Deep learning in nano-photonics: inverse design and beyond

PETER R. WIECHA¹,* , ARNAUD ARBOUET²,* , CHRISTIAN GIRARD²,* , AND OTTO L. MUSKENS³,*

¹ LAAS, Université de Toulouse, CNRS, Toulouse, France
² CEMES, Université de Toulouse, CNRS, Toulouse, France
³ Physics and Astronomy, Faculty of Engineering and Physical Sciences, University of Southampton, Southampton, UK
pwiecha@laas.fr
arbouet@cemes.fr
girard@cemes.fr
omuskens@soton.ac.uk

Compiled January 14, 2021

Deep learning in the context of nano-photonics is mostly discussed in terms of its potential for inverse design of photonic devices or nanostructures. Many of the recent works on machine-learning inverse design are highly specific, and the drawbacks of the respective approaches are often not immediately clear. In this review we want therefore to provide a critical review on the capabilities of deep learning for inverse design and the progress which has been made so far. We classify the different deep learning-based inverse design approaches at a higher level as well as by the context of their respective applications and critically discuss their strengths and weaknesses. While a significant part of the community’s attention lies on nano-photonic inverse design, deep learning has evolved as a tool for a large variety of applications. The second part of the review will focus therefore on machine learning research in nano-photonics “beyond inverse design”. This spans from physics informed neural networks for tremendous acceleration of photonics simulations, over sparse data reconstruction, imaging and “knowledge discovery” to experimental applications.

http://dx.doi.org/XX.XXXX/XX.XX.XXXXXX

CONTENTS

1 Introduction 1

2 Deep learning based nano-photonics inverse design 3
   A Surrogate model based inverse design . . . . . . . . 3
   B Direct neural network inverse design . . . . . . . . 4
   C Strategies to improve neural-network inverse design . 6
   D Heuristics vs. deep learning – a critical comparison . 9

3 Beyond inverse design 9
   A Physics informed neural networks: Solving PDEs . . 9
   B Interpretation of physical properties . . . . . . . . 10
   C Deep learning for interpretation of photonics experiments 12

4 Conclusions and perspectives 13

1. INTRODUCTION

Light-matter interaction at sub-wavelength dimensions can lead to astonishing effects like localized surface plasmon resonances which concentrate light to deeply sub-wavelength volumes [1], the appearance of optical magnetic resonances in otherwise non-magnetic media [2], the possibility to shape optical near-fields with sub-wavelength structure [3], the emergence of non-linear optical phenomena [4] or strong enhancement of quantum emitter luminescence [5] to name just a few. Those nano-scale optical effects can be exploited for a broad variety of applications, for instance in integrated quantum optics [6], for meta-materials [7] and in this context specifically for metasurfaces like flat lenses [8]. It is for example even possible to create all-optical devices which use light to solve integral equations or perform other analogue optical computing tasks [9–11].

Still, ever since the advent of nano-optics with the invention of near-field microscopy [13–15] the numerical description of many problems continues to be challenging [16]. An example is the rational design of nano-photonic structures for specific tasks which remains a general problem that often involves brute force “forward” calculations, or solving inverse scattering problems. Other challenges in nano-optics are related to experimental limitations such as the stochastic nature of single-photon emitters, fluctuating nanoscale force fields such as Brownian motion, and the diffraction limit blocking access to sub-wavelength information. Such effects often complicate the interpretation of nano-optics experiments and require the use of more sophisticated techniques for data analysis, for example combining data with prior knowledge or sparsity constraints. All these obstacles are about to be pushed significantly further by the emerging computational
An artificial neuron (AN) is simply a mathematical function which mimics the behavior of a biological neuron.

\[ f(x) = y \]

The step-like behavior of neuronal activation, which starts to fire once a threshold stimulation is exceeded, can be implemented by various nonlinear mathematical functions. A popular example is the logistic function (also called “sigmoid”), shown in the above sketch. If the scalar product of an input vector \( x \) and the neuron-intrinsic weight parameters \( w_i \) is larger than the neuron’s bias parameter \( b \), the output \( y \) is “high” (the artificial neuron fires). Else it is “low”.

An artificial neural network (ANN) is composed of several such ANs, usually arranged in “layers”. The output value of a neuron is fed into a succeeding layer of neurons. The final layer is the network output \( y \). For instance in a so-called fully connected ANN, every neuron of one layer is connected to every neuron of the following layer:

**Box 1. Artificial neurons, neural networks and their training**

Training an ANN is done via a numerical minimization of a loss function \( L \), which describes the error of the network in predicting samples of the training data. A popular loss function is the mean square error loss (MSE), in particular used for regression tasks:

\[
L(w_i, b_j) = \frac{1}{N} \sum_{i=1}^{N} (y_{train,i} - y_{ANN}(x_{train,i}))^2
\]

Training on small batches composed of random subsets of \( N \) training samples helps to “jump” out of local minima by adding a stochastic component to the procedure. One of the most common training algorithms is Stochastic Gradient Descent (“SGD”) [12].
very active research field on hardware implementations of artificial neural networks which includes – but is limited to – research efforts on photonics platforms [23, 40, 41].

In this mini-review we focus on selected key results that have recently led to breakthrough advancements in the research on inverse design of photonic nanostructures and metasurfaces. Rather than compiling an exhaustive catalogue of every single publication, we provide an overview of milestone concepts for improving deep learning inverse design fidelity, which recently allowed to bring ANNs closer to the performance of conventional optimization methods. We believe that such a summary of concepts is of particular interest for researchers in the field. We dedicate the second part of the review to an overview of original applications of deep learning in nano-photonics beyond structural inverse design. Specifically we summarize recent developments around physics informed neural networks in optics, on deep learning for knowledge discovery and explainable machine learning as well as on applications of ANNs to nano-photonics experiments.

2. DEEP LEARNING BASED NANO-PHOTONICS INVERSE DESIGN

The first part of this mini-review is dedicated to deep learning based inverse design techniques as well as to concepts to improve the inverse design model fidelity. As stated before, we do not aim to provide an exhaustive list of applications. An up-to-date and very complete overview of possible optimization targets can be found for instance in the recent reviews by Ma et al. [30] or by Jiang et al. [32].

“Conventional” inverse design methods

Before the recent rise of deep learning methods, inverse design of nano-photonic structures was often based on intuitive considerations and systematic fine-tuning (see e.g. [42, 43]). A more systematic alternative was the combination of numerical simulation methods with gradient based or heuristic optimization algorithms like simulated annealing, topology optimization or genetic algorithms [35, 44–47]. Such methods led to some remarkable success for instance in the optimization of plasmonic optical antennas [48, 49], dielectric multi-functional nanostructures [50] or metasurfaces [34, 51]. A great advantage of such methods is the possibility to include fabrication constraints or robustness conditions in the optimization procedure [50, 52].

However, heuristics coupled to numerical simulation techniques is slow and computationally expensive. Furthermore, for each new optimization target, the parameter-space needs to be searched from scratch, implying hundreds to thousands of numerical simulations. The recent advent of data-driven techniques like deep learning holds promise to accelerate the computation by many orders of magnitude and quite some remarkable progress has been made in the past few years. One can distinguish two types of approach that have gained traction. The first one replaces the forward simulation in an iterative optimization with an ANN, while the second aims to build an inverse ANN that solves the problem directly. Below we critically discuss the two approaches as well as efforts at improving the quality of results.

A. Surrogate model based inverse design

Deep learning models are particularly strong in predicting approximate solutions to direct problems like the optical response of photonic structures. A possible approach to accelerate inverse design is therefore to
use a “forward neural network” as an ultra-fast predictor together with an optimization technique. In such a case the ANN acts as a so-called surrogate model, taking the place of the much slower conventional simulation method.

**Deep learning forward solver**

ANNs have been successfully trained on the prediction of various physical quantities in nanophotonics. Early works have proposed ANNs to create phenomenological models of non-linear optical effects or of optical ionization using experimental training data [55, 56]. Recently, the idea has been picked up and it has been shown for instance that scattering and extinction spectra can be predicted with high accuracy [57] and also that the phase can be included in the predictions [58], which is important for nanostructures in metasurfaces. The prediction of far-field observables can also be extended to include proximity effects in a dense metasurface, beyond the local phase approximation. The latter has been demonstrated by including the near-field interactions with the nearest neighbor structures in the training data [59]. The prediction of physical effects is not limited to extinction, transmission or other far-field effects. It has been shown that also near-field effects can be approximated accurately, for instance around nanowires of complex shape [60].

While networks that predict an observable like the scattering cross section perform usually very well within the range of their training data, such models often generalize rather poorly to cases outside the parameter range covered by the training data. The ANN acts then as universal function approximator, but it does not develop a deeper “understanding” of the underlying physics. In order to alleviate this problem, it turned out to be helpful to provide the network with pre-processed data. For instance, instead of training an ANN with pure optical extinction spectra, So et al. [26] trained their model using a decomposition in multiple electric and magnetic dipole resonances (ED, respectively MD), to predict the optical response of multi-material multi-shell nano-spheres. The approach is illustrated in figure 1a and has also been used to inverse design multi-shell spheres for Kerker-type directional scattering. Using a metallic grating as model example, Blanchard-Dionne and Martin demonstrated that a neural network that learns light-matter interaction through a representation as multiple Lorentz-oscillators generalizes about an order of magnitude better outside the training data range, compared to a predictor network based on the raw optical spectrum [27] (see figure 1b). Instead of predicting specific physical observables such as the extinction cross section, Wiecha et al. demonstrated that a network can learn a discrete dipole approximation of the electric polarization density inside a 3D nanostructure of arbitrary shape [28].

**Forward predictor networks + evolutionary optimization**

In general, the greatest advantage of deep learning techniques as surrogate models for physics simulations is their tremendous evaluation speed. Once trained, an ANN delivers its prediction within fractions of milliseconds, which is usually orders of magnitude faster than a numerical simulation. Therefore, replacing conventional physics simulations by surrogate ANNs is a natural solution to speed-up the inverse design of photonic nano-structures via global optimization heuristics [61, 62]. This concept has recently been applied by several groups to the design of individual photonic nanostructures or metasurfaces [63–68].

However, while the approach can significantly accelerate heuristic-based inverse design, it remains an iterative approach requiring thousands of calls to the surrogate model as well as intermediate computation steps. Furthermore, the surrogate model represents only an approximation to the physical reality, introducing a systematic error. And even worse than that, it cannot be guaranteed that the surrogate model does not contain singular points of totally false solutions [69], to which the optimization algorithm may converge in the worst case scenario. Robust implementations therefore require a simulation-based fine-tuning procedure subsequent to the surrogate-based optimization run, which often relativizes the gain in speed [70, 71]. The same problem holds of course also for the here-after discussed ANN-only inverse design methods.

**B. Direct neural network inverse design**

As mentioned above, using forward ANNs as surrogate models for evolutionary optimization is computationally not the most efficient technique and bares the risk of converging to singular points of the
Let’s assume a simple toy problem: Under fixed wavelength illumination, we want to tailor the extinction coefficient of a gold nanorod by varying its length:

Already this simple problem is ambiguous: Several rod lengths can lead to the same extinction, which makes a naive ANN implementation fail in those cases.

The “tandem neural network” can stabilize the generator (G) via a physics loss, based on a pre-trained forward model (fwd). This approach however limits the inverse design to one solution per design target, rendering inaccessible possible multiple solutions to a given problem:

Mixture density networks predict several possible solutions at a time including their respective importance as Gaussian distributions. A disadvantage is that the maximum number of possible simultaneous solutions needs to be known.

Conditional generative adversarial networks (cGANs) or conditional [adversarial] autoencoders (c[A]AE) add a normal distributed latent vector (usually: “z”) to the design target (“condition”), which encodes dynamically multiple possible solutions. Adversarial models furthermore use a trained loss, a so-called discriminator network (D), which tries to distinguish generated (fake) from training samples (true).

Another model that circumvents the one-to-many problem is the conditional Generative Adversarial Network (cGAN) [54, 70, 76–78]. A cGAN takes as input not only the design target but also an additional “latent vector”, which is a normally distributed sequence of random values. The network then learns to use different values of the latent vector to address the distinct non-unique solutions. In addition to the introduction of a latent vector, a further peculiarity of cGANs is their loss function, which is a discriminator network that tries to distinguish generated solutions from real ones, and which is also subject to training. During training, the cGAN loss function hence evolves in a balanced way for the adversarial loss function to work efficiently.

A further type of one-to-many solving networks are conditional adversarial or conditional variational autoencoders (cAAEs, cVAEs) [68, 79–82]. Those are usually symmetric models that take the physical response as input, which they try to identically reconstruct at their output layer. In a conditional autoencoder, a bottleneck layer is placed in the ANN center. This bottleneck contains the design parameters on the one hand (as it is the case in a tandem-network), but on the other
C. Strategies to improve neural-network inverse design

Data-driven inverse design has the important drawback that the accuracy of the model is first of all limited by the quality of the data and an interpolation error between the data-samples is introduced by the ANN. Early works on inverse design therefore reported rather qualitative agreement, but relatively large quantitative inaccuracies. Therefore, in the recent past remarkable efforts have been put in developing methods to improve neural network inverse design. In this section we want to provide an overview over the most successful concepts. In general, two main constituents offer the largest potential for optimization: The training data and the neural network model.

Improving the data quality

As mentioned before, many ANN models do actually generalize relatively poorly to cases outside the parameter range of the training data. They act mainly as generalized function approximators, hence they interpolate very efficiently to fill the gaps in the training data, while their extrapolation capability remains limited. But also the interpolation risks to be unsatisfactory if the physical model underlying the training data has sharp features such as high quality factor resonances. If the training data does not contain a sufficient number of such resonant cases, there is a high risk that those features will be very poorly approximated by an ANN.

To tackle this problem, training data can be generated using an optimization algorithm to produce specific responses for the dataset [53, 88]. In case of many free parameters this procedure is time-consuming, therefore recently the idea of iterative training data generation has emerged [53, 66, 68, 88, 89, 94]. The principal idea is depicted in figure 3a. An initial dataset is generated traditionally via a randomized procedure, on which an inverse design ANN is trained. This network is subsequently used to construct devices based on realistic design targets, but these designs are likely to be rather mediocre as the initial ANN performs relatively poor. Now, the true physical response of these mediocre ANN designs is calculated in another run of numerical simulations, and these samples are appended to the training data. The generator ANN is then trained again on the now extended training data and the generative cycle is repeated. In this way, the neural network can literally learn from its previous mistakes and its performance on the specific design task will significantly improve. Fig. 3a shows the example of an optical cloak design problem, for which the inverse design accuracy could be improved by more than one order of magnitude thanks to iterative training [89]. To visualize the evolution of the training data quality, figure 3b shows the statistical distributions of resonator quality factors in a fully random dataset of photonic crystal cavities (left) compared to a dataset after one iteration of iterative training (right) [94]. The lack of resonant geometries in the randomly generated dataset is evident. Despite those solutions not being present in the initial dataset, the ANN managed to conjecture a certain amount of resonant cases, improving the training data for the second iteration. By repeating the procedure the training data increasingly contains resonant geometries, which consequently allows the ANN to inverse design close-to-optimum solutions. Another positive side-effect specifically in tandem networks is that iterative training simultaneously improves the accuracy of the forward network [53]. A recent work showed that an even better design performance can be
An obvious drawback of iterative procedures is their computational cost. Data generation is usually slow and the expensive network training needs to be repeated several times on increasing amounts of training samples. Several suggestions have been made to accelerate the convergence of iterative data generation in order to reduce the number of cycles. For instance by training several networks, the statistics from multiple predictions can be used to assess the quality and the uncertainty of the ANN output (“wisdom of the many” [101], see also Box 4). This information can be exploited to choose only the best new solutions for re-simulation and insertion into the expanded training data, which reduces the number of expensive physics simulations [66]. Similarly, an evolutionary optimization algorithm might be coupled to a generative ANN in the iterative cycle to further specialize the training data with regards to the anticipated optimization target [68]. A drawback of such training-data optimization strategies is a risk of over-specializing the network to optimum cases and losing its capability to generalize to arbitrary situations. Therefore care needs to be taken that the training
Box 4. “Wisdom of the many”

Wisdom of the many or also wisdom of the crowd denotes the procedure of training multiple neural networks on the same data, each ANN with random initialization. We illustrate the idea by the example of an optical spectrum predictor network:

While this approach adds a significant computational cost (training several networks), the mean $\mu$ of $N$ independent predictions provides a $\sqrt{N}$ times smaller statistical error compared to using a single ANN. Furthermore, the standard deviation $\sigma$ of multiple predictions can be used to assess the credence of the ANN output.

Physics model based loss function

A similar, yet somehow more radical concept is to not use a fixed set of training data at all but instead implement a loss function based on a physical model within the framework of the machine learning toolkit. Such an approach has been illustrated recently by the example of inverse designing multi-layer thin-film stacks for specific reflection and transmission spectra [98]. As highlighted by a red box on the right in figure 3f, a transfer-matrix method (TMM) has been implemented directly in the deep learning toolkit as a loss function. In consequence, error backpropagation is possible through the TMM solver, and the network can be trained without an explicit dataset. The loss function in this so-called “GLOnet” is used to optimize the transmission and reflection spectra of a multi-layer stack with respect to a design target. It is worth mentioning that the GLOnet learns to optimize a single design target, hence in principle the training of the network takes the place of a conventional global optimization algorithm run (hence its name “GLOnet”). The authors of Ref. [98] claim that the training dynamics allow their GLOnet to ideally adapt its optimization scheme to each problem, resulting in better and faster convergence compared to hard-coded optimizers. The same authors have generalized their concept to a somehow more flexible inverse network called “conditional GLOnet”, using an iterative training scheme instead of a fully differentiable physics loss function. For the training, gradients of the design efficiency are calculated via adjoint simulations and re-injected for backpropagation through the network [102]. The conditional GLOnet is conceptually similar to a Pareto optimization in which a set of optimum solutions for a multi-objective problem is calculated [103]. While the specific solving of a single problem is intentional in Refs. [98, 102], as already mentioned before over-specialization is an inherent danger of all iterative data-generation methods.

Another concept to replace the dataset by a direct evaluation of a physics model has been demonstrated for the Helmholtz equation, by developing a loss function which directly evaluates this partial differential equation (PDE). Such ANN model is called a “physics informed neural network” (PINN). In case of a Helmholtz-PINN, the network learns to directly solve the wave equation in frequency domain. The inverse design target is then implemented as a boundary condition matching problem [90, 104]. As in the GLOnet case, also such PINN inverse design requires a new training run for each optimization target. PINNs will be discussed in more detail later in this review.

Sophisticated ANN models

The second main lever allowing for performance optimization of inverse design ANNs is the neural network model itself. It has been proven helpful to adopt recent findings in the research on optimum network layout for deep learning. For instance, if applicable the “U-Net” architecture [105] offers much better training convergence and
Apart from optimizing the network model and generating training data was recently adapted to nano-photonics [96, 110] (see also figure 3d). ANNs can be applied in addition to the above named network architectures. The concept is based on representing the design parameters in a “modal” representation as multiple Gaussian distributions, where each of the Gaussian distributions describes a possible solution to an ambiguous problem (see also figure 3d). This concept has been proposed some time ago for microwave device inverse design [108, 109] and was recently adapted to nano-photonics [96, 110] (see also figure 3d). The advantage is that the network can in principle deliver all possible solutions together with a weight for their respective priorities. A drawback of the approach is that the approximate number of non-unique solutions needs to be known in advance.

Another recent proposition to optimize inverse networks specifically for noisy situations like in experiments is the implementation of concepts from machine learning based image denoising [111]. As shown in figure 3e, Hu et al. added artificial noise on training data and could demonstrate that a denoising network based inverse ANN offers a very robust performance even when trained on very noisy data [97]. This opens promising perspectives for experimental applications.

Reformatting the input data

Apart from optimizing the network model and generating training data of high quality, the format of the inputs and outputs of a neural network can play a decisive role in whether the ANN manages to “understand” the data or not. An example is illustrated in figure 4a, where a physical problem is to be solved on a non-Cartesian coordinate domain. On 2D problems such as the here shown one, typically convolutional neural networks (CNNs) are most efficient. However, as can be seen in the leftmost panel, the imposed discretization on a square mesh is very poor. This holds in particular for the domain borders. Gao et al. [99] proposed therefore to apply a transformation of the coordinate system from the physics domain to the CNN reference domain prior training. As illustrated in figure 4b by the example of solving the heat equation, this additional pre-processing allows to successfully apply ANNs to very complex non-uniform physical domains.

The problem of discretization can also be alleviated by applying a topology encoding procedure, for instance via Fourier transformation [100]. The idea is illustrated in figure 4c-d. Such encoding can allow not only to describe geometries with odd shapes without restrictions due to discretization, it allows furthermore to condense the information to a low-dimensional space, which is helpful to reduce ANN complexity and furthermore advantageous in preventing overfitting.

Other concepts

Further possibilities to improve the quality of ANN based inverse design are to use the ANN only as a first step for a rough estimate, and apply a conventional iterative approach in a subsequent refinement step. Heuristic optimization algorithms usually benefit strongly from a good initial guess [70]. Another recent proposition is to use a forward neural network purely as an ultra-fast physics predictor to construct a huge look-up table [112]. Using a well trained forward network, a look-up table can be created which covers the entire parameter space at a very fine resolution, impossible to achieve with conventional numerical methods. Appropriate solutions to specific problems can subsequently be searched in this database. Transfer learning has also been recently applied to nano-optics problems to improve ANN performance if only small amounts of data exist [113]. For instance experimental data is often expensive, but the situation can be improved by training an ANN first on simulated data, and subsequently specialize the pre-trained network via transfer learning on the experimental dataset [114].

D. Heuristics vs. deep learning – a critical comparison

It is of utmost importance to emphasize that a data-driven inverse design technique can never outperform an iterative method if it is based on the same simulation model used for training data generation. At least not if no time constraint is set for the iterative optimization. Well trained and optimized data-driven ANNs usually produce errors in the order of a few percent [28, 60]. Furthermore, it is virtually impossible to completely suppress outliers in the network predictions [69]. At the singular points the error of the ANN can be orders of magnitude higher. It is thus a delicate task to assess whether a prediction is valid or rather the result of a singularity in the ANN.

While recently some sophisticated training techniques were presented that are capable to train ANNs for performances similar to conventional inverse optimization, they are either still considerably constrained or the high accuracy has a severe impact on the computational cost. Examples are physics-loss based inverse ANNs or networks based on progressive-complexity training schemes [95, 98]. The model described in Ref. [98] for example is constrained to a simple transfer-matrix description of a multi-layer system as well as to the inverse design of a single optimization target.

The fact that ANNs always introduce an additional error is inherent to the data-driven nature of machine learning, which implies that a ML model can never outperform the accuracy of the simulations used to create the dataset or the model defining the training loss. On the other hand, once trained ANN techniques can offer extreme speed-up of the inverse design, generally many orders of magnitude faster than iterative approaches based on numerical simulations. It is not unusual that milliseconds stand against hours or even days. This is a marvellous advantage and often well worth to accept the reduced accuracy of ANN based techniques. In daily applications a few percent error might actually not even matter too much, in particular when compared to the typical magnitude of inaccuracies in fabrication.

On the other hand, concerning the inverse design speed it is important to remember that the ultra-fast predictions require a fully trained neural network. This implies the computationally highly demanding data generation as well as the very expensive training of the ANN. In many situations, conventional global optimization is in sum actually computationally cheaper. In conclusion, deep learning based inverse design is mainly interesting for applications which require a large number of repetitions of similar design tasks, or that rely on ultimate speed for the design generation.

3. BEYOND INVERSE DESIGN

The second part of this review is dedicated to applications of deep learning in nano-photonics “beyond inverse design”. We give an overview on physics informed neural networks, we present recent work on ANNs for physics interpretation and knowledge discovery as well as experimental applications.

A. Physics informed neural networks: Solving PDEs

Most machine learning applications in physics aim to predict derived observables such as transmittance or extinction cross sections. In contrast, the idea of physics informed neural networks (PINNs) is to train an ANN to directly predict the solution of a partial differential equation ([P]DE). While this would be also possible using a dataset
of pre-calculated solutions, the particularity of PINNs is that instead of using a loss function for data-comparison like MSE, the PINN-loss implements an explicit evaluation of the PDE. In consequence, no pre-calculated training data is required for training. For the PINN-loss, the PDE derivatives of the ANN-predicted observables are directly implemented in the respective deep learning toolkit. Thus, the PINN-loss can be seen as a consistency check for the predicted solution. Because modern deep learning toolkits offer powerful automatic differentiation functionalities, error backpropagation through the PINN-loss remains possible and the ANN can be efficiently trained without data.

This concept has been first proposed in 2019 by Raissi et al. [119] and has since then attracted a great deal of attention across countless research communities in physics such as fluid mechanics [119, 120], thermodynamics [121] or geophysics [107]. Compared to data-based ANNs, the accuracy of PINNs is in general significantly higher. On the other hand, because PINNs evaluate the underlying PDE “point-by-point”, they are usually slower than conventional data-based models. Since the latter work on physical observables, it is easier to predict higher-dimensional data-structures at a time, making better use of the massively parallel computing architectures of modern GPUs. Nevertheless, PINNs are usually orders of magnitude faster than numerical PDE solvers.

Applications to nano-photonics are still scarce. Recently Moseley et al. demonstrated that PINNs are capable to accurately solve the wave equation in time domain [107]. An example is shown in figure 5a, demonstrating seismic wave propagation through an inhomogeneous medium at successive snapshots in time. As can be seen, the PINN is capable to predict the evolution of the wave propagation even in a complex environment. While Ref. [107] treats shock waves in geophysics, the problem is conceptually identical to the wave equation in electrodynamics.

Depicted in figure 5b, Fang and Zhan recently demonstrated that a PINN can accurately solve the Helmholtz equation, describing wave propagation in frequency domain [104]. They found that sinusoidal activation functions are the most adequate choice to solve a differential equation with time-harmonic solutions. By formulating the inverse design as a boundary condition matching problem, it was possible to use the Helmholtz-PINN for the design of an optical cloak, as illustrated in the bottom of figure 5b. A similar frequency-domain PINN has been proposed for the homogenization of optical metamaterials [90]. A disadvantage of PINNs is that the environment needs to be defined at the training stage and hence a new network needs to be trained if the boundary conditions change. Each PINN-based inverse design therefore involves a new training procedure, comparable with conventional iterative techniques, which is evidently much slower than “direct” inverse-ANN models. Conceptually related to PINNs is also the so-called “GLOnet”, which is discussed in more detail above [98] (see also figure 3f).

B. Interpretation of physical properties

In this section we will review recent approaches to extract information and correlations from deep learning models in order to reveal physical insights.

There is on the one hand the possibility to use deep learning models for dimensionality reduction. Figure 6a shows a work of Kiarashinejad et al. in which the number of discrete values in reflectance spectra from a set of electrodynamical simulations is reduced from 200 to 2 via an unsupervised autoencoder ANN. In a second step, the non-convex hull of the compressed responses is calculated, which represents the region in the 2D compressed space containing all encoded points. This region allows to assess the range of accessible physical responses within the allowed design parameters, the method is hence helpful to identify the physical limitations of specific nanostructure models. Note that the full physical response of any point in the reduced dimensionality space can be reconstructed using the decoder part of the autoencoder, also those points that were not present in the training data. This means that feasible and non-feasible responses can be analyzed in the original response space (under the assumption that the neural network generalizes well to out-of-training situations). Note that autoencoders are unsupervised ANN models, which are known to require relatively few data for training. This facilitates the application of the technique to experimental data.

In a similar approach, the impact of variations of individual design parameters on the latent space can be studied. Those parameters whose variations have large (respectively little) impact on the latent space contribute strongly (respectively weakly) to the optical response [116, 126, 127]. The latent space is indicated by yellow highlighted neurons in Fig. 6b, top right. The impact of physical parameters on these weights is illustrated in the bottom right of figure 6b. Iten et al. [117] extended the encoder-decoder ANN for interpretable physics via an approach inspired by humans’ interpretation and modelling of physical observations. The concept is depicted in Fig. 6c, where the motion of a mass is observed as function of time $x(t)$. To implement this concept in an ANN the authors append a condition to the latent-vector at the bottleneck of an encoder-decoder ANN (see Fig. 6d). This condition is here called a question, the example in Fig. 6c uses the time $t'$ for which the ANN shall predict the position of the moving mass ($= \text{the answer}$). In the context of nano-photonics the question could be an optical spectrum of a nanostructure. The “answer”...
Fig. 6. Examples of “knowledge discovery” through machine learning. a) The feasibility of a physical response by a defined geometric model can be assessed by a dimensionality reduction through an autoencoder neural network and subsequent non-convex hull determination. Adapted from [115], copyright (2019) the authors. b) Study of the impact of the number of bottleneck neurons $N$ (left spectra) as well as of nanostructure design variations on the activation of the bottleneck neurons (W1-W4 in case $N = 4$, yellow neurons in the top right panel). This analysis allows to assess the physical importance of individual design parameters and reveals information about the complexity of the optical response. Adapted with permission from [116], copyright (2019) John Wiley and Sons. (c-d) By mimicking the human approach of interpreting and modelling physical observations (c), a conditional encoder-decoder network (d) can be used to discover implicit physics concepts from data. Reprinted with permission from [117], copyright (2020) APS. e) Exploiting the high speed of a physics predictor network permits a systematic analysis of the achievable phase and intensity variations in metasurface constituent design. Adapted from [118], copyright (2020) Optical Society of America.

As already mentioned before, another way to gain insight in physical processes through a machine learning analysis is to use a physical parametrization of the training data, such that the neural network explicitly returns a physical quantity. As shown in figure 1a and 1b, extinction spectra can for example be pre-processed in a modal decomposition, such as a superposition of electric and magnetic dipole resonances [26] or as a decomposition in Lorentzian resonance profiles [27]. Once trained, the respective neural networks deliver an explicit interpretation of the predicted spectra.

In another recent work, so-called explainable machine learning has been used to assess the importance of constituent parts of a nanostructure with respect to its optical response, as well as to identify those parts of the structure that contribute only weakly to the light-matter interaction [128]. Such information is important for the design of fabrication-robust nanostructures, but also for applications in which sub-constituents of high impact on the nanostructure’s optical response...
need to be identified, e.g., for switchable optical antennas. Another recent work proposes interpretable machine learning models like decision trees and random forests to understand the physical mechanisms behind inverse design results [129].

C. Deep learning for interpretation of photonics experiments

The last section of this review is dedicated to recent applications of deep learning in nano-photonics experiments.

Deep learning has proven to enable unprecedented statistical evaluation of large and complicated data, which was formerly impossible with conventional methods. It has been demonstrated for instance...
that ANN models can learn from huge microscopy datasets to optically characterize 2D materials such as graphene or transition-metal dichalcogenides (TMDs) [130] or to automatically localize and classify nano-scale defects [131] or to track particles in 3D space using holographic microscopy [132]. Deep learning was also successfully applied for the ultra-fast analysis of single molecule emission patterns [133] as well as for the experimental reconstruction of quantum states for quantum optics tomography [134].

By training an ANN on large amounts of experimental optical scattering spectra from complex photonic nanostructures, recently an optical information storage concept has been proposed, able to push the data-density beyond the optical diffraction limit [121]. The principle is depicted in figure 7a. Digital information is encoded in silicon nanostructures, which are designed such that each nanostructure encoding a specific bit-sequence possesses a unique scattering spectrum. Visible light scattering is subsequently interpreted by an artificial neural network (Fig. 7b). Training on experimental data renders this read-out robust against fabrication imperfections and instrumental noise. Therein, the ANN is the key ingredient to allow high readout accuracies from distorted data (Fig. 7c). Deep learning can be used for various further experimental classification tasks in nano-optics. For instance, as depicted in figure 7d it has been recently demonstrated that an ANN can learn to classify different species of Anthrax spores from holographic microscopy images [122]. The confusion rates in the individual classes (Fig. 7e) allow furthermore to calibrate and control electrically reconfigurable photonic circuits by interpreting the ANN reconstructions. Since data-driven approaches always bare the risk of being biased towards the training data, care must be taken when detrimentally over-specialized inverse networks, care must be taken when interpreting the ANN reconstructions. Since data-driven approaches always bare the risk of being biased towards the training data, a neural network might for instance detect a learned statistical distribution even in pure noise.

Deep learning can be applied not only to data analysis, but is also increasingly used to control real time experimental feedback systems. Recent examples touching the field of nano-photonics are mainly found in AI-stabilized microscopy. ANNs can be applied for instance to real-time image enhancement [163], microcopy stabilizing feed-back systems [20, 164] or to conduct sparse data acquisition schemes for the acceleration of scanning microscopy systems via compressive sensing [165]. ANNs have been also applied to controlling laser mode-locking stabilization systems [166–168]. So far the direct application of ANNs to experimental hardware for nano-photonics is still scarce, but the research is in an early stage. A recent work proposed for instance to calibrate and control electrically reconfigurable photonic circuits by deep learning algorithms [169, 169]. Another example is a pioneering work of Selle et al. [56] which proposed to use ANNs coupled to a femtosecond laser pulse shaper for real-time control of the light-matter interaction in nano-structures or molecules. We expect a very rapid development of applications in this direction in the near future, in particular real-time critical applications like sensing [170] will hugely benefit from the tremendous acceleration potential of ANNs.

4. CONCLUSIONS AND PERSPECTIVES

In conclusion, in this mini-review we discussed the most recent developments in deep learning methods applied to nano-photonics. In the
first section we focused on ANN-driven nano-photonic inverse design methods and discussed concepts to improve the design quality of inverse ANNs in comparison with conventional optimization techniques. In the second part we discussed applications of deep learning in nano-photronics “beyond inverse design”, spanning from physics-informed neural networks over ANNs for physical knowledge extraction to data interpretation and experimental applications.

We would like to emphasize that despite their latest remarkable success and their undeniable great potential, artificial neural networks are “black boxes”. It is extremely hard, mostly even impossible to understand how a neural network generates its predictions. It has been demonstrated at many occasions that even the most sophisticated ANNs, trained on the most carefully assembled datasets contain singular points at which their predictions diverge. Another noteworthy danger of data-driven techniques is that they bare a considerable risk to be biased with respect to their training data, like an incident where google’s image-tagging algorithm learned implicit racism from its training data [171]. We therefore appeal to the reader to keep in mind that, simply speaking, “what you put in is what you get out”. In consequence the ANN models are only the second most important ingredient to deep learning. The essential element is first of all the training data. Unfortunately it is often understated and not discussed with sufficient emphasis that high quality training data is of utmost importance. By reviewing techniques that aim at improving the training data quality, we tried to arouse some awareness in this respect. Another important aspect in this context is the amount of training data, required to train a well performing and generalizing ANN. Unfortunately in many problems which would be naturally suited for deep learning applications, training data is scarce or very expensive to generate. Additionally, the more general a problem for an ANN is, the more training data is usually required for a good prediction fidelity. Last but not least, adapting an ANN model to a new problem often requires the entire training data to be generated from scratch, which might even be the case for minor modifications. These aspects can create considerable computational barriers for broad and flexible applications of ANNs.

Deep learning techniques in the context of nano-photronics have experienced a tremendous amount of attention in the past few years and research has literally exploded. ANNs have enabled manifold applications which formerly seemed strictly impossible. As discussed above, a prominent example are data-driven ultra-fast solvers for various inverse problems, for which conventional methods are computationally extremely expensive and slow. We expect that further groundbreaking applications will be developed in the near future. For instance very promising progress has been made in the field of Quantum machine learning [172], which aims at using deep learning concepts to push the capabilities and interpretability of quantum computing systems. In this context, machine learning algorithms recently have autonomously proposed designs for non-trivial quantum optics experiments [173–175]. We expect that deep learning will continue to produce exciting pioneering results. We also anticipate that deep learning techniques will become a common numerical tool, regularly employed for the daily use.

Acknowledgments

We thank the NVIDIA Corporation for the donation of a Quadro P6000 GPU used for this research. This work was supported by the German Research Foundation (DFG) through a research fellowship (WI 5261/1-1). The authors acknowledge the CALMIP computing facility (grant p20010). OM acknowledges support through EPSRC grant EP/M009122/1.

Disclosures

The authors declare no conflicts of interest.

REFERENCES

1. P. Mühlschlegel, H.-J. Eisler, O. J. F. Martin, B. Hecht, and D. W. Pohl, “Resonant Optical Antennas,” Science 308, 1607–1609 (2005).
2. A. I. Kuznetsov, A. E. Miroshnichenko, M. L. Brongersma, Y. S. Kivshar, and B. Luk’yanchuk, “Optically resonant dielectric nanostructures,” Science 354 (2016).
3. C. Girard, “Near fields in nanostructures,” Reports on Prog. Phys. 68, 1883–1933 (2005).
4. M. Kauranen and A. V. Zayats, “Nonlinear plasmonics,” Nat. Photonics 6, 737–748 (2012).
5. G. C. des Francs, J. Barthes, A. Bouhelier, J. C. Weeber, A. Dereux, A. Cuche, and C. Girard, “Plasmonic Purcell factor and coupling efficiency to surface plasmons. Implications for addressing and controlling optical nanostructures,” J. Opt. 18, 094005 (2016).
6. J. Wang, F. Sciarrotta, A. Laing, and M. G. Thompson, “Integrated photonic quantum technologies,” Nat. Photonics 14, 273–284 (2020).
7. J. B. Pendry, “Negative Refraction Makes a Perfect Lens,” Phys. Rev. Lett. 85, 3966–3969 (2000).
8. P. Genevet, F. Capasso, F. Aletta, M. Khorasaninejad, and R. Devlin, “Recent advances in planar optics: From plasmonic to dielectric meta-surfaces,” Optica. 4, 139–152 (2017).
9. N. M. Estakhri, B. Edwards, and N. Engheta, “Inverse-designed metasstructures that solve equations,” Science, 363, 1333–1338 (2019).
10. W. R. Clements, P. C. Humphreys, B. J.Metcalf, W. S. Kolthammer, and I. A. Walmsley, “Optimal design for universal multiphoton interferometers,” Optics. 3, 1460–1465 (2016).
11. F. Zangeneh-nejad, D. L. Scoumas, A. Alu, and R. Fleury, “Analogue computing with metamaterials,” Nat. Rev. Mater. pp. 1–19 (2020).
12. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning (MIT Press, 2016).
13. E. A. Ash and G. Nicholls, “Super-resolution Aperture Scanning Microscope,” Nature, 237, 510–512 (1972).
14. D. W. Pohl, W. Denk, and M. Lanz, “Optical stethoscopy: Image recording with resolution λ/20,” Appl. Phys. Lett. 44, 651–653 (1984).
15. E. Betzig, A. Harootunian, A. Lewis, and M. Isaacson, “Near-field diffraction by a slit: Implications for superresolution microscopy,” Appl. Opt. 25, 1890–1900 (1986).
16. B. Gallinet, J. Butet, and O. J. F. Martin, “Numerical methods for nanophotonics: Standard problems and future challenges,” Laser & Photonics Rev. 9, 577–603 (2015).
17. Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, 521, 436–444 (2015).
18. S. Chan and E. L. Siegel, “Will machine learning end the viability of radiology as a thriving medical specialty?” The Br. J. Radiol. 92, 20180416 (2018).
19. A. S. Lundervold and A. Lundervold, “An overview of deep learning in medical imaging focusing on MRI,” Zeitschrift für Medizinische Physik 29, 102–127 (2019).
20. D. A. Cirovic, “Feed-forward artificial neural networks: Applications to spectroscopy,” TRAC Trends Anal. Chem. 16, 148–155 (1997).
21. T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, B. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, “Language Models are Few-Shot Learners,” arXiv:2005.14165 [cs] (2020).
22. C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning,” arXiv:1602.07261 [cs] (2016).
23. X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, “All-optical machine learning using diffractive deep neural networks,” Science 361, 1004–1008 (2018).
24. T. W. Hughes, I. A. D. Williamson, M. Minkov, and S. Fan, “Wave physics as an analog recurrent neural network,” Sci. Adv. 5, eaay6946 (2019).
25. D. Mengu, Y. Rivenson, and A. Ozcan, “Scale-, shift- and rotation-invariant diffractive optical networks,” arXiv:2010.12747 [physics]
learning approach toward scalable imaging through scattering media,” Optica 5, 1181–11819 (2018).

156. B. Rahmani, D. Loterie, G. Konstantinou, D. Psaltis, and C. Moser, “Multimode optical fiber transmission with a deep learning network,” Light, Sci. & Appl. 7, 69 (2018).

157. G. D. Bruce, L. O’Donnell, M. Chen, M. Facchin, and K. Dholakia, “Femtometer-resolved simultaneous measurement of multiple laser wavelengths in a speckle wavemeter,” Opt. Lett. 45, 1926–1929 (2020).

158. S. Popoff, G. Lerosey, M. Fink, A. C. Boccara, and S. Gigan, “Image transmission through an opaque material,” Nat. Commun. 1, 81 (2010).

159. R. French, S. Gigan, and O. L. Muskens, “Snapshot fiber spectral imaging using speckle correlations and compressive sensing,” Opt. Express 26, 32302–32316 (2018).

160. H. Pinkard, Z. Phillips, A. Babakhani, D. A. Fletcher, and L. Waller, “Deep learning for single-shot autofocus microscopy,” Optica 6, 794–797 (2019).

161. Z. A. Kudyshev, S. I. Bogdanov, T. Isacsson, A. V. Kildishev, A. Boltasseva, and V. M. Shalaev, “Rapid Classification of Quantum Sources Enabled by Machine Learning,” Adv. Quantum Technol. 3, 2000067 (2020).

162. C. You, M. A. Quiroz-Juárez, A. Lambert, N. Bhusal, C. Dong, A. Perez-Leija, A. Javaid, R. d. J. León-Montiel, and O. S. Magaña-Loaiza, “Identification of light sources using machine learning,” Appl. Phys. Rev. 7, 021404 (2020).

163. Y. Rivenson, H. Ceylan Koydemir, H. Wang, Z. Wei, Z. Ren, H. Günday, Y. Zhang, Z. Göröcs, K. Liang, D. Tseng, and A. Ozcan, “Deep Learning Enhanced Mobile-Phone Microscopy,” ACS Photonics 5, 2354–2364 (2018).

164. X. Li, J. Dong, B. Li, Y. Zhang, Y. Zhang, A. Veeraraghavan, and X. Ji, “Fast confocal microscopy imaging based on deep learning,” in 2020 IEEE International Conference on Computational Photography (ICCP), 2020, pp. 1–12.

165. J. M. Ede and R. Beanland, “Partial Scanning Transmission Electron Microscopy with Deep Learning,” Sci. Reports 10, 8332 (2020).

166. S. L. Brunton, X. Fu, and J. N. Kutz, “Self-Tuning Fiber Lasers,” IEEE J. Sel. Top. Quantum Electron. 20, 464–471 (2014).

167. J. N. Kutz and S. L. Brunton, “Intelligent Systems for Stabilizing Mode-Locked Lasers and Frequency Combs: Machine Learning and Equation-Free Control Paradigms for Self-Tuning Optics,” Nanophotonics. 4, 459–471 (2015).

168. T. Baumeister, S. L. Brunton, and J. N. Kutz, “Deep learning and model predictive control for self-tuning mode-locked lasers,” JOSA B 35, 617–626 (2018).

169. A. Youssef, R. J. Chapman, A. Peruzzo, C. Ferrie, and M. Tomamichel, “Modeling and control of a reconfigurable photonic circuit using deep learning,” Quantum Sci. Technol. 5, 025001 (2020).

170. B. Wang, J. C. Cancilla, J. S. Torrecilla, and H. Haick, “Artificial Sensing Intelligence with Silicon Nanowires for Ultrasensitive Detection in the Gas Phase,” Nano Lett. 14, 933–938 (2014).

171. “Google says sorry for racist auto-tag in photo app,” Guardian. (2015).

172. M. Schuld, I. Sinayskiy, and F. Petruccione, “An introduction to quantum machine learning,” Contemp. Phys. 56, 172–185 (2015).

173. M. Krenn, M. Malik, R. Fickler, R. Lapkiewicz, and A. Zeilinger, “Automated Search for new Quantum Experiments,” Phys. Rev. Lett. 116, 090405 (2016).

174. A. A. Melnikov, H. P. Nautrup, M. Krenn, V. Dunjko, M. Tiersch, A. Zeilinger, and H. J. Briegel, “Active learning machine learns to create new quantum experiments,” Proc. Natl. Acad. Sci. 115, 1221–1226 (2018).

175. M. Krenn, M. Erhard, and A. Zeilinger, “Computer-inspired quantum experiments,” Nat. Rev. Phys. 2, 649–661 (2020).