



Open Access

Applications of Neural Networks for Spectrum Prediction and Inverse Design in the Terahertz Band

IEEE Photonics Journal

An IEEE Photonics Society Publication

Volume 12, Number 5, October 2020

Jianfeng Li Yingzhan Li Yi Cen Chao Zhang Tao Luo Daquan Yang





DOI: 10.1109/JPHOT.2020.3022053





Applications of Neural Networks for Spectrum Prediction and Inverse Design in the Terahertz Band

Jianfeng Li,¹ Yingzhan Li¹,¹ Yi Cen¹,² Chao Zhang¹,¹ Tao Luo¹,¹ and Daquan Yang¹

¹School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
²School of Information Engineering, Minzu University of China, Beijing 100081, China

DOI:10.1109/JPHOT.2020.3022053 This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

Manuscript received July 31, 2020; revised August 26, 2020; accepted August 31, 2020. Date of publication September 7, 2020; date of current version September 16, 2020. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 11974058; in part by the Fundamental Research Funds for the Central Universities under Grant 2018XKJC05; in part by Fund of the State Key Laboratory of Information Photonics and Optical Communications under Grant IPOC2019ZT03, BUPT, China; in part by the National Key Research and Development Program of China under Grants 2016YFF0201003 and 2019YFC1709202; and in part by the Research Fund for Young Teachers of Minzu University of China under Grant 2020QNPY110. Corresponding authors: Yi Cen; Daquan Yang (e-mail: yi_cen@126.com; ydq@bupt.edu.cn).

Abstract: Terahertz wave has attracted significant attention in recent years, and terahertz devices have been applied in various fields. However, the complicated and time-consuming spectrum prediction and structure design issues have hindered the widespread application of terahertz science. In this work, we propose a new method to use neural networks to predict the reflection spectrum in the terahertz band, and more importantly, design a micro-nano structure with an on-demand optical response. To verify the effectiveness, we select a terahertz metasurface as an example for discussion. After the neural networks are trained, the spectrum prediction can achieve high precision, and the neural network also has encouraging performance when solving the design problem of micro-nano structure. Furthermore, we conclude that we can choose structure design neural networks to improve accuracy. Our work demonstrates that such a data-driven neural network can be applied to study the prediction and design problem of metasurface in the terahertz band and provide more opportunities for the terahertz devices in the future.

Index Terms: Advanced optics design, metamaterials, technologies for computing, THZ optics.

1. Introduction

Terahertz (THz) wave refers to the electromagnetic wave with the frequency range of 0.1–10 THz. As a unique electromagnetic spectrum, THz has attracted significant attention due to the combination of the advantages of microwave and infrared waves [1], [2]. Over the past decades, researches on the THz band have proliferated, and THz devices have been applied in various fields [3], such as security, medical inspection, radio astronomy, high-speed wireless communications, remote sensing, biomedical imaging, and high-resolution spectroscopy [4]–[10]. In order to expand the application of THz devices further, we should conduct more researches to handle various



Fig. 1. (a) Schematic diagram of the metasurface. (b) Structure diagram and parameters of the metasurface unit cell.

challenges. One challenge is how to effectively predict the spectrum for a given micro-nano structure, while another rougher challenge is how to design a micro-nano structure based on the on-demand spectrum. The former is referred to as forward prediction, and the latter is known as inverse design, which has gained wide attention and great achievements [11]–[13] recently. For the forward prediction, a complex theoretical model should be constructed first, and then the calculation involving Maxwell equations relies heavily on iterative and numerical simulations, which is time-consuming and affects the accuracy of calculation [14]. Typical methods for the inverse design are genetic algorithm and adjoint method [15], [16], both of which require vast trial and error with the increasing complexity of micro-nano structure. These issues hinder the widespread application of THz science. Therefore, more convenient approaches should be discussed to eliminate these obstacles.

Recently, deep learning, as a data-driven method, has become a promising solution to these optical problems [14]-[25]. It is well-known that deep learning can uncover the unknown relations among a massive number of variables by representing and generalizing complex functions or data [14], and has shown high power in physics [26]–[28], medical science [17], speech recognition [29]. decision making [30], chemistry [31], microscopy [32], material science[33] and other fields. As the most widely used component in the deep learning architecture [14], neural network (NN) is one of the most suitable choices to solve the prediction and design problem between the micro-nano structure and the corresponding spectrum. For example, C. C. Nadell et al. demonstrated a model of complex all-dielectric metasurface systems with deep neural networks [18]. J. Peurifoy et al. used NNs to solve nanophotonic inverse design problems [16]. G. Carleo et al. applied the NNs to deal with the quantum many-body problem [19]. Meanwhile, a well-trained NN can solve spectrum prediction and inverse design problem in seconds, which is unprecedented using conventional methods. Therefore, NNs can be greatly applied to the performance evaluation of THz metasurfaces. As a result, it can be speculated that this method can promote the development of THz science. Here, we propose a novel method to use NNs to achieve THz spectrum prediction and inverse design and explore its performance.

2. Data Simulation and NNs

To verify the effectiveness of NNs, we select a metasurface shown in Fig. 1(a) to consider its geometry structure and reflection spectrum. The metasurface marked as blue areas in Fig. 1(a) consists of zirconia microspheres with dielectric constant of 40 and loss tangent of 0.15 and a square polyimide substrate with dielectric constant of 3.6 and loss tangent of 0.01 [7]. In addition, the green areas in Fig. 1(a) are light sources. CST MWS is employed to get simulation data. In the simulation, the periodic boundary condition is employed, and the unit cell of the metasurface is shown in Fig. 1(b). The metasurface can be well-determined by three parameters, the period of the metasurface, the height of the substrate, and the radius of the sphere, called p, h, and r, respectively. Noted that the parameter p and r must match (2.1):

$$2r \le p \tag{2.1}$$



Fig. 2. Schematic of the forward network and inverse network.

Explicitly, we select period *p* varies from 150 μ m to 225 μ m with a step of 5 μ m. Similarly, height *h* varies from 10 μ m to 80 μ m with a step of 5 μ m, and radius *r* varies from 30 μ m to 109 μ m with a step of 1 μ m. Then 15 000 sets of structure parameters are obtained and fed to the CST MWS to generate 15 000 spectra sampled at points between 0.4 THz and 0.9 THz every 0.002 THz. Then we randomly select 3000 pairs as the validation set and the others as the training set of NNs. It is worth mentioning that generating the data set requires a long time, but it is deserving. Because once trained, the NNs can complete the prediction in a few seconds, which is much faster than conventional calculations.

As shown in Fig. 2, for the forward prediction NN (called forward network), the input should consist of 3 points describing the structure of metasurface, and the corresponding output should be composed of 251 points representing the reflection spectrum. When the input and output are reversed, it is the inverse design NN (called inverse network) we proposed above.

The parameter weights of NNs are initialized to truncate normal distribution while biases are initialized to zero generally. We select Rectified Linear Unit (RELU) as the activation function and Adam of great adaptation as the optimizer of NNs. The network can avoid overfitting by regularization methods. The loss function that measures the distance between actual and predicted values and drives the optimization of NNs is the Root Mean Square Error (RMSE) function described as (2.2), and the Mean Absolute Percentage Error (MAPE) function described as (2.3) is used to evaluate the accuracy of NNs. The smaller RMSE and MAPE, the more accurate NNs.

$$RMSE = \sqrt{\frac{\sum_{n} (y_{actual} - y_{predicted})^2}{n}}$$
(2.2)

$$MAPE = \frac{100\%}{n} \sum_{n} \left| \frac{y_{actual} - y_{predicted}}{y_{actual}} \right|$$
(2.3)

For the data set, the structure parameters range from 10 μ m to 225 μ m, and the spectrum value is in the range of 0 to 1. In order to get better performance of NNs, we normalize the structure parameters to a distribution with mean 0 and variance 1 so that they are on the same scale as the spectrum value. Besides, we find that if we normalize the spectrum value in the same way, which means that the spectrum value is normalized to a distribution with mean 0 and variance 1, the NNs will obtain higher accuracy. To illustrate this conclusion, we train two forward networks with three middle layers and 500 neurons per layer but fed with different training sets. One NN is fed with normalized structure parameters and normalized spectra. As shown in Fig. 3(a), when other hyper-parameters are set, we compare the MAPE value of the validation set and find the NN fed all normalized training data (red curve) has lower MAPE, which means higher accuracy. Meanwhile, remember to anti-normalize after predicting.



Fig. 3. Comparison of MAPE value of forward NNs on whether to normalize the spectrum data set.



Fig. 4. (a) Structure diagram of the forward network. (b)–(e) Comparison of the simulated spectra and the predicted spectra from the trained forward network.

3. Results and Analysis

3.1 Forward Prediction

For the forward prediction, we construct a fully connected network with three middle layers and 500 neurons per layer, as shown in Fig. 4(a). Based on the analysis above, we train the forward NN and find the NN has an excellent effect. Fig. 4(b)–(e) show the examples of the predicted reflection spectra from the trained forward network. It should be noticed that the data in this paper are selected from the validation set employed to illustrate the effect of trained NNs. The simulated spectra (blue curve) and the predicted spectra (red curve) agree with each other well, and as

a result, the blue curve is almost covered by the red curve in Fig. 4(b)–(e). It suggests that the NN has mastered the complex relationship of metasurface and reflect spectrum, rather than just saving the corresponding matchup of training data. Specifically, the RMSE and MAPE values on the training data are 0.0095 and 1.6647%, respectively, while the values on the validation set are 0.0103 and 1.7222%.

Through the analysis above, it has concluded that the forward network has high precision in predicting reflection spectrum. In addition to high precision, it only takes 7.592 seconds to predict 15 000 spectra for the trained forward NN with a 2.1-GHz AMD Ryzen 5 processor. In other words, the NN can calculate spectra at a rate of 1948 spectra every second. As for the conventional methods, it still needs 98 seconds to generate one spectrum, even using the simulation software to get a faster result. By comparing, the speed of the trained forward NN is 190 904 times faster than the conventional method. In summary, the forward NN shows a great efficiency when predicting the reflection spectrum.

3.2 Inverse Design

After evaluating the effect of forward NN, to research the performance of NN on the inverse design, we build a general fully connected NN (called GN for convenience) with three layers and 500 neurons per layer. Moreover, as shown in Fig. 5(a), a trained forward network will be connected behind to generate a second-predicted reflection spectrum to explore the accuracy of the inverse network further. Considering the inverse design NN will only generate three structure parameters, we doubt whether a one-layer NN with fewer neurons can solve the inverse design problem. Therefore, we build a relatively simplified network (called SN) with only one middle layer composed of 100 neurons, connected to a trained forward NN, as shown in Fig. 5(b) to verify the assumption.

In order to evaluate the performance of the two inverse networks, the actual structure parameters (green columnar) and predicted structure parameters by the GN (blue columnar) and SN (red columnar) are shown in Fig. 5(c)-(d). It can be figured out the predicted structure parameters from the trained GN and SN are all pretty close to the actual value, while the GN has a smaller error. Through calculating, the MAPE value of GN on the training set is 1.3190%, and on the validation set is 1.7905%, while the value of SN is 5.2645%, 5.7643%, respectively. Furthermore, we put the predicted structure parameters in Fig. 5(c)-(d) into the trained forward network and then obtain the second-predicted spectra. As shown in Fig. 5(e)-(f), the slight difference between the second-predicted spectra (blue curve and red curve) and simulated spectra (green curve) proves the excellent performance of both forward network and the two inverse networks. The shaded area at the bottom of Fig. 5(e)-(f) shows the absolute value of the difference in the second-predicted and simulated spectra shown on the right vertical axis defined as (3.1):

$$\left| \boldsymbol{R}_{\text{simulated}} - \boldsymbol{R}_{\text{sec ond}-\text{predicted}} \right|$$
(3.1)

As can be seen, the structure parameters from GN will generate a more accurate spectrum.

Next, we focus on the size of the two networks. The number of parameters in GN is 628 503 calculated by $(251 \times 500 + 500) + (500 \times 500 + 500) \times 2 + (500 \times 3 + 3)$, while in SN is 25 503 calculated by $(251 \times 100 + 100) + (100 \times 3 + 3)$, which is 24 times smaller than that of GN. Intuitively, it takes 633 seconds for the GN to complete 10 000 epochs, but just 124 seconds for the SN. In other words, 80% of the time will be saved if we choose SN when training. Furthermore, it only takes 5.211 seconds for the trained SN to generate 15 000 sets of structure parameters, avoiding a lot of trial and error.

We have concluded that the SN has lower accuracy but a simpler structure and quicker training, while GN shows a higher accuracy, complicated structure and slower training speed. Considering 10 μ m error in the sample preparation is acceptable, the SN is also of great significance. Taken together, due to the high flexibility of NNs, we can choose GN to pursue high precision or choose SN for quicker training and less computing resources.



Fig. 5. (a) Structure diagram of the GN connected with a trained forward network. (b) Structure diagram of the SN connected with a trained forward network. (c)–(d) Comparison of the actual structure parameters and the predicted structure parameters from the trained inverse networks. (e)–(f) Comparison of the simulated spectra and the second-predicted spectra. The second-predicted spectra marked as \mathbf{R} * in (a)–(b) are generated from the trained forward network fed with the predicted structure parameters in (c)–(d).

3.3 Optimization for Inverse Network

We have concluded that the one-layer NN with fewer neurons has lower accuracy. A possible explanation for this might be that due to its shallow depth and fewer neurons, the one-layer NN cannot grasp the majority of corresponding vital relationships, and thus has a slightly inferior effect when representing and generalizing complex functions. Based on the analysis, we doubt whether the one-layer NN can have better performance by adding critical points to the NN manually. In order to explore the critical information of reflection spectrum, we select ten sets of structure parameters with the same p and h but different r to generate ten spectra shown in Fig. 6(a). It shows there are continuous changes and similar trends between those spectra, indicating the importance of spectrum inflection points. Then we query whether adding the inflection points to the network can improve the accuracy of inverse networks, and the conclusion is positive.



Fig. 6. (a) Ten reflection spectra generated by the same structure parameter p and h but different r. (b) Schematic diagram of marking inflection points.



Fig. 7. (a) Structure diagram of the one-layer inverse network fed with integrated spectra. The value of subscript character n depends on the vector I. (b) Comparison of MAPE value of inverse network on whether being fed with integrated spectra.

To put this hypothesis to the test, we firstly mark the inflection points with three methods as vector I^A , I^B , and I^C , respectively. As demonstrated in Fig. 6(b), method one is to divide one spectrum into 26 segments, and each of the first 25 segments contains ten points, while the final segment contains only one point as there are 251 points in one spectrum. The segment will be noted as 1 if there exists an inflection point, otherwise noted as 0, and then vector I^A will be obtained. Method two is similar to method one, and the only difference is that the spectrum will be divided into 51 segments with the first 50 segments containing five points and the final segment containing only one point. Then similar to I^A , the 51 segments in I^B will be noted as 1 or 0. Method three is to note the inflection points directly. For example, the spectrum in Fig. 6(b) has five inflection points, which are the 47th, 101st, 146th, 176th, and 217th points, respectively, described as a vector I^C firstly. Considering there are 13 inflection points in a spectrum at most, the vector I^A , I^B , and I^C are combined with the original spectrum into an integrated spectrum with 264, 277, or 302 points.

Next, the integrated spectra will be fed to a new SN with just one middle layer consisting of only 100 neurons to generate structure parameters, as shown in Fig. 7(a). Note that the subscript character n in Fig. 7(a) can be 26, 51, or 13, depending on whether the vector I is I^A , I^B , or I^C . To get a more persuasive result, we keep the hyper-parameters constant and then feed the network with three different integrated spectra and the original spectrum having no infection to compare the MAPE value of the validation set at different learning rates. Note that all the four networks have a pretty good performance with the selected constant hyper-parameters. As shown in Fig. 7(b),

it indicates that all the three integrated spectra have a lower MAPE in most cases, leading to a decrease of 0.6% generally. Furthermore, we put the three integrated spectra into a three-layer NN with 500 neurons per layer and find a decrease of 0.2% on MAPE value. In conclusion, adding the inflection points to the network manually indeed improves the accuracy of inverse networks and performs more effectively on the fewer-neurons network.

4. Conclusion

In this paper, we propose a novel method to use NNs to solve the forward prediction and inverse design in the THz metasurface. The forward network for the spectrum prediction has a high precision with an error of 1.72% and can predict 1948 spectra per second. Meanwhile, we can choose inverse networks with different complexity to satisfy different demands with acceptable error due to the flexibility of NNs. These findings suggest that the NN is an effective tool when solving the prediction and design problem of metasurface in the THz band, and are expected to expand the application of THz devices in the future.

References

- H. Cai *et al.*, "Multifunctional hybrid metasurfaces for dynamic tuning of terahertz waves," *Adv. Opt. Mater.*, vol. 6, no. 14, Jul. 2018, Art. no. 1800257.
- [2] E. Yang, C. Zhang, X. Li, and C. Lan, "InSb-enhanced thermally tunable terahertz silicon metasurfaces," IEEE Access, vol. 7, pp. 95087–95093, Aug. 2019.
- [3] B. Li, Y. Zeng, B. Chen, and C. H. Chan, "Terahertz frequency-selective surface with polarization selection and conversion characteristics," *IEEE T. THz. Sci. Techn.*, vol. 9, no. 5, pp. 510–519, Sep. 2019.
- [4] D. A. Bandurin, "Resonant terahertz detection using graphene plasmons," Nat. Commun., vol. 9, no. 1, Dec. 2018, Art. no. 5392.
- [5] A. J. L. Adam, "Review of near-field terahertz measurement methods and their applications," J. Infrared Millim. Teh., vol. 32, no. 8, pp. 976–1019, Jul. 2011.
- [6] D. Yang, C. Zhang, K. Bi, and C. Lan, "High-throughput and low-cost terahertz all-dielectric resonators made of polymer/ceramic composite particles," *IEEE Photon. J.*, vol. 11, no. 1, Feb. 2019, Art no. 5900408.
- [7] D. Yang, C. Zhang, X. Ju, Y. Ji, and C. Lan, "Multi-resonance and ultra-wideband terahertz metasurface absorber based on micro-template-assisted self-assembly method," *Opt. Express*, vol. 28, no. 2, pp. 2547–2556, Jan. 2020.
- [8] S. Gui, J. Li, and Y. Pi, "Security imaging for multi-target screening based on adaptive scene segmentation with terahertz radar," *IEEE Sens. J.*, vol. 19, no. 7, pp. 2675–2684, Apr. 2019.
- [9] X. Yang, Y. Pi, T. Liu, and H. Wang, "Three-dimensional imaging of space debris with space-based terahertz radar," IEEE Sens. J., vol. 18, no. 3, pp. 1063–1072, Feb. 1, 2018.
- [10] Q. Xia and J. M. Jornet, "Expedited neighbor discovery in directional terahertz communication networks enhanced by antenna side-lobe information," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7804–7814, Aug. 2019.
- [11] S. Molesky, Z. Lin, A. Y. Piggott, W. Jin, J. Vucković, and A. W. Rodriguez, "Inverse design in nanophotonics," Nat. Photon., vol. 12, no. 11, pp. 659–670, Nov. 2018.
- [12] Z. Lin, V. Liu, R. Pestourie, and S. G. Johnson, "Topology optimization of freeform large-area metasurfaces," Opt. Express, vol. 27, no. 11, pp. 15765–15775, May 2019.
- [13] R. Pestourie, C. Pérez-Árancibia, Z. Lin, W. Shin, F. Capasso, and S. G. Johnson, "Inverse design of large-area metasurfaces," Opt. Express, vol. 26, no. 26, pp. 33732–33747, Dec. 2018.
- [14] M. Wei, F. Cheng, and Y. Liu, "Deep-learning-enabled on-demand design of chiral metamaterials," ACS Nano, vol. 12, no. 6, pp. 6326–6334, Jun. 2018.
- [15] C. Liu, Y. Tan, E. Am, and Z. Yu, "Training deep neural networks for the inverse design of nanophotonic structures," ACS Photon., vol. 5, no. 4, pp. 1365–1369, Feb. 2018.
- [16] J. Peurifoy et al., "Nanophotonic particle simulation and inverse design using artificial neural networks," Sci. Adv., vol. 4, no. 6, Jun. 2018, Art. no. eaar4206.
- [17] Y. Long, J. Ren, Y. Li, and H. Chen, "Inverse design of photonic topological state via machine learning," Appl. Phys. Lett., vol. 114, no. 18, May 2019, Art. no. 181105.
- [18] C. C. Nadell, B. Huang, J. M. Malof, and W. J. Padilla, "Deep learning for accelerated all-dielectric metasurface design," Opt. Express, vol. 27, no. 20, pp. 27523–27535, Sep. 2019.
- [19] G. Carleo and M. Troyer, "Solving the quantum many-body problem with artificial neural networks," Science, vol. 355, no. 6325, pp. 602–606, Feb. 2017.
- [20] T. Zhang et al., "Efficient spectrum prediction and inverse design for plasmonic waveguide systems based on artificial neural networks," Photon. Res., vol. 7, no. 3, pp. 368–380, Mar. 2019.
- [21] M. Meem et al., "Large-area, high-numerical-aperture multi-level diffractive lens via inverse design," Optica, vol. 7, no. 3, pp. 252–253, Mar. 2020.
- [22] R. Hussain et al., "An ultra-compact particle size analyser using a CMOS image sensor and machine learning," Light-Sci. Appl., vol. 9, no. 1, Feb. 2020, Art. no. 21.

- [23] J. Hao, L. Zheng, D. Yang, and Y. Guo, "Inverse design of photonic crystal nanobeam cavity structure via deep neural network," presented at the Asia Commun. Photon. Conf., OSA Tech. Dig. Ser. Opt. Soc. of America, Chengdu, China, Nov. 2-5, 2019, Paper M4A. 296.
- [24] I. Malkiel, M. Mrejen, A. Nagler, U. Arieli, L. Wolf, and H. Suchowski, "Deep learning for the design of nano-photonic structures," presented at the IEEE Int. Conf. Comput. Photography., 2018, pp. 1–14.
- [25] L. Xu *et al.*, "Enhanced light-matter interactions in dielectric nanostructures via machine learning approach," *Adv. Photon.*, vol. 2, no. 2, Apr. 2020, Art. no. 026003.
- [26] P. B. Wigley et al., "Fast machine-learning online optimization of ultra-cold-atom experiments," Sci. Rep., vol. 6, May 2016, Art. no. 25890.
- [27] P. Baldi, P. Sadowski, and D. Whiteson, "Searching for exotic particles in high-energy physics with deep learning," *Nat. Commun.*, vol. 5, Jul. 2014, Art. no. 4308.
- [28] J. Carrasquilla and R. G. Melko, "Machine learning phases of matter," *Nat. Phys.*, vol. 13, pp. 431–434, May 2017.
- [29] G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Proc. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.
- [30] D. Silver et al., "Mastering the game of Go with deep neural networks and tree search," Nature, vol. 529, pp. 484–489, Jan. 2016.
- [31] G. B. Goh, N. O. Hodas, and A. Vishnu, "Deep learning for computational chemistry," J. Comput. Chem., vol. 38, no. 16, pp. 1291–1307, Jun. 2017.
- [32] Y. Rivenson, Z. Gorocs, H. Gunaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," Optica, vol. 4, no. 11, pp. 1437–1443, Nov. 2017.
- [33] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, and C. Kim, "Machine learning in materials informatics: Recent applications and prospects," *NPJ Comput. Mater.*, vol. 3, no. 54, Dec. 2017, Art. no. 54.