Progress In Electromagnetics Research M, Vol. 119, 143–160, 2023

AI-Based Fast Design for General Fiber-to-Waveguide Grating Couplers

Zhenjia Zeng[1], Qiangsheng Huang[2, *], and Sailing He[1, 2, *]

Abstract—Utilizing deep learning to replace numerical simulation solvers for electromagnetic wave propagation is a promising approach for the rapid design of photonic devices. However, to realize the advantages of deep learning for rapid design, it is essential to apply it to a general device structure. In this study, we propose a method that employs deep learning to assist in fast design of a general grating coupler structure. We use a modified 1D-ResNet18(1D-MR18) to predict the coupling efficiency of various grating couplers at different wavelengths. After comparing and selecting the optimal combination of learning rate, activation functions, and batch normalization size, the 1D-MR18 demonstrates remarkable accuracy ($MSE : 2.18 \times 10^{-5}$, $R^2 : 0.969$, $MAE : 0.003$). By integrating the 1D-MR18 with the adaptive particle swarm algorithm, we can efficiently design periodic and nonuniform grating couplers that meet various functional requirements, including single-wavelength grating couplers, multi-wavelength grating couplers, and robust grating couplers. The time for designing a single device is no more than 2 minutes, and the shortest is only 17 seconds. This novel approach of employing deep learning for the fast and efficient design from standard photonic device structures offers valuable insights and guidance for photonic devices design.

1. INTRODUCTION

Grating is an optical element composed of a series of parallel trenches and teeth arranged in a precisely periodic manner, where the spacing and shape of these trenches and teeth can be precisely controlled to disperse the incident light or control the diffraction angle of incident light [1]. A grating can couple the light efficiently from a fiber to an SOI (Silicon-on-Insulator) waveguide [1]. Edge couplers [2] and grating couplers [3] are two common devices used for coupling light between optical fibers and integrated photonic circuits. Compared to edge couplers, grating couplers are more compact in shape, have lower fabrication costs, and offer greater tolerance in the alignment of optical fibers during coupling, enabling wafer-scale optical testing and chip manufacturing [4, 5].

In the design of photonic devices, traditional approaches typically involve starting with standard geometric structures and conducting extensive parameter scans or utilizing optimization algorithms such as Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) for iterative refinement to find the optimal structural parameters [6, 7]. This process requires complex Maxwell's numerical solvers which are highly time-consuming and inefficient [7]. Numerical solvers necessitate meticulous meshing and iterative solving of the device structure, demanding intricate modeling procedures and significant computational resources. As the device volume increases or when fine meshing is required, the computational resources needed become even more substantial, consequently prolonging the simulation time [8, 9]. Conducting numerous numerical simulations to achieve the optimal device structure becomes practically infeasible [7, 10].

Received 27 July 2023, Accepted 30 September 2023, Scheduled 4 October 2023

* Corresponding authors: Sailing He (sailing@zju.edu.cn), Qiangsheng Huang (qiangsheng@zju.edu.cn).

[1] National Engineering Research Center for Optical Instruments, Center for Optical and Electromagnetic Research, Zhejiang University, Hangzhou 310058, China. [2] Shanghai Institute for Advanced Study, Zhejiang University, Shanghai 201203, China.

Deep learning is a data-driven methodology, primarily relying on the construction of multi-layered neural networks to learn feature representations from data, thereby modeling the mapping relationships or hidden rules between input and output data within a dataset [11]. Deep learning has achieved significant breakthroughs in various domains, including speech recognition [12], image recognition [13], natural language processing [14], and has even surpassed human performance levels in many applications [15]. In the field of photonic devices design, there is also a growing interest in leveraging deep learning methods. Deep learning can effectively capture the highly nonlinear relationships between the geometric structure of photonics devices and their optical responses based on experimentally or computationally obtained datasets [9, 16]. These datasets can be generated in a parallel manner and applied to the design of photonics devices with structurally similar but functionally distinct characteristics, thus circumventing the need for extensive numerical simulations with each new design. Presently, deep learning has found widespread application in the rapid design of photonics devices, such as metasurfaces [17], nanoparticles [18], membrane stacks [19], and integrated waveguide components [20].

In the design of grating couplers, studies [21, 22] have explored the use of deep neural networks as a replacement for numerical simulators to achieve more efficient designs. However, the neural network architectures and grating structures employed in these studies were relatively simplistic, limiting the exploration to smaller parameter spaces and achieving only limited functionalities of the couplers [21, 22]. The true advantage of employing neural networks as substitutes for numerical simulators lies in their universality. Only when a trained neural network is capable of designing diverse devices with different functionalities, can it effectively distribute the time cost of generating datasets and training models [23]. Non-uniform gratings offer higher design degrees of freedom, enabling not only higher coupling efficiency but also satisfying the design requirements for different bandwidths [24, 25]. Therefore, they present a standard structure for designing different functional grating couplers and have been widely adopted in the design of highly efficient, large-bandwidth grating couplers [24, 25].

In this paper, we propose a novel approach utilizing deep neural networks to assist in grating coupler design from standard grating coupler structures. We employ one-dimensional nonuniform grating couplers as the fundamental geometric structures and introduce two parameters, incident angle and etch depth, into the parameter space. The grating coupler structure we adopt exhibits excellent versatility, making it applicable to the design of grating couplers with diverse functionalities. Upon sufficient training, the deep neural network can model the highly nonlinear relationship between the parameter space of one-dimensional nonuniform coupling gratings and their optical characteristics, replacing numerical solvers for rapid computation of the grating coupler's optical properties. This precise solving approach can be accomplished within fractions of a second, enabling rapid optimization of grating couplers. We apply this method to the fast design of various grating couplers with different functionalities, such as single-wavelength grating couplers, multi-wavelength grating couplers, and robust grating couplers.

2. DESIGN PARAMETERS OF GRATING COUPLERS

The simplest grating coupler is achieved using a uniform grating. However, due to the directionality of the grating and mode mismatch, the coupling efficiency of the uniform grating can theoretically reach only about 80% at best, making it challenging for multi-wavelength or large bandwidth designs [26]. Higher coupling efficiency, wider bandwidth, or additional functionalities can only be achieved through alternative approaches, such as nonuniform gratings [24], multi-layer gratings [27], gratings with multiple etch depths [28], and unconventional geometric shapes of gratings [29]. Among these options, nonuniform gratings stand out for their simplicity, ease of fabrication, and high design flexibility, making them suitable for various functional grating coupler designs [30]. Grating couplers are often implemented on the Silicon-on-Insulator (SOI) platform, which offers strong optical confinement and compatibility with Complementary Metal-Oxide-Semiconductor (CMOS) processes, making it an excellent structure for designing grating couplers [31].

In this work, we adopt a nonuniform grating coupler based on the SOI platform as our fundamental structure. Figure 1 illustrates the two-dimensional cross-sectional schematic of the adopted grating coupler. The parameters of our grating coupler are referenced from the configuration in Reference [21],

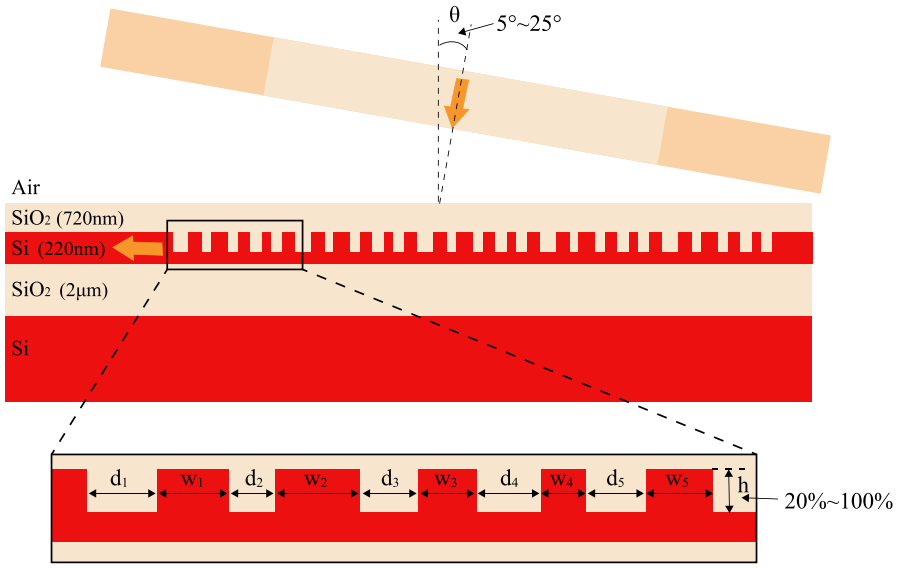

Figure 1. Two-dimensional cross-sectional schematic of a nonuniform grating coupler.

with a grating length of 12 μm, a buried oxide layer thickness of 2 μm, and a fixed Si thickness of 220 nm. On the surface of the SOI, we cover a layer of SiO_2 with a thickness of 720 nm. A previous study [32] indicates that adding a Poly-Silicon layer to the grating can enhance the coupling efficiency of incident light to the output, improving the overall performance of the grating coupler. A TE-mode Gaussian beam is incident into the grating from an SMF-28 optical fiber, with the fiber positioned at the center of the grating. The incident angle is θ, and the Gaussian beam's mode field diameter is 10.4 μm. The nonuniform grating is defined to have 25 trench widths and 25 tooth widths. To reduce the design space and ease the burden on the neural network fitting process, we divide the nonuniform grating into 5 sections, each having the same structure with a width of 2.4 μm. Consequently, for the entire nonuniform grating, we consider 10 design parameters per grating period (5 trench widths d_1, d_2, ..., d_5, and 5 tooth widths w_1, w_2, ..., w_5), etch depth ratio h, and incident angle θ.

We proceeded to generate the dataset required for training the neural network model based on the adopted nonuniformly coupled grating structure. In this work, we limit the incident angle between 5 to 25 degrees. This range is chosen because at very small incident angles, the input light undergoes second-order Bragg diffraction in the coupling grating, reducing the coupling efficiency and affecting the resonant wavelength and dispersion characteristics of the coupling grating [33]. Very large incident angles can lead to total reflection in the SiO_2 cladding and also result in mode mismatch of the incident light. To maintain a unified scale for the input data, we normalize the incident angle by dividing it by the maximum allowed value of 25 degrees, thus restricting θ to the range of 0.2 to 1.0. Regarding the etch depth ratio, h, we only consider coupling gratings with etch depths between 20% to 100%, as too shallow etch depths lead to minimal light coupling into the grating. Therefore, the range of h is set between 0.2 to 1.0. For the trench width and tooth width, considering the resolution of electron beam lithography, we impose a minimum width of 100 nm for both the trench and tooth. To maintain a constant total length of the grating while adhering strictly to the minimum feature size limitation, we adopt a reparameterization method for generating the trench width and tooth width randomly, yet satisfying some constrains. We assume that the grating has a total length of L, with M trenches and M teeth, and minimum feature sizes w_m and d_m. We introduce random latent space variables s_{w_1}, ..., s_{w_M}, s_{d_1}, ..., s_{d_M} to generate random trench width and tooth width, respectively. According to

$$w_i = \frac{s_{w_i}}{\sum_{j=1}^{M} s_{w_j} + s_{d_j}} (L - Mw_m - Md_m) + w_m, \quad i = 1, 2, \cdots, M \tag{1}$$

$$d_i = \frac{s_{d_i}}{\sum_{j=1}^{M} s_{w_j} + s_{d_j}} (L - Mw_m - Md_m) + d_m, \quad i = 1, 2, \cdots, M \tag{2}$$

where w_i represents the trench width, and d_i represents the tooth width. Due to

$$L - Mw_m - Md_m \geq 0, \quad s_{w_i} \geq 0, \quad s_{d_j} \geq 0 \tag{3}$$

we have $w_i \geq w_m$ and $d_i \geq d_m$. Additionally,

$$\sum_{i=1}^{M} w_i + d_i = \frac{\sum_{i=1}^{M} s_{w_i} + s_{d_i}}{\sum_{j=1}^{M} s_{w_j} + s_{d_j}} (L - Mw_m - Md_m) + Md_m + Mw_m = L \tag{4}$$

Therefore, after applying the reparameterization method, each generated grating has a fixed total length, denoted by L, and the trench width and tooth width are strictly larger than the minimum feature sizes w_m and d_m. In this study, L is set to 2.4 µm, and both w_m and d_m are 100 nm. This approach effectively prevents significant variations in grating length during dataset generation, which could hinder the convergence of the neural network training process and ensures strict control of the grating's minimum feature sizes. Commercial software Lumerical FDTD is employed for generating the dataset by random sampling within the constrained parameter space and latent space. In the field of optical communication, common central wavelengths for grating couplers are 1310 nm and 1550 nm. Considering the bandwidth requirements, we set the incident light wavelength in the range of 1200 nm to 1700 nm and discretize it into 200 points uniformly in frequency space. Since the optical response of the grating coupler is greatly influenced by the widths of the teeth and trenches, we use "precise volume average" mesh refinement method, which is an auto nonuniform method and can provides more sensitive meshing to small geometric variations [34]. All the boundary conditions are set to "PML", which can absorb electromagnetic waves with minimal reflections (as if in a homogenous space) [35, 36]. The parameters of the grating coupler and the corresponding coupling efficiencies for each wavelength are saved as input and output for the neural network, respectively. Therefore, our input is a 12-dimensional vector, while the output is a 200-dimensional vector. We generated a total of 303074 data sets and divided them in an 8 : 2 ratio, with 242459 sets used for training and 60615 sets for testing.

3. DEEP LEARNING MODEL

In the grating coupler design, the mapping between structural parameters and coupling efficiency at different wavelengths is a highly complex and highly nonlinear function. Previously, we have employed simple fully connected and convolutional neural network to modeling the inner mapping, but both were challenging to accurately establish the relationship between the structural parameters and coupling efficiency. Therefore, we adopted a deeper neural network architecture to capture the implicit complex relationship between the structural parameters and coupling efficiency.

The deep model structure used in this study, as shown in Figure 2(a), is based on an 18-layer ResNet. ResNet is a classical deep convolutional neural network architecture introduced by Kaiming He et al. in 2015 [37]. The key feature of ResNet is the incorporation of residual blocks, which mitigates the degradation problem in deep neural networks [37, 38]. This degradation issue refers to the phenomenon where increasing network depth leads to an increase in test error [39]. The main structure of ResNet comprises multiple residual blocks, each containing multiple convolutional layers, batch normalization layers, activation functions, and a shortcut connection. The shortcut connection within the residual block allows the neural network to skip some layers and directly propagate the input signal to the subsequent layers, thus alleviating the vanishing gradient and exploding gradient problems in deep neural networks [38].

Unlike the original ResNet-18, where inputs are two-dimensional, our input is a one-dimensional vector. Consequently, we replaced all 2D convolutional kernels in the network with 1D convolutions and

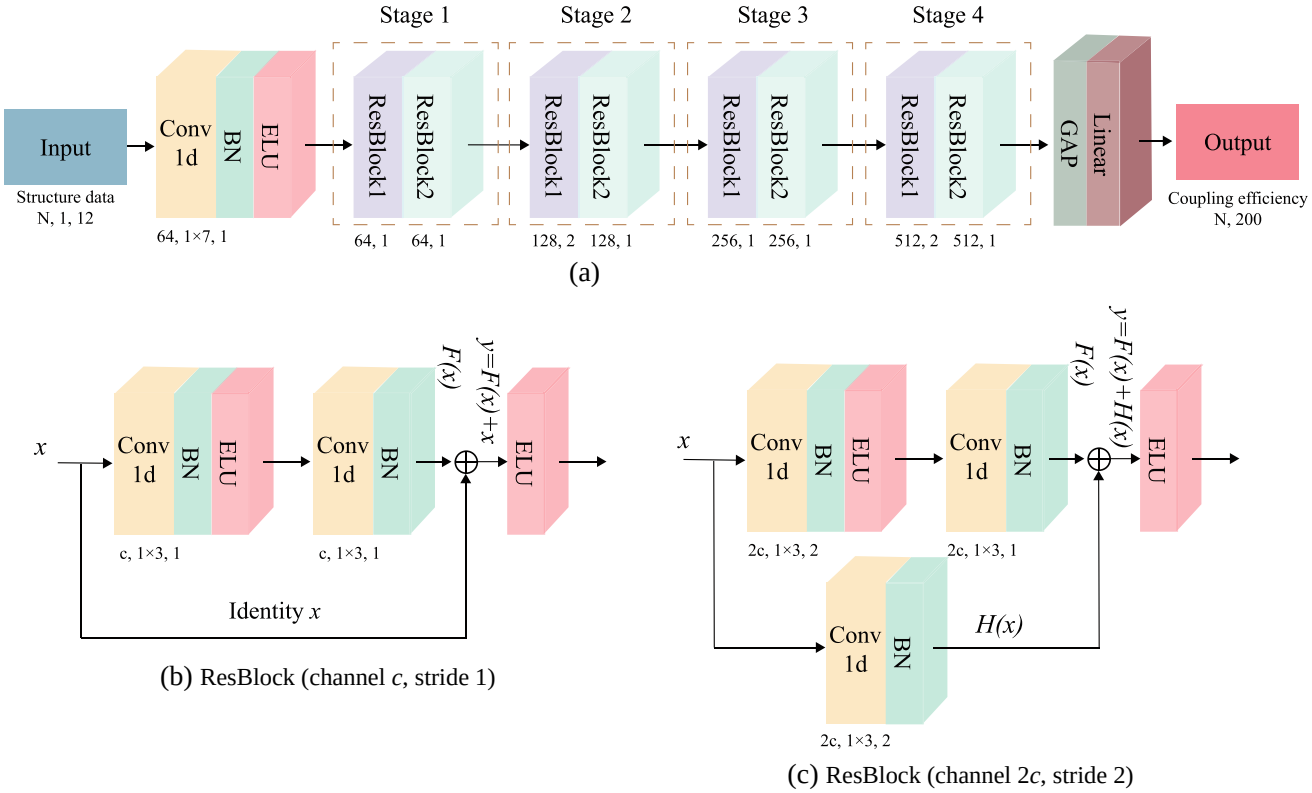

Figure 2. 1D-MR18. (a) Architecture of 1D-MR18; (b) Residual block structure when the input and output dimensions are the same; (c) Residual block structure when the input and output dimensions differ.

modified the shortcut connections within the residual blocks to be 1D connections. As our input consists of continuous data rather than the original image data used in ResNet-18, the information content in the data is substantial. Therefore, the initial larger stride convolutions and pooling in ResNet-18 may lead to information loss in the input. To address this, we removed the max-pooling layer from ResNet-18 and modified the convolutional stride in the first convolutional layer and the third residual block to be 1. Although the original ResNet-18 used the ReLU function as the activation function, in recent years, new activation functions have been continuously proposed and have shown better performance than the ReLU function in experiments, such as ELU [40], SELU [41], SiLU [42], etc. During the model pre-training, we trained the model using different activation functions under the same conditions and compared their performance, which we will discuss in the model training section. The experimental results showed that the model with ELU as the activation function exhibited superior performance. We adopted ELU as the activation function for our model.

In the residual blocks, we employed two sets of convolutional layers, batch normalization layers, activation function layers, and a shortcut connection structure. Depending on the output dimensionality, we used two different residual structures, as depicted in Figure 2(b) and Figure 2(c). Their corresponding output relationships are as follows:

$$y = F(x) + x \tag{5}$$

$$y = F(x) + H(x) \tag{6}$$

where x represents the input to the residual block; y denotes the output of the residual block; $F(x)$ refers to the output of the residual block without the shortcut connection; and $H(x)$ represents the output of the shortcut connection when the input and output dimensions differ. After the final linear layer, we removed the original activation function in ResNet because our problem is a continuous spatial nonlinear regression problem, and the presence of the activation function in the last layer would affect

the continuity of the model output. As the model is a modification of ResNet18, we refer to it as 1D-MR18.

4. MODEL TRAINING AND EVALUATION

We constructed our model using Python 3.8 and PyTorch 1.10 and conducted training on a machine equipped with a NVIDIA GeForce GTX 1080 Ti GPU running Ubuntu. The model's input is a one-dimensional structural vector of length 12, containing the incident angle of the grating coupler, etch depth ratio, as well as the widths of 5 trench widths and 5 tooth width that constitute the periodic grating. The output is the coupling efficiency discretized into 200 points between wavelengths ranging from 1200 nm to 1700 nm using Lumerical FDTD's built-in adaptive mesh technology. We adopted NAdam algorithm as the optimization algorithm for our model. NAdam is an extension of the Adam algorithm, incorporating the Nesterov momentum term. In comparison to Adam, NAdam can accelerate the convergence speed and enhance the performance of the model [43]. Before the formal training, we considered factors that could impact the model's predictive performance, namely the learning rate, activation functions, and batch size. We denoted MSE_t to represent the average loss function value of the model on the test set. In addition to the loss function, we also evaluated our model's predictive performance using R^2, a statistical metric used to assess the goodness of fit of the regression model to the observed data and provides more informative insights compared to MSE [44]. R^2 is defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2} \tag{7}$$

Here, y_i represents the ground truth, which is the coupling efficiency simulated using FDTD. $\hat{y}_i$ represents the predicted value, which is the model's predicted coupling efficiency. $\bar{y}_i$ denotes the average of the ground truth values and N is the number of samples, which is the discrete number of wavelengths. We denote the average R^2 on the test set as R_t^2, which measures the goodness of fit of the model's predictions to the observed coupling efficiency.

4.1. Learning Rate Comparison

The learning rate is a crucial factor affecting model performance. The choice of learning rate directly influences the convergence speed and final performance of the model. If the learning rate is too large, the model may diverge during training and fail to converge to an optimal solution [45]. Conversely, if the learning rate is too small, the model may converge slowly or get trapped in local minima, making it difficult to escape [45]. To improve the model's ability to escape local minima and achieve faster convergence, we employed a grid search approach to explore the learning rate space. Initially, we trained 1D-MR18 using equally spaced learning rates on a logarithmic scale, such as 0.01, 0.001, 0.0001, and 0.00001, and compared their performance in terms of MSE_t and R_t^2. Based on this preliminary analysis, we determined that the optimal learning rate lies within the range of 0.01 to 0.0001. Next, we conducted a logarithmic search within the range of 0.01 to 0.0001 on the learning rate axis. Figure 3 illustrates the model's performance under different learning rates. Both MSE_t and R_t^2 tend to reach their minimum or maximum values within the range of 0.00025 to 0.0025, and the differences within this range are relatively small. As a result, we selected a learning rate of 0.0005 from the interval of 0.00025 to 0.0025 as the final learning rate for the formal training.

4.2. Activation Function Comparison

The original ResNet18 framework employs the ReLU activation function. However, in recent years, new activation functions have been continuously proposed and experimentally demonstrated to outperform ReLU in terms of performance, such as ELU [40], SELU [41], SiLU [42], etc. To achieve superior

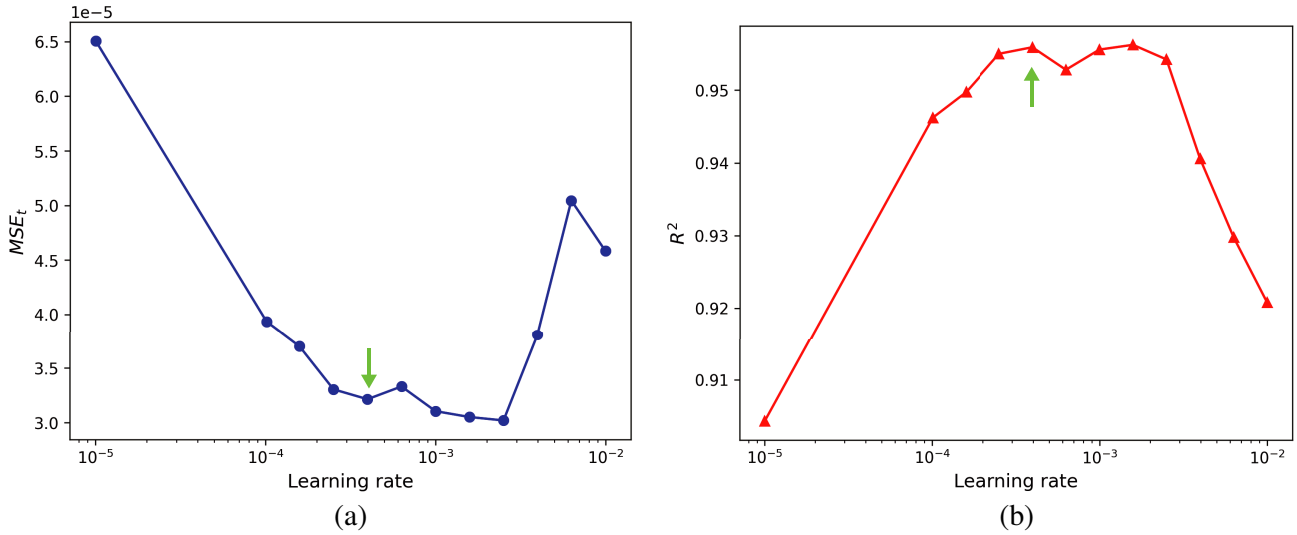

Figure 3. Comparison of the model performance under different learning rates. (a) MSE_t under various learning rates; (b) R_t^2 under various learning rates.

model performance, this study compares the performance of 1D-MR18 utilizing these newly introduced activation functions under the same training conditions. The results are presented in Figure 4. It is evident that the 1D-MR18 using the ELU activation function exhibits the best performance, both in terms of MSE_t and R_t^2. Consequently, we adopt the ELU function as the activation function in our network framework.

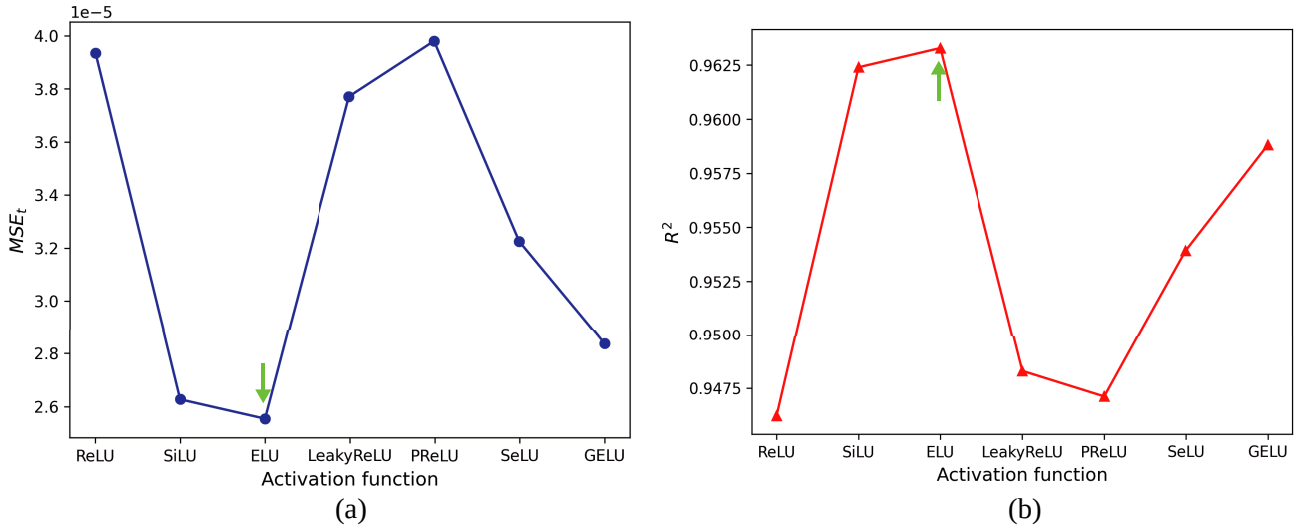

Figure 4. Comparison of the model performance under different activation functions. (a) MSE_t under various activation functions; (b) R_t^2 under various activation functions.

4.3. Batch Size Comparison

Batch Normalization is a regularization technique used in deep neural networks [46]. It normalizes the input data for each mini-batch and introduces learnable parameters to adjust the normalized results. By normalizing the data for each mini-batch, Batch Normalization ensures that the inputs to each layer have similar distributions, reducing internal covariate shift and accelerating model convergence [46].

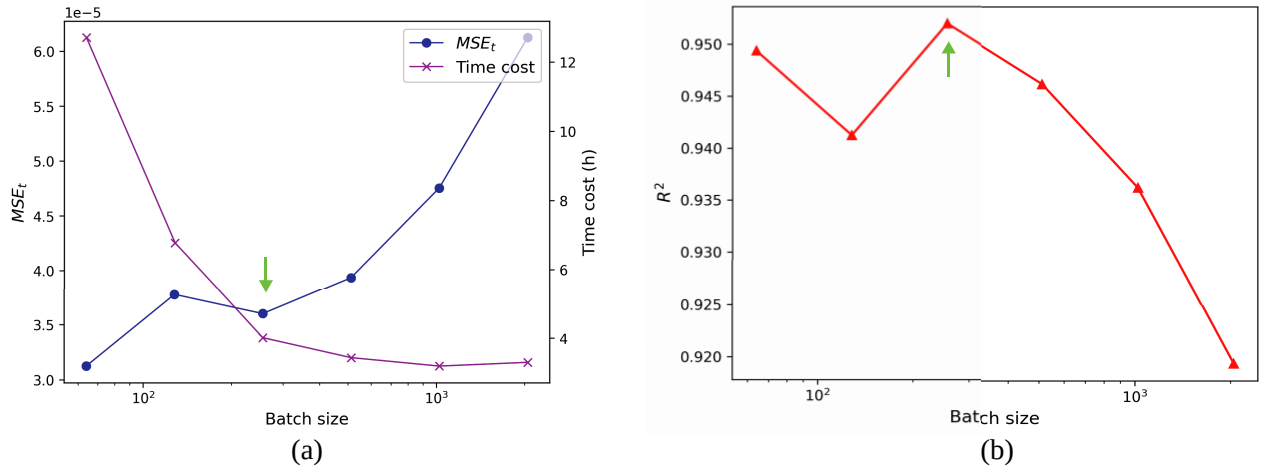

Figure 5. Comparison of model performance under different batch sizes. (a) MSE_t and training time loss under various batch sizes; (b) R_t^2 under various batch sizes.

Selecting an appropriate batch size is a crucial hyperparameter tuning process for both the effectiveness of Batch Normalization and the overall training performance of the model. Larger batch sizes can provide more accurate estimates of mean and variance for Batch Normalization, reduce training time, and enhance training stability [47, 48]. However, they may increase the model's runtime memory footprint and diminish the model's ability to handle variations in gradients, leading to a decline in model performance [47, 48]. Smaller batch sizes can reducing runtime memory and improve the model's generalization performance, but they may prolong training time and reduce the model's resilience to noise [49]. We conducted training using different batch sizes and compared their MSE_t, R_t^2, and training time, as shown in Figure 5. From the perspective of MSE_t and R_t^2, when the batch size is small, the performance of the model varies only slightly. As the batch size increases, the performance of the model gradually declines. However, considering the training time, smaller batch sizes result in exponentially increasing training time. To strike a balance between performance and training time, we selected a batch size of 256 as the optimal parameter for the formal training of the model.

4.4. Model Evaluation

We conducted formal training based on the selected hyperparameters as discussed above. The aforementioned discussion results can provide reasonably good choices for hyperparameters, but a more fine-grained hyperparameter search may lead to even better performance. The variation of the loss function during the training process is illustrated in Figure 6(a). After 600 epochs of training, the loss functions for both the training and test sets no longer decrease, indicating that the model has reached sufficient stability. At this stage, the model's MSE_t is 2.18×10^{-5}, and R_t^2 is 0.969, demonstrating that the model's predictions on unknown data are closely aligned with the true values. We also performed statistical analysis on the model's average prediction error on the test set. The calculation for the average prediction error is as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{8}$$

where y_i represents the ground truth values, which is the coupling efficiency obtained from FDTD simulations, and $\hat{y}_i$ represents the predicted values, which is the model's predicted coupling efficiency. N is the number of samples, which is the discrete number of wavelengths and in this study is 200. The statistical results of the model's average prediction error on the test set are presented in Figure 6(b). The model's prediction error is mostly within the range of 0 to 0.01, with an overall average prediction error of 0.003. This indicates that our model's predictions are very close to the true values. Since the

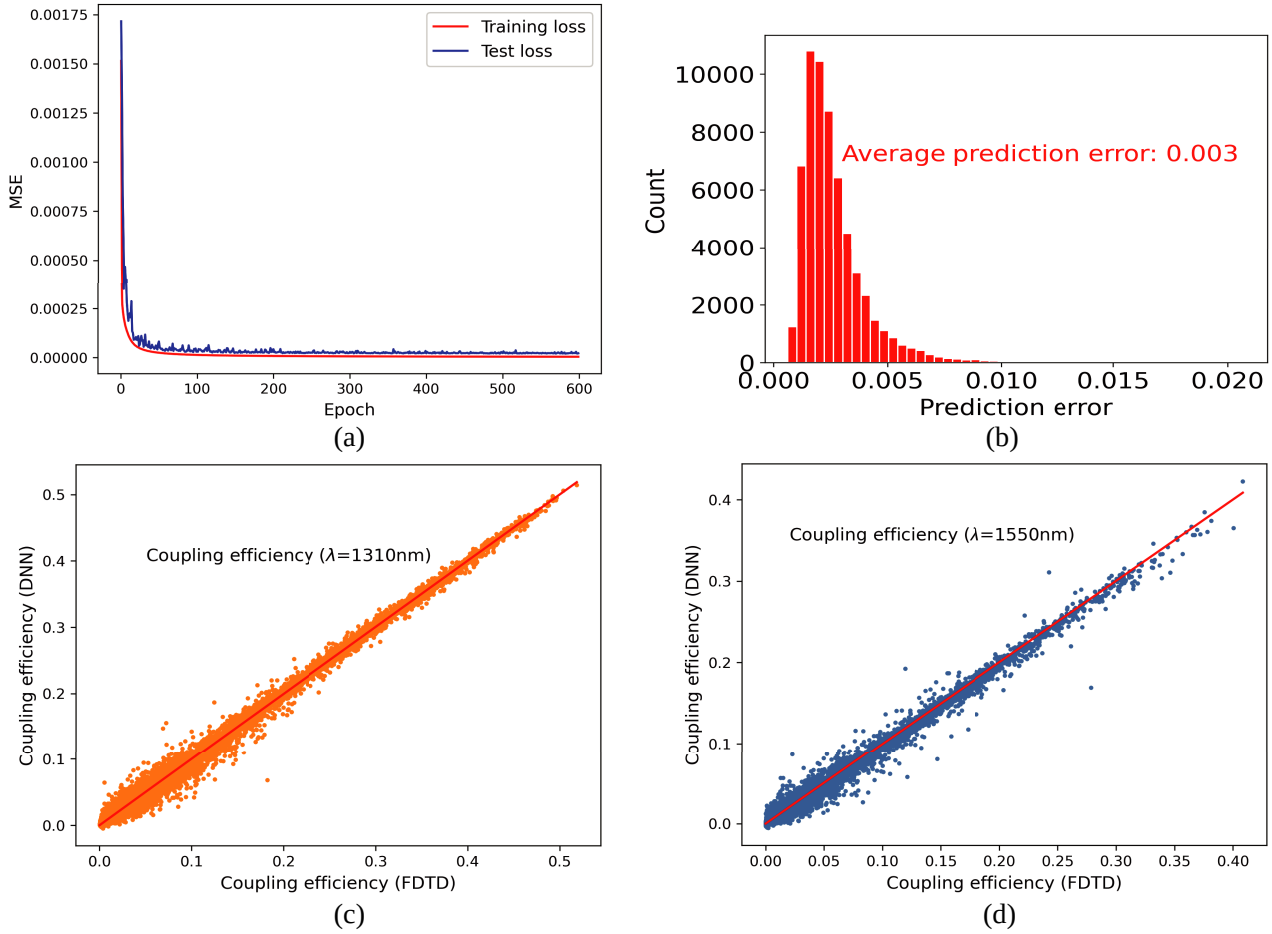

Figure 6. (a) Variation curve of the loss function during the training process; (b) Statistical plot of the prediction error of the model on the test set; (c) Correlation between the model's predictions and the actual values at 1310 nm; (d) Correlation between the model's predictions and the actual values at 1550 nm.

common operating wavelengths for the coupling grating are 1310 nm and 1550 nm, we also compared the correlation between the model's predictions and the ground truth values at these two wavelengths, as shown in Figure 6(c) and Figure 6(d). All data points are closely aligned around the $y = x$ line, indicating that our model's predictions are in good agreement with the true values obtained from FDTD simulations.

5. GRATING COUPLER DESIGN

We then proceed to use the trained deep learning model for grating coupler design. Our design process is illustrated in Figure 7(a), where we employ an Adaptive Particle Swarm Optimization (APSO) algorithm [50] to search the design space of the device. APSO is a variant of the Particle Swarm Optimization algorithm that incorporates adaptive mechanisms to dynamically adjust particle swarm parameters, thereby enhancing search efficiency [50]. During this process, our deep learning model replaces the numerical model to calculate the Figure of Merit (FOM) value for each individual, and the results are fed back into the APSO algorithm for iterative optimization. The comparison between our design method and traditional design methods is presented in Figure 7(a) and Figure 7(b). Notably, our method takes an average of only 0.06 s to compute the FOM for a particle individual, while using the numerical simulator FDTD requires 19 s. Consequently, our approach achieves a speedup

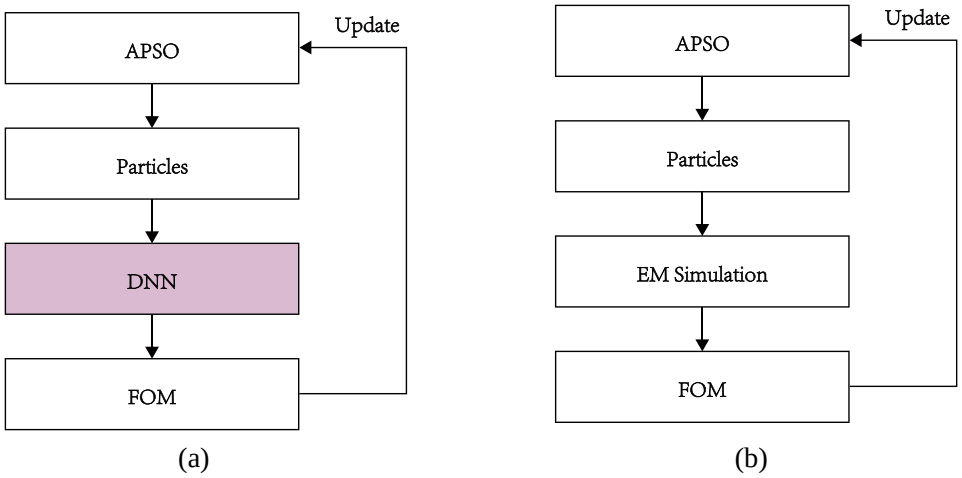

Figure 7. (a) A global optimization design process using the deep neural network (DNN); (b) A global optimization design process using a numerical simulation method (FDTD) for electromagnetic wave propagation.

of approximately 300 times in terms of solution speed. Since we employ a general-purpose grating coupler structure, our deep learning model can be applied to the design of grating couplers for various functionalities.

5.1. Single-Wavelength Grating Coupler Design

Single-wavelength couplers are common designs for grating couplers, aiming to achieve maximum coupling efficiency for a specific wavelength [26]. The primary function of these grating couplers is to couple light of a single wavelength between optical fibers and integrated platforms. Higher coupling efficiency helps reduce signal loss and improves the transmission performance of the system. This feature finds crucial applications in areas such as optical communication [51], optical sensing [52], and LiDAR [5]. The commonly used operating wavelengths for grating couplers are 1310 nm and 1550 nm. Using our design methodology, we optimize grating couplers for both of them. The objective function we employ aims to maximize the coupling efficiency at the target wavelength, formulated as follows:

$$FOM = Eff_{obj} \tag{9}$$

where, Eff_{obj} represents the coupling efficiency at the target wavelength.

The design results are presented in Figure 8, showing that the grating couplers designed for the four operating wavelengths exhibit high coupling efficiency. Our designed single-wavelength couplers do not outperform existing literature results, where the highest coupling efficiency achieved for a grating coupler at the operating wavelength of 1550 nm is 72.8% [3]. This discrepancy could be attributed to the consideration of only 1/5 of the overall device parameters and the restriction on the total length of the grating in our design, whereas adopting more flexible device structures could potentially yield higher coupling efficiency grating couplers. Compared to traditional design methods, our approach demonstrates significant speed advantages. The design of the grating couplers for the two operating wavelengths can be completed within one minute, with each of them taking only 21 seconds. In contrast, traditional optimization methods typically require several hours to several tens of hours for similar designs [6, 7]. The corresponding parameters of each grating coupler in Figure 8 are presented in Table 1 (θ and h have been converted back to angles and etch depths respectively for data visualization).

5.2. Multi-Wavelength Grating Coupler Design

Multi-wavelength grating couplers are designed to achieve efficient coupling of multiple wavelengths, transmitting them simultaneously into the same waveguide. By using these multi-wavelength couplers, different wavelengths of light signals can be transmitted within the same optical fiber, contributing to

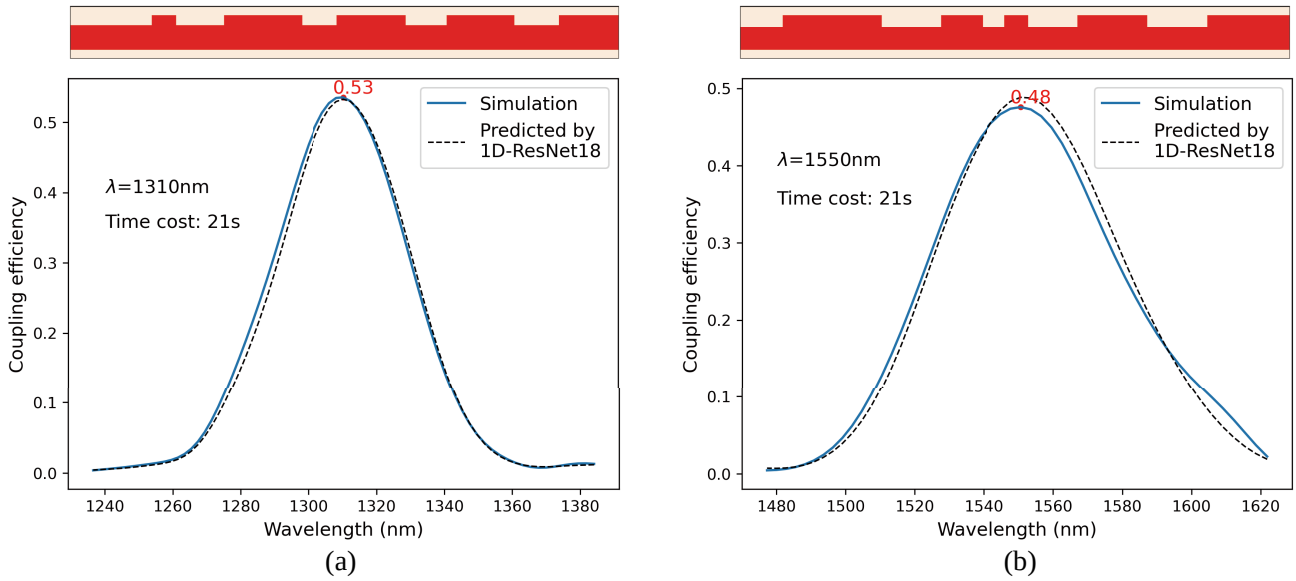

Figure 8. Design results for single-wavelength grating coupler; (a) $\lambda = 1310\,\text{nm}$; (b) $\lambda = 1550\,\text{nm}$.

Table 1. Corresponding parameters of each optimized single-wavelength grating coupler in Fig. 8.

λ (nm)	θ (°)	h (nm)	d_1 (μm)	w_1 (μm)	d_2 (μm)	w_2 (μm)	d_3 (μm)	w_3 (μm)	d_4 (μm)	w_4 (μm)	d_5 (μm)	w_5 (μm)
1310	8.4	61	0.356	0.109	0.21	0.339	0.151	0.3	0.184	0.29	0.199	0.262
1550	5.5	79	0.182	0.431	0.266	0.176	0.1	0.1	0.218	0.308	0.261	0.358

increased communication bandwidth and improved system transmission capacity [53]. The optimization objective for multi-wavelength grating couplers typically seeks to maximize the coupling efficiency at multiple wavelengths while maintaining similar coupling efficiencies across different operating wavelengths. To achieve these two objectives, our design employs the following optimization objective function:

$$FOM = \sum_{i=1}^{K} Eff_{obj_i} - \beta \cdot \sum_{i=1}^{K} \left| Eff_{obj_i} - \overline{Eff}_{obj} \right| \tag{10}$$

where K represents the number of operating wavelengths to be optimized; Eff_{obj_i} denotes the coupling efficiency at the i-th operating wavelength; $\overline{Eff}_{obj}$ represents the average coupling efficiency across all target wavelengths; and β is a weight adjustment factor.

We have determined through experimentation that β is equal to 1 and selected four sets of dual-wavelength couplers from commonly used wavelength bands for design, as shown in Figure 9. The results demonstrate the effectiveness of our method in tailoring designs according to specific requirements. The coupling efficiency for dual-wavelength operation is generally lower than single-wavelength operation, indicating that satisfying the requirements for dual-wavelength devices demands more stringent device structures with higher degrees of freedom in device parameters. The dual-wavelength designs presented in Figure 9 can also be rapidly designed in approximately one minute, showcasing the superior speed advantages of our model-based design approach. The corresponding parameters of each grating coupler in Figure 9 are presented in Table 2 (θ and h have been converted back to angles and etch depths respectively for data visualization).

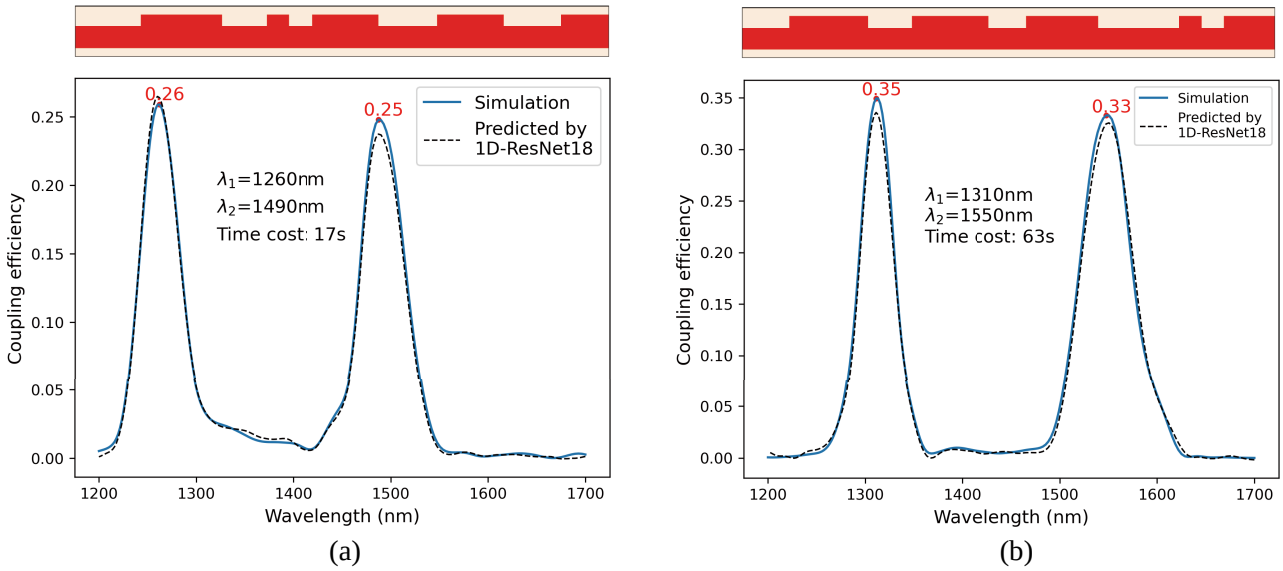

Figure 9. Design results for dual-wavelength grating couplers; (a) $\lambda_1 = 1260\,\text{nm}$, $\lambda_2 = 1490\,\text{nm}$; (b) $\lambda_1 = 1310\,\text{nm}$, $\lambda_2 = 1550\,\text{nm}$.

Table 2. Corresponding parameters of each optimized dual-wavelength grating coupler in Figure 9.

λ (nm)	θ (°)	h (nm)	d_1 (μm)	w_1 (μm)	d_2 (μm)	w_2 (μm)	d_3 (μm)	w_3 (μm)	d_4 (μm)	w_4 (μm)	d_5 (μm)	w_5 (μm)
1260 1490	13.1	78	0.294	0.367	0.2	0.1	0.104	0.299	0.265	0.297	0.267	0.208
1310 1550	5.4	81	0.214	0.352	0.2	0.346	0.165	0.327	0.367	0.1	0.1	0.229

5.3. Robust Grating Coupler Design

During the etching process of grating couplers, the manufactured device structures may deviate from the nominal design due to the limitations of etching technology precision [54]. These deviations result in certain parameter variations rather than exact replicas of the intended design [54]. In practical applications, if a grating coupler is effective only at a specific operating wavelength, significant changes in coupling efficiency when deviating from that wavelength can render the designed grating coupler impractical [55]. Therefore, ensuring robustness in the designed grating couplers is of paramount importance. Considering robustness during the design process effectively addresses the issues arising from limitations in etching technology precision, thereby enhancing the reliability and applicability of the devices.

To address the etching precision issue in grating couplers, we treat the etching accuracy error as corrosion or expansion of the grating teeth. During the design process, we integrate the original device's coupling efficiency with the target coupling efficiency of the dilated or eroded device and calculate the difference with the target coupling efficiency of the original device. Let the original device's structural parameters be denoted as $S_{original}$, the structural parameters after dilating the grating teeth by a nm as S_{dilate_a}, and the structural parameters after eroding the grating teeth by a nm as S_{erode_a}. The relationship is expressed as:

$$S_{dilate_a}: \ \theta = \theta,\ h = h,\ w_i \leftarrow w_i - a,\ d_i \leftarrow d_i + a,\ i = 1, 2, \ldots, M \tag{11}$$

$$S_{erode_a}: \ \theta = \theta,\ h = h,\ w_i \leftarrow w_i + a,\ d_i \leftarrow d_i - a,\ i = 1, 2, \ldots, M \tag{12}$$

The corresponding objective function is as follows:

$$FOM = Eff_{obj}(S_{original}) - \beta \cdot \left|Eff_{obj}(S_{original}) - Eff_{obj}(S_{dilate_a})\right| - \beta \cdot \left|Eff_{obj}(S_{original}) - Eff_{obj}(S_{erode_a})\right| \quad (13)$$

where $Eff_{obj}(S_{original})$ represents the target coupling efficiency of the original device's structure; $Eff_{obj}(S_{dilate_a})$ represents the target coupling efficiency of the device with the grating teeth dilated by a nm; $Eff_{obj}(S_{erode_a})$ represents the target coupling efficiency of the device with the grating teeth eroded by an nm; and β is a weight adjustment factor.

To address the significant decrease in efficiency caused by deviations in the incident angle, we integrate the target coupling efficiency of the original incident angle with the target coupling efficiency when deviating from the original incident angle, and calculate the difference with the target coupling efficiency at the original incident angle. Let the structural parameters of the device when deviating left

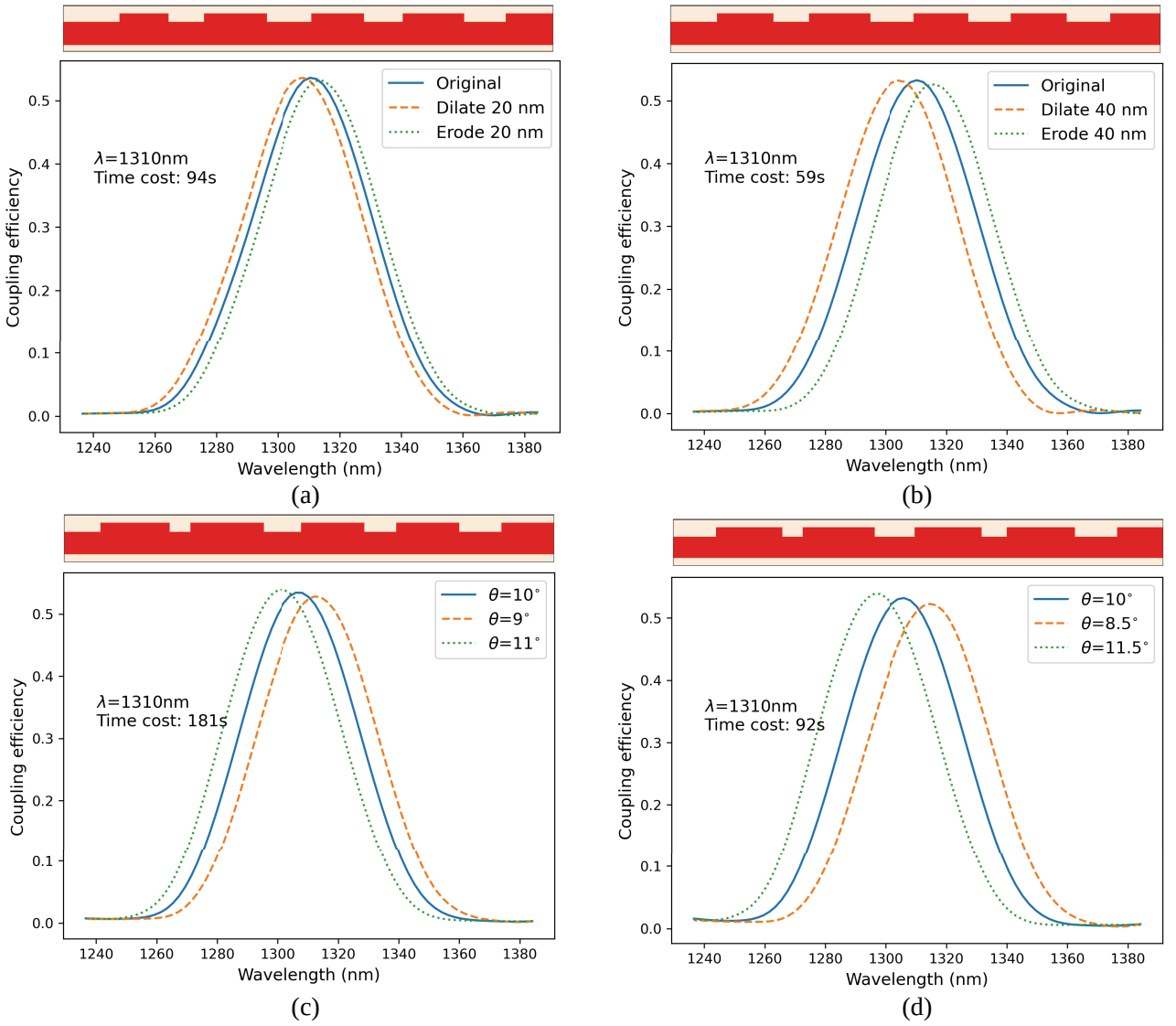

Figure 10. Designs result for robust grating couplers. (a) $\lambda = 1310$ nm, dilating or eroding by 20 nm; (b) $\lambda = 1310$ nm, dilating or eroding by 40 nm; (c) $\lambda = 1310$ nm, the incident angle is set to 10 degrees with deviations of 1 degree left and right; (d) $\lambda = 1310$ nm, the incident angle is set to 10 degrees with deviations of 1.5 degrees left and right.

from the angle α be denoted as $S_{-\alpha}$, and the structural parameters of the device when deviating right from the angle α be denoted as S_{α}. The relationship is expressed as:

$$S_{-\alpha}: \ \theta \leftarrow \theta - \alpha,\ h = h,\ w_i = w_i,\ d_i = d_i,\ i = 1, 2, \ldots, M \tag{14}$$

$$S_{\alpha}: \ \theta \leftarrow \theta + \alpha,\ h = h,\ w_i = w_i,\ d_i = d_i,\ i = 1, 2, \ldots, M \tag{15}$$

The corresponding objective function is as follows:

$$FOM = \mathit{Eff}_{obj} - \beta \cdot \left|\mathit{Eff}_{obj} - \mathit{Eff}_{obj_{-\alpha}}\right| - \beta \cdot \left|\mathit{Eff}_{obj} - \mathit{Eff}_{obj_{\alpha}}\right| \tag{16}$$

where Eff_{obj} represents the target coupling efficiency at the original incident angle; $\mathit{Eff}_{obj_{-\alpha}}$ represents the coupling efficiency when deviating left by α degrees from the original incident angle; $\mathit{Eff}_{obj_{\alpha}}$ represents the coupling efficiency when deviating right by α degrees from the original incident angle; and β is a weight adjustment factor.

We have determined through experimentation that β is equal to 1 and have designed grating couplers for the working wavelength of 1310 nm with three different scenarios: dilating or eroding the grating by 20 nm, dilating or eroding the grating by 40 nm, and setting the incident angle at 10 degrees while deviating left and right by 1 degree, and deviating left and right by 1.5 degrees. As shown in Figure 10, the coupling efficiency curves for dilating or eroding by 20 nm and 40 nm closely overlap with the original curve. For the incident angle deviations of 1 degree and 1.5 degrees, although the coupling curves exhibit slight deviations from the original coupling efficiency curve, these variations are within an acceptable range. The coupling efficiencies at the working wavelength remain relatively close to the original values. This indicates that our method is capable of designing grating couplers effectively, even in the presence of precision errors and incident angle deviations. The corresponding parameters of each grating coupler in Figure 10 are presented in Table 3 (θ and h have been converted back to angles and etch depths respectively for data visualization).

Table 3. Corresponding parameters of each optimized robust grating coupler in Figure 10.

λ (nm)	θ (°)	h (nm)	d_1 (μm)	w_1 (μm)	d_2 (μm)	w_2 (μm)	d_3 (μm)	w_3 (μm)	d_4 (μm)	w_4 (μm)	d_5 (μm)	w_5 (μm)
1310	8.8	58	0.276	0.239	0.178	0.303	0.22	0.257	0.19	0.298	0.207	0.23
1310	8.9	58	0.23	0.271	0.178	0.297	0.214	0.266	0.211	0.272	0.215	0.246
1310	10	63	0.177	0.339	0.1	0.363	0.186	0.305	0.16	0.307	0.206	0.256
1310	10	65	0.21	0.321	0.1	0.352	0.199	0.324	0.129	0.33	0.209	0.227

6. SUMMARY

Using deep learning as a replacement for numerical solvers for electromagnetic wave propagation is a promising approach for rapid photonics device design. However, deep learning often requires significant computational resources and time-consuming operations, such as dataset preparation and model training. Thus, to fully exploit the advantages of fast design using deep learning, it should be applied to a general device structure. In this study, we propose a universal periodic nonuniform grating structure and employ a modified 1D-ResNet18 (1D-MR18) neural network for rapid design. We begin by discussing and comparing the impacts of learning rate, activation functions, and batch normalization size on the 1D-MR18 and determine its optimal parameter settings. We evaluate the predictive performance of 1D-MR18 for grating coupler efficiency using metrics such as MSE, R^2, MAE, and correlations between the predicted values of 1D-MR18 at 1310 nm and 1550 nm and the truth values. The results demonstrate that 1D-MR18 accurately predicts grating coupler efficiency with low MSE (as low as 2.18×10^{-5}), high R^2 (up to 0.969), and low MAE (as low as 0.0003). We combine 1D-MR18 with the APSO Algorithm for designing general periodic nonuniform grating structures. This approach efficiently designs grating couplers with various functional requirements, including single-wavelength couplers, multi-wavelength grating couplers, and robust grating couplers. The time cost for designing a single device ranges from a maximum of 2 minutes to as short as 17 seconds. Although the

designed grating coupler efficiency may not surpass existing literature due to structural constraints, it meets the demands of both scientific research and industrial production. Moreover, considering more flexible grating coupler structures would give higher coupling efficiency, indicating significant potential for our method in grating coupler design. Our rapid and efficient approach, applying deep learning to general-purpose photonic devices, offers insights and guidance for silicon photonics device design.

ACKNOWLEDGMENT

The authors are grateful to Dr. Julian Evans, Dr. Zhipeng Hu, Dr. Tingbiao Guo, and Lei Wang of Zhejiang University for valuable discussions.

FUNDING

This research was funded by the National Key Research and Development Program of China (2022YFB2804100), the Special Development Fund of Shanghai Zhangjiang Science City, and National Natural Science Foundation of China (11621101).

REFERENCES

1. Nevière, M. and E. Popov, *Light Propagation in Periodic Media: Differential Theory and Design*, CRC Press, 2002.
2. Mu, X., S. Wu, L. Cheng, and H. Fu, "Edge couplers in silicon photonic integrated circuits: A review," *Applied Sciences*, Vol. 10, No. 4, 1538, 2020.
3. Marchetti, R., C. Lacava, A. Khokhar, X. Chen, I. Cristiani, D. J. Richardson, G. T. Reed, P. Petropoulos, and P. Minzioni, "High-efficiency grating-couplers: Demonstration of a new design strategy," *Scientific Reports*, Vol. 7, 16670, Nov. 2017.
4. Chrostowski, L. and M. Hochberg, *Silicon Photonics Design: From Devices to Systems*, Cambridge University Press, 2015.
5. Cheng, L., S. Mao, Z. Li, Y. Han, and H. Y. Fu, "Grating couplers on silicon photonics: Design principles, emerging trends and practical issues," *Micromachines*, Vol. 11, No. 7, 666, 2020.
6. Djavid, M., S. A. Mirtaheri, and M. S. Abrishamian, "Photonic crystal notch-filter design using particle swarm optimization theory and finite-difference time-domain analysis," *J. Opt. Soc. Am. B*, Vol. 26, 849–853, Apr. 2009.
7. Shen, L., Z. Ye, and S. He, "Design of two-dimensional photonic crystals with large absolute band gaps using a genetic algorithm," *Phys. Rev. B*, Vol. 68, 035109, Jul. 2003.
8. Assefa, S., F. Xia, and Y. A. Vlasov, "Reinventing germanium avalanche photodetector for nanophotonic on-chip optical interconnects," *Nature*, Vol. 464, 80–84, Mar. 2010.
9. Peurifoy, J., 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, eaar4206, 2018.
10. Wu, Y.-D., "High efficiency multi-functional all-optical logic gates based on MIM plasmonic waveguide structure with the Kerr-type nonlinear nano-ring resonators," *Progress In Electromagnetics Research*, Vol. 170, 79–95, 2021.
11. LeCun, Y., Y. Bengio, and G. Hinton, "Deep learning," *Nature*, Vol. 521, 436–444, May 2015.
12. Zhang, Z., J. Geiger, J. Pohjalainen, A. E.-D. Mousa, W. Jin, and B. Schuller, "Deep learning for environmentally robust speech recognition: An overview of recent developments," *ACM Trans. Intell. Syst. Technol.*, Vol. 9, Apr. 2018.
13. Pak, M. and S. Kim, "A review of deep learning in image recognition," *2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT)*, 1–3, 2017.
14. Otter, D. W., J. R. Medina, and J. K. Kalita, "A survey of the usages of deep learning for natural language processing," *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 32, No. 2, 604–624, 2021.

15. Ho-Phuoc, T., "CIFAR10 to compare visual recognition performance between deep neural networks and humans," CoRR, arXiv:1811.07270, 2018.
16. Shan, T., M. Li, S. Xu, and F. Yang, "Phase synthesis of beam-scanning reflectarray antenna based on deep learning technique," *Progress In Electromagnetics Research*, Vol. 172, 41–49, 2021.
17. An, S., B. Zheng, M. Y. Shalaginov, H. Tang, H. Li, L. Zhou, J. Ding, A. M. Agarwal, C. Rivero-Baleine, M. Kang, K. A. Richardson, T. Gu, J. Hu, C. Fowler, and H. Zhang, "Deep learning modeling approach for metasurfaces with high degrees of freedom," *Opt. Express*, Vol. 28, 31932–31942, Oct. 2020.
18. So, S., J. Mun, and J. Rho, "Simultaneous inverse design of materials and structures via deep learning: Demonstration of dipole resonance engineering using core-shell nanoparticles," *ACS Applied Materials & Interfaces*, Vol. 11, 24264–24268, Jul. 2019.
19. Jiang, J. and J. A. Fan, "Multiobjective and categorical global optimization of photonic structures based on resnet generative neural networks," *Nanophotonics*, Vol. 10, No. 1, 361–369, 2021.
20. Jiang, H., J. Chen, K. Tang, H. Yan, and R. Hao, "Exploring high-performance slow light grating waveguides by means of deep learning," *IEEE Photonics Technology Letters*, Vol. 34, No. 20, 1112–1115, 2022.
21. Miyatake, Y., N. Sekine, K. Toprasertpong, S. Takagi, and M. Takenaka, "Computational design of efficient grating couplers using artificial intelligence," *Japanese Journal of Applied Physics*, Vol. 59, No. SG, SGGE09.1–SGGE09.8, 2019.
22. Tu, X., W. Xie, Z. Chen, M.-F. Ge, T. Huang, C. Song, and H. Y. Fu, "Analysis of deep neural network models for inverse design of silicon photonic grating coupler," *Journal of Lightwave Technology*, Vol. 39, No. 9, 2790–2799, 2021.
23. Tahersima, M. H., 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, 1368, Feb. 2019.
24. Taillaert, D., P. Bienstman, and R. Baets, "Compact efficient broadband grating coupler for silicon on-insulator waveguides," *Opt. Lett.*, Vol. 29, 2749–2751, Dec. 2004.
25. Na, N., H. Frish, I.-W. Hsieh, O. Harel, R. George, A. Barkai, and H. Rong, "Efficient broadband silicon-on-insulator grating coupler with low backreflection," *Opt. Lett.*, Vol. 36, 2101–2103, Jun. 2011.
26. Orobtchouk, R., A. Layadi, H. Gualous, D. Pascal, A. Koster, and S. Laval, "High-efficiency light coupling in a submicrometric silicon-on-insulator waveguide," *Appl. Opt.*, Vol. 39, 5773–5777, Nov. 2000.
27. Michaels, A. and E. Yablonovitch, "Inverse design of near unity efficiency perfectly vertical grating couplers," *Opt. Express*, Vol. 26, 4766–4779, Feb. 2018.
28. Chen, X., D. J. Thomson, L. Crudginton, A. Z. Khokhar, and G. T. Reed, "Dual-etch apodised grating couplers for efficient fibre-chip coupling near 1310 nm wavelength," *Opt. Express*, Vol. 25, 17864–17871, Jul. 2017.
29. Sánchez-Postigo, A., J. G. Wangüemert-Pérez, J. M. Luque-González, I. Molina-Fernández, P. Cheben, C. A. Alonso-Ramos, R. Halir, J. H. Schmid, and A. Ortega-Moñux, "Broadband fiber-chip zero-order surface grating coupler with 0.4 dB efficiency," *Opt. Lett.*, Vol. 41, 3013–3016, Jul. 2016.
30. Su, L., R. Trivedi, N. V. Sapra, A. Y. Piggott, D. Vercruysse, and J. Vučković, "Fully-automated optimization of grating couplers," *Opt. Express*, Vol. 26, 4023–4034, Feb. 2018.
31. Rahim, A., T. Spuesens, R. Baets, and W. Bogaerts, "Open-access silicon photonics: Current status and emerging initiatives," *Proceedings of the IEEE*, Vol. 106, No. 12, 2313–2330, 2018.
32. Roelkens, G., D. V. Thourhout, and R. Baets, "High efficiency silicon-on-insulator grating coupler based on a poly-silicon overlay," *Opt. Express*, Vol. 14, 11622–11630, Nov. 2006.

33. Taillaert, D., W. Bogaerts, P. Bienstman, T. Krauss, P. Van Daele, I. Moerman, S. Verstuyft, K. De Mesel, and R. Baets, "An out-of-plane grating coupler for efficient butt-coupling between compact planar waveguides and single-mode fibers," *IEEE Journal of Quantum Electronics*, Vol. 38, No. 7, 949–955, 2002.
34. Taflove, A., S. C. Hagness, and M. Piket-May, "Computational electromagnetics: The finite difference time-domain method," *The Electrical Engineering Handbook*, Vol. 3, No. 15, 629–670, 2005.
35. Bérenger, J.-P., *Perfectly Matched Layer (PML) for Computational Electromagnetics*, Springer Nature, 2022.
36. Gedney, S. D. and B. Zhao, "An auxiliary differential equation formulation for the complex frequency shifted PML," *IEEE Transactions on Antennas and Propagation*, Vol. 58, No. 3, 838–847, 2010.
37. He, K., X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016.
38. He, K., X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," *Computer Vision — ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, eds., 630–645, Springer International Publishing, Cham, 2016.
39. Glorot, X. and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Y. W. Teh and M. Titterington, eds., Vol. 9 of Proceedings of Machine Learning Research, PMLR, 249–256, Chia Laguna Resort, Sardinia, Italy, May 13–15, 2010.
40. Clevert, D.-A., T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUS)," arXiv:1511.07289, 2015.
41. Klambauer, G., T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-normalizing neural networks," *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds., Vol. 30, Curran Associates, Inc., 2017.
42. Elfwing, S., E. Uchibe, and K. Doya, "Sigmoid-weighted linear units for neural network function approximation in reinforcement learning," *Neural Networks*, Vol. 107, 3–11, Special issue on deep reinforcement learning, 2018.
43. Dozat, T., "Incorporating nesterov momentum into adam," *ICLR 2016*, 2016.
44. Chicco, D., M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation," *PeerJ Computer Science*, Vol. 7, e623, 2021.
45. Smith, L. N., "A disciplined approach to neural network hyper-parameters: Part 1 — Learning rate, batch size, momentum, and weight decay," CoRR, abs/1803.09820, 2018.
46. Ioffe, S. and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *Proceedings of the 32nd International Conference on Machine Learning*, F. Bach and D. Blei, eds., Vol. 37 of Proceedings of Machine Learning Research, PMLR, 448–456, Lille, France, Jul. 7–9, 2015.
47. Goyal, P., P. Dollár, R. B. Girshick, P. Noordhuis, L. Wesolowski, A. Kyrola, A. Tulloch, Y. Jia, and K. He, "Accurate, large minibatch SGD: Training imagenet in 1 hour," CoRR, abs/1706.02677, 2017.
48. Masters, D. and C. Luschi, "Revisiting small batch training for deep neural networks," CoRR, abs/1804.07612, 2018.
49. Smith, S. L., P. Kindermans, and Q. V. Le, "Don't decay the learning rate, increase the batch size," CoRR, abs/1711.00489, 2017.
50. Zhang, Z., Y. Jiang, S. Zhang, S. Geng, H. Wang, and G. Sang, "An adaptive particle swarm optimization algorithm for reservoir operation optimization," *Applied Soft Computing*, Vol. 18, 167–177, 2014.

51. Wang, J., C. Gui, C. Li, Q. Yang, D. Gao, J. Sun, and X. Zhang, "On-chip demultiplexing of polarization and wavelength multiplexed of DM/OQAM 64/128-QAM signals using silicon 2D grating coupler and microring resonators," *Optical Fiber Communication Conference*, Th2A.48, Optica Publishing Group, 2014.
52. Li, H.-Y., W.-C. Hsu, K.-C. Liu, Y.-L. Chen, L.-K. Chau, S. Hsieh, and W.-H. Hsieh, "A low cost, label-free biosensor based on a novel double-sided grating waveguide coupler with sub-surface cavities," *Sensors and Actuators B: Chemical*, Vol. 206, 371–380, 2015.
53. Xu, L., X. Chen, C. Li, and H. K. Tsang, "Bi-wavelength two dimensional chirped grating couplers for low cost WDM pon transceivers," *Optics Communications*, Vol. 284, No. 8, 2242–2244, 2011.
54. Zhou, X. and H. K. Tsang, "Photolithography fabricated sub-decibel high-efficiency silicon waveguide grating coupler," *IEEE Photonics Technology Letters*, Vol. 35, No. 1, 43–46, 2023.
55. Duan, L., C. Xu, S. Zhong, H. Zhou, and J. Duan, "Optical power auto-alignment method with eugenics sorting for enhancing the alignment speed and robustness of fiber-grating couplers," *Opt. Express*, Vol. 30, 39544–39560, Oct. 2022.