

Research Article

Simei Mao, Lirong Cheng, Houyu Chen, Xuanyi Liu, Zihan Geng, Qian Li and Hongyan Fu*

Multi-task topology optimization of photonic devices in low-dimensional Fourier domain via deep learning

<https://doi.org/10.1515/nanoph-2022-0361>

Received June 21, 2022; accepted September 27, 2022;

published online October 4, 2022

Abstract: Silicon photonics enables compact integrated photonic devices with versatile functionalities and mass manufacturing capability. However, the optimization of high-performance free-form optical devices is still challenging due to the complex light-matter interaction involved that requires time-consuming electromagnetic simulations. This problem becomes even more prominent when multiple devices are required, typically requiring separate iterative optimizations. To facilitate multi-task inverse design, we propose a topology optimization method based on deep neural network (DNN) in low-dimensional Fourier domain. The DNN takes target optical responses as inputs and predicts low-frequency Fourier components, which are then utilized to reconstruct device geometries. Removing high-frequency components for reduced design degree-of-freedom (DOF) helps control minimal features and speed up training. For demonstration, the proposed method is utilized for wavelength filter design. The trained DNN can design multiple filters instantly and concurrently with high accuracy. Totally

different targets can also be further optimized through transfer learning on existing network with greatly reduced optimization rounds. Our approach can be also adapted to other free-form photonic devices, including a waveguide-coupled single-photon source that we demonstrate to prove generalizability. Such DNN-assisted topology optimization significantly reduces the time and resources required for multi-task optimization, enabling large-scale photonic device design in various applications.

Keywords: deep learning; inverse design; silicon photonics; topology optimization.

1 Introduction

Integrated photonics has attracted enormous attention for its merits of low power consumption and high bandwidth. What is more, silicon photonics is compatible with complementary metal oxide semiconductor (CMOS) process, which enables massive fabrication of integrated photonic devices at low cost. Silicon photonic devices and systems have been widely used for optical communications [1], optical computations [2] and quantum optics [3]. Unlike electric circuits which realize different functions through the combination of basic electronic components, a single photonic device can achieve versatile functions by designing device structures. However, due to the complex light-matter interaction, the design of specific-functional devices is not explicit. Especially for those devices with irregular structures, their performance evaluations have to rely on time-consuming numerical electromagnetic (EM) simulations [4], such as the three-dimensional finite-difference-time-domain (3D FDTD) method.

To facilitate the design silicon photonic devices, various inverse design methods based on EM simulations have been proposed, like heuristic algorithms and gradient-descent algorithms [5, 6]. Heuristic algorithms, such as genetic algorithms [7–9] and particle swarm optimization [10–12], generate a group of initial design parameters and

*Corresponding author: Hongyan Fu, Tsinghua-Berkeley Shenzhen Institute and Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China, E-mail: hyfu@sz.tsinghua.edu.cn

Simei Mao, Lirong Cheng, Houyu Chen, Xuanyi Liu and Zihan Geng, Tsinghua-Berkeley Shenzhen Institute and Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China, E-mail: maosm19@mails.tsinghua.edu.cn (S. Mao), clr18@mails.tsinghua.edu.cn (L. Cheng), chenhouy21@mails.tsinghua.edu.cn (H. Chen), liuxuany20@mails.tsinghua.edu.cn (X. Liu), geng.zihan@sz.tsinghua.edu.cn (Z. Geng). <https://orcid.org/0000-0002-0486-9171> (S. Mao)

Qian Li, School of Electronic and Computer Engineering, Peking University, Shenzhen 518055, China, E-mail: liqian@pkusz.edu.cn

update them towards the direction with larger figure-of-merit (FOM). Heuristic algorithms are applicable when the design degree-of-freedom (DOF) is fairly low (less than hundreds), as the required simulations at each epoch are supposed to be larger than the DOF at each iteration. The gradient-descent algorithms like the adjoint method are suitable for designing devices with large DOF (over thousands), where all design parameters could be updated with two simulations [13–15]. However, the gradient-descent algorithms are prone to fall into local optimal [6, 16]. The minimal feature size control is also a big issue, especially for topology optimization (TO) where all pixels in structure are taken as DOF. Some researchers try to solve it by adding a penalty term to loss function [17], applying a filter to remove small features at each iteration [18] or enlarging the size of each pixel to above critical values [19]. Recently, the re-parameterization methods have also been proposed to control the minimal feature size [20].

The above iterative algorithms are more suitable for the single-task optimization. When the target is varied, the whole tedious optimization procedure has to be performed again. Thankfully, deep neural networks (DNNs) bridging the relationship between input and output have been successfully applied for the design of silicon photonic devices [21–27]. However, there are two challenges for DNN-based inverse design methods. Firstly, the DNN requires a great size of prepared training data samples with high performance, usually a thousand times larger than DOF, which is extremely time-consuming to generate. Melati et al. proposed to use machine learning to map the high-dimensional original design space to sub-space with lower dimensionality [28]. However, this method also requires lots of training data samples for this mapping process. Liu et al. proposed a transform-domain-based encoding method to compress real space image into sparse representation in Fourier domain without the dependence of data [29]. However, it uses rigorous Fourier domain processing so that conjugate symmetry is preserved in the post-processing Fourier spectra. Even though the images are restored rigorously, some potential images, are sacrificed due to the selective choice of low-frequency components. Secondly, the trained DNN has limited generalizability. If a target is too far away from the training data set, the network may lose its ability for prediction. Jiang et al. proposed a simulator-based DNN training method for global optimization of 1D metagratings [16, 30], where no prepared training data sample is required. Nevertheless, it is limited to optimization of 1-D periodic structures, whose design DOF is as low as 25 and the dimension of optical target is only 2. What's more, the EM simulation

time for each device is less than one second, already allowing quick optimization by other heuristic algorithms as demonstrated previously [5–15]. In contrast, efficient optimization of 2D free-form device is in greater need, due to the long simulation time and high DOF of 2D free-form structures.

In this work, we propose a DNN-assisted multi-task topology optimization method in Fourier domain with minimal feature size control for 2D free-form devices. The DNN maps optical responses to low-frequency components in Fourier domain, while the topology optimization serves as a supervisor to guide the training of DNN. The device structures can be reconstructed by inverse fast Fourier transform (IFFT) from Fourier domain. Here, we only take the low-frequency components as DOF instead of all pixels in real space, because images in real space contain high-frequency components that contribute to unwanted noise and details. In this way, minimal features in the device structure can be reduced. Another advantage of this Fourier transform domain processing is that it reduces redundant DOF and makes network training easier to converge. For proof-of-concept demonstration, we apply this novel method for the design of integrated wavelength filters on silicon-on-insulator (SOI) platform. Multiple randomly generated filtering targets can be achieved with high accuracy using our multi-task optimization. The trained DNN can be used to predict the wavelength filters with high accuracy instantly. What is more, the trained DNN can then be used to further design other filters with totally different spectra through transfer learning. The proposed method can also be used for multi-stage optimization, with simple DNNs requiring reduced computation resources at each stage. To further demonstrate the generality of our inverse design procedure, we adapt it to the design of on-chip single photon source with high waveguide coupling efficiency achieved.

2 Theoretical analysis

2.1 Topology optimization with adjoint method

For a target optical response \mathbf{y} like the low-pass, high-pass, band-pass or band-stop spectra, the image pattern \mathbf{X} needs to be specified through optimization. The proposed device structure has 201×101 pixels in total, i.e., the design DOF of the device is 20,301, which is hard to optimize by heuristic algorithms. Fortunately, topology optimization with adjoint method is capable of processing such design

with numerous DOF. The optical response \mathbf{y} is the analytical function of electric fields \mathbf{E} distributions at the output port, given as

$$\mathbf{y} = f_1(\mathbf{E}), \quad (1)$$

where the electric field distributions \mathbf{E} in the output area is determined by the light-material interaction in the design area. For a normalized injection light source, the output electric field \mathbf{E} is varied with the permittivity ϵ at the design area, given as

$$\mathbf{E} = f_2(\epsilon). \quad (2)$$

The function f_2 is an implicit function, which transfers the light-material interaction from the design structure to the output area. The derivative of \mathbf{E} with respect to permittivity ϵ can be calculated by the adjoint method [31] as

$$\frac{d\mathbf{E}}{d\epsilon} = 2\epsilon_0 dV \text{Re}[\mathbf{E}_{\text{fwd}} \cdot \mathbf{E}_{\text{adj}}], \quad (3)$$

where ϵ_0 is the vacuum permittivity and dV is the volume of each pixel. \mathbf{E}_{fwd} and \mathbf{E}_{adj} are electric fields at the design area calculated by forward and inverse simulations. Supposing the permittivity of silica and silicon are ϵ_1 and ϵ_2 , the ϵ_{ij} is represented by the grayscale value of pixel x_{ij} in the design area as

$$\epsilon_{ij} = \epsilon_1 + \frac{x_{ij} + 1}{2}(\epsilon_2 - \epsilon_1). \quad (4)$$

For a target optical response, the device structure \mathbf{X} at the design area can be updated by gradient decent. The gradient of optical response with respect to the geometry structure is

$$\frac{d\mathbf{y}}{d\mathbf{X}} = \frac{d\mathbf{y}}{d\mathbf{E}} \cdot \frac{d\mathbf{E}}{d\epsilon} \cdot \frac{d\epsilon}{d\mathbf{X}}. \quad (5)$$

The terms $d\mathbf{y}/d\mathbf{E}$ and $d\epsilon/d\mathbf{X}$ are analytical, while the non-analytical middle term $d\mathbf{E}/d\epsilon$ can be calculated by adjoint method with two simulations. In this way, all the parameters \mathbf{X} can be updated with a learning rate α as

$$\mathbf{X}^{(\text{new})} = \mathbf{X}^{(\text{old})} + \alpha \frac{d\mathbf{y}}{d\mathbf{X}}. \quad (6)$$

2.2 Minimal feature size control in Fourier domain

Even though the proposed device has a DOF of 201×101 , not all of the combinations are able to be fabricated. In other words, some pixels do not contribute significantly to final device performance. A popular way to control the minimal feature size is to use a filter to remove small features throughout optimization [18], which can be realized either in real space or in frequency domain. In space domain, the original random image is convolved with a kernel to filter out noises from the original image. Then, a threshold function is applied to binarize the filtered image.

As a result, isolated features are reduced. In frequency domain, the high-frequency components represent the rapid changes in an image, i.e., the noise and details, while the low-frequency components represent the main part of the image. If high frequency components of an image are filtered out, it would also be smoothed. As shown in Figure 1(a), the original image is converted to the Fourier domain \mathbf{F} by fast Fourier transform (FFT). Then, a mask \mathbf{m} which only allows the low-frequency components to pass is multiplied with \mathbf{F} . By applying inverse fast Fourier transform (IFFT) and an activation function, an image with most pixels gathered is also obtained, while the convolution processing in space domain is replaced by the multiplication in Fourier domain.

Further, instead of taking the whole image in the real space as the design parameters, we utilize the low-frequency components exclusively to reconstruct the image. As shown in Figure 1(b), the low-frequency components \mathbf{F}_L is padded with zero to ensure the size of the frequency spectra \mathbf{F}_S is the same as the image in real space. After IFFT and binarization, the reconstructed structure is the same as Figure 1(a), while the design parameters are the low-frequency components \mathbf{F}_L other than the whole image \mathbf{X} the real space. This processing method can be both used for controlling the minimal feature size as well as reducing the redundant DOF.

2.3 DNN-assisted topology optimization in Fourier domain

Topology optimization is an efficient way of optimizing device with large DOF. However, based on gradient descent algorithm, it is prone to fall into local optimal. Besides, it only works for single-task optimization. Fortunately, DNNs bridging the device geometry with optical response can be used to compensate the disadvantages of traditional topology optimization. Traditionally, DNNs are data samples dependent. If the DOF is too high, the network has to be scaled up and the required number of training data samples will amount. As shown in Figure 2, we propose a new DNN-assisted topology optimization method which efficiently combines the advantages of topology optimization and deep neural network. Since only the low-frequency components in Fourier domain are considered, the minimal feature size can be controlled easily while the redundant DOF is reduced to make the network easier to converge in training.

The DNN consists of 4 fully-connected layers with 100, 512, 1024 and 870 neurons in each layer. The input of the DNN is the target optical response \mathbf{y} . The activation functions of the middle layers are LeakyReLU.

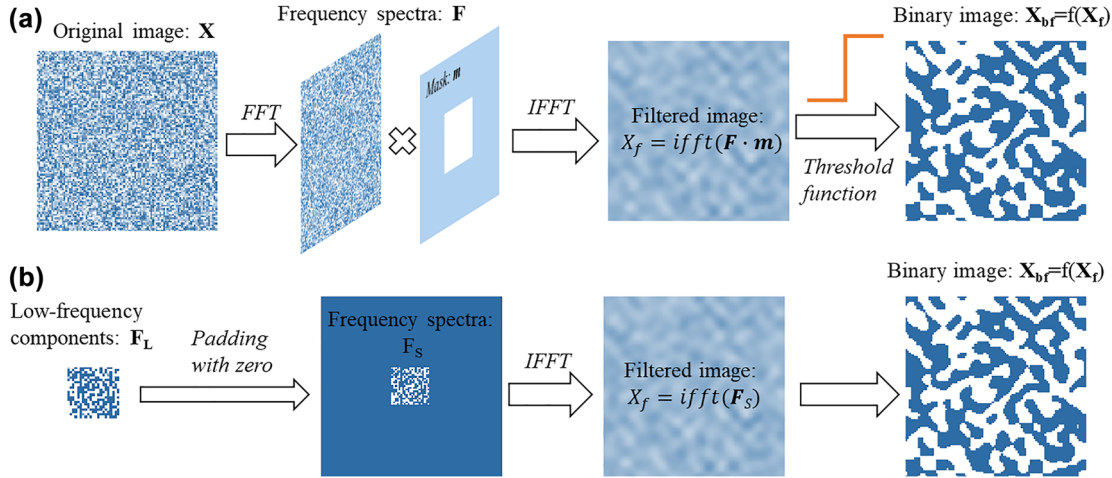


Figure 1: Minimal feature size control methods. (a) Control the minimal feature size by filtering out the high frequency components in the Fourier domain. (b) Utilizing the low-frequency components to reconstruct the real structure.

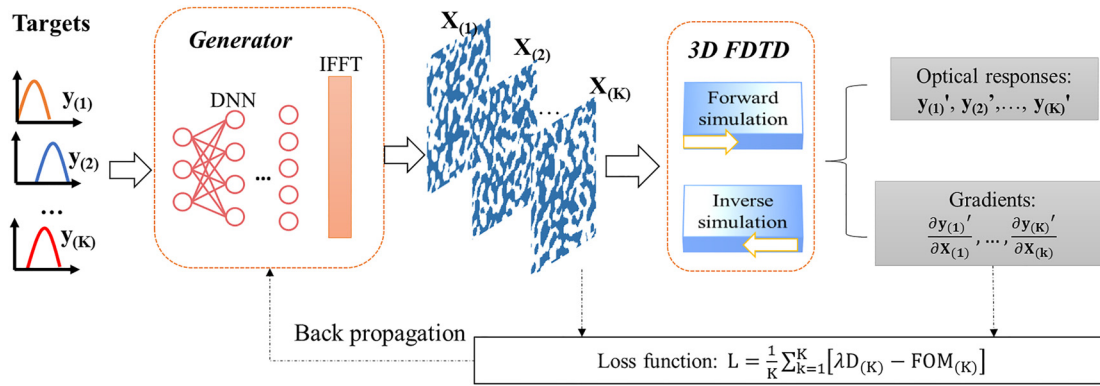


Figure 2: Schematic of the proposed DNN-assisted topology optimization in Fourier domain for multi-task optimization.

The output \mathbf{O} of the DNN is separated into \mathbf{O}_1 and \mathbf{O}_2 , which represents the real and the imaginary parts of low-frequency components in Fourier domain, respectively. By combing them as complex numbers and reshaping them to two-dimensional data, the low-frequency components \mathbf{F}_L in Fourier domain are obtained. The high-frequency components \mathbf{F}_H in Fourier domain are replaced by zero to get the full spectra \mathbf{F} . After the IFFT, we can get the image \mathbf{I} in space, which is further activated by Tanh function to obtain the device structure \mathbf{X} . The predicted device structure \mathbf{X} is simulated by a 3D FDTD solver to calculate the actual optical response \mathbf{y}' . The FOM of k th predicted device is defined as

$$\text{FOM}_{(k)} = 1 - \left| \int_{\lambda=\lambda_{\min}}^{\lambda_{\max}} \frac{|y_{(k)}(\lambda) - y'_{(k)}(\lambda)|^p}{\lambda_{\max} - \lambda_{\min}} d\lambda \right|^{\frac{1}{p}}, \quad (7)$$

where λ is the working wavelength. The parameter p is a normalization constant, which is set as 2 in our optimization for the consideration of calculation speed and avoiding large prediction error.

Apart from the FOM, the discreteness, i.e., the state of binarization, is also an important metric of final generated device. The discreteness of k th device is defined as

$$D_{(k)} = -|X_{(k)}| \cdot (2 - X_{(k)}). \quad (8)$$

When all the pixels in \mathbf{X} are binarized with value of -1 or 1 , D gets the smallest value as -1 . Therefore, the loss function of K generated devices can be expressed as

$$L = \frac{1}{K} \sum_{k=1}^K (\lambda D_{(k)} - \text{FOM}_{(k)}), \quad (9)$$

where the hyper-parameter K is the number of generated devices at each training epochs. It is also referred to as the mini-batch size for the network training [32]. The set of

batch size is problem-related. For simulator-based network training, the required time for the whole training procedure mainly comes from simulation. Therefore, the choose of batch size is determined by the available computation resources to ensure fastest parallel simulation. For our computer, we set K to be 12. Another hyper-parameter λ is used to control the preference of performance or binarization progress. It is assigned a gradually increasing value as

$$\lambda = \frac{n}{N}, \quad (10)$$

where n is the current n th training epoch and N is the total training epochs. The hyper-parameter λ is a small value at the beginning to make sure the device achieves high performance and then it is increased to ensure the optimized device is totally binary.

With above loss function L , the DNN can be trained by backpropagation. The gradients of L with respect to parameters in DNN are calculated as

$$\begin{aligned} \frac{\partial L}{\partial w_m} &= \frac{1}{K} \sum_{k=1}^K \frac{\partial L}{\partial X_{(k)}} \cdot \frac{\partial X_{(k)}}{\partial w_m} \\ &= \frac{1}{K} \sum_{k=1}^K \left(\lambda \frac{\partial D_{(k)}}{\partial X_{(k)}} - \frac{\partial FOM_{(k)}}{\partial X_{(k)}} \right) \cdot \frac{\partial X_{(k)}}{\partial w_m}. \end{aligned} \quad (11)$$

Where the first term $\partial X_{(k)} / \partial w_m$ is analytical and the third term $\partial X_{(k)} / \partial w_m$ determined by the DNN is also analytical. The non-analytical second term can be calculated as

$$\begin{aligned} \frac{\partial FOM_{(k)}}{\partial X_{(k)}} &= - \left| \int_{\lambda=\lambda_{\min}}^{\lambda_{\max}} \frac{|y(\lambda) - y'(\lambda)|^p}{\lambda_{\max} - \lambda_{\min}} d\lambda \right|^{\frac{1}{p}-1} \cdot \int_{\lambda=\lambda_{\min}}^{\lambda_{\max}} \\ &\quad \times \frac{|y(\lambda) - y'(\lambda)|^{p-1} \cdot \text{sign}[y(\lambda) - y'(\lambda)]}{\lambda_{\max} - \lambda_{\min}} \\ &\quad \times \frac{\partial y(\lambda)}{\partial X_{(k)}} d\lambda. \end{aligned} \quad (12)$$

where the term $\partial y(\lambda) / \partial X$ can be calculated by Eq. (5) with the adjoint method.

3 Multi-task parallel optimization of wavelength filters

3.1 Problem setup

Integrated wavelength filters which selectively transmit certain wavelengths and block the others are essential for optical signal processing. The most widely used filters are low-pass, high-pass, band-pass and band-stop

filters [33–36]. However, these proposed empirical-based structures in large footprint are for specific functions, while versatile wavelength filters are required for many applications.

To increase the diversity of wavelength filters for different application scenarios, we propose a compact and general structure on a 220 nm thick standard SOI platform with silica cladding. As shown in Figure 3, with full spectrum input, the device will output various optical responses when the middle area is patterned with different images. The middle slab area is a $4 \times 2 \mu\text{m}^2$ rectangle where each pixel is filled with either silicon or silica. The size of each pixel is $20 \times 20 \text{ nm}^2$ for the balance of simulation accuracy and optical efficiency. The middle slab area connects with two 500 nm wide waveguides for input and output.

For mathematical representation, the geometry structure of the middle area is presented as a binary matrix

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,201} \\ \vdots & \ddots & \vdots \\ x_{101,1} & \cdots & x_{101,201} \end{bmatrix} \quad (13)$$

where x_{ij} represents the pixel of i th row and j th column in the design area. The value of x_{ij} is 1 or -1 when it is filled with silicon or silica. The wavelength range of the proposed device is 1260–1360 nm, where 100 points are evenly sampled from the spectra. The optical response of the device is recorded as

$$\mathbf{y} = [y_1, y_2, \dots, y_{100}] \quad (14)$$

where y_k represents the transmission efficiency at the k th wavelength point.

3.2 Setting the size of Fourier components

In our proposed method, the size of low-frequency components, also referred as the DOF, determines how many details should be included for optimization. To investigate the effects of changing DOF, five wavelength filters are designed with DOFs of 15×7 , 21×11 , 29×15 , 37×19 and 57×29 , respectively. The final FOMs of optimized wavelength filters with different DOFs are shown in Figure 4(a), where inset images depict optimized structures. The solid curves in Figure 4(b) demonstrate optical responses of optimized wavelength filters with different DOFs, while the dashed line represents the optical target. Except for the DOF of 15×7 , the rest devices are with similar optical responses. We also notice that when the DOF is too large such as 57×29 , the performance will decrease. This

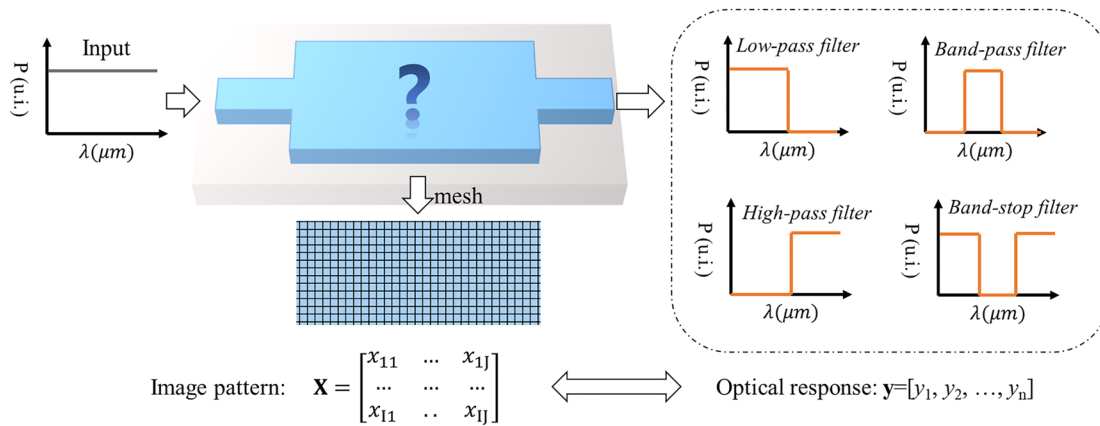


Figure 3: Schematic of the wavelength filter on SOI platform. Various transmission spectra can be achieved with different patterns in the middle slab region.

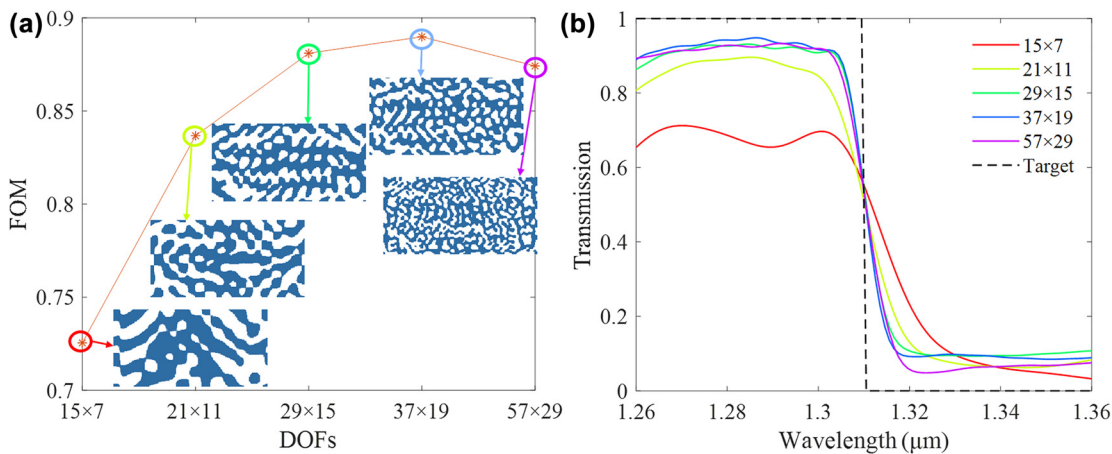


Figure 4: The investigation of impacts for different DOFs. (a) Optimized wavelength filters with five different DOFs as 15×7 , 21×11 , 29×15 , 37×19 and 49×25 . (b) Optical responses of five optimized wavelength filters with different DOFs, where the solid lines represent the simulated results and the dashed line represent the optical target.

is because the optimization is prone to fall into local optimal when the DOF is too large. Further increasing DOF will not improve the performance significantly, while the complexity of the geometry will increase dramatically. Therefore, we set the DOF to 29×15 for the rest of our optimization.

The choice of DOF, or in our case, the number of Fourier components, is based on trial and error. Quantitative estimation of optimization bounds as proposed recently [37] is promising to facilitate our design approach, although the application of such method in problems involving 3D FDTD simulations is still challenging [38]. It would be interesting to further explore the bound estimation of 3D problems in our future work, as a guidance to our inverse design procedures.

3.3 Multi-task optimization results

To ensure the diversity of samples during training process, wavelength filters with random threshold wavelengths and random filter widths are generated as the targets for each epoch. For the balance of network training efficiency and simulation speed, 12 randomly generated targets are input to the DNN for parallel calculation. The training results are shown in Figure 5(a), where the yellow curve and purple curve represent the mean FOM and mean binary state at each training epoch, respectively. As the targeted spectra at each epoch are quite different, the training curves have many ripples. After 711 training epochs, the training curve stabilized. The mean FOM is converged to 0.82 and the mean binary state is close to -1 .

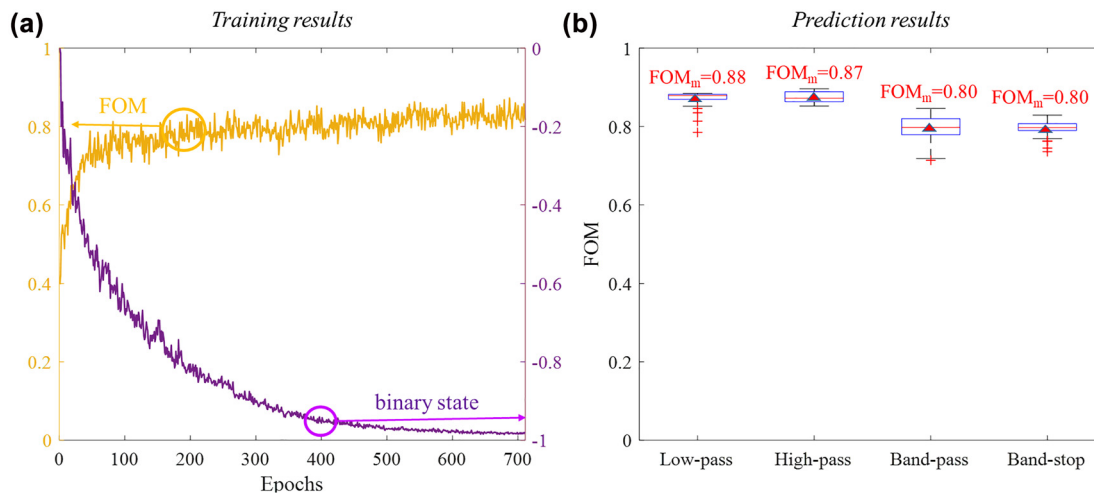


Figure 5: Multi-task optimization results of wavelength filters. (a) Mean FOM (yellow curve) and mean binary state (purple curve) of randomly generated optical targets at each training epoch. (b) Boxplot of the FOMs of predicted devices by the trained DNN for randomly generated optical targets.

Overall, it takes 17,064 simulations for this multi-task network training. It takes less than two weeks to finish the multi-task network training on our computer (Intel(R) Xeon(R) Platinum 8171 M CPU). The training time can be further reduced with more computing resources, as parallel simulations are supported for our proposed multi-task optimization method. It can also be reduced by limiting the number of random targets at each epoch, depending on the scale of the devices required for optimization.

To test the performance of the trained DNN, 100 low-pass, high-pass, band-pass and band-stop wavelength filters are randomly generated as the targets. The trained DNN predicts their corresponding device structures within a second. To measure the performance of generated devices, these predicted structures are further simulated by 3D FDTD method to calculate their actual transmission spectra. The FOMs of four kinds of predicted wavelength filters are shown in the box plot in Figure 5(b). The mean FOMs of low-pass and high-pass filters are 0.88 and 0.87, while the mean FOMs of band-pass and band-stop filters are 0.80. Besides, for all the randomly generated optical targets, the trained DNN could predict their corresponding devices with FOMs of larger than 0.72, even for the worst cases.

For clearer illustration, 12 predicted wavelength filters are shown in Figure 6. The dashed lines represent the target spectra, while the solid lines represent the real transmission spectra of the devices predicted by our trained DNN. The inset images of each subplot figure are the predicted device structures. Figure 6 suggests that the spectra of the generated devices are very close to design targets. More

importantly, for our method the device performance can be further improved by continuously fine-tuning the DNN, while for traditional data-driven DNNs the network will lose their ability for prediction if the target is too far away from the prepared training dataset.

4 Discussions

4.1 Achieving additional design targets by transfer learning

For most DNN-based inverse design, if a targeted optical response is quite different from the training data samples, the performance of predicted device is not guaranteed. However, with our method, completely new targets can be optimized based on the current network without the need of an optimization round from scratch, as the DNN pre-trained by multi-task optimization can be used just like transfer learning. For example, spectra with two filtering windows never exist in our training targets. As such target is too far away from the original training samples, the spectrum of predicted device resembles the original training samples with only one filtering window, as illustrated by red dashed line in Figure 7(b). As shown in Figure 7(a), the pre-trained network is further trained with the new design target. Since the pre-trained DNN has already learned to binarize the predicted structure, we can find the discreteness is always close to one in all training epochs. After 100 epochs of further training with new optical response, the FOM can be improved from

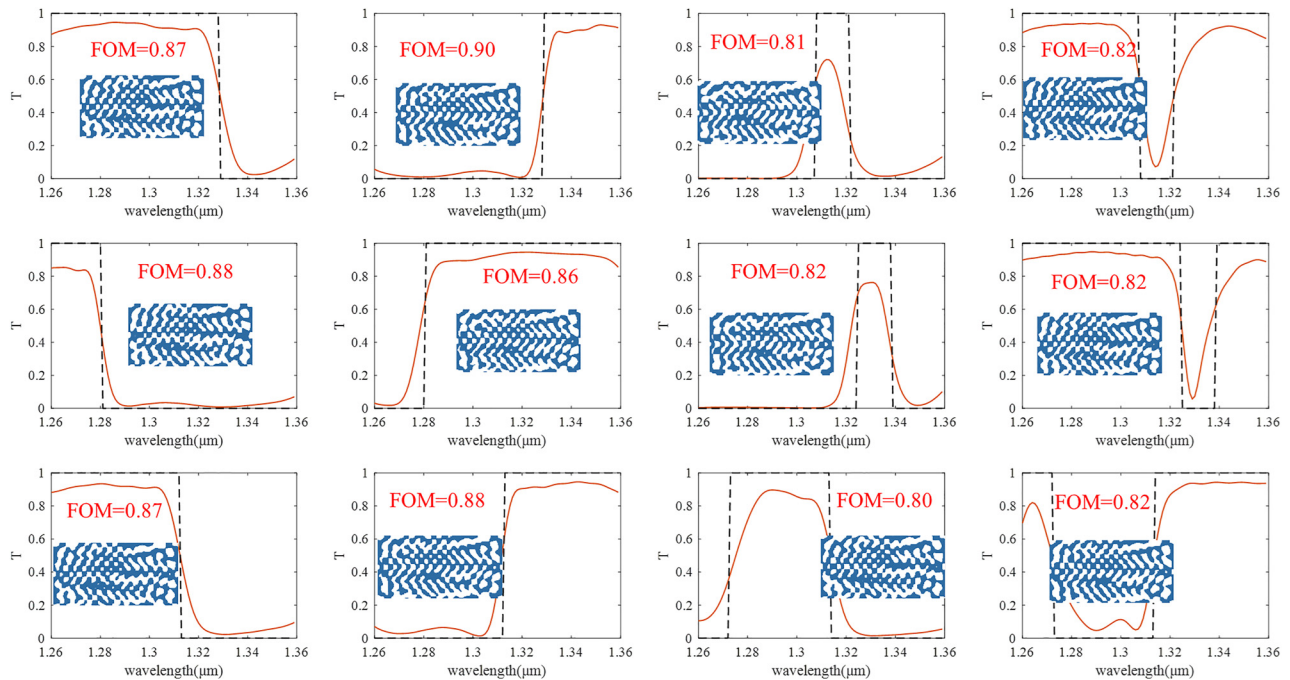


Figure 6: Illustration of 12 predicted wavelength filters. The dashed lines represent the randomly generated optical targets, while the solid lines represent the actual optical responses of predicted devices by the trained DNN. The inset images of each subplot figure show the predicted device structures.

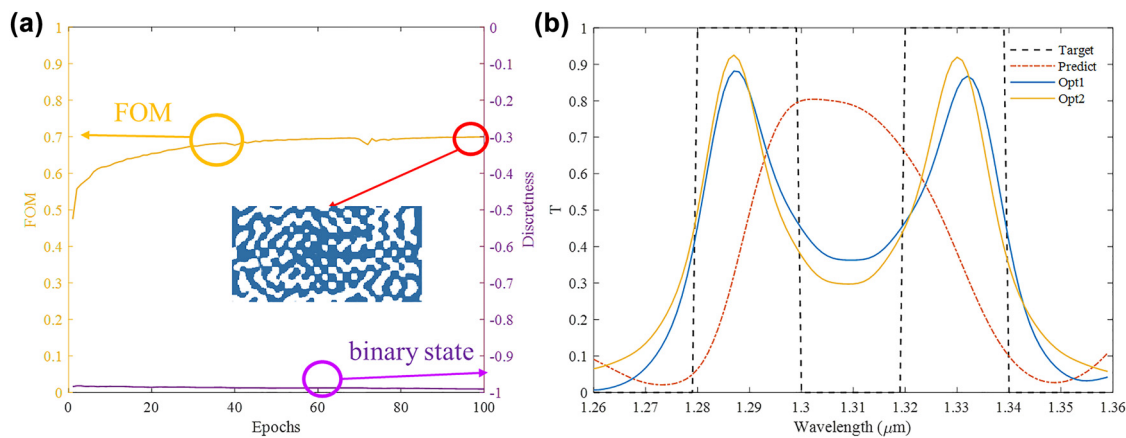


Figure 7: Optimization of a new wavelength filter with two opening windows. (a) Optimization procedure with pre-trained DNN, where yellow line represents the FOM and purple line represents the discreteness. The inset figure shows the further optimized structure. (b) Black and orange dash lines represent the target and predicted optical spectra, while blue (Opt1) and yellow (Opt2) solid lines represent optimized spectra with and without pre-trained DNN.

0.46 to 0.70. The optimized spectrum denoted by blue solid line in Figure 7(b) is much similar to our target. For comparison, we also optimized such an optical filter without the pre-trained DNN. The optimized FOM is also around 0.7 and the spectrum is denoted as yellow solid line in Figure 7(b), which is quite similar to the optimized results by further learning with pre-trained DNN. However, it takes 400 epochs to get such optimization result without

the pre-trained DNN. It suggests even though certain optical target is quite different from the training samples, it can be further optimized with significantly reduced training epochs using our pre-trained generator.

4.2 Multi-stage optimization

Our proposed method can also be extended for multi-stage optimization. When the design DOF is large, reusing the

findings of simulations with fewer Fourier components in simulations with more Fourier components can be an effective way to reduce the complexity of the single DNNs and thus speed up the convergence of network training. For example, we implement the multi-stage optimization in our optical wavelength filters design as shown in Figure 8. In stage 1, the optical target (band pass filter spectrum) is mapped to the 15×7 low-frequency components in Fourier space by DNN_1 . The optimized structure in stage 1 is shown as \mathbf{X}_1 and the optimized spectrum is represented in \mathbf{y}_1 as the orange curve. Due to the lack of enough DOF, the FOM is only 0.67. In stage 2, the optimized low-frequency components in stage 1 are maintained. Meanwhile, the optical target is mapped to the outer layer Fourier components by DNN_2 , which is then combined with those optimized low-frequency components to form the enlarged 29×15 components to reconstruct the final device structure. After such optimization and component expansion, the final optimized structure is shown as \mathbf{X}_2 , which includes more details compared to \mathbf{X}_1 . The optimized wavelength spectrum in stage 2 is represented as the orange curve in \mathbf{y}_2 , whose FOM is increased to 0.80.

4.3 Adaptation to the optimization of an on-chip single-photon source

Apart from the design of wavelength filters, the proposed method can also be extended for the optimization for other nanophotonic devices. For example, Omer et al.

demonstrated a topology-optimized waveguide coupler for single-photon source [39]. It can also be optimized by our DNN-based optimization method. The embedded Si_3N_4 -hBN hybrid cavity on a SiO_2 substrate is shown in Figure 9(a). The hybrid $2 \times 2 \mu\text{m}^2$ square area is the design area. The mesh accuracy here is also set as $20 \times 20 \text{ nm}^2$ for the balance of computation speed and simulation accuracy. The size of Fourier components is set to be 15×15 to enable enough design DOF and to filter out those small features. As the optical target in this case is only 1, the generator is supposed to map 1 input to $2 \times 15 \times 15$ outputs. During training, it is prone to have the gradient exploding problem. To solve this problem, the 1 optical target is represented by a vector with 100 dimensions through an embedding layer first. Then the network and other hyper-parameters are set to be the same as Section 3. As it is a single-task optimization, the training curve is smooth as shown in Figure 9(b). The coupling efficiency is converged to 0.95 and the optimized device is totally discrete after 200 training epochs. The inset figures in Figure 9(b) illustrate the device structures at different training epochs. Along the optimization procedure, the generated device structure evolves toward the shape of a Bragg-reflector-like structure, as such structure could enable highly efficient field redirection towards the waveguide. The electrical fields of the optimized device structure are shown in Figure 9(c), where light is well coupled to the waveguide. This example suggests that our proposed method can be generalized to other device optimization problems, including those with high field intensities.

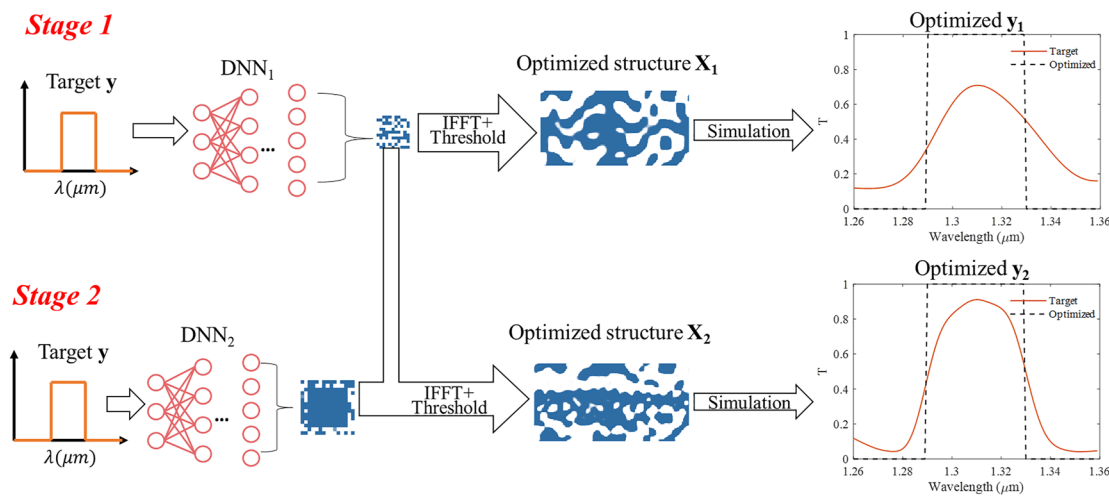


Figure 8: Schematic of multi-stage optimization procedure. In stage 1, the optical target is mapped to the low-frequency components in Fourier space by a simple DNN_1 . In stage 2, the optical target is mapped to outer Fourier components by another simple DNN_2 . The right pictures show the optimized structures \mathbf{X}_1 , \mathbf{X}_2 and their corresponding wavelength spectra \mathbf{y}_1 , \mathbf{y}_2 .

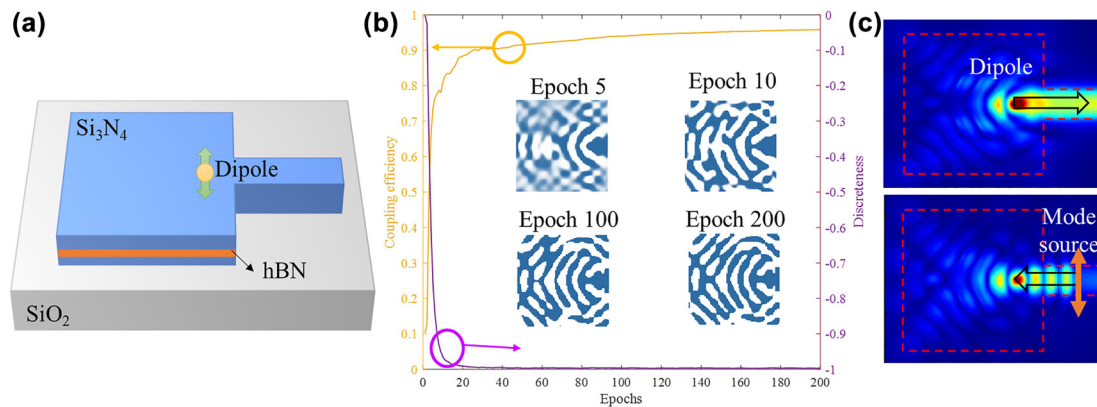


Figure 9: Optimization of single photon emission. (a) The schematic of embedded Si_3N_4 -hBN hybrid cavity on a SiO_2 substrate [39], where the square area is also the design area. (b) Optimization procedure for high coupling efficiency with our method, where the inset figures show the generated structure at different training epochs. (c) Simulated electric fields with dipole input (upper one) and mode source input (lower one) of the optimized device structure.

5 Conclusions

In conclusion, we proposed a DNN-assisted topology optimization method in Fourier domain for the design of integrated wavelength filters. The targeted optical responses are input to the DNN, where the low-frequency component components in Fourier domain are output. After zero-padding, IFFT and activation function, the device structure patterns in real space can be reconstructed. By forward and adjoint 3D FDTD simulations, the loss gradients of generated device structures are calculated, which can be used for the training of DNN. Taking the low-frequency Fourier components as DOF helps control the minimal feature size of generated devices. It can also reduce the redundant DOF to make the DNN prone to converge. This DNN-driven topology optimization method can be used for concurrent optimizations of multiple wavelength filters, where randomly generated wavelength filters are input as the targets for network training. After 711 training epochs, the mean FOM is converged to around 0.82. The trained DNN can also be further utilized for the design of totally different wavelength filters through transfer learning, where fewer optimization epochs would be required compared to optimization from scratch. What is more, the proposed method can also be used for multi-stage optimization with simpler DNNs at each stage. To demonstrate the generalizability of our method, we also adapt it to the design of waveguide-coupled single-photon source with high coupling efficiency. Our proposed method paves a new way for the design of free-form and compact nano-photonic devices.

Author contributions: Simei Mao and Lirong Cheng contribute equally to this work. All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: This work was supported in part by Guangdong Basic and Applied Basic Research Foundation under Grant 2021A1515011450 and in part by Shenzhen Science and Technology Innovation Commission under Grant JCYJ20180507183815699.

Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

Data availability: Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

- [1] D. Thomson, A. Zilkie, J. E. Bowers, et al., "Roadmap on silicon photonics," *J. Opt.*, vol. 18, 2016, Art no. 073003.
- [2] T. Ferreira de Lima, B. J. Shastri, A. N. Tait, M. A. Nahmias, and P. R. Prucnal, "Progress in neuromorphic photonics," *Nanophotonics*, vol. 6, pp. 577–599, 2017.
- [3] X. Qiang, X. Zhou, J. Wang, et al., "Large-scale silicon quantum photonics implementing arbitrary two-qubit processing," *Nat. Photonics*, vol. 12, pp. 534–539, 2018.
- [4] S. Mao, L. Cheng, C. Zhao, F. N. Khan, Q. Li, and H. Fu, "Inverse design for silicon photonics: from iterative optimization algorithms to deep neural networks," *Appl. Sci.*, vol. 11, p. 3822, 2021.
- [5] K. Yao, R. Unni, and Y. Zheng, "Intelligent nanophotonics: merging photonics and artificial intelligence at the nanoscale," *Nanophotonics*, vol. 8, pp. 339–366, 2019.

- [6] S. Molesky, Z. Lin, A. Y. Piggott, et al., “Inverse design in nanophotonics,” *Nat. Photonics*, vol. 12, pp. 659–670, 2018.
- [7] Z. Yu, H. Cui, and X. Sun, “Genetic-algorithm-optimized wideband on-chip polarization rotator with an ultrasmall footprint,” *Opt. Lett.*, vol. 42, pp. 3093–3096, 2017.
- [8] Z. Liu, X. Liu, Z. Xiao, et al., “Integrated nanophotonic wavelength router based on an intelligent algorithm,” *Optica*, vol. 6, pp. 1367–1373, 2019.
- [9] S. Mao, L. Cheng, C. Zhao, and H. Fu, “Ultra-broadband and ultra-compact polarization beam splitter based on a tapered subwavelength-grating waveguide and slot waveguide,” *Opt. Express*, vol. 29, pp. 28066–28077, 2021.
- [10] Y. Zhang, S. Yang, A. Lim, et al., “A compact and low loss Y-junction for submicron silicon waveguide,” *Opt. Express*, vol. 21, pp. 1310–1316, 2013.
- [11] W. Chen, B. Zhang, P. Wang, et al., “Ultra-compact and low-loss silicon polarization beam splitter using a particle-swarm-optimized counter-tapered coupler,” *Opt. Express*, vol. 28, pp. 30701–30709, 2020.
- [12] H. Guan, Y. Ma, R. Shi, et al., “Ultracompact silicon-on-insulator polarization rotator for polarization-diversified circuits,” *Opt. Lett.*, vol. 39, pp. 4703–4706, 2014.
- [13] L. Cheng, S. Mao, Z. Chen, Y. Wang, C. Zhao, and H. Fu, “Ultra-compact dual-mode mode-size converter for silicon photonic few-mode fiber interfaces,” *Opt. Express*, vol. 29, pp. 33728–33740, 2021.
- [14] 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,” *Nat. Photonics*, vol. 9, pp. 374–377, 2015.
- [15] S. Mao, L. Cheng, C. Zhao, and H. Fu, “Coarse wavelength division (de)multiplexer based on cascaded topology optimized wavelength filters,” *Proc. CLEO*, pp. 1–2, 2021. https://doi.org/10.1364/cleo_at.2021.jw1a.62.
- [16] J. Jiang and J. A. Fan, “Global optimization of dielectric metasurfaces using a physics-driven neural network,” *Nano Lett.*, vol. 19, pp. 5366–5372, 2019.
- [17] D. Verduyn, N. V. Sapra, L. Su, R. Trivedi, and J. Vuckovic, “Analytical level set fabrication constraints for inverse design,” *Sci. Rep.*, vol. 9, p. 8999, 2019.
- [18] A. M. Hammond, A. Oskooi, S. G. Johnson, and S. E. Ralph, “Photonic topology optimization with semiconductor-foundry design-rule constraints,” *Opt. Express*, vol. 29, pp. 23916–23938, 2021.
- [19] K. Wang, X. Ren, W. Chang, L. Lu, D. Liu, and M. Zhang, “Inverse design of digital nanophotonic devices using the adjoint method,” *Photon. Res.*, vol. 8, pp. 528–533, 2020.
- [20] E. Khoram, X. Qian, M. Yuan, and Z. Yu, “Controlling the minimal feature sizes in adjoint optimization of nanophotonic devices using b-spline surfaces,” *Opt. Express*, vol. 28, pp. 7060–7069, 2020.
- [21] S. So, T. Badloe, J. Noh, J. Bravo-Abad, and J. Rho, “Deep learning enabled inverse design in nanophotonics,” *Nanophotonics*, vol. 9, pp. 1041–1057, 2020.
- [22] A. M. Hammond and R. M. Camacho, “Designing integrated photonic devices using artificial neural networks,” *Opt. Express*, vol. 27, pp. 29620–29638, 2019.
- [23] Y. Tang, K. Kojima, T. Koike-Akino, et al., “Generative deep learning model for inverse design of integrated nanophotonic devices,” *Laser Photon. Rev.*, vol. 14, 2020, Art no. 2000287.
- [24] Y. Ren, L. Zhang, W. Wang, et al., “Genetic-algorithm-based deep neural networks for highly efficient photonic device design,” *Photon. Res.*, vol. 9, pp. 247–252, 2021.
- [25] S. Mao, L. Cheng, F. N. Khan, et al., “Inverse design of high-dimensional nanostructured 2×2 optical processors based on deep convolutional neural networks,” *J. Lightw. Technol.*, vol. 40, pp. 2926–2932, 2022.
- [26] Y. Long, J. Ren, Y. Li, and H. Chen, “Inverse design of photonic topological state via machine learning,” *Appl. Phys. Lett.*, vol. 114, 2019, Art no. 181105.
- [27] D. Gostimirovic and W. N. Ye, “An Open-Source Artificial neural network model for polarization-insensitive silicon-on-insulator subwavelength grating couplers,” *IEEE J. Sel. Top. Quant. Electron.*, vol. 25, pp. 1–5, 2019.
- [28] D. Melati, Y. Grinberg, M. Kamandar Dezfouli, et al., “Mapping the global design space of nanophotonic components using machine learning pattern recognition,” *Nat. Commun.*, vol. 10, p. 4775, 2019.
- [29] Z. Liu, Z. Zhu, and W. Cai, “Topological encoding method for data-driven photonics inverse design,” *Opt. Express*, vol. 28, no. 4, pp. 4825–4834, 2020.
- [30] J. Jiang and J. A. Fan, “Simulator-based training of generative neural networks for the inverse design of metasurfaces,” *Nanophotonics*, vol. 9, pp. 1059–1069, 2019.
- [31] C. M. Lalau-Keraly, S. Bhargava, O. D. Miller, and E. Yablonovitch, “Adjoint shape optimization applied to electromagnetic design,” *Opt. Express*, vol. 21, pp. 21693–21701, 2013.
- [32] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, Cambridge, MIT Press, 2016.
- [33] P. Xu, Y. Zhang, S. Zhang, Y. Chen, and S. Yu, “Scaling and cascading compact metamaterial photonic waveguide filter blocks,” *Opt. Lett.*, vol. 45, pp. 4072–4075, 2020.
- [34] X. B. Xu, X. Guo, W. Chen, et al., “Flat-top optical filter via the adiabatic evolution of light in an asymmetric coupler,” *Phys. Rev. A*, vol. 100, 2019, Art no. 023809.
- [35] E. S. Magden, N. Li, M. Raval, et al., “Transmissive silicon photonic dichroic filters with spectrally selective waveguides,” *Nat. Commun.*, vol. 9, p. 3009, 2018.
- [36] Q. Huang, K. Jie, Q. Liu, Y. Huang, Y. Wang, and J. Xia, “Ultra-compact, broadband tunable optical bandstop filters based on a multimode onedimensional photonic crystal waveguide,” *Opt. Express*, vol. 24, pp. 20542–20553, 2016.
- [37] S. Molesky, P. Chao, J. Mohajan, W. Reinhart, H. Chi, and A. W. Rodriguez, “T-operator limits on optical communication:

- metaoptics, computation, and input-output transformations,” *Phys. Rev. Res. Int.*, vol. 4, no. 1, 2022, Art no. 013020.
- [38] P. Chao, B. Strekha, R. Kuate Defo, S. Molesky, and A. W. Rodriguez, “Physical limits in electromagnetism,” *Nat. Rev. Phys.*, vol. 4, pp. 543–559, 2022.
- [39] O. Yesilyurt, Z. A. Kudyshev, A. Boltasseva, V. M. Shalaev, and A. V. Kildishev, “Efficient topology-optimized couplers for on-chip single-photon sources,” *ACS Photonics*, vol. 8, no. 10, pp. 3061–3068, 2021.