally efficient than alternative approaches, particularly those that rely on multiple optimization runs and random starting points. Additionally, generalizability tests performed on Si_3N_4 waveguides showed an average of 74% device performance improvement. Thus, we have demonstrated that our XAI-based approach provides an automated and systematic solution for an electromagnetic optimization algorithm to escape local minima and achieve greater device performance. Looking ahead, integrating conventional optimization and data-driven machine learning will likely prove a fruitful direction for inverse design and physics discovery in photonic systems.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsphotonics.1c01636.

AutoML training results, details of the boundary extraction algorithm, random starting design analysis, local minimum analysis, early stop analysis, reduced dataset analysis, random change comparison, generalizability or material alternative assessment, and overfitting analyses (PDF)

AUTHOR INFORMATION

Corresponding Author

Aaswath P. Raman − *Department of Materials Science and Engineering and California NanoSystems Institute, University of California, Los Angeles, Los Angeles, California 90095, United States;* orcid.org/0000-0003-0106-8341; Email: aaswath@ucla.edu

Authors

Christopher Yeung − *Department of Materials Science and Engineering, University of California, Los Angeles, Los Angeles, California 90095, United States; NG Next,*

https://doi.org/10.1021/acsphotonics.1c01636
ACS Photonics 2022, 9, 1577−1585

Northrop Grumman Corporation, Redondo Beach, California 90278, United States; orcid.org/0000-0002-8964-1044

David Ho − *Department of Materials Science and Engineering, University of California, Los Angeles, Los Angeles, California 90095, United States*

Benjamin Pham − *Department of Materials Science and Engineering, University of California, Los Angeles, Los Angeles, California 90095, United States*

Katherine T. Fountaine − *NG Next, Northrop Grumman Corporation, Redondo Beach, California 90278, United States*

Zihan Zhang − *Department of Mathematics and Department of Statistics, University of California, Los Angeles, Los Angeles, California 90095, United States*

Kara Levy − *Department of Materials Science and Engineering, University of California, Los Angeles, Los Angeles, California 90095, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsphotonics.1c01636

Funding

This work was supported by the Sloan Research Fellowship from the Alfred P. Sloan Foundation, the UCLA Hellman Fellows Award and the DARPA Young Faculty Award (W911NF2110345).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work used computational and storage services associated with the Hoffman2 Shared Cluster provided by UCLA Institute for Digital Research and Education's Research Technology Group.

■ REFERENCES

(1) Sandborn, P.; Quack, N.; Hoghooghi, N.; Chou, J. B.; Ferrara, J.; Gambini, S.; Behroozpour, B.; Zhu, L.; Boser, B.; Chang-Hasnain, C.; Wu, M. C. *Linear frequency chirp generation employing optoelectronic feedback loop and integrated silicon photonics*; CLEO; IEEE, 2013; pp 1−2.

(2) Meng, Y.; Chen, Y.; Lu, L.; Ding, Y.; Cusano, A.; Fan, J. A.; Hu, Q.; Wang, K.; Xie, Z.; Liu, Z.; Yang, Y.; Liu, Q.; Gong, M.; Xiao, Q.; Sun, S.; Zhang, M.; Yuan, X.; Ni, X. Optical meta-waveguides for integrated photonics and beyond. *Light: Sci. Appl.* **2021**, *10*, 1−44.

(3) Krutova, I. A.; Saygin, M. Y.; Dyakonov, I. V.; Kulik, S. P. Optimized low-loss integrated photonics silicon-nitride Y-branch splitter. *AIP Conf. Proc.* **2020**, *2241*, No. 020027.

(4) Shen, Y.; Harris, N. C.; Skirlo, S.; Prabhu, M.; Baehr-Jones, T.; Hochberg, M.; Sun, X.; Zhao, S.; Larochelle, H.; Englund, D.; Soljacic, M. Deep learning with coherent nanophotonic circuits. *Nat. Photonics* **2017**, *11*, 441.

(5) Harris, N. C.; Carolan, J.; Bunandar, D.; Prabhu, M.; Hochberg, M.; Baehr-Jones, T.; Fanto, M. L.; Smith, A. M.; Tison, C. C.; Alsing, P. M.; Englund, D. Linear programmable nanophotonic processors. *Optica* **2018**, *5*, 1623.

(6) Miller, O. D. *Photonic Design: From Fundamental Solar Cell Physics to Computational Inverse Design*; University of California: Berkeley, 2012; pp 1−137.

(7) Li, W.; Meng, F.; Chen, Y.; Li, Y.; Huang, X. Topology optimization of photonic and phononic crystals and metamaterials: a review. *Adv. Theory Simul.* **2019**, *2*, No. 1900017.

(8) Campbell, S. D.; Sell, D.; Jenkins, R. P.; Whiting, E. B.; Fan, J. A.; Werner, D. H. Review of numerical optimization techniques for meta-device design. *Opt. Mater. Express* **2019**, *9*, 1842−1863.

(9) Molesky, S.; Lin, Z.; Piggott, A. Y.; Jin, W.; Vucković, J.; Rodriguez, A. W. Inverse design in nanophotonics. *Nat. Photonics* **2018**, *12*, 659−670.

(10) Sell, D.; Yang, J.; Doshay, S.; Yang, R.; Fan, J. A. Large-angle, multifunctional metagratings based on freeform multimode geometries. *Nano Lett.* **2017**, *17*, 3752−3757.

(11) Matzen, R.; Jensen, J. S.; Sigmund, O. Topology optimization for transient response of photonic crystal structures. *J. Opt. Soc. Am. B* **2010**, *27*, 2040−2050.

(12) Deng, Y.; Korvink, J. G. Topology optimization for three-dimensional electromagnetic waves using an edge element-based finite-element method. *Proc. R. Soc. A* **2016**, *472*, No. 20150835.

(13) Shen, B.; Wang, P.; Polson, R.; Menon, R. Integrated metamaterials for efficient and compact free-space-to-waveguide coupling. *Opt. Express* **2014**, *22*, 27175−27182.

(14) Jiang, J.; Chen, M.; Fan, J. A. Deep neural networks for the evaluation and design of photonic devices. *Nat. Rev. Mater.* **2021**, *6*, 679−700.

(15) Fan, J. A. Freeform metasurface design based on topology optimization. *MRS Bull.* **2020**, *45*, 196−201.

(16) Jiang, J.; Fan, J. A. Multiobjective and categorical global optimization of photonic structures based on ResNet generative neural networks. *J. Nanophotonics* **2021**, *10*, 361−369.

(17) Yeung, C.; Tsai, R.; Pham, B.; King, B.; Kawagoe, Y.; Ho, D.; Liang, J.; Knight, M. W.; Raman, A. P. Global Inverse Design across Multiple Photonic Structure Classes Using Generative Deep Learning. *Adv. Opt. Mater.* **2021**, *9*, No. 2100548.

(18) Ma, W.; Liu, Z.; Kudyshev, Z. A.; Boltasseva, A.; Cai, W.; Liu, Y. Deep learning for the design of photonic structures. *Nat. Photonics* **2021**, *15*, 77−90.

(19) Nalep, J. *Smart Delivery Systems: Solving Complex Vehicle Routing Problems*; Elsevier, 2020; pp 1−276.

(20) Kim, W. J.; O'Brien, J. D. Optimization of a two-dimensional photonic-crystal waveguide branch by simulated annealing and the finite-element method. *J. Opt. Soc. Am. B* **2004**, *21*, 289−295.

(21) Zhao, C.; Zhang, J. Binary plasmonics: launching surface plasmon polaritons to a desired pattern. *Opt. Lett.* **2009**, *34*, 2417−2419.

(22) Fayyaz, Z.; Mohammadian, N.; Salimi, F.; Fatima, A.; Tabar, M. R. R.; Avanaki, M. R. N. Simulated annealing optimization in wavefront shaping controlled transmission. *Appl. Opt.* **2018**, *57*, 6233−6242.

(23) Vilone, G.; Longo, L. Explainable Artificial Intelligence: a Systematic Review. 2020, arXiv:2006.00093 [cs], May 29. https://arxiv.org/abs/2006.00093 (accessed May 15, 2020).

(24) Lundberg, S. M.; Nair, B.; Vavilala, M. S.; Horibe, M.; Eisses, M. J.; Adams, T.; Liston, D. E.; Low, D. K.; Newman, S.; Kim, J.; Lee, S. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat. Biomed. Eng.* **2018**, *2*, 749−760.

(25) Arrieta, A. B.; Díaz-Rodríguez, N.; Ser, J. D.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; Chatila, R.; Herrera, F. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf Fusion* **2020**, *58*, 82−115.

(26) Yeung, C.; Tsai, J.; King, B.; Kawagoe, Y.; Ho, D.; Knight, M. W.; Raman, A. P. Elucidating the Behavior of Nanophotonic Structures through Explainable Machine Learning Algorithms. *ACS Photonics* **2020**, *7*, 2309−2318.

(27) Lundberg, S.; Lee, S. A Unified Approach to Interpreting Model Predictions. 2017, arXiv:1705.07874 [cs], May 22. https://arxiv.org/abs/1705.07874 (accessed June 11, 2020).

(28) Lalau-Keraly, C. M.; Bhargava, S.; Miller, O. D.; Yablonovitch, E. Adjoint shape optimization applied to electromagnetic design. *Opt. Express* **2013**, *21*, 21693−21701.

(29) Osher, S.; Sethian, J. A. Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. *J. Comput. Phys.* **1988**, *79*, 12−49.

(30) Jin, H.; Song, Q.; Hu, X. Auto-Keras: An Efficient Neural Architecture Search System. 2018, *arXiv:1806.10282 [cs]*, June 27. https://arxiv.org/abs/1806.10282 (accessed June 11, 2020).
(31) Zhang, Y.; Husko, C.; Lefrancois, S.; Rey, I. H.; Krauss, T. F.; Schröder, J.; Eggleton, B. J. Non-degenerate two-photon absorption in silicon waveguides: analytical and experimental study. *Opt. Express* **2015**, *23*, 17101−17110.
(32) Dirani, H. E.; Youssef, L.; Petit-Etienne, C.; Kerdiles, S.; Grosse, P.; Monat, C.; Pargon, E.; Sciancalepore, C. Ultralow-loss tightly confining Si_3N_4 waveguides and high-Q microresonators. *Opt. Express* **2019**, *27*, 30726−30740.
(33) Baets, R.; Subramanian, A. Z.; Clemmen, S.; Kuyken, B.; Bienstman, P.; Le Thomas, N.; Roelkens, G.; Van Thourhout, D.; Helin, P.; Severi, S. Silicon Photonics: silicon nitride versus silicon-on-insulator. *Opt. Fiber Commun. Conf* **2016**, 1−3.
(34) Jin, C.; Ge, R.; Netrapalli, P.; Kakade, S. M.; Jordan, M. I. How to escape saddle points efficiently. *ICML* **2017**, 1−9.
(35) Mangalathu, S.; Hwang, S.; Jeon, J. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. *Eng. Struct.* **2020**, *219*, No. 110927.
(36) Ogami, C.; Tsuji, Y.; Seki, H.; Kawano, H.; To, H.; Matsumoto, Y.; Hosono, H. An artificial neural network−pharmacokinetic model and its interpretation using Shapley additive explanations. *CPT: Pharmacometrics Syst. Pharmacol.* **2021**, 760−768.
(37) Zhang, Y.; Zhang, X.; Zhu, W. ANC: Attention Network for COVID-19 Explainable Diagnosis Based on Convolutional Block Attention Module. *Comp. Mod. Eng. Sci.* **2021**, *127*, 1037−1058.
(38) Sharma, V.; Dyreson, C. COVID-19 Screening Using Residual Attention Network an Artificial Intelligence Approach. 2020, arXiv:2006.16106 [eess], June 26, https://arxiv.org/abs/2006.16106 (accessed July 25, 2020).
(39) Grewal, M.; Srivastava, M. M.; Kumar, P.; Varadarajan, S. RADNET: Radiologist Level Accuracy Using Deep Learning for Hemorrhage Detection in CT Scans. 2017, arXiv.1710.04934 [cs], Oct. 13. https://arxiv.org/abs/1710.04934 (accessed August 5, 2020).
(40) Yeung, C.; Tsai, J. M.; King, B.; Pham, B.; Ho, D.; Liang, J.; Knight, M. W.; Raman, A. P. Multiplexed supercell metasurface design and optimization with tandem residual networks. *J. Nanophotonics* **2021**, *10*, 1133−1143.

https://doi.org/10.1021/acsphotonics.1c01636
ACS Photonics 2022, 9, 1577−1585