

ARTICLE

Open Access

Photonic-dispersion neural networks for inverse scattering problems

Tongyu Li^{1,2}, Ang Chen², Lingjie Fan^{1,2}, Minjia Zheng^{1,2}, Jiajun Wang^{1,2}, Guopeng Lu², Maoxiong Zhao^{1,2}, Xinbin Cheng³, Wei Li⁴, Xiaohan Liu^{1,5}, Haiwei Yin², Lei Shi^{1,2,5} and Jian Zi^{1,5}

Abstract

Inferring the properties of a scattering objective by analyzing the optical far-field responses within the framework of inverse problems is of great practical significance. However, it still faces major challenges when the parameter range is growing and involves inevitable experimental noises. Here, we propose a solving strategy containing robust neural-networks-based algorithms and informative photonic dispersions to overcome such challenges for a sort of inverse scattering problem—reconstructing grating profiles. Using two typical neural networks, forward-mapping type and inverse-mapping type, we reconstruct grating profiles whose geometric features span hundreds of nanometers with nanometric sensitivity and several seconds of time consumption. A forward-mapping neural network with a parameters-to-point architecture especially stands out in generating analytical photonic dispersions accurately, featured by sharp Fano-shaped spectra. Meanwhile, to implement the strategy experimentally, a Fourier-optics-based angle-resolved imaging spectroscopy with an all-fixed light path is developed to measure the dispersions by a single shot, acquiring adequate information. Our forward-mapping algorithm can enable real-time comparisons between robust predictions and experimental data with actual noises, showing an excellent linear correlation ($R^2 > 0.982$) with the measurements of atomic force microscopy. Our work provides a new strategy for reconstructing grating profiles in inverse scattering problems.

Introduction

Inverse scattering problems (ISPs) arise in many fields of science and engineering such as computed tomography^{1,2}, fiber Bragg gratings³, and optical metrology^{4–7}. A typical ISP, is composed of three parts: a set of scattering objectives, a set of light responses and a measurement operator. For scattering objectives, one should make a parameter space whose elements are arrays of parameters, describing the scatters' geometries and components; for light responses, a data space is needed whose elements correspond to the measured optical responses of scatters in the far field, such as reflectance spectra. As the connection of

these two sets, a measurement operator characterizes the mapping from parameter space to data space. To solve ISPs, namely inferring an element of the parameter space from that of the data space, it executes the inversion of the measurement operator—inversion operator. Two key properties of the inversion operator are its injectivity and stability⁸. Injectivity requires the acquired data to uniquely characterize the parameters, and stability is closely related to the measurement noises.

Many algorithms and measuring techniques have been developed to solve ISPs with good injectivity and stability. In terms of algorithms, the genetic algorithm⁹ and library approach¹⁰ stand out with their understandability and feasibility. However, the existing algorithms are usually time-consuming due to the global optimization of a huge parameter space. Recently, neural networks^{11,12} (NNs) have offered a new perspective to solve inverse problems^{13–17}; for instance, by inverse structure design^{18–27}. But when applied to practical ISPs, the performance of

Correspondence: Lei Shi (lshi@fudan.edu.cn) or Jian Zi (jzi@fudan.edu.cn)

¹State Key Laboratory of Surface Physics, Key Laboratory of Micro- and Nano-Photonics Structures (Ministry of Education) and Department of Physics, Fudan University, Shanghai 200433, China

²Shanghai Engineering Research Center of Optical Metrology for Nano-fabrication (SERCOM), Shanghai 200433, China

Full list of author information is available at the end of the article

These authors contributed equally: Tongyu Li, Ang Chen.

© The Author(s) 2021



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

NN-based algorithms suffers from inevitable measuring noises, showing low stability. As for measuring techniques, high-throughput measuring methodologies such as Mueller matrix ellipsometry²⁸ are of great practical importance to provide adequate information for mapping algorithms. Redundant information ensures the injectivity of mapping, but it brings in sort of extra vibration instability when detecting multi-dimensional signals by mechanical modules. Thus, it is still a challenge to perform a rapid stable high-throughput measurement by a single-shot imaging technique.

To obtain a technique with both injectivity and stability, we develop a high-throughput Fourier-optics-based angle-resolved imaging spectroscopy (ARS) embedded with robust NN-based algorithms to solve ISPs. Our solving strategies are experimentally applied to a particular ISP of reconstructing the silicon-on-insulator (SOI) grating profiles with nanometric-scale precision. We discuss two kinds of NN-based algorithms: One is the inverse-mapping algorithm, and the other is the forward-mapping-based optimization algorithm—forward-mapping algorithm. We first train an inverse-mapping NN to learn the inversion operator, directly mapping from data space to parameter space. On the other side, the forward-mapping NN is trained to learn the measurement operator from parameter space to data space, and a gradient-based optimization is further performed on the parameter space to find the optimal solution. Both algorithms are able to reconstruct the SOI grating profiles in the parameter space, whose size is orders of magnitude larger than those of the traditional methods (whose covered ranges of considered parameters are usually no more than 20 nm). The consumed mapping time is at the level of seconds, in which the inverse mapping costs < 1 s while forward mapping costs around 20 s. Considering the performance on experimental data, the forward-mapping algorithm shows more robustness to actual measurements and enables a real-time comparison of the responses after the solving process. In addition, we propose that the dispersions of a scattering objective can be used as the elements of data space. The structure information is contained in the shapes of dispersion curves labeled by wavelength (λ) and angle (θ), besides the absolute quantity of the reflection intensity, offering multi-dimensional information. Using the home-made ARS, we experimentally obtain the dispersion patterns with the all-fixed light path by single shot imaging. When armed with the NN-based algorithms, the reconstructed geometric parameters achieve a strong linear correlation ($R^2 > 0.982$) with the measurements of atomic force microscopy (AFM).

Results

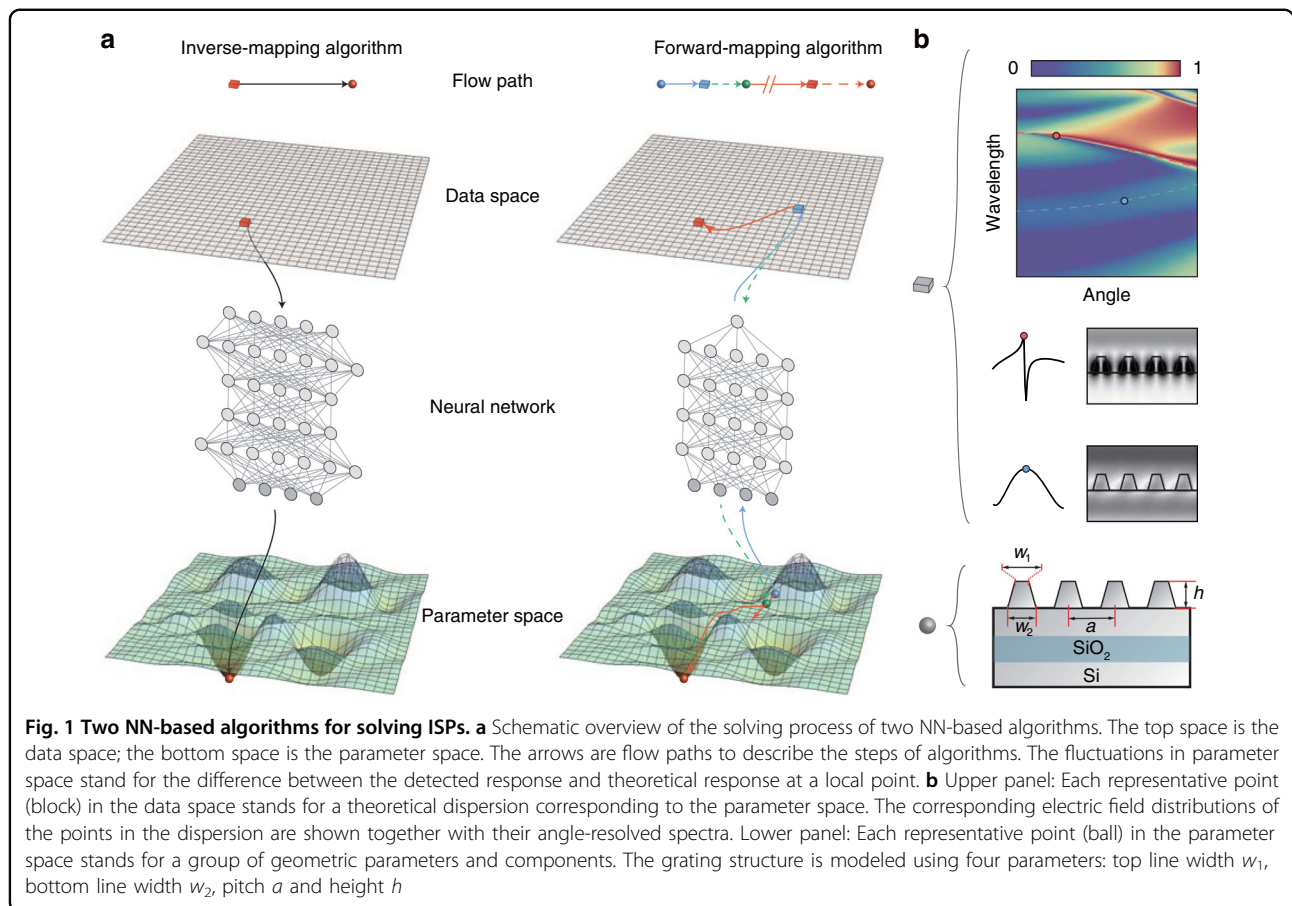
In this section, we mainly discuss the key technical innovations in both mapping algorithms and measuring

methodology, and our feasible strategy is further performed to reconstruct the SOI grating profiles from detected dispersions. For the algorithms, we focus more specifically on the details of the method using the forward-mapping algorithm in the main text, and those for the inverse-mapping algorithm are given in Supplementary Information.

Overview of the algorithms

For ISPs, the discussions are usually expanded between the parameter space and the data space, as illustrated in Fig. 1a. Each representative point (ball) in the parameter space stands for a group of geometric parameters and components; each representative point (block) in the data space stands for the detected responses corresponding to a ball in the parameter space. The aim of solving ISPs is to try to establish an inverse mapping from data space to parameter space, inferring the parameters of the scattering objectives from the given detected (light) responses. Since the core of the inverse problems is to characterize the inversion of the measurement operator, it is natural to train a NN as a recognizer performing the inverse mapping directly, called inverse-mapping algorithm as shown in the left panel of Fig. 1a. Once the NN recognizer is trained as an inverse operator, the inverse mapping can solve ISPs in a straightforward way. The solving process is quite intuitive for the inverse-mapping algorithm: a detected response of the scattering objective (red block) enters the inverse-mapping NN that is previously trained on the simulated data sets, and the prediction of the parameters is further output by NN (red ball). To practically train an inverse-mapping NN is a challenge, because in most ISP cases the objective response can be viewed as the theoretical response superimposed with a measurement noise. Although the noise compared with signals is weak, it could be amplified by the inverse-mapping NN, leading to enormous deviations of predictions.

To show such noise influence, we start with an inverse-mapping NN trained on the simulated examples without noises. On a noise-free test set, 98.7% predicted parameters had deviations of < 1 nm after 400 epochs of training. However, the performance on the test set with Gaussian noises ($\mu = 0$, $\sigma = 0.1$) became unsatisfactory (Fig. S1). A direct method to overcome that is to augment the data set during the training period, practically by adding some types of random noises, including Gaussian noise, to the theoretical responses. At this time, performance on the same noisy test set was improved greatly such that 97.8% predicted parameters had deviations of < 5 nm. It means that the robustness of inverse-mapping NNs can be enhanced by data set augmentation with the corresponding type of noise. But unfortunately, some unexpected noises in measurements still lead to unstable



results. In this regard, due to the intrinsic architecture of the inverse-mapping NN, it has no ability to further tune the predicted parameters subtly.

For another way, the NN can be trained as a generator to generate responses, called the forward-mapping algorithm, as shown in the right panel of Fig. 1a. Once the NN generator is trained as a substitute for simulation algorithms, the whole architecture of the forward mapping can be entirely analytical, enabling us to calculate the gradients of the input parameters directly with the back propagation algorithm. Specifically, the optimization process starts from a random point (blue ball) in the parameter space. Entering the NN generator along the blue arrow, the array of random initial parameters is mapped to the corresponding response (blue block). The difference between the generated response R_g and the detected response R_d (red block) is described with a cost function, for instance, mean square error $C = \sum |R_d - R_g|^2$, reflecting in the fluctuations in parameter space. The gradient can therefore be defined as $\nabla_p C$, where P stands for the parameters, as depicted by the green dash line in Fig. 1a. With the calculated gradient, the parameters (green ball) can be updated along the red arrow using some advanced gradient descent algorithms.

After repeating the above steps for a few times, the optimization process will finally find the optimal point (red ball) in the parameter space. To prevent converging to a locally optimal solution, the optimization process described above usually starts simultaneously from several initial points in the parameter space, and the candidate solution is picked out with the smallest C . So far, the gradient descent algorithm plays to its rapid convergence ability to approach the globally optimal solution. Gradient-free algorithms, such as the search algorithm or greedy algorithm, are finally performed to search for the final solution starting with the selected candidate solution. Considering the actual response with measurement noises, the gradient-free algorithm enables the forward-mapping algorithm to finely tune the parameters in parameter space to find a solution whose corresponding response is nearest to the measured one. The detailed results of the forward-mapping algorithm applied to ISPs are given in the following sections.

Data space element: photonic dispersion

Many kinds of light responses can be chosen as the elements of the data space. As an instance, Mueller matrix is usually used to describe the modulations of polarization

states of an objective. Though almost all of the structure information is contained in the elements of the Mueller matrix, yet it needs to meet tough conditions such as stable rolling cantilevers and superhigh signal-to-noise ratio. Here, we experimentally propose a new kind of light response for solving ISPs: the photonic dispersions of the grating, characterized by the wavelength–angle (λ – θ) mapping. A tremendous number of researches in nanophotonics have revealed that wealthy accessible information lies in photonic dispersions^{29–32} such as photonic band structures and iso-frequency contours. Solving inverse problems in photonic crystals with photonic band structures has been reported by recent works: Wei et al.¹⁶ established an inverse-mapping algorithm with a convolutional NN to predict the Zak phase of 1D photonic crystals precisely from input Hamiltonians. Christensen et al.³³ trained a convolutional NN and generative adversarial networks to predict and design inverse photonic crystal band structures with orders-of-magnitude speedup. In our case, a measured photonic dispersion that contains both abundant band structures features and reflectance information is used to solve ISPs. A typical dispersion of an SOI grating with s-polarized light excitation is shown in Fig. 1b. The dispersion bands depicted by the observed stripes stem from different physical mechanisms. For example, one broad dispersion band, marked as a blue dashed curve, can be interpreted as the thin-film interference, while one Fano-shaped dispersion band, marked as a red dashed curve, is caused by the coupling between guided resonances and thin-film oscillations^{34,35}. The various kinds of dispersive information can be further understood by their corresponding field distributions: Fields represented by the red point are extremely enhanced at the gratings, while those represented by the blue point are almost evenly distributed in the space. In this way, the grating structures can somehow injectively be in accordance with the dispersions labeled by (λ, θ) , since the labeled detected intensities as well as the stripe-formed shapes could be regarded as the ruler measuring how strongly detected lights can interact with the grating structures. Thus, in this work, we use photonic dispersions as our data space elements. Besides, the grating profile can be modeled as isosceles trapezoids with four parameters: top line width w_1 , bottom line width w_2 , pitch a and height h . These geometric parameters vary in a range of hundreds of nanometers, constituting a parameter space of huge sizes. Specifically, we consider the line widths between 130 nm and 330 nm with the bottom line width longer than the top, pitch between 350 nm and 550 nm, and height between 160 nm and 270 nm.

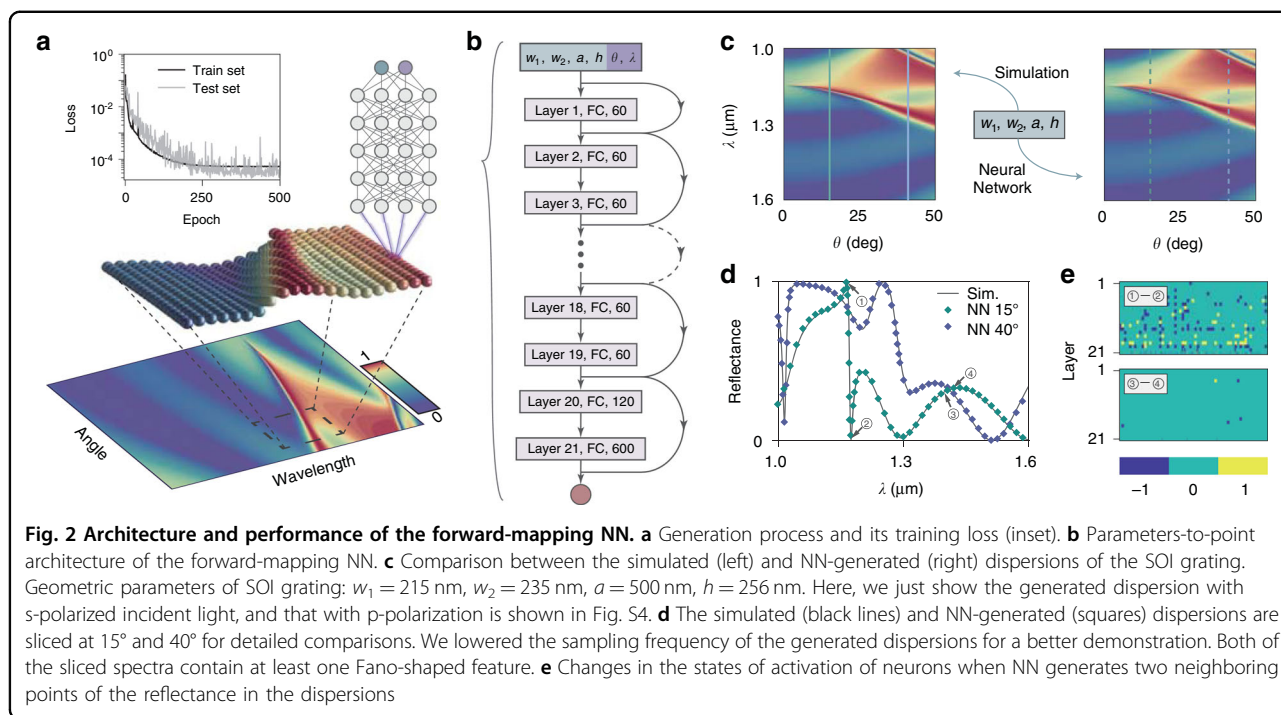
Forward-mapping NN

We observe sharp features and concretely abrupt changes at several wavelengths on the reflectance spectra,

which are quite general in scattering problems. Then the first and foremost step in the forward-mapping algorithm is to train a NN that enables to generate such sharp features with high degrees of precision. Using a NN to generate high-quality factor resonance has essential implications, since it is one of the crucial properties in nanostructure and has attracted attention to the enhancement of light–matter interactions³⁶. We firstly train a NN with the typical parameters-to-spectrum architecture, namely forward mapping the input geometric parameters to the whole spectrum in one time, to generate the dispersions. However, because of the correlations between two neighboring neurons in the output layer, we find such a NN that can only generate thin-film-interference features but fail in sharp Fano-shaped ones in a large-wavelength region (Fig. S3).

To overcome this deficiency, we develop a parameters-to-point forward-mapping NN with a different generating process, as illustrated in Fig. 2b. The word point here stands for a pixel (the ball connected with NN in Fig. 2a) on a dispersion pattern (reflectance at a single certain wavelength λ and a single certain angle θ). Besides the grating parameters, the NN inputs also include the labeled coordinates (λ, θ) of the dispersions and the output gives the corresponding reflectance. Such NN is realized by a residual fully-connected network with 21 layers: 60 neurons per layer in the former 19 layers, and 120 neurons and 600 neurons in the last two layers. Curving arrows between layers stand for the shortcuts of the residual blocks³⁷. Batch normalization is applied before each nonlinear layer³⁸. We use a rigorous coupled-wave analysis³⁹ (RCWA) method to simulate the reflectance of 5×10^7 points to set up a training set (see the Supplementary Information). Labels $(w_1, w_2, a, h, \lambda, \theta)$ of these samples form a six-dimensional space, where the reflectance data can be sampled with the Monte Carlo method. We have then trained the NN with an Adam optimizer⁴⁰ on the training set, with the training error plotted in the inset of Fig. 2a. After training, the generation process of a complete dispersion pattern is realized by the pixel-by-pixel strategy: varying generating point coordinates (λ, θ) making NN scan on the generation region to calculate the reflectance of each pixel. (In practice, the scanning process is performed in parallel.) With desired resolutions, the intervals of (λ, θ) can be tuned finely.

To verify the performance of our proposed NN, we did a comparison between the simulated and NN-generated dispersions (Fig. 2c). The slices of two dispersions are further compared in Fig. 2d, showing a high accuracy of the NN generation capability. It also should be pointed out that our NN does indeed perform well in generating sharp spectra with Fano-shaped features. For instance, when moving from point 1 towards point 2 within 10 nm along the green spectrum, the reflectance has an abrupt fall with



a change of almost one. To explore why the new NN was able to generate such sharp features, we checked the changes in activation states of the neurons. We use three colors to present the state changes [Fig. 2e]: blue stands for neurons switching from active states to inactive states, yellow for an inverse way, and green for those keeping the same states. We chose four points on the green spectrum for interpretation: Each of the groups (1,2) and (3,4) is 10 nm away from the other for its wavelength label, with (1,2) accounting for Fano-shaped dispersions and (3,4) for thin-film dispersions. In (1,2), small wavelength change makes more and more neurons switch their states of activation along the data flow, like the falling dominoes. For the layer next to the output, 25.7% of the neurons switch their states, resulting in an abrupt output change. As a contrast, with only fragmentary neurons switching the states, the outputs of (3,4) have only few changes.

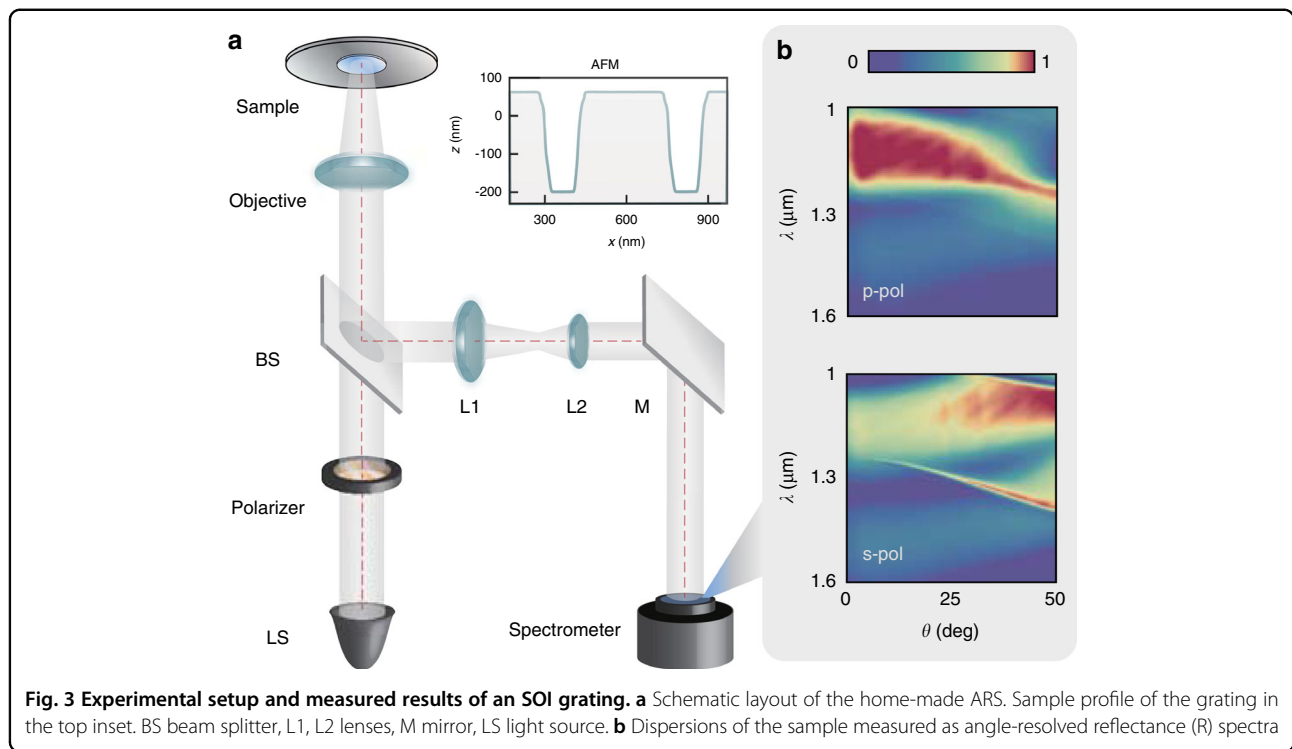
The parameters-to-point forward-mapping NN has advantages on several aspects. First, it has less time-consumption in generating photonic dispersions compared to the available electromagnetic simulation algorithms. For example, it will cost almost 10 min for the RCWA method to simulate only one photonic dispersion of an SOI grating with 200×51 pixels (interval wavelength is 3 nm) while just 2.82 s are needed to generate equally sized dispersions for 200 samples by our proposed NN. The second advantage lies in that it can be trained with very few data sets. The inverse-mapping NN usually needs a data set containing 60,000 dispersions (14 GB), still a quite small-volume set in comparison with the

library approach. As for our proposed NN case, a 3 GB data set has been already enough. Lastly, the parameters-to-point forward-mapping NN is 200 times smaller than the inverse-mapping one due to its slender architecture, taking up only 1 MB of memory.

Technology platform

In this section, we introduce the developed measuring methodology to measure the dispersions of SOI gratings. These grating structures are fabricated by using electron-beam lithography on the device layer of an SOI wafer (the thicknesses of the device layer and the buried oxide layer: ~ 270 nm and ~ 1 μ m) with nominally geometric parameters (line widths: 175–300 nm, periodicity: 400–500 nm). The structure areas ($400 \mu\text{m} \times 400 \mu\text{m}$) are much larger than the periodicity, which shield the influence of the grating boundary. The surface topography of an SOI grating measured by AFM is plotted in the top right inset of Fig. 3a.

Taking the angle-resolved reflectance spectra as a holistic characterization of the photonic band dispersion requires both broad spectral imaging and full angular incidences of the incident light. To meet such harsh requirements, we build our ARS based on Fourier analysis, as shown in Fig. 3a. Using Köhler illumination, the sample is illuminated at the front focal plane of the microscope by a halogen lamp. Passing through the incident linear polarizer and focused by an objective lens, the incident beam is convergent and linearly polarized. Then, the reflected beam is Fourier transformed by the same objective lens and imaged to an imaging



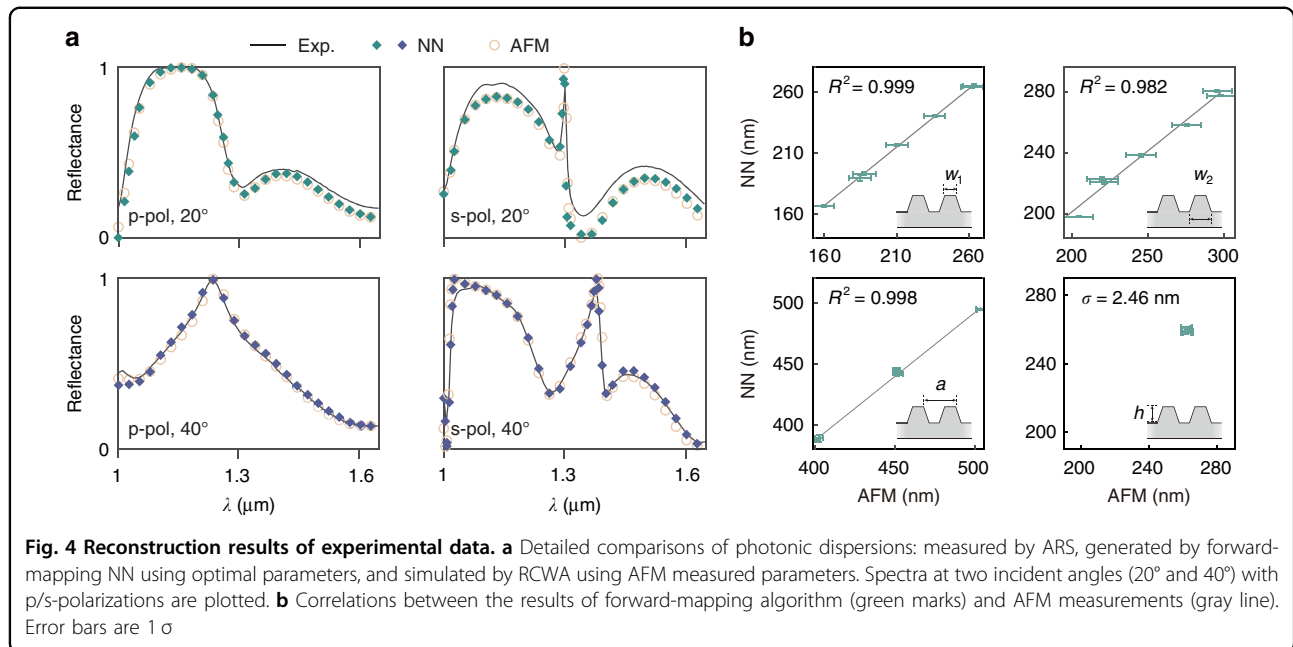
spectrometer. We can finally observe the dispersions through the two-dimensional charged-coupled-device camera. Here, the objective is 0.95 NA with $\times 100$ magnification, achieving incident angles of up to 50° in near-infrared light. The spot diameter is approximately $100 \mu\text{m}$, and measuring range of the spectrometer is from 1 to $1.6 \mu\text{m}$. The measured dispersions with p/s-polarized incidence are shown in Fig. 3b, where distinct dispersion bands including sharp Fano-shape features can be clearly observed. The multi-angle detection can be practically realized by the objective lens instead of a mechanical module, enabling us to obtain a dispersion pattern by a single shot³¹. The short measurement procedure and informative dispersions make ARS as a high-through put measuring methodology. In addition, all the optical elements are fixed during the measurement, avoiding additional mechanical noises, which is indispensable since the static light paths offer feasibilities for further calibrations.

Reconstruction results

To validate our forward-mapping algorithm, we first reconstruct SOI grating profiles from their simulated dispersions. 1,000 p/s-polarized pairs of dispersions with different geometric parameters are calculated by RCWA simulation, making the data set. Following the procedure explained in the subsection ‘Overview of algorithms’, we note that in the parameters-to-point NN, the gradient of parameters can be expressed as $\nabla_{\mathbf{p}} C = \sum \nabla_{\mathbf{p}} C_i / m$, where i stands for a pixel on the dispersion i and m for the

number of pixels. To test the robustness of our algorithm, reconstructions are performed on both noise-free and noisy data, where noisy data are generated by adding Gaussian noises ($\mu = 0, \sigma = 0.2$) to simulate noise-free ones. Statistically, deviations δ between ground truths and obtained optimal parameters are shown in Fig. S6. We see that the δ for every geometric parameter tends to gather around zero, giving nanometric sensitivity. Results on noise-free and noisy data are similar, validating our algorithm’s robustness. Meanwhile, the 20-s time costs per sample reconstruction makes it commercially available to in-line measurements. Here, deviations in noise-free data are mainly caused by the tiny differences between generated and simulated data. Reconstruction results of other noises are shown in Fig. S7.

We next turn to the experimental data acquired by ARS. A comparison can be immediately made just after the solving process. One of the reconstruction results is shown in Fig. 4a: Slices of the measured dispersions considered as optimization targets are marked as black lines, in comparison with the corresponding generated dispersions plotted as colored square markers. To verify the generated results, we further use RCWA to simulate the dispersions based on the parameters from AFM measurements, given by yellow ring markers. A good agreement among these three lines tells an excellent performance of our algorithm on the measured data. The complete dispersions and more comparisons of slices are shown in Fig. S9.



Correlations between our forward-mapping NN predicted and AFM measured geometric parameters are shown in Fig. 4b. The reconstruction results of three geometry parameters (the pitch, the top and bottle line widths) for seven SOI gratings are plotted versus the corresponding AFM measurement data, while those of the height are separately shown due to the nearly same height for sample fabrications. NN predictions of the first three parameters achieve a strong linear correlation ($R^2 > 0.982$) with the AFM measurements, and those of the height are also well reconstructed with mean deviation < 2.46 nm. It should be noted that the AFM measurements are an average within the local regions, since the volume of the probe in AFM is not negligible.

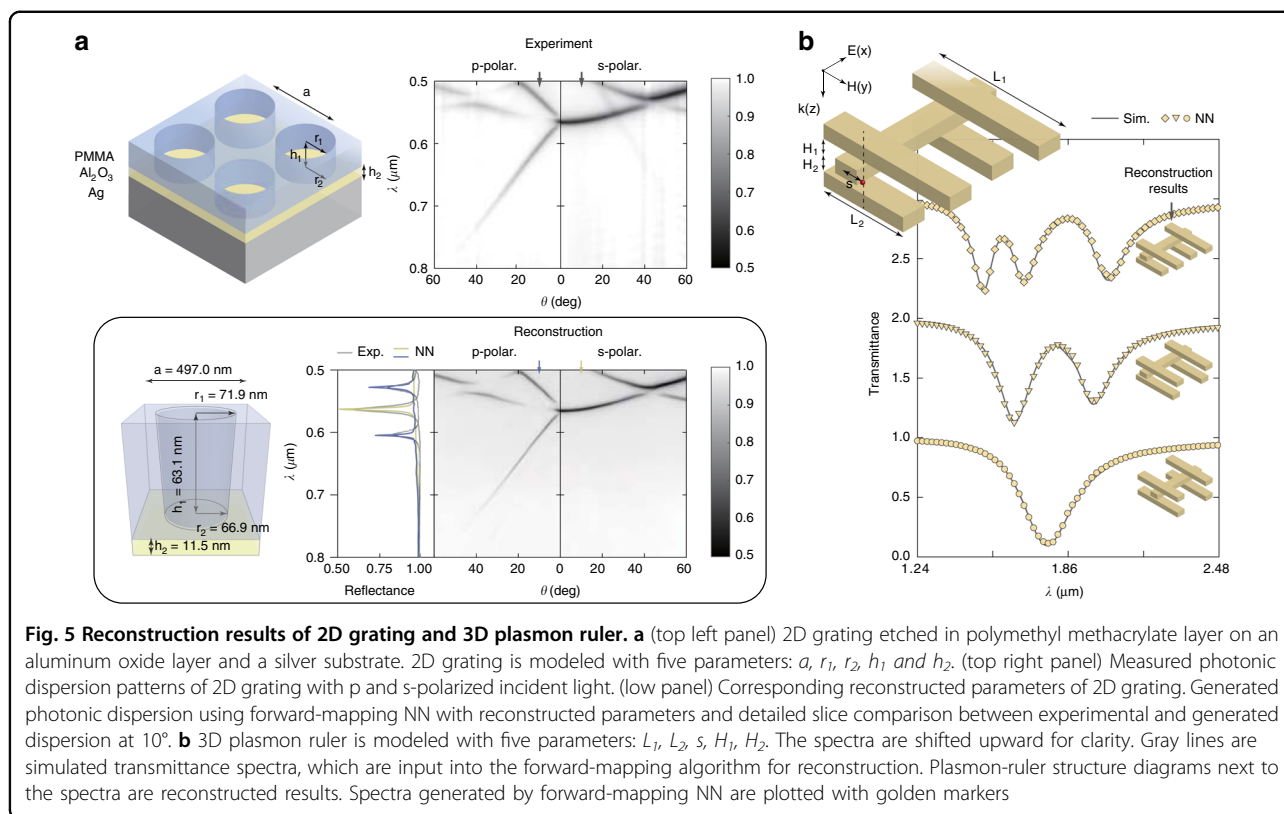
As for the inverse-mapping algorithm, despite its high solving speed (< 1 s) and excellent accuracy (98.7% predicted parameters with the deviation < 1 nm) on simulated data set, the performance on the experimental data is indeed inferior to that of the forward-mapping algorithm (see the Supplementary Information). Given it is impossible to make a data set involving all possible kinds of noises, inverse-mapping NN will always face some unexpected noise. Therefore, experiments are most likely to be mapped to some deviated parameters near the optimal solution, which could not be further finely tuned due to a lack of optimization process. In a word, the inverse-mapping algorithm has its limited advantages. However, if a parameter space is extremely large, forward mapping and inverse mapping may complement mutually as follows: Inverse-mapping NN is able to map the experimental responses to a point in the parameter space near the optimal solution at a high speed, which can be utilized as the

initial parameters for the forward-mapping NN to further finely tune to find the optimal solution. At this time, the inverse-mapping NN helps to give a rational starting point instead of a random one, and the forward-mapping NN ensures a robust optimization process in turn.

Discussion

In this section, we are going to discuss four issues. Firstly, having no contribution to the far field, rapidly attenuated near-field signals are a typical barrier in reconstruction of sub-wavelength structures. Although directly solving for the surface topography without near-field information is hard, alternatively, we can determine grating structures by searching for a solution with a minimum cost function value in parameter space with prior knowledge of the suitable model. As to the parameter space, the more the information obtained in one measurement, the more distinct the difference that appears in the parameter space. A distinct difference in parameter space is helpful for the algorithm to find the minimum cost function, since the large gradient guides the algorithm to reach the unique optimal solution effectively. Due to a large NA objective lens, we can acquire the spectrum information from multiple-angles in one measurement. It is shown that with the wider angle range considered, the topography of the parameter space goes from flat to steep (see the Fig. S11). Besides, with the prior knowledge of the model, potential non-unique solutions are excluded which makes the parameter space have a unique optimal point.

Secondly, we discussed the influences of the noise. Identical noises were generated in the Fano region and the



non-Fano region to view the difference in reconstruction results (Fig. S8). It is interesting to compare the perturbation caused by Gaussian blur and bias Gaussian blur. Large parameter deviations only occur when the Fano region is convoluted with a bias Gaussian kernel. For a Gaussian kernel, the convolution only smoothed the peak but did not change the peak position. It shows that Fano-shape dispersion has robustness to the perturbation on the amplitude of the peak. For a bias Gaussian kernel, it led to a shift in peak position, which led to a large variation in the cost function. Note that the location of these peaks is determined by our well-calibrated spectrometer. Thus, the peak position shift should be viewed as a measurement error instead of noises. From another perspective, it shows a high sensitivity of Fano-shape dispersion since only a little parameter deviation will cause a large peak shift.

Thirdly, parameter separation can be visualized by mapping the parameter space. The topography of a pitch-height plane is bowl-like which is easy for the gradient descent algorithm to find the minimum value. The topography of the w_1 – w_2 space is canyon-like, whose separation is not good as the pitch-height plane. A gradient descent with momentum and a search algorithm were introduced in the algorithm to improve the convergence behavior which yielded the same result every time with no more than 0.1 nm deviation from different initial parameters.

In measurement, changing the azimuth and increasing the acceptance angle may be two feasible schemes to directly change the topography of the w_1 – w_2 space (Fig. S12).

Finally, due to the pixel-by-pixel generation strategy of parameters-to-point architecture, the proposed NN can be flexibly migrated to other models for solving multiple ISPs. To demonstrate its migration ability, we trained forward-mapping NNs on data sets of 2D gratings⁴¹ and 3D plasmon-ruler structures⁴² for reconstruction, as shown in Fig. 5. In Fig. 5a, we fabricated a 2D grating on the polymethyl methacrylate layer and measured its photonic dispersion patterns with p- and s-polarized incident light. Reconstruction results obtained by using the forward-mapping algorithm are shown in the lower panel of Fig. 5a. In Fig. 5b, we demonstrate the reconstruction of 3D plasmon-ruler structures from simulated transmittance spectra using the forward-mapping algorithm. (Details of 2D grating and 3D plasmon-ruler structure reconstruction are shown in Fig. S14 and Fig. S15.) Since the plasmon-ruler structure has wide potential applications in monitoring macromolecular transformations, the combination of 3D plasmon rulers and the proposed algorithm will pave the road to usage of plasmon rulers in biological and soft-matter systems⁴².

In brief, we have developed a new feasible strategy containing an NN-based algorithm and high-throughput ARS for solving ISPs. It reconstructs SOI gratings with

nanometric sensitivity and seconds-level time consumption, covering a wide-range parameter space. Through adopting a parameters-to-point architecture, forward-mapping NN is able to generate photonic dispersions containing sharp Fano-shaped features with high precision, strong robustness, and small volume, which ensures the injectivity and stability for grating reconstructions. It also shows high efficiency when combining the hybrid optimization algorithms, making it available for the industrial in-line data process. The proposed algorithm can also be flexibly migrated to solve ISPs with other models. Furthermore, the ability to acquire the experimental dispersions in a single shot by the Fourier-based ARS is another unique technique for increasing the detecting efficiency. Our strategy has made good predictions against actual noises in accord with the AFM measurements, but with nondestructive nature, which means it could provide a versatile methodology to reconstruct grating profiles as well as other ISPs.

Materials and methods

Training and simulations

The training of NN was performed using a single server with a NVidia Tesla V100 graphics card and Intel(R) Xeon (R) Gold 6230 central processing unit. It costs ~ 10 h for generating the data set of a forward-mapping NN and ~5 days for that of an inverse-mapping NN. The training process of forward-mapping NN costs ~ 8 h and that of an inverse-mapping NN costs ~ 6 h.

AFM measurement

Samples were measured with a commercial AFM (Dimension Icon, Bruker) in tapping mode. The AFM was calibrated using the standard artifacts. A high-aspect-ratio tip probe (TESPA-HAR, Bruker) was used to characterize the trench on the sample. The profile is scanned with 512 pixels/line. For each grating on the sample, the geometrical parameters were analyzed to obtain the average from five different scanned profiles with SPIP software.

Acknowledgements

We thank Prof. Xiaoping Liu, Prof. Kun Ding, and Dr. Wenzhe Liu for helpful discussions. The work was supported by the China National Key Basic Research Program (2016YFA0301103, 2016YFA0302000 and 2018YFA0306201) and the National Science Foundation of China (11774063, 11727811, 91750102 and 91963212). A.C. was supported by Shanghai Rising-Star Program (20QB1402200). L.S. was further supported by the Science and Technology Commission of Shanghai Municipality (19XD1434600, 2019SHZDZX01, and 19DZ2253000).

Author details

¹State Key Laboratory of Surface Physics, Key Laboratory of Micro- and Nano-Photonics Structures (Ministry of Education) and Department of Physics, Fudan University, Shanghai 200433, China. ²Shanghai Engineering Research Center of Optical Metrology for Nano-fabrication (SERCOM), Shanghai 200433, China. ³Institute of Precision Optical Engineering, School of Physics Science and Engineering, Tongji University, Shanghai 200092, China. ⁴National Institute of Metrology, Beijing 100029, China. ⁵Collaborative Innovation Center of Advanced Microstructures, Nanjing University, Nanjing 210093, China

Author contributions

T.L., A.C., L.S., and J.Z. conceived the basic idea for this work. T.L. and A.C. designed the experiments. T.L. developed the algorithms, performed RCWA simulations, and performed the optical experiments. G.L. and M.Z. constructed ARS. J. Wang fabricated 2D grating samples. W.L. performed AFM measurements. L.S. and J.Z. supervised the research and the development of the manuscript. T.L. and A.C. wrote the draft of the manuscript, and all authors took part in the discussion and revision and approved the final copy of the manuscript.

Data availability

The data that support the findings of this study are available from the authors on reasonable request; see author contributions for specific data sets.

Conflict of interest

A.C., G.L., and H.Y. have financial interest in Ideaoptics Instruments Co., Ltd. The remaining authors declare no competing interests.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41377-021-00600-y>.

Received: 30 November 2020 Revised: 10 July 2021 Accepted: 13 July 2021
Published online: 27 July 2021

References

- Xu, M. H. & Wang, L. V. Universal back-projection algorithm for photoacoustic computed tomography. *Phys. Rev. E* **71**, 016706 (2005).
- Goy, A. et al. High-resolution limited-angle phase tomography of dense layered objects using deep neural networks. *Proc. Natl Acad. Sci. USA* **116**, 19848–19856 (2019).
- Li, H. P. et al. Advances in the design and fabrication of high-channel-count fiber Bragg gratings. *J. Lightwave Technol.* **25**, 2739–2750 (2007).
- Novikova, T. et al. Metrology of replicated diffractive optics with Mueller polarimetry in conical diffraction. *Opt. Express* **15**, 2033–2046 (2007).
- Kim, Y. N. et al. Device based in-chip critical dimension and overlay metrology. *Opt. Express* **17**, 21336–21343 (2009).
- Liu, S. Y. et al. Mueller matrix imaging ellipsometry for nanostructure metrology. *Opt. Express* **23**, 17316–17329 (2015).
- Qin, J. et al. Deep subwavelength nanometric image reconstruction using Fourier domain optical normalization. *Light. Sci. Appl.* **5**, e16038 (2016).
- Bal, G. *Introduction to Inverse Problems* (Columbia University, 2012).
- Froemming, N. S. & Henkelman, G. Optimizing core-shell nanoparticle catalysts with a genetic algorithm. *J. Chem. Phys.* **131**, 234103 (2009).
- Paz, V. F. et al. Solving the inverse grating problem by white light interference Fourier scatterometry. *Light. Sci. Appl.* **1**, e36 (2012).
- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
- Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* (MIT Press, 2016).
- Li, Y. Z., Xue, Y. J. & Tian, L. Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media. *Optica* **5**, 1181–1190 (2018).
- Li, S. et al. Imaging through glass diffusers using densely connected convolutional networks. *Optica* **5**, 803–813 (2018).
- Liu, Z. W. et al. Superhigh-resolution recognition of optical vortex modes assisted by a deep-learning method. *Phys. Rev. Lett.* **123**, 183902 (2019).
- Wei, B. et al. Machine prediction of topological transitions in photonic crystals. *Phys. Rev. Appl.* **14**, 044032 (2020).
- Huang, L., Xu, L. & Miroshnichenko, A. E. *Deep Learning Enabled Nanophotonics. Advances and Applications in Deep Learning* (IntechOpen, 2020).
- Piggott, A. Y. et al. Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer. *Nat. Photonics* **9**, 374–377 (2015).
- Peurifoy, J. et al. Nanophotonic particle simulation and inverse design using artificial neural networks. *Sci. Adv.* **4**, eaar4206 (2018).
- Liu, D. J. et al. Training deep neural networks for the inverse design of nanophotonic structures. *ACS Photonics* **5**, 1365–1369 (2018).
- Malkiel, I. et al. Plasmonic nanostructure design and characterization via Deep Learning. *Light. Sci. Appl.* **7**, 60 (2018).
- Liu, Z. C. et al. Generative model for the inverse design of metasurfaces. *Nano Lett.* **18**, 6570–6576 (2018).

23. Tao, Z. L. et al. Optical circular dichroism engineering in chiral metamaterials utilizing a deep learning network. *Opt. Lett.* **45**, 1403–1406 (2020).
24. Molesky, S. et al. Inverse design in nanophotonics. *Nat. Photonics* **12**, 659–670 (2018).
25. Ma, W. et al. Deep learning for the design of photonic structures. *Nat. Photonics* **15**, 77–90 (2021).
26. Hughes, T. W. et al. Wave physics as an analog recurrent neural network. *Science Advances* **5**, eaay6946 (2019).
27. Jiang, J. Q. & Fan, J. A. Global optimization of dielectric metasurfaces using a physics-driven neural network. *Nano Lett.* **19**, 5366–5372 (2019).
28. Novikova, T. et al. Application of Mueller polarimetry in conical diffraction for critical dimension measurements in microelectronics. *Appl. Opt.* **45**, 3688–3697 (2006).
29. Sakoda, K. *Optical Properties of Photonic Crystals* 2nd edn. (Springer-Verlag, 2005).
30. Hsu, C. W. et al. Bound states in the continuum. *Nat. Rev. Mater.* **1**, 16048 (2016).
31. Zhang, Y. W. et al. Observation of polarization vortices in momentum space. *Phys. Rev. Lett.* **120**, 186103 (2018).
32. Zhang, Y. W. et al. Momentum-space imaging spectroscopy for the study of nanophotonic materials. *Sci. Bull.* **66**, 824–838 (2021).
33. Christensen, T. et al. Predictive and generative machine learning models for photonic crystals. *Nanophotonics* **9**, 4183–4192 (2020).
34. Fan, S. H. & Joannopoulos, J. D. Analysis of guided resonances in photonic crystal slabs. *Phys. Rev. B* **65**, 235112 (2002).
35. Miroshnichenko, A. E., Flach, S. & Kivshar, Y. S. Fano resonances in nanoscale structures. *Rev. Mod. Phys.* **82**, 2257 (2010).
36. Xu, L. et al. Enhanced light–matter interactions in dielectric nanostructures via machine-learning approach. *Adv. Photonics* **2**, 026003 (2020).
37. He, K. M. et al. in *Proc. of 2016 IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, 2016).
38. Ioffe, S. & Szegedy, C. Batch normalization: accelerating deep network training by reducing internal covariate shift. Preprint at <https://arxiv.org/abs/1502.03167> (2015).
39. Liu, V. & Fan, S. H. S^4 : a free electromagnetic solver for layered periodic structures. *Computer Phys. Commun.* **183**, 2233–2244 (2012).
40. Kingma, D. & Ba, J. Adam: a method for stochastic optimization. Preprint at <https://arxiv.org/abs/1412.6980> (2014).
41. Wang, J. J. et al. Routing valley exciton emission of a WS_2 monolayer via delocalized Bloch modes of in-plane inversion-symmetry-broken photonic crystal slabs. *Light. Sci. Appl.* **9**, 148 (2020).
42. Liu, N. et al. Three-dimensional plasmon rulers. *Science* **332**, 1407–1410 (2011).