MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

Deep Neural Networks for Inverse Design of Nanophotonic Devices

Kojima, Keisuke; TaherSima, Mohammad; Koike-Akino, Toshiaki; Jha, Devesh; Tang, Yingheng; Wang, Ye; Parsons, Kieran

TR2021-001 January 23, 2021

Abstract

Deep learning is now playing a major role in designing photonic devices, including nanostructured photonics. In this paper, we investigate three models for designing nanophonic power splitters with multiple splitting ratios. The first model is a forward regression model, wherein the trained deep neural network (DNN) is used within the optimization loop. The second is an inverse regression model, in which the trained DNN constructs a structure with the desired target performance given as input. The third model is a generative network, which can randomly produce a series of optimized designs for a target performance. Focusing on the nanophotonic power splitters, we show how the devices can be designed by these three types of DNN models.

IEEE Journal of Lightwave Technology

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

Deep Neural Networks for Inverse Design of Nanophotonic Devices

Keisuke Kojima, *Senior Member, IEEE, Fellow, OSA,* Mohammad H. Tahersima, *Member, IEEE, Member, OSA*
Toshiaki Koike-Akino, *Senior Member, IEEE, Senior Member, OSA,* Devesh K. Jha, *Member, IEEE,*
Yingheng Tang, *Student Member, IEEE,* Ye Wang, *Senior Member, IEEE,*
and Kieran Parsons, *Senior Member, IEEE, Member, OSA*

(Invited Paper)

***Abstract*—Deep learning is now playing a major role in designing photonic devices, including nanostructured photonics. In this paper, we investigate three models for designing nanophonic power splitters with multiple splitting ratios. The first model is a forward regression model, wherein the trained deep neural network (DNN) is used within the optimization loop. The second is an inverse regression model, in which the trained DNN constructs a structure with the desired target performance given as input. The third model is a generative network, which can randomly produce a series of optimized designs for a target performance. Focusing on the nanophotonic power splitters, we show how the devices can be designed by these three types of DNN models.**

***Index Terms*—deep learning, neural networks, nanophotonics, integrated photonics, inverse design, generative neural networks.**

I. Introduction

Subwavelength nanostructured materials can be used to control incident electromagnetic fields into specific transmitted and reflected wavefronts. Recent nanophotonic devices have used such complex structures to enable novel applications in optics, integrated photonics, sensing, and computational metamaterials in a compact and energy-efficient form [1]–[12]. Recently, a compherensive review article [13] on silicon-based miniaturized structures including nanophotonic devices was published. Nevertheless, optimizing nanostructures with a large number of possible combinations of parameters is a challenging task in practice.

It is partly because finite-difference time-domain (FDTD) methods for computing the electromagnetic field distribution often require long simulation time, several minutes to hours depending on the size and the mesh of the photonic device, in order to estimate the optical transmission response. Even though there are several techniques to accelerate FDTD simulations [14], for designing nanostructures to achieve a target transmission profile, a large number of FDTD simulations needs to be carried out in practice, e.g., in a meta-heuristic manner.

Manuscript received July 17, 2020; revised xxx, xx, xxxx.

K. Kojima, M. Tahersima, T. Koike-Akino, D. Jha, Y. Tang, Y. Wang, and K. Parsons are with Mitsubishi Electric Research Laboratories (MERL), 201 Broadway, Cambridge, MA 02139, USA (e-mail: {kojima, tahersima, koike, jha, ytang, yewang, parsons}@merl.com).

M. Tahersima is currently with Cooley LLP.

Y. Tang is also with Electrical and Computer Engineering Dept., Purdue University, West Lafeyette, IN, USA

To resolve the issue, some optimization methods such as direct binary search (DBS) [15] and particle swarm optimization (PSO) [16] have been implemented to show some success. More recently, an artificial intelligence using neural networks (NN) has been integrated in an optimization process that can accelerate optimization by reducing the required number of numerical simulations. For example, in [17], [18], we demonstrated how NNs can help to streamline the design process, although the prediction accuracy was limited, due to the shallow layer structure.

Deep learning methods are representation-learning techniques obtained by composition of non-linear models that transform the representation at the previous level into a higher and slightly more abstract level in a hierarchical manner [19]. The main idea is that by cascading a large number of such transformations, nearly arbitrary complex functions can be learned in a data-driven fashion using deep neural networks (DNN) [20]. The huge success of deep learning in modeling complex input-output relationship has attracted attention from several scientific communities such as material discovery and design [21], [22], high energy physics [23], single molecule imaging [24], medical diagnosis [25], and particle physics [26]. The optical community also started working on signal processing and network automation of optical fiber communications [27]–[30], and inverse modeling for design of nanostructured optical components using DNN [31]–[36]. Also, there have been many reports on optical implementation of artificial neural networks [37]–[43]. DNNs can be used to predict an optical response of a topology (Forward Modeling) as well as to design a topology given a desired optical response (Inverse Modeling). Another class of DNNs for designing devices and materials is a generative DNN model [31], [44]–[46].

Fig. 1 shows the trend of machine learning applications in optical communications and devices community in the past decade, similar to the plot in [29]. Here, we plot the number of articles in each year according to Google Scholar search of the keyword combinations such as; "machine learning" + "optical device" or "deep learning" + "optical device." As we can see, machine learning has been already used for optical device community for many years and increasing by an annual rate of 125%. For deep learning applications, more rapid annual increase rate of 320% can be found in the past half decade. We also observe that the optical device community is behind

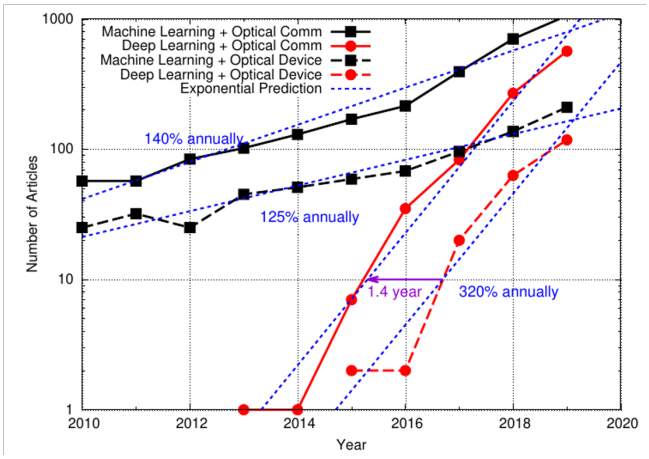

Fig. 1. Machine/deep learning trend in optical communication and optical devices applications (keyword hits on Google Scholar, excluding non-relevant ones).

the optical communication community by 1.4 years in terms of the absolute number of publications.

Inverse design of photonic structures was conventionally demonstrated using adjoint sensitivity analysis [47]–[50]. More recently, Liu *et al.* used a tandem NN architecture to learn non-unique electromagnetic scattering of alternating dielectric thin films with varying thickness [31]. Peurifoy *et al.* demonstrated DNNs to approximate light scattering of multilayer shell nanoparticles of SiO_2 and TiO_2 using a fully connected DNNs [34]. Asano and Noda provided an NN for prediction of the quality factor in two dimensional photonic crystals [36]. Hedge paired DNN with evolutionary algorithms to accelerate antireflection coating designs [51]. Banerji *et al.* used a reinforcement learning to design nanophotonic power splitters [43]. The design space for integrated photonic device is considerably larger than previously demonstrated optical scattering applications, that call for robust deeper networks such as Convolutional Neural Networks (CNN) [20].

Nanophotonic beam splitters have been widely used to equally divide the power into the output ports. Although an arbitrary split ratio can be applied in various applications such as signal monitoring, feedback circuits, or optical quantization [52], the design space is hardly explored due to design complexity. Tian *et al.* [53] demonstrated a silicon-on-insulator (SOI)-based 1×3 coupler with variable splitting ratio in a 15×15 μm^2 device footprint with 60 nm wavelength range and 80% transmission efficiency. Xu *et al.* [54] optimized positioning of squared etched pixels to achieve 80% efficiency for arbitrary ratio power dividers in 3.6×3.6 μm^2 device footprint. To design photonic power splitters with arbitrary splitting ratio, the designer often begins with an overall structure based on analytical models and fine tunes the structure using parameter sweep in numerical simulations.

This paper overviews three methods of DNNs applied to designing/optimizing nanophotonic power splitters. The general concept of using DNNs for designing/optimizing photonic devices may be readily applied to other types of devices. We demonstrate that by using deep learning methods, we could efficiently learn the design space of a broadband integrated photonic power splitter to realize a compact device.

II. Nanophotonic Power Splitters

We consider a nanostructured power splitter for integrated photonics platform throughout this paper. The goal of this device is to split the input power at a fixed splitting ratio towards two output ports, with a flat wavelength response with insertion loss as low as possible. Such a power splitter can be a building block of various types of photonic integrated circuits. To design the power splitter using DNN we chose a simple three port structure on a standard fully etched SOI platform. One input and two outputs ports having 0.5μm wide waveguides are connected using an adiabatic taper to the 2.25μm wide square power splitter design region with a connection width of 1.125μm as shown in Fig. 2.

We use Lumerical FDTD simulation to generate labeled data for training the network. We feed the training dataset into a DNN to learn the relationship between hole vectors (HV) and spectral response (SPEC) at each port. Our data contain several HV of size 20×20, labeled by its SPEC at the two output ports and reflection at the input port. Each hole is a circle with a maximum diameter of 72 nm that is easily fabricable using well-established lithography methods [50], [55]. In the HV, each hole can be represented by a binary state of 1 for etched ($n = n_{\text{Silicon}}$) and 0 for not etched ($n = n_{\text{Silica}}$). We consider variable-size holes in Section V, where the HV value is associated with hole area relative to the maximum size. In order to comply with the fabrication limit, we also put a constraint that the minimum hole diameter is 40 nm, corresponding to the HV value of $40^2/72^2 = 0.31$. Hence, if the HV value is below 0.31, there is no hole. When generating the training dataset using DBS on a computing cluster, Lumerical script is used to modify the topology of the device for the FDTD simulations. Python or Matlab automation is used to modify the device topology, when the DNN is included in the optimization loop or active learning. We use random patterned initial HVs and optimize them using heuristic optimization approaches for various optimization metrics to collect a diverse set of labeled training data for supervised learning.

This paper discusses three different methodologies to design the nanophotonic power splitters via deep learning. Firstly, we consider to use a forward model to predict SPEC given HV as shown in Fig. 3. Next, we discuss an inverse modeling to directly construct HV given a target SPEC. Finally, we address another method based on a stochastic generative model which implicitly integrates forward and inverse models.

Dispersive refractive indices of silicon and silica from literature [56] are used for all simulations for a broadband simulation in the range of 1.30μm–1.80μm, or 1.25μm–1.80μm. The fundamental transverse electric (TE) mode was used at input source and TE mode output power was recorded for transmission and reflection. We note that transverse magnetic (TM) mode output is lower than 10^{-5}, meaning there is little mode conversion due to the vertically symmetric device structure.

We use the open source machine learning framework of Tensorflow/Keras in Section III and PyTorch in Sections IV

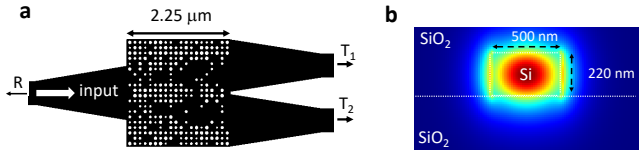

Fig. 2. Schematic of the SOI-based power splitter. a) Top view, where the size of the square region is $2.25 \times 2.25\mu\text{m}^2$, T_1 and T_2 denote the modal transmissions of output ports 1 and 2, and R denotes the reflection at input port. b) Cross-section of the input/output waveguide. By optimizing binary sequence of position of etch hole it is possible to adjust light propagation into either of the ports. In Section V, continuous variables are used to represent the hole sizes.

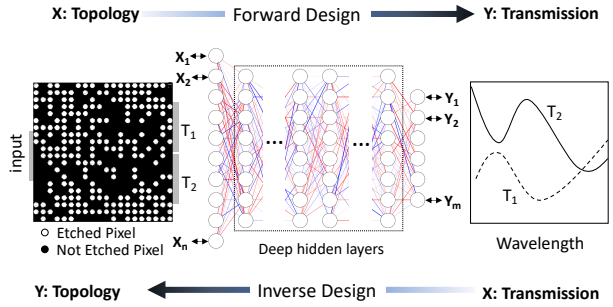

Fig. 3. We use DNN for forward and inverse modeling of nanophotonic devices. The DNN can take device topology design as input and spectral response of the metadevice as label or vice versa [57].

and V in the Python language to build and test our DNNs. For the forward modeling case, the DNN needs to be invoked prior to each FDTD run ($>$ 1,000 times for one optimization run), and Tensorflow/Keras has the advantage of shorter start-up time. On the other hand, for inverse and generative modeling, only a few start up is needed for the whole design process, and the choice of the framework is more like a historical or personal choice. The training data are generated by FDTD simulations using a high-performance computing cluster with more than 100 processors. The DNN training, testing, and subsequent FDTD simulations are conducted on a computer with a graphics processing unit (GPU) board Nvidia GeForce GTX Titan Z (12GB memory).

III. Deep Learning for Forward Modeling to Predict Optical Response

This section describes the forward regression model, using DNN to predict the transmission and reflection spectra, given the two-dimensional array (20×20) which corresponds to the binary image for the hole locations. We train a DNN to predict the transmission and reflection spectra vector which is 63-dimensional, consisting of spectral data (at 21 discrete wavelengths from 1300 to 1800 nm) for transmission at both port 1, port 2, and reflection at the input port. Once the DNN is trained, it is used as the predicted spectra within a DBS optimization loop. We initially used a fully-connected DNN with multiple layers where each layer has 100 neurons. The number of layers was considered as a hyperparameter which was optimized during the numerical experiments. However, we found that increasing the depth of the fully-connected

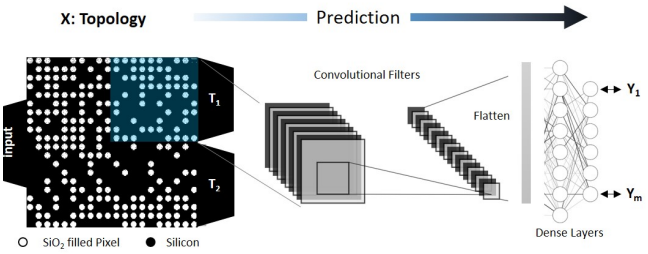

Fig. 4. Convolutional neural network (CNN) architecture [58].

DNNs did not improve the performance of the network. As one way to achieve better results for deeper networks, we used a residual network (ResNet) to train both the forward and inverse problems in [57].

Since the nanophotonic and photonic crystal designs are analogous to image processing and recognition problems, CNNs [20] are proposed to improve the forward prediction accuracy [36], [58]. The CNN architecture is shown in Fig. 4, where we treat the input data as two-dimentional images. Our network consists of three 2D convolutional layers followed by one fully connected layer. The three convolutional layers have 64, 128, and 256 filters each with kernel sizes of 10×10, 5×5, and 4×4, and strides of 1×1, 2×2, and 2×2, respectively, with the Rectified Linear Unit (ReLU) activation function. The fully connected layer has 256 neurons with the sigmoid activation function.

In the network training process, we minimized the mean-square error (MSE) as follows:

$$\begin{aligned}\text{MSE} = &\frac{1}{N}\sum_{\lambda=\lambda_{\min}}^{\lambda_{\max}}\Big[\big|T_1(\lambda)-T_1^\star(\lambda)\big|^2 \\ &+\big|T_2(\lambda)-T_2^\star(\lambda)\big|^2+\big|R(\lambda)-R^\star(\lambda)\big|^2\Big],\end{aligned} \tag{1}$$

where $T_1(\lambda)$, $T_2(\lambda)$ and $R(\lambda)$ denote the transmission at the port 1 and 2 and reflection at the input port at wavelegths λ, respectively, and N is the total number of spectral points. We let $[\cdot]^\star$ denote the corresponding target values. Here, we take a sum across uniformly sampled wavelengths λ from a minimum wavelength $\lambda_{\min}$ to a maximum $\lambda_{\max}$. In the DBS optimization process, we minimized the following loss function:

$$\text{Metric} = \big|T_1 - T_1^\star\big|^2 + \big|T_2 - T_2^\star\big|^2 + \alpha\big|R - R^\star\big|^2, \tag{2}$$

where T_1 and T_2 are the lowest transmitted power within the spectral range of 1300 nm and 1800 nm, while R is the largest reflection power. Specificaly we have

$$T_1 = \min_\lambda T_1(\lambda),\ T_2 = \min_\lambda T_2(\lambda),\ R = \max_\lambda R(\lambda), \tag{3}$$

where we chose $\alpha = 10$ as a weighting factor. We start with random patterns of 20×20 binary HVs, and generate training data by using the standard DBS, with target splitting ratios of 0 (0 : 10), 0.2 (2 : 8), 0.35 (3.5 : 6.5), and 0.5 (5 : 5), each for a few times. Here, typically 1,200 simulations are conducted for each DBS run through the course of the optimization. A total of 11,133 training data are generated by

combining all the intermediate data and removing duplicated patterns. Blue points in Fig. 5 show the training data, where the total transmission $(T_1 + T_2)$ is plotted against the splitting ratio $T_1/(T_1 + T_2)$. Due to the symmetry, the actual number of distinct training data is 22,266. We then try to design a power splitter with $T_1^\star = 0.27$ and $T_2^\star = 0.73$, starting from the three initial conditions indicated by the red circles in Fig. 5.

In the baseline DBS, we start with a lower left hole, and flip the binary value to assess the metric using the FDTD simulation. When the new metric is lower, we choose the new binary value, or if the metric is higher, we revert back to the original binary value. Then, we continue with sequential application of the same procedure to the other holes.

We use the forward DNN modeling to accelerate the DBS process. In the DNN-assisted DBS, we first train the DNN with 300 epochs using the initial training data, which takes about an hour on the computer with the GPU board. Next we hypothetically flip each of the 400 holes, and for each hole flipping, the output of the DNN is used as a prediction of the SPEC. We then pick the a new pattern with a flipped hole corresponding to the lowest metric and verify the spectra via FDTD simulation. When the actual FDTD result is better as expected, we keep the flipped value, unless otherwise, we try another hole corresponding to the next best metric and verify with FDTD simulation. We train with one epoch of the original training data and 15 epochs of the accumulated newly acquired data. This is essentially an accelerated DBS using metric values predicted by the DNN.

Fig. 6 shows a comparison between a conventional DBS and DNN-assisted DBS, plotting the metric as a function of the number of FDTD runs. It is confirmed that the DNN-assisted DBS optimizes the device structure much faster than the conventional DBS, especially at the early part of the optimization, leading to better device designs overall. For example, in the case of the DNN-assisted DBS with the first initial condition, the success rate of predicting a better HV is 45% in the first 20 iterations, and each improvement is significant. Between the 21th and 200th iterations, the success rate goes down to 11%. Note that each FDTD run takes about two minutes, while the additional training (active learning) for each FDTD run takes 20 seconds. Hence, there is an overhead of about 20% per FDTD. Fig. 7 shows an example of the optimized device via the DNN-assisted DBS. One can see that the spectral response is very flat across a wide range of wavelengths at least from 1300 to 1800 nm. The total transmittance greater than 87.5% was achieved with a low reflection below 0.5%. In this section, we verified that the forward modeling is effective to accelerate the optimization process to design high-performance devices even though it does not directly provide the optimized topology.

IV. Deep Learning for Inverse Modeling to Construct Device Topology

The forward modeling described in the previous section is generally good enough to predict the performance given device topology. However, it needs to be combined with an external optimization method such as DBS or other meta-heuristic algorithms to explore different device topology. In

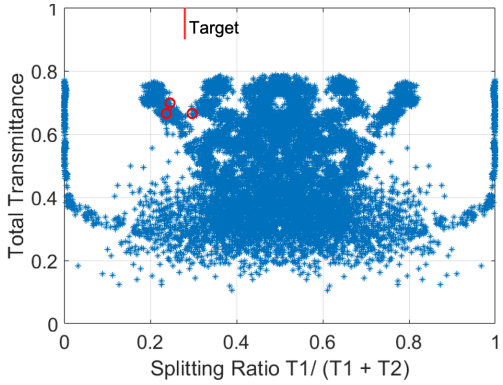

Fig. 5. Total transmittance of 22,000 training data as a function of the splitting ratio. The red circles indicate the three initial conditions used for training the forward regression model, and the line shows the target splitting.

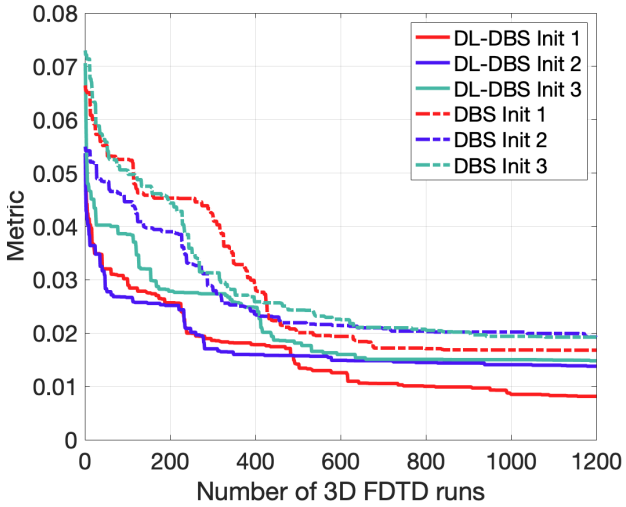

Fig. 6. Metric as a function of the number of 3D FDTD runs, for the conventional DBS and the DNN-DBS methods. Three different initial conditions are used.

contrast, the inverse modeling takes the spectral response as the input, and directly generates the device topology as the output, as shown in the bottom part of Fig. 3.

We here consider an inverse modeling DNN consisting of three fully connected layers and two 2D deconvolution layers. The dimension of the input is 63 which is the same as the SPEC data dimension. The three fully connected layers have 400, 800, and 1600 neurons, each with the ReLU activation function. The intermediate data dimension becomes 1600 which is later reshaped into $16 \times 10 \times 10$ to be fed in the deconvolution layers to generate the final reconstructed HV (a size of 20×20). These deconvolutional layers, not included in our original inverse modeling paper [57], can improve the performance.

The DNN is trained to predict HV in favor of minimizing the binary cross-entropy (BCE) loss as follows:

$$\text{BCE} = -\sum_{i=1}^{n} \Big[y_i \log x_i + (1 - y_i) \log(1 - x_i) \Big], \qquad (4)$$

where n is the maximum number of holes, y_i denotes the ith HV value of the training data, and x_i is the output of DNN

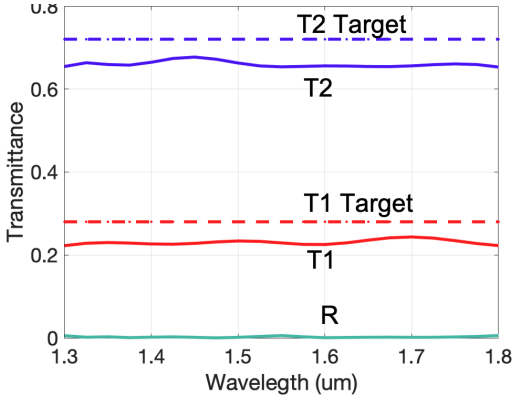

Fig. 7. Example of a $1:3$ power splitter designed by the forward regression model. For the range of 1300 nm and 1800 nm, the sum of the worst case T_1 and T_2 was 87.5%, while the worst case reflection was less than 0.5%.

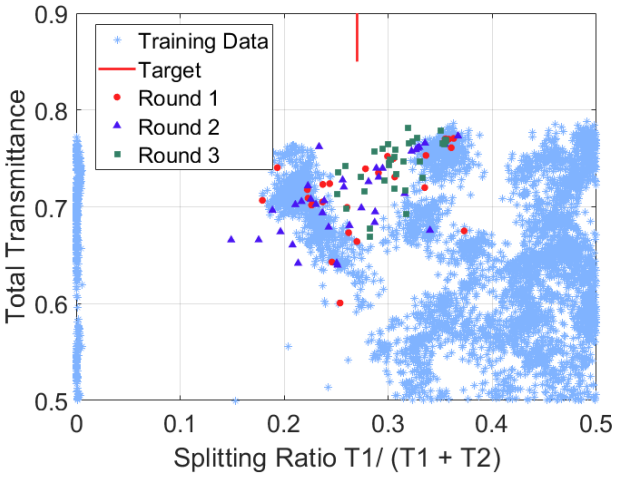

Fig. 8. Demonstration of inverse design for a power splitter with a splitting ratio of 0.27. The light blue asterisks, red circles, dark blue triangles, and green squares the training data, the first, the second, and the third round results, where the total transmittance is plotted against the splitting ratio. The red line denotes the target splitting ratio of 0.27.

as the corresponding estimate of y_i. The predicted HVs x_i can take any value from 0 to 1 from a Bernoulli distribution classifier. The classification tends to converge to either 0 or 1 as the loss reduces by increasing the number of training epochs. We train the network using the same data as used in Section III.

To test the generalization capabilities of the network, we investigate the design performance on arbitrary and unseen target case. A target splitting ratio of 0.27 with high total transmittance is taken as an example, where there is a gap in the training data, as shown in the red line in Fig. 8. In the design stage, the splitting ratios are chosen to be $0.23, 0.24, \ldots, 0.32$, and the total transmittances are $0.80, 0.82, 0.84$, and 0.86, and hence we have 40 combinations of input data set. We generate a reference table containing broadband constant transmission values for each port and use them as the input data batch for the inverse design DNN model. The output data are then quantized with a threshold of 0.5, and the binary sequence is then fed back into the FDTD solver. In the next stage, we run independent FDTD simulation to check the validity of the responses. The results are shown as the red circles in Fig. 8. Some data points fill the gap, while many are overlapping with the original training data region. This completes the first round. Then we add these 40 new data to the training data, retrain the network, and we repeat this active learning process. The results from the second round are shown as the dark blow triangles, and those from the third round are indicated by the green squares in Fig. 8.

As can be seen, this inverse modeling has the capability of generating the unseen results out of the training data, and the results further improve as more improved data points are added. In this process, the training takes about half an hour at each round on a high-performance computer with a GPU board. Once trained, the inverse design process is instantaneous (less than 1 s) for 40 devices.

V. Deep Learning for Generative Modeling to Produce Device Topology Candidates

The inverse modeling described in the previous section has a potential to generate an unseen good device structure achieving near the target SPEC. However, the generated topology by the inverse modeling DNN is usually deterministic given the desired SPEC input. Although we can still employ a stochastic variational sampling at the output nodes according to the Bernouli distribution, the variation of the topology candidates tends to be limited. It can significantly constrain the capability to explore different candidates of topology to achieve improved devices. Besides the forward and inverse modeling, another methodology based on generative modeling has been proposed, e.g., [31], [44]–[46]. The generative network can produce a series of improved designs from the training data, based on random number sampling, in a more explicit and systematic way. In the field of metamaterial researches, a few works have proposed to use such a generative network for the nanostructured pattern generation, including Generative Adversarial Networks (GAN) [31], [44] and Conditional Variational Autoencoder (CVAE) [45], [46]. In [46], we have applied generative deep learning models to integrated photonics for the first time, to our knowledge. We discuss the capability of the generative modeling method employing the improved version of CVAE in this section.

The CVAE can model the distribution of the splitter HVs associated with target spectrum characteristics. In our application, we use different hole sizes to express the appearances, which serve as the conditions of CVAE. With the variable-size holes, the generated patterns have a high potential to work better in the light guidance and make the generated devices more stable.

For device optimization, the trained decoder of the CVAE model is used with the desired condition along with a latent variable sampled from the normal distribution $\mathcal{N}(0,1)$, by which a series of HV topology candidates are generated. However, the latent variable in CVAE tends to be correlated with the condition SPEC s, which will result in degradation of the device performance for the pattern generation because random sampling at latent space can adversely impact the

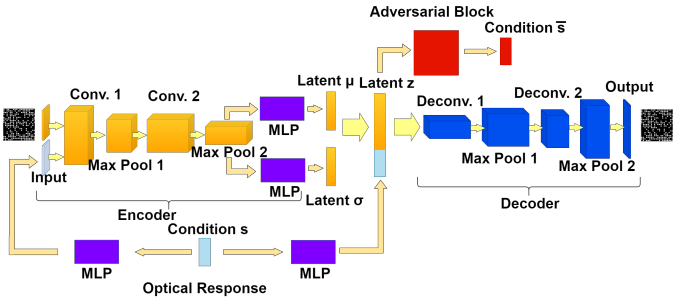

Fig. 9. The A-CVAE model structure, wherein the input is the two channels of 20×20 matrices, where the first channel is the HV of the device and the second channel is the decoded spectra data. The first convolutional layer has 16 channels with a kernel size of 3. The second convolutional layer has 32 channels with a kernel size of 3. The condition s is the transmission and reflection spectra obtained from the FDTD simulation to form a 3×21 matrix, and is fed to the decoder network through a fully connected (FC) layer. In the case of the standard CVAE, there is no adversarial block and the input from the condition s to the encoder network.

target spectra. To address such a problem of the conventional CVAE model, an adversarial block was introduced as shown in Fig. 9, where a separate branch to an adversary block is used for isolating the latent variable z from the nuisance variations s (the target SPEC) [59]–[61]. We call it Adversarial CVAE (A-CVAE).

Fig. 9 shows the detailed structure of the A-CVAE model. Here two convolutional layers are used both for the encoder and the decoder networks. The number of channels for the two layers are 16 and 32, and the max pooling stride is 2, after that there is one fully connected layer to reduce the latent variable to 63. The output is then concatenated with the SPEC performance data and feed them into the decoder. The validations are calculated by using the FDTD simulation to verify a figure of merit (FOM) of generated patterns, where the FOM is calculated by:

$$\begin{aligned}\mathsf{FOM} =& 1 - 10 \sum_{\lambda=\lambda_{\min}}^{\lambda_{\max}} \Big[\big|T_1(\lambda) - T_1^\star(\lambda)\big|^2 \\ &+ \big|T_2(\lambda) - T_2^\star(\lambda)\big|^2 + \alpha \big|R(\lambda) - R^\star(\lambda)\big|^2 \Big],\end{aligned} \tag{5}$$

where $\alpha = 4$ is used as a weighting factor to balance between the contributions from transmission and the reflection. As the SPEC performance approaches the target, the FOM increases towards 1, in which case we obtain an ideal power splitter without excess loss for $R^\star(\lambda) = 0$ and $T_1^\star(\lambda) + T_2^\star(\lambda) = 1$.

In A-CVAE, the latent z variable will be fed into an adversarial block to estimate the SPEC $\bar{s} := (\bar{s}_1, \ldots, \bar{s}_n)$. The loss function for the A-CVAE model is shown as follows:

$$\begin{aligned}\mathsf{Loss} =& -\sum_{i=1}^{n} \Big[y_i \log x_i + (1 - y_i) \log(1 - x_i) \Big] \\ &+ \frac{1}{2} \sum_{j=1}^{J} \Big[\mu_{zj}^2 + \sigma_{zj}^2 - \log(\sigma_{zj}^2) - 1 \Big] - \frac{\beta}{n} \sum_{i=1}^{n} (s_i - \bar{s}_i)^2.\end{aligned} \tag{6}$$

The loss function has three parts. The first term is the VAE reconstruction loss in the BCE criterion and the second term is the Kullback–Leibler (KL) divergence. The last term is a regularization term which is the MSE loss of the adversarial block. Since the condition information contained in the latent variable z shall be minimized, the MSE loss between s and $\bar{s}$ needs to be maximized. A complete update of the network generally requires alternating updates in two iteration cycles. The first iteration is used to backpropagate and update the CVAE model based on the loss function in (6). The second iteration is used to update the adversarial block solely based on the MSE loss between s and $\bar{s}$. The total training time using a computer with a GPU board is around 5 minutes.

After training the proposed machine learning model herein, we test the A-CVAE model by sampling random latent z to generate nanopatterned power splitter devices according to the generated HVs. In order to verify effectiveness of the generative model, we choose four different types of devices with different splitting ratios ($5:5$, $6:4$, $7:3$, $8:2$). Fig. 10 shows the comparison of the performance among the devices generated by the CVAE model and that by the A-CVAE. The FOM is calculated for 20 randomly generated devices from the trained CVAE and A-CVAE models. This figure shows that the conventional CVAE model can generally learn the distribution of the data, but it cannot beat the training data in terms of performance. With the help of the adversarial censoring, the generated devices generally have a better performance than the training data.

Four different splitting ratios are used as a target value to test the model performance (marked with dashed lines). The devices generated by A-CVAE model can fit the target splitting ratio better with excellent total transmission. The average FOMs for the CVAE and A-CVAE models are 0.771 and 0.888, respectively.

As the above results show, generative deep learning models can generate a series of improved results based on the statistical characteristics of the training data. In particular, we demonstrated that the adversarial censoring further improves the performance with high stability. Note that part of the good performances of CVAE and A-CVAE come from the fact that variable hole sizes are adopted, which could not be done in the case of DBS.

VI. Nanophotonic Power Splitter Experiment

In order to verify the validity of the simulations, we prototyped power splitters using a commercial multip project wafer (MPW) service by Applied Nanotools Inc. The wafer is processed through a standard 220 nm SOI process with SiO_2 cladding. Direct electron beam writing is used for lithography. In this particular design, the size of the square region is $2.6 \times 2.6 \mu m^2$ and the binary hole size is 90 nm in diameter, from a historical reason.

The scanning electron microscope (SEM) image of a $1:3$ beam splitter is shown in Fig. 11. The device is designed by the inverse model as described earlier using binary variables, and the holes are clearly defined.

The measurement was conducted with grating couplers, using amplified spontaneous emission (ASE) from an Erbium-doped fiber amplifier as a light source. The transmittance is

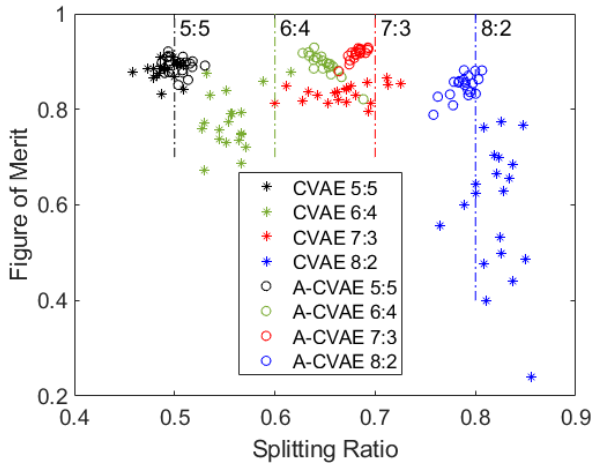

Fig. 10. Figure of merit as a function of the splitting ratio. Each color corresponds to different target splitting ratios.

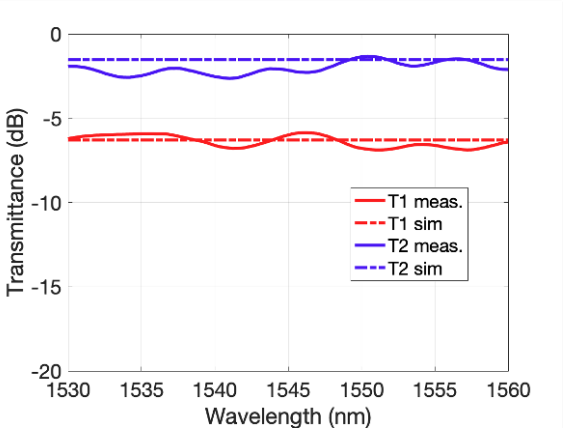

Fig. 12. Solid and dashed lines show the measured and the simulated transmittance, respectively, of a 1 : 3 power splitter. The red and blue lines show transmittance to output ports 1 and 2, respectively.

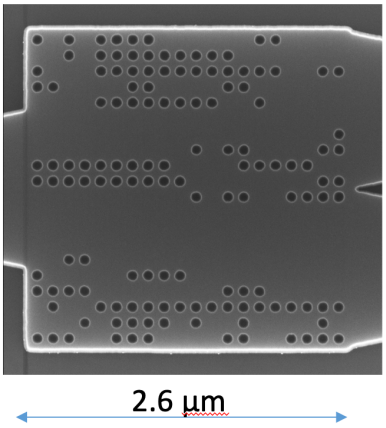

Fig. 11. An SEM image of a prototyped 1 : 3 power splitter.

defined by the ratio of the transmitted power of the device under test and a reference device having the same designs of grating couplers. The measurement wavelength range is limited by the grating couplers and the ASE source. Solid lines in Fig. 12 show the measured transmittance, and the dashed lines show the transmittance obtained from simulations. The measured values agree well with the simulated values, validating our simulation and optimization results. The cause of the small ripples in the measured transmittance is not clear at this moment.

VII. Discussion

DNNs can be used to take device structure data (shape, depth, and permittivity) to predict the optical response of the nanostructure in the forward modeling framework. In this case DNNs can be used as a viable counterpart for fast approximation of the optical response, in comparison to the use of computationally heavy FDTD simulations. Another way to use DNNs is taking an optical response to provide the user with an approximate solution of nanostructure in the inverse modeling framework, which does not rely on external meta-heuristic optimizers unlike forward modeling. Although the DNN initially needs a sufficient amount of data for the training purpose, it is possible to process several heuristic optimization metrics in parallel on a computing cluster to speed up generating the training data. We can design the nanostructured geometry in a fraction of second once the network is trained to represent the topology as optical response and vice versa. The generative model, on the other hand, implicitly integrates forward and inverse models. Once the network is trained, the generation of new designs takes practically little time. The use of adversarial censoring further improves the design capability. Overall, DNNs have the capability of learning the training data with a high generalizability.

The adjoint method is widely used in the inverse design of nanophotonic devices [8], [48], [62]. Given an ideal initial condition for parameters, the optimization process can be efficiently done in a small number of iterations (tens of FDTD simulations). However, the initial condition needs to be carefully chosen in order to obtain the optimal result or many initial conditions need to be tried. DNNs are very different. They are trained from a library (training data) of FDTD simulation results, which may come from previous imperfect optimization results with multiple target conditions (splitting ratio and bandwidth, in our case). Then the DNNs, especially the inverse and the generative models, will try to learn/generalize from the library, and generate a series of improved results for a given condition. Once the model is trained, inverse designs for multiple conditions can be generated in almost no time. Further FDTD simulations are not required, and can be used only for verification purposes.

Table I provides a comparison of power splitters using different optimization methods, including a more conventional y-junction device optimized by classical particle swarm optimization (PSO) [63], and nanophotonic splitters designed by the adjoint method [8], [62] and the fast search method (FDM) [54]. Our work has the most broad bandwidth with very small device footprint and low insertion loss.

TABLE I
COMPARISON OF SIMULATION RESULTS FOR PHOTONICS POWER SPLITTERS USING DIFFERENT OPTIMIZATION METHODS

Split ratio	Insertion loss	Bandwidth	Footprint	Method	Reference	Setup time	Computational time
$1:1$	0.09 dB	100 nm	$2 \times 2\ \mu m^2$	Adjoint	Lalau-Keraly et. al. [8]		NA
$1:1$	0.32 dB	40 nm	$2.6 \times 2.6\ \mu m^2$	Adjoint	Wang et. al. [62]		1.2 hr
$1:1$	0.13 dB	80 nm	$1.2 \times 2\ \mu m^2$	PSO	Zhang et. al. [63]		NA
$4:6$	1 dB (measured)	30 nm	$3.6 \times 3.6\ \mu m^2$	FSM	Xu et. al. [54]		120 hr
$4:6$	0.65 dB	550 nm	$2.25 \times 2.25\ \mu m^2$	A-CVAE	Our work [64]	~ 89 hr*	~ 5 min**
$3:7$	0.51 dB	550 nm	$2.25 \times 2.25\ \mu m^2$	A-CVAE	Our work [64]	~ 89 hr*	~ 5 min**

* Here the time is mostly for the data collection. The data collection is ~ 20 s $\times\ 16000 = 89$ hr.
** The training time is ~ 5 min and the new design generation time is ~ 5 s.

VIII. CONCLUSION

To design complicated nanophotonic devices with hundreds of parameters, a sophisticated design algorithm is necessary, and deep learning offers a promising solution. We demonstrated three different types of DNN methodologies, i.e., forward modeling, inverse modeling, and generative modeling, to design nanophotonic power splitters.

The forward modeling uses a DNN to predict the SPEC performances given the device topology. The DNN is integrated with an optimization method, to reduce the requirement of computationally intensive FDTD simulations. As more data are accumulated for online training, the prediction accuracy further improves.

The inverse modeling uses a DNN to directly generate the device topology given a target performance. By supplying a series of modified SPEC performance, a series of good device structure candidates is generated. This can avoid the use of external optimizer methods.

For the generative model, we used a CVAE with adversarial censoring. Once trained, the CVAE can generate a series of improved device structures given SPEC performance data as a condition along with random variables sampled from the normal distribution. We confirmed that the adversarial censoring significantly improves the design capability.

Further improvement of DNN methodologies will create a blackbox trained from a vast amount of data and generate new designs on-demand for a broad range of target conditions. The area of deep learning-assisted photonic device design has been quickly expanding, and these modeling methods will play important roles in the advancement of photonic device design activities.

REFERENCES

[1] X. Ni, Z. J. Wong, M. Mrejen, Y. Wang, and X. Zhang, "An ultrathin invisibility skin cloak for visible light," *Science*, vol. 349, no. 6254, pp. 1310–1314, 2015.

[2] A. Alù and N. Engheta, "Achieving transparency with plasmonic and metamaterial coatings," *Physical Review E*, vol. 72, no. 1, p. 016623, 2005.

[3] F. Monticone, N. M. Estakhri, and A. Alù, "Full control of nanoscale optical transmission with a composite metascreen," *Phys. Rev. Lett.*, vol. 110, p. 203903, May 2013. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.110.203903

[4] E. Arbabi, A. Arbabi, S. M. Kamali, Y. Horie, and A. Faraon, "Multi-wavelength polarization-insensitive lenses based on dielectric metasurfaces with meta-molecules," *Optica*, vol. 3, no. 6, pp. 628–633, 2016.

[5] M. Khorasaninejad, W. T. Chen, R. C. Devlin, J. Oh, A. Y. Zhu, and F. Capasso, "Metalenses at visible wavelengths: Diffraction-limited focusing and subwavelength resolution imaging," *Science*, vol. 352, no. 6290, pp. 1190–1194, 2016.

[6] A. Krasnok, M. Tymchenko, and A. Alù, "Nonlinear metasurfaces: A paradigm shift in nonlinear optics," *Materials Today*, vol. 21, no. 1, pp. 8 – 21, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S136970211730233X

[7] A. K. Azad, W. J. Kort-Kamp, M. Sykora, N. R. Weisse-Bernstein, T. S. Luk, A. J. Taylor, D. A. Dalvit, and H.-T. Chen, "Metasurface broadband solar absorber," *Scientific Reports*, vol. 6, no. 20347, 2016.

[8] C. M. Lalau-Keraly, S. Bhargava, O. D. Miller, and E. Yablonovitch, "Adjoint shape optimization applied to electromagnetic design," *Optics express*, vol. 21, no. 18, pp. 21 693–21 701, 2013.

[9] A. Motayed, G. Aluri, A. V. Davydov, M. V. Rao, V. P. Oleshko, R. Bajpai, M. E. Zaghloul, B. Thomson, B. Wen, T. Xie *et al.*, "Highly selective nanostructure sensors and methods of detecting target analytes," May 29 2018, uS Patent 9,983,183.

[10] A. Silva, F. Monticone, G. Castaldi, V. Galdi, A. Alù, and N. Engheta, "Performing mathematical operations with metamaterials," *Science*, vol. 343, no. 6167, pp. 160–163, 2014.

[11] Z. Chu, Y. Liu, J. Sheng, L. Wang, J. Du, and K. Xu, "On-chip optical attenuators designed by artifical neural networks," in *2018 Asia Communications and Photonics Conference (ACP)*. IEEE, 2018, pp. 1–3.

[12] Z. Liu, X. Liu, Z. Xiao, C. Lu, H.-Q. Wang, Y. Wu, X. Hu, Y.-C. Liu, H. Zhang, and X. Zhang, "Integrated nanophotonic wavelength router based on an intelligent algorithm," *Optica*, vol. 6, no. 10, pp. 1367–1373, 2019.

[13] M. Teng, A. Honardoost, Y. Alahmadi, S. S. Polkoo, K. Kojima, H. Wen, C. K. Renshaw, P. LiKamWa, G. Li, S. Fathpour *et al.*, "Miniaturized silicon photonics devices for integrated optical signal processors," *Journal of Lightwave Technology*, 2019.

[14] L. Su, D. Vercruysse, J. Skarda, N. V. Sapra, J. A. Petykiewicz, and J. Vučković, "Nanophotonic inverse design with SPINS: Software architecture and practical considerations," *Applied Physics Reviews*, vol. 7, no. 1, p. 011407, 2020.

[15] B. Shen, P. Wang, R. Polson, and R. Menon, "An integrated-nanophotonics polarization beamsplitter with $2.4 \times 2.4\ \mu m^2$ footprint," *Nature Photonics volume 9, 378–382*, 2015.

[16] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Proceedings of ICNN'95 - International Conference on Neural Networks*, 1995.

[17] K. Kojima, B. Wang, U. Kamilov, T. Koike-Akino, and K. Parsons, "Acceleration of FDTD-based inverse design using a neural network approach," in *Integrated Photonics Research, Silicon and Nanophotonics*. Optical Society of America, 2017, pp. ITu1A–4.

[18] M. Teng, K. Kojima, T. Koike-Akino, B. Wang, C. Lin, and K. Parsons, "Broadband SOI mode order converter based on topology optimization," in *2018 Optical Fiber Communications Conference and Exposition (OFC)*, March 2018, pp. 1–3.

[19] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Communications of the ACM*, vol. 60, no. 6. New York, NY, USA: ACM, May 2017, pp. 84–90. [Online]. Available: http://doi.acm.org/10.1145/3065386

[21] J. Ghaboussi, J. Garrett Jr, and X. Wu, "Knowledge-based modeling of material behavior with neural networks," *Journal of Engineering Mechanics*, vol. 117, no. 1, pp. 132–153, 1991.

[22] B. Sanchez-Lengeling and A. Aspuru-Guzik, "Inverse molecular design using machine learning: Generative models for matter engineering," *Science*, vol. 361, no. 6400, pp. 360–365, 2018.

[23] P. Baldi, P. Sadowski, and D. Whiteson, "Searching for exotic particles in high-energy physics with deep learning," *Nature Communications*, vol. 5, no. 4308, 2014.

[24] M. Yasui, M. Hiroshima, J. Kozuka, Y. Sako, and M. Ueda, "Automated single-molecule imaging in living cells," *Nature communications*, vol. 9, no. 1, p. 3061, 2018.

[25] Y. Jun, T. Eo, T. Kim, H. Shin, D. Hwang, S. H. Bae, Y. W. Park, H.-J. Lee, B. W. Choi, and S. S. Ahn, "Deep-learned 3D black-blood imaging using automatic labelling technique and 3d convolutional neural networks for detecting metastatic brain tumors," *Scientific reports*, vol. 8, no. 1, 2018.

[26] A. Radovic, M. Williams, D. Rousseau, M. Kagan, D. Bonacorsi, A. Himmel, A. Aurisano, K. Terao, and T. Wongjirad, "Machine learning at the energy and intensity frontiers of particle physics," *Nature*, vol. 560, no. 7716, 2018.

[27] T. Koike-Akino, "Perspective of statistical learning for nonlinear equalization in coherent optical communications," in *Advanced Photonics for Communications*. Optical Society of America, 2014, p. ST2D.2. [Online]. Available: http://www.osapublishing.org/abstract.cfm?URI=SPPCom-2014-ST2D.2

[28] D. Zibar, M. Piels, R. Jones, and C. G. Schäeffer, "Machine learning techniques in optical communication," *Journal of Lightwave Technology*, vol. 34, no. 6, pp. 1442–1452, 2015.

[29] T. Koike-Akino, Y. Wang, D. S. Millar, K. Kojima, and K. Parsons, "Neural turbo equalization to mitigate fiber nonlinearity," in *European Conference on Optical Communication (ECOC)*, 2019, p. Tu.1.B.1.

[30] D. Rafique and L. Velasco, "Machine learning for network automation: Overview, architecture, and applications (invited tutorial)," *J. Opt. Commun. Netw.*, vol. 10, no. 10, pp. D126–D143, Oct 2018. [Online]. Available: http://jocn.osa.org/abstract.cfm?URI=jocn-10-10-D126

[31] D. Liu, Y. Tan, E. Khoram, and Z. Yu, "Training deep neural networks for the inverse design of nanophotonic structures," *ACS Photonics*, vol. 5, no. 4, pp. 1365–1369, 2018.

[32] W. Ma, F. Cheng, and Y. Liu, "Deep-learning enabled on-demand design of chiral metamaterials," *ACS Nano*, vol. 12, no. 6, pp. 6326–6334, 2018.

[33] I. Malkiel, M. Mrejen, A. Nagler, U. Arieli, L. Wolf, and H. Suchowski, "Deep learning for the design of nano-photonic structures," in *2018 IEEE International Conference on Computational Photography (ICCP)*, May 2018, pp. 1–14.

[34] J. Peurifoy, Y. Shen, L. Jing, Y. Yang, F. Cano-Renteria, B. G. DeLacy, J. D. Joannopoulos, M. Tegmark, and M. Soljačić, "Nanophotonic particle simulation and inverse design using artificial neural networks," *Science Advances*, vol. 4, no. 6, 2018. [Online]. Available: http://advances.sciencemag.org/content/4/6/eaar4206

[35] Y. Sun, Z. Xia, and U. S. Kamilov, "Efficient and accurate inversion of multiple scattering with deep learning," *Optics Express*, vol. 26, no. 11, pp. 14 678–14 688, 2018.

[36] T. Asano and S. Noda, "Optimization of photonic crystal nanocavities based on deep learning," *Optics express*, vol. 26, no. 25, pp. 32 704–32 717, 2018.

[37] J. Ohta, K. Kojima, N. Y, T. Shuichi, and K. Kazuo, "Optical neurochip based on a three-layered feed-forward model," *Optics letters*, vol. 15, no. 23, pp. 1362–1364, 1990.

[38] A. N. Tait, T. F. Lima, E. Zhou, A. X. Wu, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, "Neuromorphic photonic networks using silicon photonic weight banks," *Scientific Reports*, vol. 7, no. 1, 2017.

[39] A. Mehrabian, Y. Al-Kabani, V. J. Sorger, and T. El-Ghazawi, "PCNNA: A photonic convolutional neural network accelerator," *arXiv preprint arXiv:1807.08792*, 2018.

[40] X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, "All-optical machine learning using diffractive deep neural networks," *Science*, 2018. [Online]. Available: http://science.sciencemag.org/content/early/2018/07/25/science.aat8084

[41] J. Chiles, S. M. Buckley, S. W. Nam, R. P. Mirin, and J. M. Shainline, "Design, fabrication, and metrology of 10×100 multi-planar integrated photonic routing manifolds for neural networks," *APL Photonics*, vol. 3, no. 10, 2018.

[42] T. W. Hughes, M. Minkov, Y. Shi, and S. Fan, "Training of photonic neural networks through in situ backpropagation and gradient measurement," *Optica*, vol. 5, no. 7, pp. 864–871, 2018.

[43] S. Banerji, A. Majumder, A. Hamrick, R. Menon, and B. Sensale-Rodriguez, "Machine learning enables design of on-chip integrated silicon T-junctions with footprint of 1.2μm $\times$ 1.2μm," *arXiv preprint arXiv:2004.11134*, 2020.

[44] S. An, B. Zheng, H. Tang, M. Y. Shalaginov, L. Zhou, H. Li, T. Gu, J. Hu, C. Fowler, and H. Zhang, "Multifunctional metasurface design with a generative adversarial network," *arXiv preprint arXiv:1908.04851*, 2019.

[45] W. Ma, F. Cheng, Y. Xu, Q. Wen, and Y. Liu, "Probabilistic representation and inverse design of metamaterials based on a deep generative model with semi-supervised learning strategy," *arXiv preprint arXiv:1901.10819*, 2019.

[46] Y. Tang, K. Kojima, T. Koike-Akino, Y. Wang, P. Wu, M. H. Tahersima, D. Jha, K. Parsons, and M. Qi, "Generative deep learning model for a multi-level nano-optic broadband power splitter," in *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, March 2020, p. Th1A.1.

[47] Y. Cao, S. Li, L. Petzold, and R. Serban, "Adjoint sensitivity analysis for differential-algebraic equations: The adjoint dae system and its numerical solution," *SIAM Journal on Scientific Computing*, vol. 24, no. 3, pp. 1076–1089, 2003.

[48] A. Y. Piggott, J. Lu, K. G. Lagoudakis, J. Petykiewicz, T. M. Babinec, and J. Vučković, "Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer," *Nature Photonics*, vol. 9, no. 6, pp. 374–377, 2015.

[49] L. H. Frandsen and O. Sigmund, "Inverse design engineering of all-silicon polarization beam splitters," in *Photonic and Phononic Properties of Engineered Nanostructures VI*, vol. 9756. International Society for Optics and Photonics, 2016, p. 97560Y.

[50] A. Y. Piggott, J. Petykiewicz, L. Su, and J. Vučković, "Fabrication-constrained nanophotonic inverse design," *Scientific Reports*, vol. 7, no. 1, p. 1786, 2017.

[51] R. S. Hegde, "Photonics inverse design: Pairing deep neural networks with evolutionary algorithms," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 1, pp. 1–8, 2019.

[52] Z. Kang, X. Zhang, J. Yuan, X. Sang, Q. Wu, G. Farrell, and C. Yu, "Resolution-enhanced all-optical analog-to-digital converter employing cascade optical quantization operation," *Optics Express*, vol. 22, no. 18, pp. 21 441–21 453, 2014.

[53] Y. Tian, J. Qiu, M. Yu, Z. Huang, Y. Qiao, Z. Dong, and J. Wu, "Broadband 1×3 couplers with variable splitting ratio using cascaded step-size MMI," *IEEE Photonics Journal*, vol. 10, no. 3, pp. 1–8, 2018.

[54] K. Xu, L. Liu, X. Wen, W. Sun, N. Zhang, N. Yi, S. Sun, S. Xiao, and Q. Song, "Integrated photonic power divider with arbitrary power ratios," *Optics Letters*, vol. 42, no. 4, pp. 855–858, 2017.

[55] L. Lu, M. Zhang, F. Zhou, and D. Liu, "An ultra-compact colorless 50: 50 coupler based on PhC-like metamaterial structure," in *Optical Fiber Communications Conference and Exhibition (OFC), 2016*. IEEE, 2016, pp. 1–3.

[56] E. D. Palik, "Handbook of optical constants of solids," in *Handbook of optical constants of solids*. Elsevier, 1997, pp. 429–443.

[57] M. H. Tahersima, K. Kojima, T. Koike-Akino, D. Jha, B. Wang, C. Lin, and K. Parsons, "Deep neural network inverse design of integrated photonic power splitters," *Scientific Reports*, vol. 9, no. 1, p. 1368, 2019.

[58] ——, "Nanostructured photonic power splitter design via convolutional neural networks," in *CLEO: Science and Innovations*. Optical Society of America, 2019, pp. SW4J–6.

[59] G. Lample, N. Zeghidour, N. Usunier, A. Bordes, L. Denoyer, and M. Ranzato, "Fader networks: Manipulating images by sliding attributes," in *Advances in neural information processing systems*, 2017, pp. 5967–5976.

[60] Y. Wang, T. Koike-Akino, and D. Erdogmus, "Invariant representations from adversarially censored autoencoders," *arXiv preprint arXiv:1805.08097*, 2018.

[61] O. Özdenizci, Y. Wang, T. Koike-Akino, and D. Erdoğmuş, "Transfer learning in brain-computer interfaces with adversarial variational autoencoders," in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2019, pp. 207–210.

[62] K. Wang, X. Ren, W. Chang, L. Liu, D. Liu, and M. Zhang, "Inverse design of digital nanophotonic devices using the adjoint method," *Photonics Research*, vol. 8, pp. 528–533, 2020.

[63] Y. Zhang, S. Yang, A. E.-J. Lim, G.-Q. Lo, C. Galland, T. Baehr-Jones, and M. Hochberg, "A compact and low loss y-junction for submicron silicon waveguide," *Opt. Express*, vol. 21, pp. 1310–1316, Jan 2013.

[64] Y. Tang, K. Kojima, T. Koike-Akino, Y. Wang, P. Wu, Y. Xie, M. H. Tahersima, D. K. Jha, K. Parsons, and M. Qi, "Generative deep learning model for inverse design of integrated nanophotonic devices," *Laser & Photonics Reviews*, p. 2000287, 2020.

Keisuke Kojima (S'82–M'84–SM'13) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Tokyo, Tokyo, Japan, in 1981, 1983, and 1990, respectively. He also received the M.S. degree from the University of California, Berkeley, CA, USA, in 1982. He worked for eight years at the Central Research Laboratory, Mitsubishi Electric Corp., from 1983 on the research of narrow linewidth DFB and DBR lasers, and optical neural networks. He spent nine years at AT&T/Lucent Bell Laboratories on the R&D of uncooled Fabry–Perot and DFB lasers, vertical-cavity surface-emitting lasers, passive optical network systems, and metro optical systems. He also worked at Agere Systems, Denselight Semiconductors, and TriQuint Semiconductors on optical devices and modules, and optical systems testbed. He has been with Mitsubishi Electric Research Laboratories, Cambridge, MA, USA, since 2005, where he is currently working on the design of photonic-integrated circuits and coherent optical systems using deep learning, as a Distinguished Research Scientist. He has more than 200 publications in journals and conference proceedings. He is a Fellow of the Optical Society of America, Associate Editor of IEEE/OSA Journal of Lightwave Technology, and Chair of the Boston Chapter of IEEE Photonics Society.

Toshiaki Koike-Akino (M'05–SM'11) received the B.S. degree in electrical and electronics engineering, M.S. and Ph.D. degrees in communications and computer engineering from Kyoto University, Kyoto, Japan, in 2002, 2003, and 2005, respectively. During 2006–2010 he was a Postdoctoral Researcher at Harvard University, and joined Mitsubishi Electric Research Laboratories, Cambridge, MA, USA, in 2010. His research interests include digital signal processing for data communications and sensing. He received the YRP Encouragement Award 2005, the 21st TELECOM System Technology Award, the 2008 Ericsson Young Scientist Award, the IEEE GLOBECOM'08 Best Paper Award in Wireless Communications Symposium, the 24th TELECOM System Technology Encouragement Award, and the IEEE GLOBECOM'09 Best Paper Award in Wireless Communications Symposium.

Devesh K. Jha (M'16) is currently a Principal Research Scientist at Mitsubishi Electric Research Laboratories (MERL) in Cambridge, MA, USA. He received PhD in Mechanical Engineering from Penn State in December 2016. He received M.S. degrees in Mechanical Engineering and Mathematics also from Penn State. His research interests are in the areas of Machine Learning, Robotics and Deep Learning. He is a recipient of several best paper awards including the Kalman Best Paper Award 2019 from the American Society of Mechanical Engineers (ASME) Dynamic Systems and Control Division.

Mohammad H. Tahersima (M'12) received the B.S. and Ph.D. degrees in Electronics from the Kyoto Institute of Technology, Kyoto, Japan and George Washington University, Washington, DC, USA, in 2012 and 2018, respectively. He was with Mitsubishi Electric Research Laboratories, Cambridge, MA, USA in 2018 and 2019 where he worked on applications of machine learning in photonic integrated circuit design. His research interests include optical energy conversion, light-matter interaction at low-dimensional materials, and artificial intelligence. He is also a member of the Optical Society of America, the American Association for the Advancement of Science, and the International Society for Optics and Photonics.

Yingheng Tang received the B.S. degree in electrical engineering from Purdue University in 2015, where he is currently pursuing the Ph.D. degree in electrical and computer engineering. His research interests include silicon photonics and machine learning assisted optimization for photonic device design.

Ye Wang (S'06–M'12–SM'19) received the B.S. degree in Electrical and Computer Engineering from Worcester Polytechnic Institute, Worcester, MA, in 2005, and the M.S. and Ph.D. degrees in Electrical and Computer Engineering from Boston University, Boston, MA, in 2009 and 2011, respectively. In 2012, he joined MERL, Cambridge, MA, where he is now a Principal Research Scientist. He has also contributed to standardization activity in the INCITS-M1 US National Body for Biometrics Standards and the ISO/IEC-JTC1/SC37 Subcommittee on Biometrics, where he served as a project editor. Prior to joining MERL, he was a postdoctoral researcher in the Department of Electrical and Computer Engineering at Boston University, and had worked as a Firmware Engineer at AgaMatrix, Salem, NH. His technical interests include information theory, machine learning, signal processing, communication, and data privacy/security.

Kieran Parsons (M'07–SM'09) received the B.Eng. and Ph.D. degrees in electronic and communications engineering from the University of Bristol, Bristol, U.K., in 1992 and 1996, respectively. During 1997–2002, he was with Nortel Networks, Ottawa, Canada, where he worked on wireless and long-haul optical communications system architecture and design. From 2004–2006, he worked on carrier-grade mesh WiFi RF system design at BelAir Networks (now part of Ericsson) and from 2006–2009 on 10-G PHY device development with Applied Micro, both in Kanata, Canada. In 2009, he joined Mitsubishi Electric Research Laboratories, Cambridge, MA, USA, where where he is currently Senior Team Leader and Senior Principal Research Scientist. His research interests include digital signal processing and coherent optical transmission systems.