Comparison of Different Neural Network Architectures for Plasmonic Inverse Design

Qingxin Wu, Xiaozhong Li, Wenqi Wang, Qiao Dong, Yibo Xiao, Xinyi Cao, Lianhui Wang,* and Li Gao*

ABSTRACT: The merge between nanophotonics and a deep neural network has shown unprecedented capability of efficient forward modeling and accurate inverse design if an appropriate network architecture and training method are selected. Commonly, an iterative neural network and a tandem neural network can both be used in the inverse design process, where the latter is well known for tackling the nonuniqueness problem at the expense of more complex architecture. However, we are curious to compare these two networks’ performance when they are both applicable. Here, we successfully trained both networks to inverse design the far-field spectrum of plasmonic nanoantenna, and the results provide some guidelines for choosing an appropriate, sufficiently accurate, and efficient neural network architecture.

1. INTRODUCTION

Localized surface plasmon resonance (LSPR) is a conductive electron resonance phenomenon in metal nanostructures, such as nanorods, nanospheres, nanotriangles, and nanodisks. Such plasmonic nanostructures have been widely applied in surface-enhanced Raman spectroscopy (SERS), fluorescence probes, and other chemical or biological sensors with the capability of single-molecule detection. The plasmonic resonance spectral peak, shape, and wavelength largely depend on the geometry parameters of nanostructures, material properties, and the dielectric environment. Among different plasmonic nanostructures, bow-tie nanoantennas, consisting of extremely sharp tips and sub-10 nm gaps shown in Figure 1a, are known for their significant local electric field enhancement. In addition, the plasmonic resonance spectrum is highly sensitive to the geometric changes, and such a relationship is highly nonlinear and unpredictable. As shown in Figure 1b, tiny changes cause great differences in the resonance positions and peaks of the transmission spectrum, where G = 5 nm, W = 110 nm, and L = 110 nm (black line), 120 nm (red line), and 130 nm (blue line). On the other hand, very different geometric structures can result in very similar resonance spectra, as shown in Figure 1c. Such complexity brings challenges in understanding and predicting the physical relationship between plasmonic nanostructures and their electromagnetic response.

Over the last two decades, there has been extensive research based on conventional electromagnetic simulations to investigate the relationship between the nanostructure geometry and their corresponding optical properties. It is very time consuming to examine a large parameter space and not possible to find the globally optimized structural design. To obtain desired resonance spectra, prior experience and iterative optimizations are needed. The recent development of deep learning tools, such as a deep neural network (DNN), inspired by the layered and hierarchical deep architecture of the human brain, has revolutionized the field of nanophotonic device design. DNN generally consists of a forward network and an inverse network. Its training only requires a one-time investment of adequate electromagnetic (EM) simulation data, which are made up of different geometry structures and corresponding optical properties such as response spectra. By the adjustment and optimization of hyperparameters in the training process, the error between the output value of the network and the actual value constantly reduces, and the training ends after reaching a certain accuracy. In the forward network, we input geometry structures and take the response spectra as an output. DNN can accurately capture the complicated nonlinear physical relationship between a geometry structure and EM response, thus enabling highly efficient EM response prediction. On the contrary, the inverse network takes the response spectra as input and outputs the corresponding geometry structures. In this way, the network

Received: April 23, 2021
Accepted: July 20, 2021
Published: August 30, 2021
can accomplish the inverse design of multiparametered nanophotonics in a few milliseconds that conventional methods are not capable of. So far, DNN has been successfully utilized to predict and inverse design the far-field spectra of nanodisks, nanocavity, and multilayer structures.

Commonly used inverse design DNN includes the direct inverse DNN, iterative DNN, and tandem DNN.
inverse DNN appears to be the most straightforward structure by interchanging the input and output layer of the forward network. Although the way of direct inverse DNN is ideologically feasible, it encounters great challenges in practical training such as the nonuniqueness problem, which arises when different geometry structures can produce a similar far-field spectrum; the nonunique mapping between the input and the output of DNN can result in the failure of training convergence. The nonuniqueness designing can be performed by both iterative DNN and tandem DNN. Iterative DNN is identical to the forward network with fixed trained weights and comprises gradient-descent methods based on backpropagation. The network accuracy is evaluated by the loss function and continuously optimizes the input parameters (structure) as variables to reduce the error between the output and desired response. In the actual inverse training process, different hyperparameters such as an optimizer, loss function, and the learning rate also need to be optimized. Tandem DNN was first proposed by Liu et al., which connects a pretrained forward network to an inverse network; the design accuracy is obtained when the difference between the input and the output spectra response layer is minimized. It is well known to better solve the nonuniqueness problem at the expense of more complex architecture and computation. However, the inverse design accuracy of these two networks has not been compared in specific cases when they are both applicable. Here, we test both iterative and tandem DNN's capability of learning the ultrasensitive structure—response relationship of a bow-tie nanoantenna, as shown in Figure 1b, and whether they can effectively solve the nonuniqueness problem in the inverse design process, as shown in Figure 1c. The iterative DNN and tandem DNN are shown and compared in Figure 2 to evaluate their individual effectiveness in such type of a problem. From such study, we can provide useful results and guidelines for choosing an appropriate, sufficiently accurate, and efficient DNN architecture for inverse designs.

2. RESULTS AND DISCUSSION

Convergence is defined as the trend of a mean absolute error between the true value and the predicted value with the increase in the cycle number. We adopt UP of early stopping, which is defined as the stopping of the training process when the generalization error increases in s successive strips and then sets s to 10, as a convergence criterion. The mini-batch gradient descent is also employed to update weights. The forward network stops training at 500 iterations and the final value of training loss and validation loss are 0.0258 and 0.0595, shown in Figure 3a. Overall, 100 groups of data sets, which are not involved in the training process are used as test data to reveal the accuracy and performance of NN. As shown in Figure 3b, we calculated the relative errors to evaluate the accuracy of NN and adopted the loss of the MSE method (top) and the MAE method (bottom), where the solid lines represent the mean values of 0.0364 and 0.0265. This result demonstrates the MAE method excels at outlier handling for nonlinear sensitive characteristics, as proved in our previous work that substituting the loss of MSE with MAE is effective for complex EM data such as abrupt phase change and near-field enhancement (see the Methods section). Figure 3c shows a prediction sample by the MAE method with the average error around 0.0265, which proves the effective training of DNN in capturing the complicated nonlinear physical relationship. Simultaneously, the specific distribution of the relative errors is statistically illustrated in Figure 3d.
Overall, 80 out of 100 samples (80%) have errors less than 4%, while only 6 samples have errors more than 6%.

Next, we adopt and compare two commonly used inverse networks to design a geometry according to the target spectrum. One case is the iterative NN where we directly use the trained forward network for inverse design. The other one is tandem network, which connects the pretrained forward network and the inverse network, aiming at solving the nonuniqueness problems. Four hidden layers are involved in the inverse tandem network and 250 nodes for each layer, and other settings include an RMSprop optimizer and a ReLU activation function. More details about inverse tandem NN are given in the Supporting Information (Table S2). Figure 4a clearly shows that the loss curves of the inverse tandem network remain basically stable after 1200 epochs followed using UP of early stopping to stop the training process, and the final values of training loss and validation loss are 0.0521 and 0.0682, respectively. Again, we use 100 groups of untrained data for the network accuracy test. Then, the designed geometry structures are input into the finite-difference time domain (FDTD) to simulate the transmission spectra and evaluate the similarity between the actual and designed spectrum by calculating the relative error. The results are divided into four categories according to the degree of similarity, i.e., relative error smaller than 6%, between 6 and 12%, between 12 and 15%, and larger than 15%, accounting for 68, 21, 8, and 3% of iterative DNN and 65, 26, 1, and 8% of the tandem network, shown in Figure 4b. The test data with a relative error of less than 12% account for around 90%, for both iterative NN and tandem NN. Figure 4c,d shows the error distribution of test samples for inverse tandem NN and inverse iterative DNN, respectively, where the solid lines represent the average values. Obviously, the adoption of both iterative DNN and tandem DNN has realized inverse design of most data, but certain data fall into the dead space of the design space due to forced one-to-one mapping.

Table 1. Specific Structure Parameters for the Representative Spectra of Four Examples

| example | target (nm) | tandem NN (nm) | iterative NN (nm) |
|---------|-------------|----------------|-------------------|
|         | G | W | L      | G | W | L      | G | W | L      |
| 1       | 30 | 240 | 180    | 22 | 265 | 176    | 29 | 244 | 178    |
| 2       | 10 | 220 | 150    | 14.9 | 170 | 159    | 10.7 | 215 | 151    |
| 3       | 10 | 160 | 150    | 14.1 | 129 | 159    | 11.9 | 118.3 | 154.5   |
| 4       | 7  | 150 | 150    | 6.9  | 157.5 | 150.8  | 16.3 | 136 | 158    |

Figure 5. Representative results of designed iterative DNN and the tandem network predicted spectra compared with EM simulated spectra. (a) Spectrum with a relative error of 2.88% for iterative DNN and 2.97% for the tandem network. (b) Spectrum with a relative error of 12.05% for iterative DNN and 12.01% for the tandem network. (c) Spectrum with a relative error of 16.89% for iterative DNN and 15.97% for the tandem network. (d–f) Example that explains the high-sensitivity problem and nonuniqueness problem.
main trends and features fit each other. It shows that DNN effectively captured the complex and ultrasensitive EM response features of the bow-tie nanoantenna. In addition, in the final group, the microchange, especially between the actual and designed geometries of tandem NN (ΔG = 0.1 nm, ΔW = 7.5 nm, ΔL = 0.8 nm), created a great discrepancy of transmission, again demonstrating the high-sensitivity characteristics of the bow-tie nanoantenna. Compared with the similar geometry structures designed by the iterative NN, the significant difference usually appears in tandem NN between the designed and the actual geometry structure, further illustrating its ability to solve nonunique problems. Moreover, there are some apparent outliers, as shown in Figure 4c,d. Take no. 34 (G = 40 nm, W = 110 nm, L = 90 nm) in the test data for example, its relative error is 0.33117 for iterative NN and 0.18806 for tandem NN.

Figure 5e shows the significant differences between designed and actual spectra and their specific structure parameters. By further analysis and simulations, we have found that both nonuniqueness problems and high-sensitivity problems exist in no. 34, as shown in Figure 5f. A similar spectrum appears in diverse structures, i.e., a red solid line (G = 40 nm, W = 110 nm, L = 90 nm), a black dotted line (G = 60 nm, W = 110 nm, L = 90 nm), and a blue solid line (G = 30 nm, W = 110 nm, L = 80 nm). However, by comparing Figure 5e,f, distinct spectral changes occur in the slight difference, ΔG = 0, ΔW = 3 nm, ΔL = 2 nm (between the red solid line of Figure 5e and the blue solid line of Figure 5f) and ΔG = 3 nm, ΔW = 3 nm, ΔL = 5 nm (between the blue solid line of Figure 5e and the carmine dotted line of Figure 5f).

These results show that these two different DNN architectures have similar accuracy in the inverse design results. Previously, tandem DNN has achieved great success in dealing nonunique solution problems like those in structural color designs. This is because the design parameters are only three color values and the constraints on the optimization direction are not sufficient to yield a unique solution; however, the inverse design of the far-field spectrum contains 101 discrete values and these large constraints used in the iterative inverse DNN can ensure an accurate design and tandem DNN is no longer necessary in such a problem.

3. CONCLUSIONS

In this work, we have investigated two commonly used inverse neural networks for designing the transmission spectra of a bow-tie nanoantenna. By only investing 3024 groups of data for training the networks, millions of different bow-tie nanoantennae can be designed accurately and rapidly. The whole spectra design results with a relative error less than 12% account for more than 90%, which demonstrated the performance and accuracy of the two networks. Such a tool shows great promise in SERS-based chemical and biological sensing and integrated nanophotonics devices.

The results show that although the tandem network was proposed to solve the nonuniqueness inverse design problem, but for designing far-field spectra of a bow-tie nanoantenna, it shows similar performance compared to iterative DNN. With comparable design accuracy, the training of iterative DNN is much simpler and can be adopted for similar types of problems. Therefore, different DNN architectures and training solutions should be evaluated for specific problems for high accuracy and ease of application.

4. METHODS

The generation of data sets relies on full-wave EM simulations such as the FDTD method used here, based on solving Maxwell’s equations. Its basic idea is to use the central difference quotient to replace the first-order partial derivative of the field quantity with respect to time and space, and to obtain the field distribution by simulating the propagation process of the wave through recurrence in the time domain. Previous research studies have reported that the critical geometry structure parameters influencing far-field properties of a bow-tie nanoantenna are the gap (G), width (W), and length (L), as shown in Figure 1a. Therefore, we pretest the sensitivity of spectra to these three structure parameters using the FDTD method and find a reasonable range of our training data, i.e., gap (G, 5–40 nm), width (W, 20–250 nm), and length (L, 30–200 nm), as the input data, and response spectra at wavelengths from 500 to 1000 nm, as the output data. The test details are shown in Figures S1–S3 in the Supporting Information. Other periodic boundary conditions including the period of 470 nm for x and y directions and the perfectly matched layer (PML) for the z direction for the simulations are also determined. Simultaneously, “mesh” with a maximum mesh step of 1 nm was added to increase the simulation accuracy. A total of 3024 groups of training data sets are collected and the data generation time is approximately 20 days on a workstation of CPU 2.9 GHz.

The training data were divided into three categories, i.e., 2500 groups of data for network training, 424 groups for network validation, and 100 groups of test data that were not used for training and validation. Therefore, the validation groups could help us efficiently estimate the degree of overfitting, in the process of comparing the training and validation loss, and the test results show the ability and accuracy of the DNN in ordinary situations.

Here, we first train the same forward network that is used both in the tandem DNN (Figure 2a) and iterative DNN (Figure 2b). In our previous work, substituting the loss of mean-squared error (MSE) for the mean absolute error (MAE) has been proved very effective for nonlinear and sensitive EM data such as abrupt phase change and near-field enhancement.10,31 MSE is the most common evaluation indicator of loss, widely adopted by a number of NN research studies,10,18,22,32 which can be expressed as eq (1)

$$\text{error}_{\text{MSE}} = \frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2$$

where $y_i$ is the actual value and $\hat{y}_i$ is the predicted value. For the high-sensitive nonlinear features of bow-tie nanoantenna transmission, here, there are a certain number of outliers and abnormal values. The quadratic term adopted by the MSE method would unreasonably magnify the influence of outliers and abnormal values during the judgment of loss, resulting in great loss even if convergence is possible.30 Therefore, we employ the MAE method, which uses the absolute error not the quadratic term, and thus does well in dealing with the case when certain outliers are detrimental to predicted results of all samples. This method can be described by eq (2)

$$\text{error}_{\text{MAE}} = \frac{1}{m} \sum_{i=1}^{m} |y_i - \hat{y}_i|$$
where $y_i$ is the actual value and $\hat{y}_i$ is the predicted value. The best trained neural network (NN) has five hidden layers with 300 nodes in the first layer and 450 nodes in the rest. Other settings of hyperparameters in this NN consist of a Nadam optimizer and a ReLU activation function. More details about NN can be found in the Supporting Information (Table S1).

### ASSOCIATED CONTENT

#### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.1c02165.

Optimal settings of some hyperparameters in the forward network; optimal settings of some hyperparameters in the inverse network; and one test case for the sensitivity of spectra to $G$, $W$, and $L$ structure parameters using the FDTD method (PDF)

### AUTHOR INFORMATION

**Corresponding Authors**

Lianhui Wang — State Key Laboratory for Organic Electronics and Information Displays, Institute of Advanced Materials, School of Materials Science and Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210023, China; orcid.org/0000-0001-9030-9172; Email: iamlhwang@njupt.edu.cn

Li Gao — State Key Laboratory for Organic Electronics and Information Displays, Institute of Advanced Materials, School of Materials Science and Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210023, China; orcid.org/0000-0003-2465-5784; Email: iamlgao@njupt.edu.cn

**Authors**

Qingxin Wu — State Key Laboratory for Organic Electronics and Information Displays, Institute of Advanced Materials, School of Materials Science and Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

Xiaozhong Li — School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

Wenqi Wang — State Key Laboratory for Organic Electronics and Information Displays, Institute of Advanced Materials, School of Materials Science and Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

Qiao Dong — State Key Laboratory for Organic Electronics and Information Displays, Institute of Advanced Materials, School of Materials Science and Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

Yibo Xiao — State Key Laboratory for Organic Electronics and Information Displays, Institute of Advanced Materials, School of Materials Science and Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

Xinyi Cao — State Key Laboratory for Organic Electronics and Information Displays, Institute of Advanced Materials, School of Materials Science and Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

Complete contact information is available at: https://pubs.acs.org/doi/10.1021/acsomega.1c02165

**Notes**

The authors declare no competing financial interest.

### ACKNOWLEDGMENTS

The authors acknowledge support from the National Key Research and Development Program of China (2017YFA0205302), the National Natural Science Foundation of China (61974069 and 62022043), the Jiangsu Provincial Key Research and Development Program (BE2018732), the Natural Science Foundation of Jiangsu Province BK20191379, NUPUTSF NY219008, and the NJUPT 1311 Talent Program.

### REFERENCES

(1) Haque, A.; Reza, A. W.; Kumar, N. A Novel Design of Circular Edge Bow-Tie Nano Antenna for Energy Harvesting. Frequenz 2015, 69, 491–499.

(2) Hasan, D.; Ho, C. P.; Pitchappa, P.; Lee, C. Dipolar Resonance Enhancement and Magnetic Resonance in Cross-Coupled Bow-Tie Nanoantennas: Asymmetric Arrays. ACS Photonics 2015, 2, 890–898.

(3) Wu, Y.-M.; Li, L.-W.; Liu, B. Gold Bow-Tie Shaped Aperture Nanoantenna: Wide Band Near-field Resonance and Far-Field Radiation. IEEE Trans. Magn. 2010, 46, 1918–1921.

(4) Zhan, P.; Wen, T.; Wang, Z. G.; He, Y.; Shi, J.; Wang, T.; Liu, X.; Lu, G.; Ding, B. DNA Origami Directed Assembly of Gold Bowtie Nanoantennas for Single-Molecule Surface-Enhanced Raman Scattering. Angew. Chem., Int. Ed. 2018, 57, 2846–2850.

(5) Kühn, S.; Hakanson, U.; Rogobete, L.; Sandoghdar, V. Enhancement of single-molecule fluorescence using a gold nanoparticle as an optical nanoantenna. Phys. Rev. Lett. 2006, 97, No. 017402.

(6) Xin, H.; Namgung, B.; Lee, L. P. Nanoplasmonic optical antennas for life sciences and medicine. Nat. Rev. Mater. 2018, 3, 228–243.

(7) Valsecchi, C.; Brolo, A. G. Periodic metallic nanostructures as plasmonic chemical sensors. Langmuir 2013, 29, 5638–5649.

(8) Bharadwaj, P.; Deutsch, B.; Novotny, L. Optical Antennas. Adv. Opt. Photonics 2009, 1, 438–483.

(9) Duan, H.; Fernandez-Dominguez, A. I.; Bosman, M.; Maier, S. A.; Yang, J. K. Nanoplasmonics: classical down to the nanometer scale. Nano Lett. 2012, 12, 1683–1689.

(10) Hao, E.; Schatz, G. C. Electromagnetic fields around silver nanoparticles and dimers. J. Chem. Phys. 2004, 120, 357–366.

(11) Schuck, P. J.; Fromm, D. P.; Sundaramurthy, A.; Kino, G. S.; Moerner, W. E. Improving the mismatch between light and nanoscale objects with gold bowtie nanoantennas. Phys. Rev. Lett. 2005, 94, No. 017402.

(12) Sundaramurthy, A.; Crozier, K. B.; Kino, G. S.; Fromm, D. P.; Schuck, P. J.; Moerner, W. E. Field enhancement and gap-dependent resonance in a system of two opposing tip-to-tip Au nanotriangles. Phys. Rev. B 2005, 72, No. 165409.

(13) Zhao, H.; Gao, H.; Li, B. Theory and method for large electric field intensity enhancement in the nanoantenna gap. Appl. Opt. 2019, 58, 670–676.

(14) Holger, F.; Martin, O. J. F. Engineering the optical response of plasmonic nanoantennas. Opt. Express 2008, 16, 9144–9154.

(15) de Ceglia, D.; Vincenti, M. A.; De Angelis, C.; Locatelli, A.; Haus, J. W.; Scalora, M. Role of antenna modes and field coupling and resonant cavity modes in plasmonic nanorods. Nanotechnology 2016, 27, No. 415708.
(18) Hegde, R. S. Deep learning: a new tool for photonic nanostructure design. Nanoscale Adv. 2020, 2, 1007–1023.
(19) Yao, K.; Unni, R.; Zheng, Y. Intelligent nanophotonics: merging photonics and artificial intelligence at the nanoscale. Nanophotonics 2019, 8, 339–366.
(20) He, J.; He, C.; Zheng, C.; Wang, Q.; Ye, J. Plasmonic nanoparticle simulations and inverse design using machine learning. Nanoscale 2019, 11, 17444–17459.
(21) So, S.; Badloe, T.; Noh, J.; Bravo-Abad, J.; Rho, J. Deep learning enabled inverse design in nanophotonics. Nanophotonics 2020, 9, 1041–1057.
(22) Wiecha, P. R.; Muskens, O. L. Deep Learning Meets Nanophotonics: A Generalized Accurate Predictor for Near Fields and Far Fields of Arbitrary 3D Nanostructures. Nano Lett. 2020, 20, 329–338.
(23) Li, X.; Shu, J.; Gu, W.; Gao, L. Deep neural network for plasmonic sensor modeling. Opt. Mater. Express 2019, 9, 3857–3862.
(24) Malkiel, I.; Mrejen, M.; Nagler, A.; Aneli, U.; Wolf, L.; Suchowski, H. Plasmonic nanostructure design and characterization via Deep Learning. Light: Sci. Appl. 2018, 7, No. 60.
(25) Ashalley, E.; Acheampong, K.; Besteiro, L. V.; Yu, P.; Neogi, A.; Govorov, A. O.; Wang, Z. M. Multitask deep-learning-based design of chiral plasmonic metamaterials. Photonics Res. 2020, 8, 1213–1225.
(26) Liu, D.; Tan, Y.; Khoram, E.; Lu, Z. Training Deep Neural Networks for the Inverse Design of Nanophotonic Structures. ACS Photonics 2018, 5, 1365–1369.
(27) Jiang, J.; Chen, M.; Fan, J. A. Deep neural networks for the evaluation and design of photonic devices. Nat. Rev. Mater. 2020, 1–22.
(28) Gao, L.; Li, X.; Liu, D.; Wang, L.; Yu, Z. A Bidirectional Deep Neural Network for Accurate Silicon Color Design. Adv. Mater. 2019, 31, No. 1905467.
(29) Prechelt, L. Early stopping—but when?. In Neural Networks: Tricks of the trade; Orr, G. B.; Müller, K. R., Eds.; Academic Press: Springer: Berlin, Heidelberg, 1998; pp S5–69.
(30) Wu, Q.; Li, X.; Jiang, L.; Xu, X.; Fang, D.; Zhang, J.; Song, C.; Yu, Z.; Wang, L.; Gao, L. Deep neural network for designing near- and far-field properties in plasmonic antennas. Opt. Mater. Express 2021, 11, 1907–1917.
(31) Jiang, L.; Li, X.; Wu, Q.; Wang, L.; Gao, L. Neural network enabled metasurface design for phase manipulation. Opt. Express 2021, 29, 2521–2528.
(32) Kühl, P.; Weber, M.; Lohmuller, T. Plasmonic nanoantenna arrays for surface-enhanced Raman spectroscopy of lipid molecules embedded in a bilayer membrane. ACS Appl. Mater. Interfaces 2014, 6, 8947–8952.