

Review

Deep Learning and Adjoint Method Accelerated Inverse Design in Photonics: A Review

Zongyong Pan and Xiaomin Pan *

School of Cyberspace Science and Technology, Beijing Institute of Technology, Beijing 100081, China; 3120221340@bit.edu.cn

* Correspondence: xmpan@bit.edu.cn

Abstract: For photonic applications, the inverse design method plays a critical role in the optimized design of photonic devices. According to its two ingredients, inverse design in photonics can be improved from two aspects: to find solutions to Maxwell's equations more efficiently and to employ a more suitable optimization scheme. Various optimization algorithms have been employed to handle the optimization: the adjoint method (AM) has become the one of the most widely utilized ones because of its low computational cost. With the rapid development of deep learning (DL) in recent years, inverse design has also benefited from DL algorithms, leading to a new pattern of photon inverse design. Unlike the AM, DL can be an efficient solver of Maxwell's equations, as well as a nice optimizer, or even both, in inverse design. In this review, we discuss the development of the AM and DL algorithms in inverse design, and the advancements, advantages, and disadvantages of the AM and DL algorithms in photon inverse design.

Keywords: photonics; inverse design; deep learning; adjoint method

1. Introduction

With inverse design methods, we can optimize the parameters of the integrated device, obtain a specific structure of the device, or both. Among many well-established inverse design methods, two methods of traditional inverse design originating from the research group of the University of Utah and the research group of Stanford University [1,2] have received much attention due to their high performance. In general, inverse design always combines a mathematical model with an optimization algorithm: the former describes the underlying physics, and the latter discovers the suitable shape or parameters, according to the user-predefined target performance of the device.

According to its two ingredients, inverse design can be improved from two aspects: to find solutions to Maxwell's equations more efficiently and to employ a more suitable optimization scheme. These two types of techniques can be utilized together, and inverse design has already benefited from the rapid development of computational electromagnetic algorithms for various applications. The most popular solvers for Maxwell's equations include the finite element method (FEM), the finite-difference time-domain (FDTD) method [3–6], the method of moments (MoM) [7–9], and hybrid methods [10,11].

Due to traditional computational electromagnetic frameworks becoming mature and the fast development of the various optimization algorithms, many improvements have been developed in terms of the optimization process. Based on what type of optimization scheme is employed, inverse design methods can be classified into gradient-based algorithms and non-gradient ones [12].

Non-gradient algorithms are also known as gradient-free methods. The representative methods belonging to the gradient-free scheme include evolutionary algorithms [13,14] and search algorithms (including nonlinear search methods) [15]. Evolutionary algorithms can be further divided into methods based on genetic algorithms (GA) [13,16], methods



Citation: Pan, Z.-Y.; Pan, X.-M. Deep Learning and Adjoint Method Accelerated Inverse Design in Photonics: A Review. *Photonics* **2023**, *10*, 852. <https://doi.org/10.3390/photonics10070852>

Received: 30 May 2023
Revised: 5 July 2023
Accepted: 14 July 2023
Published: 23 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

based on particle swarm optimization (PSO) [14,17], and hybrid methods tailored to specific problems [18]. GA, a typical heuristic approach, is widely used for its effectiveness, simplicity, and intuitiveness [19]. Heuristic optimization relies on a somewhat limited parameterization of the solution space and subsequent random testing of a large number of parameter sets. PSO is an optimization algorithm based on swarm intelligence, which has a strong global search ability. Except when employed alone, it is always hybridized with other search algorithms to construct a more effective optimization algorithm: for example, by mixing PSO with GA, faster and more accurate global optimization can be achieved. Due to the high computational cost of solving Maxwell's equations, these methods tend to only work for targets with a few unknowns or with relatively simple geometries, as they require testing of a large number of solutions to find a satisfactory optimization result [20]. Many applications demand the sensitivity of some parameters along with acceptable optimization results: in essence, the sensitivity is the gradients of the optimization objective function for the given optimization parameters. Although the sensitivity can be calculated by using the method of finite difference (FD) approximation, the associated computational cost increases in proportion to the product of the problem size and the number of design variables, which quickly makes the computation prohibitive.

Gradient-based optimization is probably the most widely used technique in photonic inverse design: in this scheme, the gradients of the objective function, relative to all design parameters of the device, are calculated in each iteration; the design parameters are then changed in the gradient direction to improve the performance of the device. The AM [21,22] and the level set method [23–25] are two typical methods belonging to the gradient-based optimization scheme. The level set algorithm is a numerical method based on partial differential equations, which is used to describe the contour changes of an object over a specified variable. At a high level, the major benefit offered by the level set is that it provides a systematic way to organize design possibilities. The AM is one of the most widely used algorithms in gradient-based photonic inverse design. Because the AM can calculate the derivatives over all the design parameters and because it modifies the parameters in proportion to the figure of merit (FOM) gradient using only forward and adjoint simulations at each iteration, regardless of the number of design parameters [26,27], it has been successfully applied to various photonic systems [27,28]. This is one of the major reasons why the AM has been widely adopted for the topology or shape optimization of photonic devices [28], where the number of design parameters can be very large to describe complex free-form geometries. The traditional AM was recently extended to nonlinear device modeling in the frequency domain [29]. Another attractive virtue of the AM is that it can significantly reduce the computational cost of a sensitivity calculation [30,31].

Although inverse design methods, both gradient-based and non-gradient-based, have achieved significant breakthroughs in developing new optical functions, they can encounter the difficulty of a low efficiency because they inevitably call an electromagnetic solver during the optimization: consequently, their computational costs become very high, as the design parameters to be optimized increase. In addition, gradient-based methods can easily fall into local optimal solutions, resulting in slow convergence. As a result, the inverse design approach is hampered, and other algorithms need to be developed to address these problems.

To solve the inherent difficulties of traditional inverse design methods, deep learning (DL) and its variants have emerged as an alternative solution in recent years [32]. DL, as a subset of machine learning (ML), has the potential to handle tricky high-degree-of-freedom designs [33]. With the fast development of DL and its applications in various fields, many DL networks, such as fully connected networks, convolutional networks, generative networks, and recurrent neural networks, have been utilized in photonic inverse design, and they have achieved remarkable results [34]: for example, predictive and generative models based on data-driven methods have been developed for the analysis and design of photonic crystals (PhC) [35]; in [36], a network to design free-form, all-dielectric metasurface devices was proposed by combining the conditional generative adversarial

network (CGAN) with the Wasserstein generative adversarial network (WGAN). However, there are also many limitations when using DL in photon inverse designs, such as training dataset problems and inverse design non-unique problems. In addition, different designs may be produced with almost identical spectra, which may prevent correct optimization from convergence.

The combination of DL with the AM can take advantage of both techniques and thus alleviate their inherent limitations. For example, such a hybrid method was developed in [37], where photon neural networks were efficiently trained in situ and the AM was utilized to derive the photon simulation of the backpropagation algorithm. In [27], a hybrid inverse design framework was proposed by combining adjoint optimization and automatic ML with explainable AI, where ML and AI were designed to discover geometric features that cause local minima. In [38], a hybrid design approach for electromagnetic devices was developed by directly incorporating adjoint variable computation into the generative neural network, which showed the positive effects arising from the combination of DL and AM.

The rest of this review will be organized as follows. In Section 2, the AM for inverse design is discussed. In Section 3, the DL-equipped inverse design is presented. In Section 4, the hybrid AM and DL designs are reviewed. Finally, some concluding remarks are given in Section 5.

2. AM for Inverse Design

2.1. AM

The AM can effectively calculate gradients and plays an important role in inverse design for photonic applications. A very simple example of using the AM can be found in [20], which visually illustrated the mathematical process in the context of electromagnetism. Although the AM has been mainly applied to linear photonic devices [39–41], it has indeed been extended to model nonlinear devices in the frequency domain [29]. In particular, the method was used to design compact photonic switches in a Kerr nonlinear material, where low-power and high-power pulses were routed in different directions [29].

A schematic outlining the process of applying the AM to both linear and nonlinear cases is shown in Figure 1.

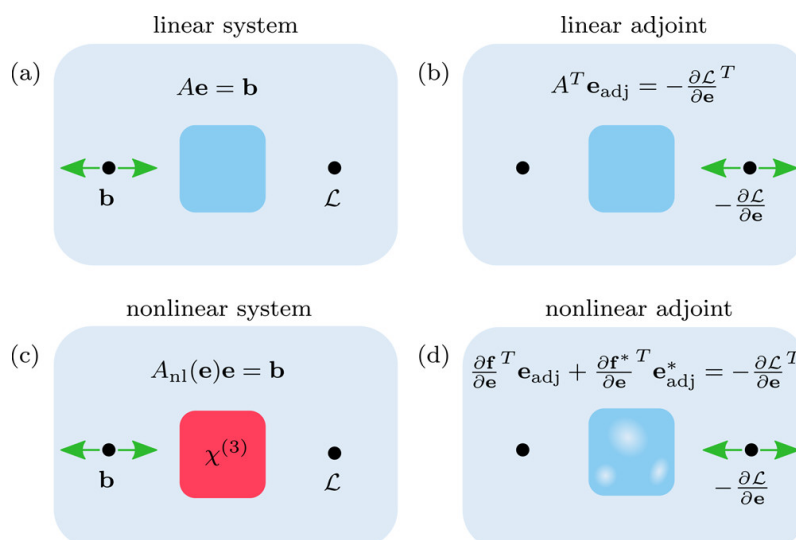


Figure 1. Illustration of the adjoint field computation for linear and nonlinear systems [29]. (a) The field intensity at a measuring point determines the objective function of a linear system, which is driven by a point source b . (b) When the measuring point acts as the source, the adjoint problem for the linear system involves the same system but, in reverse, with the source located at the measuring point. (c) The presence of Kerr nonlinearity (red) makes the system nonlinear. The electric fields are

the solutions to an equation that captures this nonlinearity. (d) In the adjoint problem for the nonlinear system, the Kerr medium is replaced by a linear region. This linear region represents a dependency on the nonlinear fields and results in a set of linear equations for the adjoint field and its complex conjugate.

2.1.1. Linear AM

Maxwell’s equations in terms of a linear system for an optical device naturally make itself compatible with the AM [39–41]. The associated compact linear system can be written as follows:

$$\mathbf{A} \cdot \mathbf{e} = \mathbf{b} \tag{1}$$

Equation (1) is a compact form, where \mathbf{A} is the system matrix, \mathbf{e} represents the electric field, and \mathbf{b} is the excitation vector. Depending on the discretization method, matrix \mathbf{A} can be sparse or dense. In particular, it is sparse when the linear system is generated using the FDTD [42] or the FEM [43] while it is dense when generated using the method of moments (MoM) [44].

Without loss of generality, a general real-valued objective function $\mathcal{L} = \mathcal{L}(\mathbf{e}(\Phi))$ is considered, where \mathbf{e} is the solution vector of Equation (1) and Φ is the vector denoting the design parameters associated with the system. In general, Φ can be any of the design parameters, such as the parameters for the underlying passive structure, as well as those for the active modulators such as the modulation strength or the modulation phases. For example, Φ can be the system object’s dielectric constant or geometric parameters. The goal of the AM is to compute the sensitivity $\frac{\partial \mathcal{L}(\mathbf{e}(\Phi))}{\partial \Phi}$. Suppose that φ is one element in Φ ; what follows shows how to compute the sensitivity by computing $\partial \mathcal{L}(\mathbf{e}(\Phi)) / \partial \varphi$,

$$\frac{\partial \mathcal{L}(\mathbf{e}(\Phi))}{\partial \varphi} = \frac{\partial \mathcal{L}}{\partial \mathbf{e}} \cdot \frac{\partial \mathbf{e}}{\partial \varphi} \tag{2}$$

In general, $\mathcal{L}(\mathbf{e}(\varphi))$ is an explicit function of \mathbf{e} . The term $\frac{\partial \mathcal{L}}{\partial \mathbf{e}}$ can be obtained analytically with negligible computational cost. The other term $\frac{\partial \mathbf{e}}{\partial \varphi}$ can be computed by taking derivatives on both sides of Equation (1) with respect to φ :

$$\frac{\partial \mathbf{A}}{\partial \varphi} \cdot \mathbf{e} + \mathbf{A} \cdot \frac{\partial \mathbf{e}}{\partial \varphi} = 0 \tag{3}$$

According to Equation (3), we have

$$\frac{\partial \mathbf{e}}{\partial \varphi} = -\mathbf{A}^{-1} \cdot \frac{\partial \mathbf{A}}{\partial \varphi} \cdot \mathbf{e} \tag{4}$$

By substituting Equation (4) into Equation (2), we obtain

$$\frac{\partial \mathcal{L}(\mathbf{e}(\varphi))}{\partial \varphi} = -\left(\frac{\partial \mathcal{L}}{\partial \mathbf{e}} \cdot \mathbf{A}^{-1}\right) \cdot \frac{\partial \mathbf{A}}{\partial \varphi} \cdot \mathbf{e} = -\mathbf{e}_{adj}^T \cdot \frac{\partial \mathbf{A}}{\partial \varphi} \cdot \mathbf{e} \tag{5}$$

where T represents transposition and \mathbf{e}_{adj} can be written as

$$\mathbf{A}^T \cdot \mathbf{e}_{adj} = -\left(\frac{\partial \mathcal{L}}{\partial \mathbf{e}}\right)^T \tag{6}$$

If \mathbf{A} is an explicit function in terms of φ , $\frac{\partial \mathbf{A}}{\partial \varphi}$ can also be analytically calculated with a negligible cost. As a result, to evaluate the sensitivity using Equation (5), the solution of the physical field \mathbf{e} and its adjoint counterpart \mathbf{e}_{adj} would like to dominate the computational cost of the whole optimization process. After these two fields are obtained, the sensitivity of the photonic device concerning any number of parameters can be evaluated with a few additional computational costs.

2.1.2. Nonlinear AM

Although the AM was originally developed to handle linear systems, it has been extended to nonlinear systems [29]. In [29], a set of parameters Φ were optimized that optimize a real-valued objective function $\mathcal{L} = \mathcal{L}(e, e^*, \Phi)$, where \mathcal{L} is, most generally, a nonlinear function of its arguments and e^* is the conjugation of e . Here, in nonlinear problems, e is the solution to a nonlinear equation:

$$\mathbf{f}(e, e^*, \Phi) = 0 \tag{7}$$

For example, Equation (7) may represent steady-state Maxwell’s equations with an intensity-dependent permittivity distribution, where e is the electric field distribution. For problems studied here, the natural choice is to take e and e^* as independent parameters, which is necessary for differentiation, as opposed to separately treating the real and imaginary parts of e [29,45]. The solution to Equation (7) may be found with any nonlinear equation solver, such as the Newton–Raphson method [46].

The goal of the optimization is to optimize the objective function concerning the design variables Φ . With that aim in mind, it is necessary to compute the sensitivity of \mathcal{L} to each element of Φ . Similar to Equation (2), the derivative of the objective function concerning a single parameter φ can be derived as

$$\frac{d\mathcal{L}}{d\varphi} = \frac{\partial\mathcal{L}}{\partial\varphi} + \frac{\partial\mathcal{L}}{\partial e} \cdot \frac{de}{d\varphi} + \frac{\partial\mathcal{L}}{\partial e^*} \cdot \frac{de^*}{d\varphi} \tag{8}$$

Equation (8) can be rewritten into the matrix form as

$$\frac{d\mathcal{L}}{d\varphi} = \frac{\partial\mathcal{L}}{\partial\varphi} + \begin{bmatrix} \frac{\partial\mathcal{L}}{\partial e} & \frac{\partial\mathcal{L}}{\partial e^*} \end{bmatrix} \cdot \begin{bmatrix} \frac{de}{d\varphi} \\ \frac{de^*}{d\varphi} \end{bmatrix} \tag{9}$$

The terms $\frac{de}{d\varphi}$ and $\frac{de^*}{d\varphi}$ can be obtained by differentiating Equation (7) as follows:

$$\frac{d\mathbf{f}}{d\varphi} = \frac{\partial\mathbf{f}}{\partial\varphi} + \frac{\partial\mathbf{f}}{\partial e} \cdot \frac{de}{d\varphi} + \frac{\partial\mathbf{f}}{\partial e^*} \cdot \frac{de^*}{d\varphi} = 0 \tag{10}$$

Combining Equation (10) with its complex conjugate yields

$$\begin{bmatrix} \frac{\partial\mathbf{f}}{\partial e} & \frac{\partial\mathbf{f}}{\partial e^*} \end{bmatrix} \cdot \begin{bmatrix} \frac{de}{d\varphi} \\ \frac{de^*}{d\varphi} \end{bmatrix} = - \begin{bmatrix} \frac{\partial\mathbf{f}}{\partial\varphi} \\ \frac{\partial\mathbf{f}^*}{\partial\varphi} \end{bmatrix} \tag{11}$$

As a result, we can rewrite Equation (9) as

$$\frac{d\mathcal{L}}{d\varphi} = \frac{\partial\mathcal{L}}{\partial\varphi} - \begin{bmatrix} \frac{\partial\mathcal{L}}{\partial e} & \frac{\partial\mathcal{L}}{\partial e^*} \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial\mathbf{f}}{\partial e} & \frac{\partial\mathbf{f}}{\partial e^*} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \frac{\partial\mathbf{f}}{\partial\varphi} \\ \frac{\partial\mathbf{f}^*}{\partial\varphi} \end{bmatrix} \tag{12}$$

In analogy with the linear case, we obtain the gradient $\frac{d\mathcal{L}}{d\varphi}$ by solving an additional linear system. To this end, a complex-valued adjoint field e_{adj} is defined as the solution to

$$(\partial\mathbf{f}/\partial e)^T \cdot e_{adj} + (\partial\mathbf{f}^*/\partial e)^T \cdot e_{adj}^* = -(\partial\mathcal{L}/\partial e)^T \tag{13}$$

where T represents transposition. The variable e_{adj} and its conjugate can be computed by solving the following system:

$$\begin{bmatrix} \frac{\partial\mathbf{f}}{\partial e} & \frac{\partial\mathbf{f}}{\partial e^*} \\ \frac{\partial\mathbf{f}^*}{\partial e} & \frac{\partial\mathbf{f}^*}{\partial e^*} \end{bmatrix}^T \cdot \begin{bmatrix} e_{adj} \\ e_{adj}^* \end{bmatrix} = - \begin{bmatrix} \frac{\partial\mathcal{L}}{\partial e} \\ \frac{\partial\mathcal{L}}{\partial e^*} \end{bmatrix} \tag{14}$$

As can be seen, the adjoint problem, which is required to determine the derivative of the objective function, is linear, even though the physical problem is nonlinear, as defined by Equation (7). Finally, the gradient of the objective function can be written as

$$\frac{d\mathcal{L}}{d\varphi} = \frac{\partial\mathcal{L}}{\partial\varphi} + 2\Re(\mathbf{e}_{adj}^T \cdot \partial\mathbf{f}/\partial\varphi) \quad (15)$$

where $\Re(\cdot)$ denotes taking the real part. In deriving Equation (14), we have used the fact that both $\mathcal{L}(\cdot)$ and φ are real [29]. In the case of multiple parameters Φ , we just need to replace $\partial\mathbf{f}/\partial\varphi$ with the matrix $\partial\mathbf{f}/\partial\Phi$. The cost of computing the gradients $\partial\mathbf{f}/\partial\Phi$ increases slowly with the number of elements in Φ since \mathbf{e}_{adj} only needs to be solved once regardless of the number of parameters. This virtue makes large-scale, gradient-based optimization possible.

2.2. Application of AM in Photonic Inverse Design

The AM can be integrated into gradient-based topology optimization, which aims to distribute materials across a given design domain to reach a predefined performance goal [47]. Although topology optimization was originally developed to solve mechanical design problems, it is now widely employed in the fields of photonic crystals, waveguides, resonators, filters, and plasma excitations in system design. In topology optimization, the material density of each element or mesh point is a design variable by parameterizing the geometry as elements like pixels. Consequently, the optimization is always featured by a large number of design parameters and complete design freedom [48]. Such flexibility during the design is often viewed as one of the great advantages of topology optimization. In [47], the design domain was taken to be larger than in the original work to show that the topology optimization can effectively utilize all available space. In [49], mechanical design problems with more than a billion design variables proved that topology optimization in any practical sense was able to provide unlimited design freedom.

In fact, the AM perfectly matches the topology optimization scheme because it can offer sensitivity information at a low computational cost, which makes a large number of design parameters possible or more efficient [50,51]. Due to this nice match, the AM has been used for topology optimization in the context of mechanical engineering and photonics engineering for decades [51–53]. In [50], the T-junction in the photonic crystal waveguide was designed. A low-loss and broadband two-mode (de-)multiplexer, optimized via topology optimization with the AM and experimentally verified, has been developed for (de-)multiplexing the fundamental and first-order transverse-electric modes in a silicon photonic wire [54]. In [55], topology optimization is used with the AM for the design of optical sub-wavelength gratings. Completely unexpected metasurface designs for challenging multi-frequency, multi-angle problems, including designs for fully coupled multi-layer structures with arbitrary per-layer patterns, can be computationally discovered via topology optimization with the AM [56]. In [57], a design method based on topology optimization is applied to optical waveguide devices.

Compared with other design methods, the AM can obtain the gradient of the objective function relative to all design parameters through two full-field simulations [58]. This optimization process is suitable for structures that simultaneously achieve various design goals and have a large parameter space [59]. As a result, except for topological optimization, the AM has also been widely applied to other types of photonic inverse design tasks, more or less, along with different acceleration techniques. In [28], large metasurfaces were optimized, where the AM is used to calculate the gradient of electric field strength relative to all design parameters. In the optimization process, multipole expansion theory (MET) was employed to accelerate the computation, which offered orders-of-magnitude-higher efficiency than the FDTD-based optimization method did. In [60], a dual-wavelength metastructure was optimized by the AM. It was shown that the proposed design method was more efficient than the traditional one used in metasurface design. In [59], a tunable mode converter filled with liquid crystals was designed using an inverse design framework

based on the AM. In this work, the AM with multiple constraints was developed along with a multi-objective function, which can be applied to any multi-state optics. The original AM cannot be directly combined with the mode extension method because the associated gradients cannot be as clearly defined as that in the finite difference method, finite element time domain, or finite element method. To overcome this difficulty, automatic differentiation techniques were employed to generalize the adjoint variable method to arbitrary computational graphs [61].

It should be noted that, depending on the type of analysis, the additional computational cost per constraint sensitivity may range from a few percent for cases where extra right-hand sides are generated for an already factorized system, to almost 100% for the transient problem cases.

As discussed in Section 2.3, the AM is a local optimization method. In [62], to alleviate this drawback, the AM was integrated with global optimization, i.e., particle swarms, where a 6-times-improved angle and 6-times-improved efficiency were reached compared with current state-of-the-art devices.

Inherently, the conventional AM can be hardly applied to the inverse design of digital devices because one cannot calculate the gradient of a digital pattern [63]. In [63], an efficient inverse design of “digital” subwavelength nanophotonic devices was realized by employing the AM, where a single-mode 3 dB power divider and a dual-mode demultiplexer were designed. Compared with the direct-binary search (DBS) brute-force method, the AM can improve the design efficiency by nearly five times. And, the performance optimization can reach approximately the same level. The digital AM is a hybrid of topology optimization and the brute-force method, which improves the inverse design efficiency of high-performance digital subwavelength nanophotonic devices.

Although the AM was originally developed in terms of linear problems, it is nowadays indeed extended to nonlinear cases. For example, the AM equipped with topology optimization was employed to design one-dimensional nonlinear nanophotonic structures [47], where the third-order instantaneous Kerr material nonlinearity was considered. The strong nonlinearity in topology optimization-based computational design problems makes it difficult to solve the system directly [64]. Therefore, the iterative approach is widely adopted. In the iterative approach, the descent direction can be defined based on the adjoint derivative, where the state and adjoint variables are obtained by solving the partial differential equations and corresponding adjoint equations, respectively.

2.3. Limitations of AM in Photon Inverse Design

Although the AM has been successfully applied to a variety of photonic systems for optimization design, it is still limited by some inherent drawbacks. As it is known, the AM is inherently a local optimization method since it heavily relies on gradient-based information [65]. Thus, it shares the same drawbacks as the gradient-based methods. In particular, the AM, more generally speaking, gradient-based optimization algorithms, is susceptible to becoming stuck in local minimum valleys or saddle points because the design space for electromagnetic structures is predominantly nonconvex [31,66,67]. A remedy is to run the optimization process multiple times, typically by using random starting points. With the remedy, a single optimization target can always be derived [68]. Of course, a nice starting point can be selected if a region of high-performance devices is known in advance to avoid multiple runs of the optimization process.

The AM is feasible for very large problems because the computation of the sensitivities for an arbitrary number of design variables constitutes only a solution process of one extra linear system. However, the storage of the gradient information for the whole optimization process may become a main challenge. A remedy is to obtain the gradient information on the fly instead of storing it [51]. Similar to other traditional inverse design methods, adjoint topology optimization still has inherent drawbacks such as time-consuming numerical calculations or simulation processes [69].

The explicit computation of gradients in the AM is sometimes challenging for certain photonics problems [70]. The applications of the AM can be affected by a set of constraints that must be satisfied for specific engineering applications. For the associated optimizations, spatial filters, thresholding steps, or additional merit function terms are often necessary. For instance, the adjoint-based shape optimizations are mostly constrained to specific geometries such as spherical or ellipsoidal Mie scatterers [71]. They rely on approximating transmission coefficient gradients with polynomial proxy functions [72], approximating finite differences [73], or using methods like the level set method [20] or surface integrals over scatterer boundaries [60]. These operations can disrupt gradient computation [26,65,74]. To make it worse, for general photonic devices, which have complex light–matter interactions, their performance metrics and design parameters are often not well described by analytical forms or proxy functions. Also, in the case of rigorous coupled-wave analysis (RCWA) [75], the transmission and reflection properties of a structure depend on a global scattering matrix with elements that are functions of the eigenmodes of each layer. While it is possible to derive adjoint fields in RCWA, changing variables or scatterer geometries requires additional derivation steps for different cases, leading to extra work when switching parameterizations [70].

If the constraint is taken to be ‘hard’ and so must be satisfied at all stages of the optimization procedure, we need to know both the value of the constraint function and its linear sensitivity to the design variables. The latter requires an extra adjoint calculation. The more hard constraints exist, the more extra adjoint calculations are required. This type of constraint, therefore, undermines the computational cost benefits of the AM. If the number of hard constraints becomes as large as the number of design variables, the benefit can entirely be lost.

In [31], limitations of the discrete AM were discussed in the context of computational fluid dynamics. Suppose the objective function has a least-squares form as

$$\mathcal{L}(e(\Phi)) = \frac{1}{2} \sum_n (e_n(\Phi) - E_n)^2 \tag{16}$$

The gradient can be computed by

$$\frac{\partial \mathcal{L}}{\partial \varphi} = \sum_n \frac{\partial \mathcal{L}}{\partial e} \cdot \frac{\partial e}{\partial \varphi} (e_n(\Phi) - E_n) \tag{17}$$

The second-order derivative can be written as

$$\frac{\partial^2 \mathcal{L}}{\partial \varphi_i \partial \varphi_j} \approx \sum_n \left(\frac{\partial \mathcal{L}}{\partial e} \frac{\partial e}{\partial \varphi_i} \right) \left(\frac{\partial \mathcal{L}}{\partial e} \frac{\partial e}{\partial \varphi_j} \right) \tag{18}$$

where it is assumed that $e_n(\Phi) - E_n$ is small. The direct linear approach described using Equations (16)–(18) gives the approximate Hessian matrix, leading to very rapid convergence for the optimization iteration. In contrast, the AM provides no information on the Hessian, so optimization methods such as BFGS [76], which build up an approximation to the Hessian, take more steps to converge than the direct linear approach for least-squares applications. Of course, for the case with a large number of design variables, the AM may still be more efficient, since the cost of each step is significantly higher when the sensitivities are evaluated directly.

3. DL for Photonic Inverse Design

Although the inverse design methods including those combined with the AM have been widely employed in photonic research, the process can still be time and computationally intensive [77] because, in each iterative step during the optimization, Maxwell’s equations should be repeatedly solved in terms of a new set of parameters. Since fabrication techniques allow for more complex three-dimensional designs, photonic designs can be

more complex and the ranges of the search parameters become larger. The associated optimization process becomes increasingly resource-intensive [78–80].

Inspired by the fast development of DL, people have combined the DL techniques with inverse design [33,66,81–83]. At present, DL has been developed rapidly in the field of photonic device inverse design, which can be more efficient than traditional iterative optimization methods.

The interaction of nonlinear optics and DL domains has revealed great potential in recent years [84], both for the understanding of physical systems, e.g., to speed up the study of nonlinear pulse propagation [85] or to improve nonlinear effect compensation [86] in fibers and for the development of photonic-based hardware to accelerate DL calculations relying on third [87] or second [88]-order nonlinear effects.

Figure 2 shows some DL networks applied to photonic inverse design. DL networks are featured by their capability in capturing and modeling highly nonlinear data relationships. A typical DL architecture is a multilayer stack of simple modules, all (or most) of which are subject to learning, and many of which compute non-linear input–output mappings [89]. With multiple non-linear layers, a DL network can effectively predict non-linear optical (NLO) phenomena such as a photonic device (NLO crystal) or a photonic technique (two-photon excitation microscopy). Therefore, DL can mimic nonlinear physics-based relationships, e.g., those between photonic-system geometries and their electromagnetic responses; they provide a fresh perspective on the forward and inverse problems.

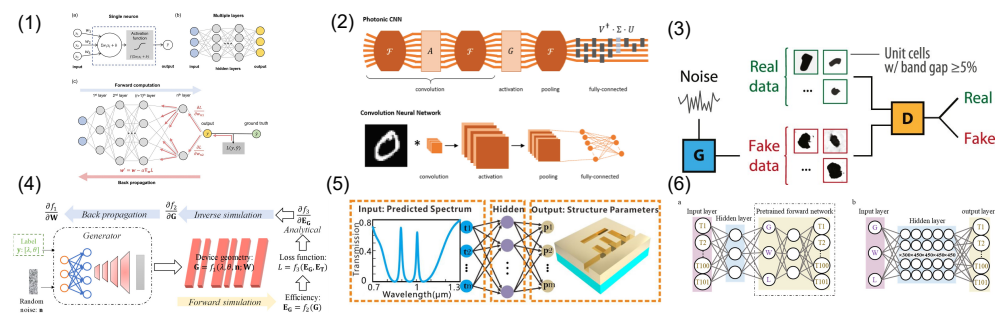


Figure 2. Photon inverse design and DL networks. (1) Schematic illustration of a neural network: (a) A single neuron calculates a weighted sum of inputs and adds a bias term, followed by a nonlinear activation function. (b) A fully connected, multiple-layered neural network. (c) Schematic of neural network training [90]. (2) Schematic of a convolutional neural network (CNN) architecture and an equivalent implementation of a photonics CNN [91]. (3) Conditional GANs facilitate image-to-image translation of photonic features [35]. (4) Merging DNN to optimization algorithms [12]. (5) Diagram of the ANNs applied in the inverse design and performance optimization problems [92]. (6) Comparison of tandem network and iterative DNN: (a) Architecture of the tandem network. (b) Architecture of iterative DNN [93].

When applied to the inverse design, DL can be used as an efficient solver to provide a fast solution to Maxwell’s equations. To this end, DL models are trained with nonlinear activation functions and backpropagation to intelligently learn nonlinear relationships between input parameters and output electric fields. The well-trained neural models, including DL ones, always offer the possibility of finding solutions outside the boundaries of the training data. That is, the models can transfer knowledge, an approach of the Maxwell’s equations known as “transfer learning” [94], which is generally not available in the traditional electromagnetic solvers.

The DL models can also be a nice optimizer for the inverse design, especially for high-dimensional systems. As discussed previously, the AM is inherently a local optimization scheme that suffers from the lack of global information. According to the successful applications of DL in various fields, DL can be more robust than the AM in recovering global information. Additionally, in the absence of prior knowledge of the parameter landscape, many problems become difficult to be optimized as the number of parameters increases.

However, DL can overcome this difficulty to some extent due to its nice generalizability after the neural networks are well trained.

3.1. DL Networks Applied to Photonic Inverse Design

Deep neural networks are often built on top of fully connected networks (FCNs) and convolutional neural networks (CNNs). FCNs can be viewed as the most primitive type of neural networks, which are composed of multiple layers of neurons, each being connected to all other ones on the adjacent layer. The fully connected nature offers FCNs the ability to simulate complex transformations. The computation between two adjacent layers can be mathematically described by a matrix–vector product or its tensor counterpart. Suppose that the number of neurons is N for the two adjacent layers, and the computational complexity of the associated layers reaches $O(N^2)$. The high complexity prevents FCN in practice from having a large number of layers and neurons. CNNs are an improved alternative to FCNs where the time-consuming matrix–vector operations (or its tensor counterpart) are replaced with the convolution operations. The invariance of the translation of the input tensor in CNNs enhances the ability of the networks to capture features from image/audio data with strong spatial/temporal correlation. When it comes to sequential data, recurrent neural networks (RNNs) [95] become the most commonly used models, which can be unambiguously generalized to sequentially connected neurons.

3.1.1. Fully Connected Networks (FCNs)

In [92], the plasmonic waveguide-coupled with cavities structure (PWCCS) was optimized by employing a FCN, as shown in Figure 3. The effectiveness of the FCN in optimizing the PWCCS was verified via the selection of Fano resonance derived from PWCCS and plasmon-induced transparency effects. In this work, the genetic algorithm was employed to design network structures and to guide the selection of hyperparameters for the FCN. Such an approach not only enables the high-precision inverse design of PWCCS but also optimizes some key performance indicators of the transmission spectrum. In [96], the inverse design of the edge state of the curved wave topology was realized via FCNs. In this work, two neural networks were constructed for the forward and backward predictions of bandgap width and geometric parameters. The dataset used to train a feed-forward artificial neural network was generated using the plane wave expansion method (PWE) with a sweeping of the arrangement radii.

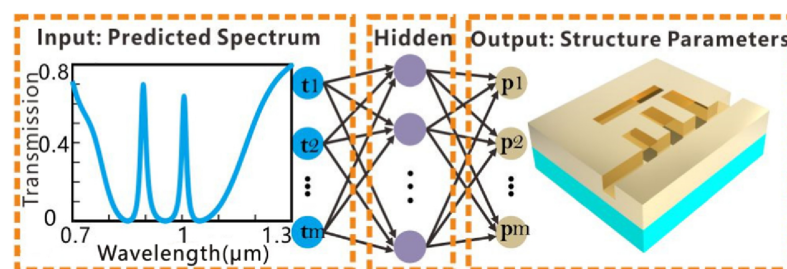


Figure 3. Diagram of the ANNs applied in the inverse design and performance optimization problems [92].

3.1.2. Convolutional Neural Networks (CNNs)

As it is known, CNNs can capture the local correlation of spatial information in the input data. This property is highly desired in photonics applications as the devices are always represented by high-dimensional spatial data. Consequently, CNNs and their improved variations have attracted many researchers who applied them to the solution of inverse design problems. For example, it has been proved that CNNs had the potential to solve the problem of the inverse design of thin film metamaterials, probing the parameter space of a given material and thickness library globally, which can be difficult and expensive for traditional methods [97]. The full generalization ability of neural networks, especially

CNNs, to generate systems with the desired spectral response to accommodate various input design parameters has been revealed.

In [98], the CNN was employed to be an inverse design tool to achieve high numerical accuracy in plasma element surfaces, which proved that the CNN was an excellent tool for the design. More specifically, the CNNs are capable of identifying peaks and troughs of the spectrum at the expense of low computational cost. Key geometric parameters can reach an accuracy as high as ± 8 nm. In this work, the comparison of CNNs and FCNs shows that CNNs had higher generalization capabilities. Additionally, it has been demonstrated that batch normalization can improve the performance of CNNs.

3.1.3. Recurrent Neural Networks (RNNs)

RNNs tackle problems associated with sequential data such as sentences and audio signals. The network receives sequential data one at a time and incrementally generates new data series. It was demonstrated in [99] that, for photonic design, RNNs were suitable to model optical signals or spectra in the time domain with a specific line shape originating from various modes of resonance. In [100], RNNs were implemented to analyze optical signals and to equalize noise in high-speed fiber transmission. In [77], RNNs were utilized to find the correlation within 2D cross-sectional images of plasmonic structures, where the results showed that the network was able to predict the absorption spectra from the given input structural images. It was revealed in [101] that the performance of RNNs can be enhanced by adopting advanced varieties of RNNs, such as Long Short-Term Memory (LSTM) [102] and gated recurrent unit (GRU). In [103], traditional neural networks, RNNs with LSTM and an RNN with GRU were used to construct the inverse design of MPF. As indicated by the results in [103], according to the training results, it was found that GRU-based RNNs were the most effective in predicting frequency response with the highest accuracy. The study in [104] demonstrated that, in combination with CNNs, RNNs were also utilized to enhance the approximation of the optical responses of nanostructures that are illustrated in images. As revealed in [99], network systems hybridizing CNNs and RNNs present a promising method for modeling and designing photonic devices with unconventional spatiotemporal properties of light.

3.1.4. Deep Neural Networks (DNNs)

A DNN is a network where the number of hidden layers is much greater than 1. The DNNs can be divided into discriminant neural networks and generative ones. In this subsection, we focus on the discriminant DNNs and delay the discussion on the generative ones in the following Section 3.1.5.

As pointed out in [105], neural networks can be used in two different ways. The first method takes the configuration of the device to be designed including the structural parameter (such as the geometrical shape of a nanostructure) as the input and the predicted electromagnetic response of the device (such as transmission spectra or differential scattering cross-section) as the output. These neural networks can be used to replace the computationally expensive electromagnetic simulations in the optimization loop, greatly reducing the design time. In [105], the DNNs belonging to this method were denoted by forward-modeling networks because they compute electromagnetic response from the device. Essentially, a forward-modeling DNN is an electromagnetic solver. The second type of neural network, named after inverse-design networks, takes the electromagnetic response as the input and directly outputs the configuration of the device. The inverse-design networks act as an inverse operator that converts the electromagnetic response to the configuration of the device. However, one significant difficulty in training inverse-design DNNs arises from a fundamental property of the inverse scattering problem: the same electromagnetic response can be produced using many different designs. Different from the previous remedy to this difficulty where the training dataset was divided into distinct groups, where there was a unique design within each group corresponding to each response, the approach of cascading an inverse-design network with forward modeling

was proposed in [105]. The obtained network was named the tandem DNN. Numerical experiments in [105] indicated that the tandem DNNs can be trained by datasets containing nonunique electromagnetic scattering instances.

In [93], the so-called iterative DNN was proposed, where trained weights of the forward modeling network [105] were fixed and gradient-descent methods based on back-propagation were employed. The iterative DNN and tandem DNNs were compared in terms of the transmission spectrum design of bow nanoantennas, where the results demonstrated that the two types of DNN architectures reached comparable performance.

In [106], three models were investigated for designing nanophotonic power splitters with multiple splitting ratios. The first model employed the DNN as a forward-modeling network to predict the spectral response (SPEC) given hole vectors (HV). The second one utilized the DNN as an inverse-design network to construct HV given a target SPEC. The third one was based on a stochastic generative model which implicitly integrated forward and inverse-design networks. In addition, a bidirectional network consisting of the forward and inverse-design networks was proposed [107]. The forward-modeling network was trained subsequently to the inverse-design network, which was used to predict the geometry from the transmission spectrum.

In [108], DNN models were used as both the electromagnetic solvers and the optimizer. By building the DNN with an FDTD solver, the proposed method can be used to efficiently design a silicon photonic grating coupler, one of the fundamental silicon photonic devices with a wavelength-sensitive optical response.

To facilitate multi-tasks inverse design, a topology optimization method based on the DNN in the low-dimensional Fourier domain was proposed in [109]. The DNN took target optical responses as inputs and predicted low-frequency Fourier components, which were then utilized to reconstruct device geometries. By removing high-frequency components for reduced design degrees of freedom (DoFs), the minimal features were controlled and the training was sped up. In [110], a forward-modeling DNN was paired with evolutionary algorithms, where the DNN was used only for preselection and initialization. The method utilizes a global evolutionary search within the forward model space and leverages the huge parallelism of modern GPUs for fast inversion.

A well-organized tutorial was given in [111], where the process of deep inverse learning applied to AEM problems was given in a step-by-step approach. In [111], common pitfalls of training and evaluation of DL models were discussed and a case study of the inverse design of a GaSb thermophotovoltaic cell was given.

3.1.5. Deep Generative Models

Generative models generally seek to learn the statistical distribution of data samples rather than the mapping from input to output data. In this respect, they can be viewed as the complement techniques of the conventional optimization and inverse design [35]. In generative models, the joint distribution of the input and output is employed to optimize a certain objective in a probabilistically generative manner instead of determining the conditional distribution and thus decision boundaries.

Generative adversarial networks (GANs) are the typical deep generative models built using generators and discriminators, which have been used for the design and optimization of dielectric and metallic metasurfaces due to their ability to generate massive nanostructures efficiently [112,113].

In [114], a conditional deep CGAN was employed as an inverse-design network, which was trained on colored images encoded with a range of material and structural parameters. It was demonstrated that, in response to target absorption spectra, the CGAN can identify an effective metasurface in terms of its class, material properties, and overall shape. In [115], a meta-heuristic optimization framework, along with a resistant autoencoder (AE), was designed and demonstrated to greatly improve the optimization search efficiency for meta-device configurations featuring complex topologies.

In [36], a network combining the CGAN with the WGAN was proposed to design free-form, all-dielectric metasurface devices. This work indicated that the proposed approach stabilized the training process and gave the network the ability to handle synthetic metasurface design problems. It also demonstrated in this work that GAN-based approaches were the preferred solution to multifunctional inverse design problems.

In [12], research was conducted on inverse models that utilize unsupervised generative neural networks to exploit the benefits of not requiring training data samples. However, time-consuming electromagnetic solvers must be integrated into the generative neural networks to guarantee compliance with the laws of physics. After training, unsupervised generative neural networks can also generate new geometric parameters based on the optical response of the target without the need for any electromagnetic solver [12].

3.2. Limitations of DL in Photon Inverse Design

Although DL is attractive in inverse design photonics, there are still limitations with DL algorithms. These limitations include the DoFs, the local minimum, the difficulty in preparing the DL training dataset, the nonunique solutions, and difficulties associated with generalization.

3.2.1. Degrees of Freedom

In topological optimization, the number of DoFs can be very large, which always makes the DL very hard to train. The relationship between geometric parameters and optical performance is very similar to the mathematical non-deterministic polynomial difficulty (NP difficulty) problem [116] and very hard to define using explicit functions. As shown in Figure 4 from [12], the designer extracts the geometric parameter x from a device structure, such as an empirical structure, the QR code one, or an irregular one.

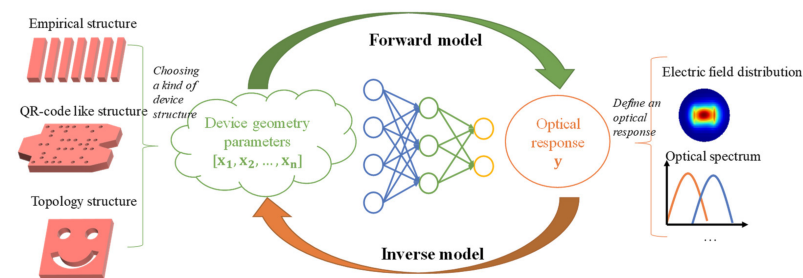


Figure 4. Schematic of DNN-assisted silicon photonic device design approach [12].

What ML method can be employed in inverse design is related to the DoFs of the photonic structure. For the case with a relatively small number DoFs, many DL networks can be employed in the inverse design. However, the DoFs continue to grow to thousands or more, and the huge dimension of the optimization space prohibits methods that require large amounts of data or a lot of simulation iterations. In this case, the generative models, such as GANs and VAEs, can be leveraged to reduce the dimensionality of the design structure and optical response [33,117].

Additionally, the DoFs of photonic devices could potentially impact the NLO phenomenon. For example, due to the varying interlayer interactions associated with different twist angles, the twisting DoFs have been widely applied to engineer the bands of van der Waals layered structures. In that work, the twist-angle-dependent second-harmonic generation (SHG) from twisted bilayer graphene samples was demonstrated along with their correlation with the evolving hybrid band structure [118]. Nonlinear optics will probably be increasingly important since people can reach a better understanding of the underlying physics although the design involves many DOFs, i.e., many photonic DOFs, such as spatial modes, frequencies, and polarizations, and many design parameters, such as the space- and time-dependent distributions of refractive index and loss/gain in photonic structures [119]. Furthermore, reconfigurable control of many modal DOFs can be achieved

with spatial light modulators or digital micromirror devices, which are now available with resolutions (i.e., number of pixels) close to 10^7 [119]. In general, the number of design DOFs routinely available to optical scientists and engineers now exceeds the technology from past decades by 3 to 5 orders of magnitude [119].

3.2.2. Training Dataset

DL often requires large and high-quality training datasets, which are generally unavailable in practice. The large amount of data required for DL could not be readily available and is usually created using simulation methods such as RCWA, FEM, and FDTD, which are time- and computation-expensive [33,90,120]. In [35], a dataset of 20,000 two-dimensional photonic crystal unit cells and their associated band structures was built, enabling the training of supervised learning models. Using these datasets, a high-accuracy CNN for band structure prediction was demonstrated, with orders-of-magnitude speedup compared with conventional theory-driven solvers. In [107], a DNN was trained using a dataset composed of 15,000 randomly generated device layouts containing eight geometrical parameters. In [121], to train a forward network model for prediction of electromagnetic response, a dataset of 90,000 samples was generated using S^4 software (version 2) [122], where the lattice constant was set between 50 nm and 150 nm and the maximum thickness of each dielectric was limited to 150 nm.

An effective way is to co-construct large datasets of various optical designs with the efforts of the optical community [33,123] by avoiding the repeated generation of simulation data. To use generative models instead of real datasets is also a nice alternative [32]. Another solution is to develop networks suitable for small training datasets. In [27], an approach based on explainable artificial intelligence was developed for the inverse design, demonstrating that their method works equally well on smaller datasets. In [124], an efficient tandem neural network with a backpropagation optimization strategy was developed to design one-dimensional photonic crystals with specific bandgaps using a small dataset.

In [125], methods to generate datasets with high quality were discussed to overcome the problem arising from the lack of training data. The optimization algorithm in [126] to generate training data may perform less efficiently as it is time-consuming for the case of many free parameters. The concept of iterative training data generation has emerged as the remedy, which allows the neural network to learn from previous mistakes and to significantly improve its performance on specific design tasks. However, iterative procedures can be computationally expensive since data generation is slow and the network may need to be re-trained several times on increasing amounts of training samples [115]. The solution is to reduce the amount of re-training by accelerating convergence by assessing the quality and uncertainty of the ANN output from multiple predictions. In addition, transfer learning has also been applied in nano-optics problems to enhance the performance of the ANN when the available data are limited [127].

3.2.3. Nonuniqueness Problem

Since different designs may produce almost identical spectra, this may prevent ML algorithms from converging correctly. Many researchers have developed various solutions to this problem. In [77,121], a method adding forward modeling networks to the inverse design DNN architecture was employed, which can be viewed as one of the most common ways to overcome the nonunique problem. In [128], an approach modeling the design parameters as multimodal probability distributions rather than discrete values was proposed, which can also overcome the nonunique problems. In [129], a physics-based preprocessing step to solve the size mismatch problem and to improve spectral generalizability was proposed. Numerical experiments in this work indicated that the proposed approach can provide accurate prediction capabilities outside the training spectrum. In [124], an algorithm called adaptive batch normalization was proposed, which was a simple and effective method that accelerates the convergence speed of many neural networks.

3.2.4. Local Minimum

Similar to the AM, DL also to some extent relies on gradient information to carry out the optimization. Thus, DL inverse design can also suffer from local minimum problems. In [130], an approach to overcome the local minimum problem was proposed by constructing a combined loss function that also fits the Fourier representation of the target function. Physically, the Fourier representation contains information about the target function's oscillatory behavior, which corresponds directly to specific parameters in the input space for many resonant photonic components. In [33], the generative models were incorporated into traditional optimization algorithms to help avoid local minimum problems. Because generative models are essentially transforming raw data into another representation, the local minimum problem in the original parameter space may be eliminated, which can alleviate the local minimum problem if optimized in a sparse representation. In [131], deep generative neural networks based on global optimization networks (GLOnets) were configured to perform categorical global optimization of photonic devices. This work showed that, compared with traditional algorithms, GLOnets can find the global optimal several orders of magnitude faster.

3.2.5. Generalization Ability

The lack of generalization can also limit DL in inverse design [132]. DL is like a black box, making it challenging to understand and interpret its results and reliability, especially when dealing with imperfect datasets or data generated by adversarial methods. Due to the challenges of validating and replicating DL results, a set of community-wide recommendations in biology for DL reporting and validation (data, optimization, model, and evaluation) was recently published. The given recommendations help to handle questions that need to be addressed when reporting DL results and are also largely applicable to the fields of optics and photonics [91].

3.2.6. Problem of Rerunning

When the optimization goal is modified, a new inverse optimization process needs to be rerun, maybe from scratch. Since a single run of traditional inverse design optimization is always time- and computation-expensive as it may require hundreds of rounds of simulations, rerunning will consume a large amount of computing resources [77]. Moreover, when the optimization goal is modified, it is necessary to recreate the dataset if DL is employed, which is large in amount and not readily available. In the AM, the change of optimization goal requires additional derivation steps for different cases, leading to extra work when switching parameterizations [70].

4. Hybridization of AM and DL For Inverse Design

The use of DL can overcome some limitations of the AM, and vice versa. Many recent efforts have been made to hybrid DL with the AM [132]. It was demonstrated that adjoint optimizations are capable of achieving state-of-the-art performance and are orders of magnitude more computationally efficient than alternative optimization methods, including DL [133–135]. As pointed out in Section 2.3, the AM is local in nature and therefore limited by corresponding limitations. The hybridization of DL with the AM can overcome this problem to some extent [27]. As shown in Figure 5 from [27], an inverse design framework was proposed by combining adjoint optimization, automatic ML, and explainable artificial intelligence.

Redefining adjoint-based optimization as training that generates neural networks is often applied to physical systems that can take advantage of gradients to improve performance where training is performed by calculating the forward and adjoint electromagnetic simulations of outputted devices [38]. Recently, GANs and adversarial autoencoders (AAEs) have been coupled with adjoint topology optimization techniques for optimizing the diffractive dielectric gratings [38,115,136]. It has been previously shown that integrating the AAE network with a conventional adjoint topology optimization formalism can result

in a speedup of approximately 4900 times for thermal emitter optimization as opposed to utilizing the conventional topology optimization method [120]. In [136], robustness against geometric erosion and dilation was achieved for the devices by conducting 30 adjoint-based topology optimization iterations on the 50 most effective GAN-generated devices. As demonstrated in [63], GANs and VAEs can achieve complex topologies by using image-based representations, and utilizing the AM can be an effective way of enhancing their performance. Additionally, generative models trained on physically informed loss or using the AM during training can also be improved by subsequent optimization-based refinement steps [137]. In [38], a network based on a conditional generative neural network and the AM was developed, which is capable of producing ensembles of highly efficient topology-optimized metasurfaces operating across a range of parameters.

In [37], a method for performing backpropagation in an ANN based on a photonic circuit was proposed, where the AM was used to derive the photonic of the backpropagation algorithm. In [138], a framework named DeepAdjoint was proposed. As a general, open-source, multi-objective “all-in-one” global photonics inverse design application framework, the method in ML optimization pipelines is made simpler and improved by hybridizing pre-trained deep generative networks with the AM. The framework employs GANs as an inverse-design network, which can predict device class, material properties (such as refractive index and Drude plasma frequency), and the nanoscale geometric structuring (such as planar topology and layer thickness) of metal–insulator–metal metasurfaces simultaneously.

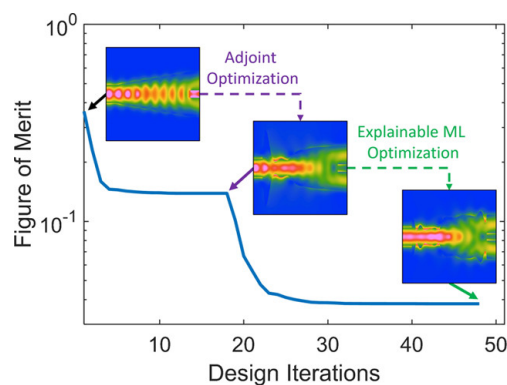


Figure 5. Photonic inverse design with the AM and ML [27].

In [139], the so-called neural-adjoint (NA) was developed. NA trains a neural network to approximate the input–output relationship and then, starting from different random locations, uses gradient descent towards locally optimal values with the AM. In [140], the NA inverse design method was studied for the inverse design of all-dielectric metasurfaces. The NA method is an effective method for predicting the high-dimensional total dielectric metasurface geometry required to generate the desired infrared absorption spectrum, even without any specialized knowledge in the field.

In [141], a recurrent neural AM based on NA was proposed. The results indicate that the proposed method is efficient in designing optical multilayer films for specific spectrum filters. In general, the hybridization of the AM and DL has been proven beneficial in handling photonic nonlinear phenomena. For example, the NLO phenomenon was more properly handled by techniques like Conditional GLOnets [38], which employed physics-driven gradients to iteratively enhance the nonlinear mapping between inputs and device layout, akin to the optimization process based on the AM.

5. Conclusions

The AM is now widely employed in the fields of photonic crystals, waveguides, resonators, filters, and plasma excitations in system design. Compared with other design methods, the AM can obtain the gradient of the objective function relative to all design

DoFs through two full-field simulations, which makes it suitable for structures that simultaneously achieve various design goals and have a large parameter space. Except for topological optimization, the AM has been applied to many other types of inverse design tasks of photons, maybe with different acceleration techniques. It is also extended to solve nonlinear optimization problems. DL technology plays an important role in optical reverse design, which can significantly reduce simulation time; improve efficiency and accuracy; and is of great significance for the design, manufacturing, and application of optical devices. Different from the AM, DL can be an efficient solver of Maxwell's equations as well as a nice optimizer. This makes DL techniques more widely employable in inverse design tasks. However, both the AM and DL have their own drawbacks. The AM is inherently a local optimization method since it heavily relies on gradient-based information. The optimization can suffer from becoming stuck in local minimum valleys or saddle points. The memory usage of the AM can also prevent it from very large problems. At the same time, DL may encounter the problems of local minima, difficulty in obtaining training datasets, the existence of non-unique solutions, and troubles in generalization. Fortunately, some of the difficulties can be solved or alleviated by combining DL with the AM.

Author Contributions: Conceptualization, Z.P. and X.P.; writing—original draft preparation, Z.P.; writing—review and editing, Z.P. and X.P.; supervision, X.P.; funding acquisition, X.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partly supported by NSFC under grant 62171033.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Shen, B.; Wang, P.; Polson, R.; Menon, R. An Integrated-Nanophotonics Polarization Beamsplitter with $2.4 \times 2.4 \mu\text{m}^2$ Footprint. *Nat. Photonics* **2015**, *9*, 378–382. [[CrossRef](#)]
2. Piggott, A.Y.; Lu, J.; Lagoudakis, K.G.; Petykiewicz, J.; Babinec, T.M.; Vučković, J. Inverse Design and Demonstration of a Compact and Broadband On-Chip Wavelength Demultiplexer. *Nat. Photonics* **2015**, *9*, 374–377. [[CrossRef](#)]
3. Peurifoy, J.; Shen, Y.; Jing, L.; Yang, Y.; Cano-Renteria, F.; DeLacy, B.G.; Joannopoulos, J.D.; Tegmark, M.; Solja, M. Nanophotonic Particle Simulation and Inverse Design Using Artificial Neural Networks. *Sci. Adv.* **2018**, *4*, eaar4206. [[CrossRef](#)]
4. Qiu, T.; Shi, X.; Wang, J.; Li, Y.; Qu, S.; Cheng, Q.; Cui, T.; Sui, S. Deep Learning: A Rapid and Efficient Route to Automatic Metasurface Design. *Adv. Sci.* **2019**, *6*, 1900128. [[CrossRef](#)] [[PubMed](#)]
5. Bor, E.; Babayigit, C.; Kurt, H.; Staliunas, K.; Turdnev, M. Directional invisibility by genetic optimization. *Opt. Lett.* **2018**, *43*, 5781–5784. [[CrossRef](#)] [[PubMed](#)]
6. Fu, P.H.; Lo, S.C.; Tsai, P.C.; Lee, K.L.; Wei, P.K. Optimization for Gold Nanostructure-Based Surface Plasmon Biosensors Using a Microgenetic Algorithm. *ACS Photonics* **2018**, *5*, 2320–2327. [[CrossRef](#)]
7. Solís, D.M.; Obelleiro, F.; Taboada, J.M. Surface Integral Equation-Domain Decomposition Scheme for Solving Multiscale Nanoparticle Assemblies with Repetitions. *IEEE Photonics J.* **2016**, *8*, 1–14. [[CrossRef](#)]
8. Majérus, B.; Butet, J.; Bernasconi, G.D.; Valapu, R.T.; Lobet, M.; Henrard, L.; Martin, O.J.F. Optical Second Harmonic Generation from Nanostructured Graphene: A Full Wave Approach. *Opt. Express* **2017**, *25*, 27015–27027. [[CrossRef](#)]
9. Yu, D.M.; Liu, Y.N.; Tian, F.L.; Pan, X.M.; Sheng, X.Q. Accurate Thermoplasmonic Simulation of Metallic Nanoparticles. *J. Quant. Spectrosc. Radiat. Transf.* **2017**, *187*, 150–160. [[CrossRef](#)]
10. Pan, X.M.; Xu, K.J.; Yang, M.L.; Sheng, X.Q. Prediction of Metallic Nano-Optical Trapping Forces by Finite Element-Boundary Integral Method. *Opt. Express* **2015**, *23*, 6130–6144. [[CrossRef](#)]
11. Pan, X.M.; Gou, M.J.; Sheng, X.Q. Prediction of Radiation Pressure Force Exerted on Moving Particles by the Two-Level Skeletonization. *Opt. Express* **2014**, *22*, 10032–10045. [[CrossRef](#)]
12. Mao, S.; Cheng, L.; Zhao, C.; Khan, F.N.; Li, Q.; Fu, H.Y. Inverse Design for Silicon Photonics: From Iterative Optimization Algorithms to Deep Neural Networks. *Appl. Sci.* **2021**, *11*, 3822. [[CrossRef](#)]
13. Chen, W.; Zhang, B.; Wang, P.; Dai, S.; Liang, W.; Li, H.; Fu, Q.; Li, J.; Li, Y.; Dai, T.; et al. Ultra-Compact and Low-Loss Silicon Polarization Beam Splitter Using a Particle-Swarm-Optimized Counter-Tapered Coupler. *Opt. Express* **2020**, *28*, 30701. [[CrossRef](#)]

14. Mao, S.; Cheng, L.; Mu, X.; Wu, S.; Fu, H.Y. Ultra-Broadband Compact Polarization Beam Splitter Based on Asymmetric Etched Directional Coupler. In Proceedings of the 14th Pacific Rim Conference on Lasers and Electro-Optics (CLEO PR 2020), Sydney, Australia, 2–6 August 2020; Optica Publishing Group: Sydney, Australia, 2020; p. C12H_1.
15. Wang, Q.; Ho, S.T. Ultracompact Multimode Interference Coupler Designed by Parallel Particle Swarm Optimization with Parallel Finite-Difference Time-Domain. *J. Light. Technol.* **2010**, *28*, 1298–1304. [[CrossRef](#)]
16. Sanchis, P.; Villalba, P.; Cuesta, F.; Håkansson, A.; Griol, A.; Galán, J.V.; Brimont, A.; Martí, J. Highly Efficient Crossing Structure for Silicon-on-Insulator Waveguides. *Opt. Lett.* **2009**, *34*, 2760. [[CrossRef](#)] [[PubMed](#)]
17. Zhang, Y.; Yang, S.; Lim, A.E.J.; Lo, G.Q.; Galland, C.; Baehr-Jones, T.; Hochberg, M. A Compact and Low Loss Y-junction for Submicron Silicon Waveguide. *Opt. Express* **2013**, *21*, 1310. [[CrossRef](#)]
18. Tanemura, T.; Balram, K.C.; Ly-Gagnon, D.S.; Wahl, P.; White, J.S.; Brongersma, M.L.; Miller, D.A.B. Multiple-Wavelength Focusing of Surface Plasmons with a Nonperiodic Nanoslit Coupler. *Nano Lett.* **2011**, *11*, 2693–2698. [[CrossRef](#)]
19. Fu, P.H.; Huang, T.Y.; Fan, K.W.; Huang, D.W. Optimization for Ultrabroadband Polarization Beam Splitters Using a Genetic Algorithm. *IEEE Photonics J.* **2019**, *11*, 1–11. [[CrossRef](#)]
20. Lalau-Keraly, C.M.; Bhargava, S.; Miller, O.D.; Yablonovitch, E. Adjoint Shape Optimization Applied to Electromagnetic Design. *Opt. Express* **2013**, *21*, 21693. [[CrossRef](#)] [[PubMed](#)]
21. Borel, P.I.; Harpøth, A.; Frandsen, L.H.; Kristensen, M.; Shi, P.; Jensen, J.S.; Sigmund, O. Topology optimization and fabrication of photonic crystal structures. *Opt. Express* **2004**, *12*, 1996–2001. [[CrossRef](#)]
22. Jensen, J.S.; Sigmund, O. Systematic design of photonic crystal structures using topology optimization: Low-loss waveguide bends. *Appl. Phys. Lett.* **2004**, *84*, 2022–2024. [[CrossRef](#)]
23. Kao, C.Y.; Osher, S.; Yablonovitch, E. Maximizing band gaps in two-dimensional photonic crystals by using level set methods. *Appl. Phys. B Lasers Opt.* **2005**, *81*, 235–244. [[CrossRef](#)]
24. Burger, M. A framework for the construction of level set methods for shape optimization and reconstruction. *Interfaces Free Bound.* **2003**, *5*, 301–329. [[CrossRef](#)]
25. Burger, M.; Osher, S.J. A survey on level set methods for inverse problems and optimal design. *Eur. J. Appl. Math.* **2005**, *16*, 263–301. [[CrossRef](#)]
26. Molesky, S.; Lin, Z.; Piggott, A.Y.; Jin, W.; Vucković, J.; Rodriguez, A.W. Inverse Design in Nanophotonics. *Nat. Photonics* **2018**, *12*, 659–670. [[CrossRef](#)]
27. Yeung, C.; Ho, D.; Pham, B.; Fountaine, K.T.; Zhang, Z.; Levy, K.; Raman, A.P. Enhancing Adjoint Optimization-Based Photonic Inverse Design with Explainable Machine Learning. *ACS Photonics* **2022**, *9*, 1577–1585. [[CrossRef](#)]
28. Zhang, D.; Liu, Z.; Yang, X.; Xiao, J.J. Inverse Design of Multifunctional Metasurface Based on Multipole Decomposition and the Adjoint Method. *ACS Photonics* **2022**, *9*, 3899–3905. [[CrossRef](#)]
29. Hughes, T.W.; Minkov, M.; Williamson, I.A.D.; Fan, S. Adjoint Method and Inverse Design for Nonlinear Nanophotonic Devices. *ACS Photonics* **2018**, *5*, 4781–4787. [[CrossRef](#)]
30. Garza, E.; Sideris, C. Fast Inverse Design of 3D Nanophotonic Devices Using Boundary Integral Methods. *ACS Photonics* **2022**, *10*, 824–835. [[CrossRef](#)]
31. Giles, M.B.; Pierce, N.A. An Introduction to the Adjoint Approach to Design. *Flow Turbul. Combust.* **2000**, *65*, 393–415. [[CrossRef](#)]
32. Tanriover, I.; Lee, D.; Chen, W.; Aydin, K. Deep Generative Modeling and Inverse Design of Manufacturable Free-Form Dielectric Metasurfaces. *ACS Photonics* **2023**, *10*, 875–883. [[CrossRef](#)]
33. Liu, Z.; Zhu, D.; Raju, L.; Cai, W. Tackling Photonic Inverse Design with Machine Learning. *Adv. Sci.* **2021**, *8*, 2002923. [[CrossRef](#)] [[PubMed](#)]
34. Mengü, D.; Rahman, M.S.S.; Luo, Y.; Li, J.; Kulce, O.; Ozcan, A. At the intersection of optics and deep learning: Statistical inference, computing, and inverse design. *Adv. Opt. Photonics* **2022**, *14*, 209–290. [[CrossRef](#)]
35. Christensen, T.; Loh, C.; Picek, S.; Jakobović, D.; Jing, L.; Fisher, S.; Ceperic, V.; Joannopoulos, J.D.; Soljačić, M. Predictive and Generative Machine Learning Models for Photonic Crystals. *Nanophotonics* **2020**, *9*, 4183–4192. [[CrossRef](#)]
36. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A. Improved Training of Wasserstein GANs. *arXiv* **2017**, arXiv:1704.00028.
37. Hughes, T.W.; Minkov, M.; Shi, Y.; Fan, S. Training of Photonic Neural Networks through in Situ Backpropagation and Gradient Measurement. *Optica* **2018**, *5*, 864. [[CrossRef](#)]
38. Jiang, J.; Fan, J.A. Global Optimization of Dielectric Metasurfaces Using a Physics-Driven Neural Network. *Nano Lett.* **2019**, *19*, 5366–5372. [[CrossRef](#)]
39. Wang, J.; Shi, Y.; Hughes, T.; Zhao, Z.; Fan, S. Adjoint-based optimization of active nanophotonic devices. *Opt. Express* **2018**, *26*, 3236–3248. [[CrossRef](#)]
40. Georgieva, N.; Glavic, S.; Bakr, M.; Bandler, J. Feasible adjoint sensitivity technique for EM design optimization. *IEEE Trans. Microw. Theory Tech.* **2002**, *50*, 2751–2758. [[CrossRef](#)]
41. Lu, J.; Vučković, J. Nanophotonic computational design. *Opt. Express* **2013**, *21*, 13351–13367. [[CrossRef](#)]
42. Yee, K. Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media. *IEEE Trans. Antennas Propag.* **1966**, *14*, 302–307. [[CrossRef](#)]
43. Jin, J. *The Finite Element Method in Electromagnetics*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2014.
44. Obayya, S. *Computational Photonics*; Wiley: Hoboken, NJ, USA, 2011.

45. Kreutz-Delgado, K. The Complex Gradient Operator and the CR-Calculus. *arXiv* **2009**, arXiv:0906.4835.
46. Press, W.H. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*; Cambridge University Press: Cambridge, UK, 2007.
47. Elesin, Y.; Lazarov, B.; Jensen, J.; Sigmund, O. Design of Robust and Efficient Photonic Switches Using Topology Optimization. *Photonics Nanostruct.-Fundam. Appl.* **2012**, *10*, 153–165. [[CrossRef](#)]
48. Rasmus, E. Christiansen and Ole Sigmund. Inverse design in photonics by topology optimization: Tutorial. *J. Opt. Soc. Am. B* **2021**, *38*, 496–509. [[CrossRef](#)]
49. Aage, N.; Andreassen, E.; Lazarov, B.S.; Sigmund, O. Giga-voxel computational morphogenesis for structural design. *Nature* **2017**, *550*, 84–86. [[CrossRef](#)] [[PubMed](#)]
50. Jensen, J.S.; Sigmund, O. Topology Optimization of Photonic Crystal Structures: A High-Bandwidth Low-Loss T-junction Waveguide. *J. Opt. Soc. Am. B* **2005**, *22*, 1191. [[CrossRef](#)]
51. Jensen, J.S.; Sigmund, O. Topology optimization for nano-photonics. *Laser Photonics Rev.* **2011**, *5*, 308–321. [[CrossRef](#)]
52. Bendsøe, M.P.; Kikuchi, N. Generating optimal topologies in structural design using a homogenization method. *Comput. Methods Appl. Mech. Eng.* **1988**, *71*, 197–224. [[CrossRef](#)]
53. Tortorelli, D.A.; Michaleris, P. Design sensitivity analysis: Overview and review. *Inverse Probl. Eng.* **1994**, *1*, 71–105. [[CrossRef](#)]
54. Frellesen, L.F.; Ding, Y.; Sigmund, O.; Frandsen, L.H. Topology Optimized Mode Multiplexing in Silicon-on-Insulator Photonic Wire Waveguides. *Opt. Express* **2016**, *24*, 16866. [[CrossRef](#)]
55. Niederberger, A.C.R.; Fattal, D.A.; Gauger, N.R.; Fan, S.; Beausoleil, R.G. Sensitivity Analysis and Optimization of Sub-Wavelength Optical Gratings Using Adjoint. *Opt. Express* **2014**, *22*, 12971. [[CrossRef](#)] [[PubMed](#)]
56. Lin, Z.; Liu, V.; Pestourie, R.; Johnson, S.G. Topology optimization of freeform large-area metasurfaces. *Opt. Express* **2019**, *27*, 15765–15775. [[CrossRef](#)] [[PubMed](#)]
57. Tsuji, Y.; Hirayama, K.; Nomura, T.; Sato, K.; Nishiwaki, S. Design of optical circuit devices based on topology optimization. *IEEE Photonics Technol. Lett.* **2006**, *18*, 850–852. [[CrossRef](#)]
58. Shrestha, P.K.; Chun, Y.T.; Chu, D. A High-Resolution Optically Addressed Spatial Light Modulator Based on ZnO Nanoparticles. *Light. Sci. Appl.* **2015**, *4*, e259. [[CrossRef](#)]
59. Zhou, H.; Liao, K.; Su, Z.; Li, T.; Geng, G.; Li, J.; Wang, Y.; Hu, X.; Huang, L. Tunable On-Chip Mode Converter Enabled by Inverse Design. *Nanophotonics* **2023**, *12*, 1105–1114. [[CrossRef](#)]
60. Mansouree, M.; Kwon, H.; Arbabi, E.; McClung, A.; Faraon, A.; Arbabi, A. Multifunctional 2.5D Metastructures Enabled by Adjoint Optimization. *Optica* **2020**, *7*, 77. [[CrossRef](#)]
61. Minkov, M.; Williamson, I.A.D.; Andreani, L.C.; Gerace, D.; Lou, B.; Song, A.Y.; Hughes, T.W.; Fan, S. Inverse Design of Photonic Crystals through Automatic Differentiation. *ACS Photonics* **2020**, *7*, 1729–1741. [[CrossRef](#)]
62. Chung, H.; Miller, O.D. Tunable Metasurface Inverse Design for 80% Switching Efficiencies and 144° Angular Deflection. *ACS Photonics* **2020**, *7*, 2236–2243. [[CrossRef](#)]
63. Wang, K.; Ren, X.; Chang, W.; Lu, L.; Liu, D.; Zhang, M. Inverse Design of Digital Nanophotonic Devices Using the Adjoint Method. *Photonics Res.* **2020**, *8*, 528. [[CrossRef](#)]
64. Deng, Y.; Liu, Z.; Song, C.; Wu, J.; Liu, Y.; Wu, Y. Topology Optimization-Based Computational Design Methodology for Surface Plasmon Polaritons. *Plasmonics* **2015**, *10*, 569–583. [[CrossRef](#)]
65. Sell, D.; Yang, J.; Doshay, S.; Yang, R.; Fan, J.A. Large-Angle, Multifunctional Metagratings Based on Freeform Multimode Geometries. *Nano Lett.* **2017**, *17*, 3752–3757. [[CrossRef](#)] [[PubMed](#)]
66. Jiang, J.; Chen, M.; Fan, J.A. Deep neural networks for the evaluation and design of photonic devices. *Nat. Rev. Mater.* **2021**, *6*, 679–700. [[CrossRef](#)]
67. Jin, C.; Ge, R.; Netrapalli, P.; Kakade, S.M.; Jordan, M.I. How to Escape Saddle Points Efficiently. *arXiv* **2017**, arXiv:1703.00887.
68. Fan, J.A. Freeform metasurface design based on topology optimization. *MRS Bull.* **2020**, *45*, 196–201. [[CrossRef](#)]
69. Deng, L.; Xu, Y.; Liu, Y. Hybrid inverse design of photonic structures by combining optimization methods with neural networks. *Photonics Nanostruct.-Fundam. Appl.* **2022**, *52*, 101073. [[CrossRef](#)]
70. Colburn, S.; Majumdar, A. Inverse design and flexible parameterization of meta-optics using algorithmic differentiation. *Commun. Phys.* **2021**, *4*, 65. [[CrossRef](#)]
71. Zhan, A.; Fryett, T.K.; Colburn, S.; Majumdar, A. Inverse design of optical elements based on arrays of dielectric spheres. *Appl. Opt.* **2018**, *57*, 1437–1446. [[CrossRef](#)] [[PubMed](#)]
72. Bayati, E.; Pestourie, R.; Colburn, S.; Lin, Z.; Johnson, S.G.; Majumdar, A. Inverse Designed Metalenses with Extended Depth of Focus. *ACS Photonics* **2020**, *7*, 873–878. [[CrossRef](#)]
73. Backer, A.S. Computational inverse design for cascaded systems of metasurface optics. *Opt. Express* **2019**, *27*, 30308–30331. [[CrossRef](#)] [[PubMed](#)]
74. Wang, E.W.; Sell, D.; Phan, T.; Fan, J.A. Robust design of topology-optimized metasurfaces. *Opt. Mater. Express* **2019**, *9*, 469–482. [[CrossRef](#)]
75. Moharam, M.G.; Gaylord, T.K. Rigorous coupled-wave analysis of planar-grating diffraction. *J. Opt. Soc. Am.* **1981**, *71*, 811–818. [[CrossRef](#)]
76. Gill, P.E.; Murray, W.; Wright, M.H. *Practical Optimization*; Academic Press: London, UK, 1981.
77. So, S.; Badloe, T.; Noh, J.; Bravo-Abad, J.; Rho, J. Deep Learning Enabled Inverse Design in Nanophotonics. *Nanophotonics* **2020**, *9*, 1041–1057. [[CrossRef](#)]

78. Bodaghi, M.; Damanpack, A.; Hu, G.; Liao, W. Large deformations of soft metamaterials fabricated by 3D printing. *Mater. Des.* **2017**, *131*, 81–91. [[CrossRef](#)]
79. Sanchis, L.; Håkansson, A.; López-Zanón, D.; Bravo-Abad, J.; Sánchez-Dehesa, J. Integrated optical devices design by genetic algorithm. *Appl. Phys. Lett.* **2004**, *84*, 4460–4462. [[CrossRef](#)]
80. Su, L.; Piggott, A.Y.; Sapra, N.V.; Petykiewicz, J.; Vučković, J. Inverse Design and Demonstration of a Compact On-Chip Narrowband Three-Channel Wavelength Demultiplexer. *ACS Photonics* **2018**, *5*, 301–305. [[CrossRef](#)]
81. Silver, D.; Schrittwieser, J.; Simonyan, K.; Antonoglou, I.; Huang, A.; Guez, A.; Hubert, T.; Baker, L.; Lai, M.; Bolton, A.; et al. Mastering the game of go without human knowledge. *Nature* **2017**, *550*, 354–359. [[CrossRef](#)] [[PubMed](#)]
82. Krasikov, S.; Tranter, A.; Bogdanov, A.; Kivshar, Y. Intelligent metaphotonics empowered by machine learning. *Opto-Electron. Adv.* **2022**, *5*, 210147. [[CrossRef](#)]
83. Wang, Z.; Xiao, Y.; Liao, K.; Li, T.; Song, H.; Chen, H.; Uddin, S.M.Z.; Mao, D.; Wang, F.; Zhou, Z.; et al. Metasurface on Integrated Photonic Platform: From Mode Converters to Machine Learning. *Nanophotonics* **2022**, *11*, 3531–3546. [[CrossRef](#)]
84. Gigli, C.; Saba, A.; Ayoub, A.B.; Psaltis, D. Predicting nonlinear optical scattering with physics-driven neural networks. *APL Photonics* **2023**, *8*, 026105. [[CrossRef](#)]
85. Salmela, L.; Tsipinakis, N.; Foi, A.; Billet, C.; Dudley, J.M.; Genty, G. Predicting ultrafast nonlinear dynamics in fibre optics with a recurrent neural network. *Nat. Mach. Intell.* **2021**, *3*, 344–354. [[CrossRef](#)]
86. Fan, Q.; Zhou, G.; Gui, T.; Lu, C.; Lau, A.P.T. Advancing theoretical understanding and practical performance of signal processing for nonlinear optical communications through machine learning. *Nat. Commun.* **2020**, *11*, 3694. [[CrossRef](#)]
87. Teğin, U.; Yıldırım, M.; Oğuz, İ.; Moser, C.; Psaltis, D. Scalable optical learning operator. *Nat. Comput. Sci.* **2021**, *1*, 542–549. [[CrossRef](#)]
88. Wright, L.G.; Onodera, T.; Stein, M.M.; Wang, T.; Schachter, D.T.; Hu, Z.; McMahon, P.L. Deep physical neural networks trained with backpropagation. *Nature* **2022**, *601*, 549–555. [[CrossRef](#)] [[PubMed](#)]
89. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
90. Yun, J.; Kim, S.; So, S.; Kim, M.; Rho, J. Deep Learning for Topological Photonics. *Adv. Phys. X* **2022**, *7*, 2046156. [[CrossRef](#)]
91. Alagappan, G.; Ong, J.R.; Yang, Z.; Ang, T.Y.L.; Zhao, W.; Jiang, Y.; Zhang, W.; Png, C.E. Leveraging AI in Photonics and Beyond. *Photonics* **2022**, *9*, 75. [[CrossRef](#)]
92. Zhang, T.; Wang, J.; Liu, Q.; Zhou, J.; Dai, J.; Han, X.; Zhou, Y.; Xu, K. Efficient Spectrum Prediction and Inverse Design for Plasmonic Waveguide Systems Based on Artificial Neural Networks. *Photonics Res.* **2019**, *7*, 368. [[CrossRef](#)]
93. Wu, Q.; Li, X.; Wang, W.; Dong, Q.; Xiao, Y.; Cao, X.; Wang, L.; Gao, L. Comparison of Different Neural Network Architectures for Plasmonic Inverse Design. *ACS Omega* **2021**, *6*, 23076–23082. [[CrossRef](#)]
94. Olivas, E.S.; Guerrero, J.D.M.; Martinez-Sober, M.; Magdalena-Benedito, J.R.; Serrano, L. *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques: Algorithms, Methods, and Techniques*; IGI Global: Hershey, PA, USA, 2009.
95. Mikolov, T.; Karafiát, M.; Burget, L.; Černocký, J.; Khudanpur, S. Recurrent neural network based language model. In Proceedings of the Interspeech 2010, Chiba, Japan, 26–30 September 2010; pp. 1045–1048. [[CrossRef](#)]
96. He, L.; Wen, Z.; Jin, Y.; Torrent, D.; Zhuang, X.; Rabczuk, T. Inverse Design of Topological Metaplates for Flexural Waves with Machine Learning. *Mater. Des.* **2021**, *199*, 109390. [[CrossRef](#)]
97. Lininger, A.; Hinczewski, M.; Strangi, G. General Inverse Design of Layered Thin-Film Materials with Convolutional Neural Networks. *ACS Photonics* **2021**, *8*, 3641–3650. [[CrossRef](#)]
98. Lin, R.; Zhai, Y.; Xiong, C.; Li, X. Inverse Design of Plasmonic Metasurfaces by Convolutional Neural Network. *Opt. Lett.* **2020**, *45*, 1362. [[CrossRef](#)]
99. Ma, W.; Liu, Z.; Kudyshev, Z.A.; Boltasseva, A.; Cai, W.; Liu, Y. Deep Learning for the Design of Photonic Structures. *Nat. Photonics* **2021**, *15*, 77–90. [[CrossRef](#)]
100. Zhou, Q.; Yang, C.; Liang, A.; Zheng, X.; Chen, Z. Low computationally complex recurrent neural network for high speed optical fiber transmission. *Opt. Commun.* **2019**, *441*, 121–126. [[CrossRef](#)]
101. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; Association for Computational Linguistics: Cedarville, OH, USA, 2014; pp. 1724–1734.
102. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
103. Lee, M.C.; Yu, C.H.; Yao, C.K.; Li, Y.L.; Peng, P.C. A Neural-network-based Inverse Design of the Microwave Photonic Filter Using Multiwavelength Laser. *Opt. Commun.* **2022**, *523*, 128729. [[CrossRef](#)]
104. Sajedian, I.; Kim, J.; Rho, J. Finding the optical properties of plasmonic structures by image processing using a combination of convolutional neural networks and recurrent neural networks. *Microsyst. Nanoeng.* **2019**, *5*, 27. [[CrossRef](#)] [[PubMed](#)]
105. Liu, D.; Tan, Y.; Khoram, E.; Yu, Z. Training Deep Neural Networks for the Inverse Design of Nanophotonic Structures. *ACS Photonics* **2018**, *5*, 1365–1369. [[CrossRef](#)]
106. Kojima, K.; Tahersima, M.H.; Koike-Akino, T.; Jha, D.K.; Tang, Y.; Wang, Y.; Parsons, K. Deep Neural Networks for Inverse Design of Nanophotonic Devices. *J. Light. Technol.* **2021**, *39*, 1010–1019. [[CrossRef](#)]

107. Malkiel, I.; Mrejen, M.; Nagler, A.; Arieli, U.; Wolf, L.; Suchowski, H. Plasmonic Nanostructure Design and Characterization via Deep Learning. *Light Sci. Appl.* **2018**, *7*, 60. [[CrossRef](#)] [[PubMed](#)]
108. Tu, X.; Xie, W.; Chen, Z.; Ge, M.F.; Huang, T.; Song, C.; Fu, H.Y. Analysis of Deep Neural Network Models for Inverse Design of Silicon Photonic Grating Coupler. *J. Light. Technol.* **2021**, *39*, 2790–2799. [[CrossRef](#)]
109. Mao, S.; Cheng, L.; Chen, H.; Liu, X.; Geng, Z.; Li, Q.; Fu, H. Multi-task topology optimization of photonic devices in low-dimensional Fourier domain via deep learning. *Nanophotonics* **2023**, *12*, 1007–1018. [[CrossRef](#)]
110. Hegde, R.S. Photonics Inverse Design: Pairing Deep Neural Networks with Evolutionary Algorithms. *IEEE J. Select. Top. Quantum Electron.* **2020**, *26*, 1–8. [[CrossRef](#)]
111. Deng, Y.; Ren, S.; Malof, J.; Padilla, W.J. Deep Inverse Photonic Design: A Tutorial. *Photonics Nanostruct.-Fundam. Appl.* **2022**, *52*, 101070. [[CrossRef](#)]
112. So, S.; Rho, J. Designing nanophotonic structures using conditional deep convolutional generative adversarial networks. *Nanophotonics* **2019**, *8*, 1255–1261. [[CrossRef](#)]
113. An, S.; Zheng, B.; Tang, H.; Shalaginov, M.Y.; Zhou, L.; Li, H.; Kang, M.; Richardson, K.A.; Gu, T.; Hu, J.; et al. Multifunctional Metasurface Design with a Generative Adversarial Network. *Adv. Opt. Mater.* **2021**, *9*, 2001433. [[CrossRef](#)]
114. Yeung, C.; Tsai, R.; Pham, B.; King, B.; Kawagoe, Y.; Ho, D.; Liang, J.; Knight, M.W.; Raman, A.P. Global Inverse Design across Multiple Photonic Structure Classes Using Generative Deep Learning. *Adv. Opt. Mater.* **2021**, *9*, 2100548. [[CrossRef](#)]
115. Kudyshev, Z.A.; Kildishev, A.V.; Shalaev, V.M.; Boltasseva, A. Machine Learning-Assisted Global Optimization of Photonic Devices. *Nanophotonics* **2020**, *10*, 371–383. [[CrossRef](#)]
116. Paz, A.; Moran, S. Non deterministic polynomial optimization problems and their approximations. *Theor. Comput. Sci.* **1981**, *15*, 251–277. [[CrossRef](#)]
117. Doersch, C. Tutorial on Variational Autoencoders. *arXiv* **2021**, arXiv:1606.05908.
118. Yang, F.; Song, W.; Meng, F.; Luo, F.; Lou, S.; Lin, S.; Gong, Z.; Cao, J.; Barnard, E.S.; Chan, E.; et al. Tunable Second Harmonic Generation in Twisted Bilayer Graphene. *Matter* **2020**, *3*, 1361–1376. [[CrossRef](#)]
119. Wright, L.G.; Renninger, W.H.; Christodoulides, D.N.; Wise, F.W. Nonlinear multimode photonics: Nonlinear optics with many degrees of freedom. *Optica* **2022**, *9*, 824–841. [[CrossRef](#)]
120. Hegde, R.S. Deep learning: A new tool for photonic nanostructure design. *Nanoscale Adv.* **2020**, *2*, 1007–1023. [[CrossRef](#)] [[PubMed](#)]
121. Qiu, C.; Wu, X.; Luo, Z.; Yang, H.; Wang, G.; Liu, N.; Huang, B. Simultaneous inverse design continuous and discrete parameters of nanophotonic structures via back-propagation inverse neural network. *Opt. Commun.* **2021**, *483*, 126641. [[CrossRef](#)]
122. Liu, V.; Fan, S. S4: A free electromagnetic solver for layered periodic structures. *Comput. Phys. Commun.* **2012**, *183*, 2233–2244. [[CrossRef](#)]
123. Jiang, J.; Lupoiu, R.; Wang, E.W.; Sell, D.; Paul Hugonin, J.; Lalanne, P.; Fan, J.A. MetaNet: A New Paradigm for Data Sharing in Photonics Research. *Opt. Express* **2020**, *28*, 13670. [[CrossRef](#)]
124. Chen, Y.; Zhu, J.; Xie, Y.; Feng, N.; Liu, Q.H. Smart Inverse Design of Graphene-Based Photonic Metamaterials by an Adaptive Artificial Neural Network. *Nanoscale* **2019**, *11*, 9749–9755. [[CrossRef](#)]
125. Wiecha, P.R.; Arbouet, A.; Girard, C.; Muskens, O.L. Deep learning in nano-photonics: Inverse design and beyond. *Photonics Res.* **2021**, *9*, B182–B200. [[CrossRef](#)]
126. Sheverdin, A.; Monticone, F.; Valagiannopoulos, C. Photonic Inverse Design with Neural Networks: The Case of Invisibility in the Visible. *Phys. Rev. Appl.* **2020**, *14*, 024054. [[CrossRef](#)]
127. Qu, Y.; Jing, L.; Shen, Y.; Qiu, M.; Soljačić, M. Migrating Knowledge between Physical Scenarios Based on Artificial Neural Networks. *ACS Photonics* **2019**, *6*, 1168–1174. [[CrossRef](#)]
128. Unni, R.; Yao, K.; Zheng, Y. Deep Convolutional Mixture Density Network for Inverse Design of Layered Photonic Structures. *ACS Photonics* **2020**, *7*, 2703–2712. [[CrossRef](#)]
129. Tanriover, I.; Hadibrata, W.; Aydin, K. Physics-Based Approach for a Neural Networks Enabled Design of All-Dielectric Metasurfaces. *ACS Photonics* **2020**, *7*, 1957–1964. . [[CrossRef](#)]
130. Lenaerts, J.; Pinson, H.; Ginis, V. Artificial Neural Networks for Inverse Design of Resonant Nanophotonic Components with Oscillatory Loss Landscapes. *Nanophotonics* **2020**, *10*, 385–392. [[CrossRef](#)]
131. Jiang, J.; Fan, J.A. Multiobjective and Categorical Global Optimization of Photonic Structures Based on ResNet Generative Neural Networks. *Nanophotonics* **2020**, *10*, 361–369. [[CrossRef](#)]
132. Wang, Q.; Makarenko, M.; Burguete Lopez, A.; Getman, F.; Fratolocci, A. Advancing Statistical Learning and Artificial Intelligence in Nanophotonics Inverse Design. *Nanophotonics* **2022**, *11*, 2483–2505. [[CrossRef](#)]
133. Miller, O.D. *Photonic Design: From Fundamental Solar Cell Physics to Computational Inverse Design*; University of California: Berkeley, CA, USA, 2012.
134. Li, W.; Meng, F.; Chen, Y.; Li, Y.F.; Huang, X. Topology Optimization of Photonic and Phononic Crystals and Metamaterials: A Review. *Adv. Theory Simulations* **2019**, *2*, 1900017. [[CrossRef](#)]
135. Campbell, S.D.; Sell, D.; Jenkins, R.P.; Whiting, E.B.; Fan, J.A.; Werner, D.H. Review of numerical optimization techniques for meta-device design. *Opt. Mater. Express* **2019**, *9*, 1842–1863. [[CrossRef](#)]
136. Jiang, J.; Sell, D.; Hoyer, S.; Hickey, J.; Yang, J.; Fan, J.A. Free-Form Diffractive Metagrating Design Based on Generative Adversarial Networks. *ACS Nano* **2019**, *13*, 8872–8878. [[CrossRef](#)]

137. Hooten, S.; Beausoleil, R.G.; Van Vaerenbergh, T. Inverse Design of Grating Couplers Using the Policy Gradient Method from Reinforcement Learning. *Nanophotonics* **2021**, *10*, 3843–3856. [[CrossRef](#)]
138. Yeung, C.; Pham, B.; Tsai, R.; Fountaine, K.T.; Raman, A.P. DeepAdjoint: An All-in-One Photonic Inverse Design Framework Integrating Data-Driven Machine Learning with Optimization Algorithms. *ACS Photonics* **2023**, *10*, 884–891. [[CrossRef](#)]
139. Ren, S.; Padilla, W.; Malof, J.M. Benchmarking deep inverse models over time, and the neural-adjoint method. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 38–48.
140. Deng, Y.; Ren, S.; Fan, K.; Malof, J.M.; Padilla, W.J. Neural-Adjoint Method for the Inverse Design of All-Dielectric Metasurfaces. *Opt. Express* **2021**, *29*, 7526. [[CrossRef](#)]
141. Zhang, D.; Bao, Q.; Chen, W.; Liu, Z.; Wei, G.; Xiao, J.J. Inverse Design of an Optical Film Filter by a Recurrent Neural Adjoint Method: An Example for a Solar Simulator. *J. Opt. Soc. Am. B* **2021**, *38*, 1814. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.