

Optics Letters

Inverse design of plasmonic metasurfaces by convolutional neural network

RONGHUI LIN,  YANFEN ZHAI, CHENXIN XIONG, AND XIAOHANG LI* 

King Abdullah University of Science and Technology (KAUST), Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, Thuwal 23955-6900, Saudi Arabia

*Corresponding author: xiaohang.li@kaust.edu.sa

Received 21 January 2020; accepted 3 February 2020; posted 6 February 2020 (Doc. ID 387404); published 4 March 2020

Artificial neural networks have shown effectiveness in the inverse design of nanophotonic structures; however, the numerical accuracy and algorithm efficiency are not analyzed adequately in previous reports. In this Letter, we demonstrate the convolutional neural network as an inverse design tool to achieve high numerical accuracy in plasmonic metasurfaces. A comparison of the convolutional neural networks and the fully connected neural networks show that convolutional neural networks have higher generalization capabilities. We share practical guidelines for optimizing the neural network and analyzed the hierarchy of accuracy in the multi-parameter inverse design of plasmonic metasurfaces. A high inverse design accuracy of ± 8 nm for the critical geometrical parameters is demonstrated. © 2020 Optical Society of America

<https://doi.org/10.1364/OL.387404>

The deep neural networks (DNNs) have been proven to be powerful in solving complicated problems in various fields ranging from biology [1], chemistry [2], physics [3], to geology [4]; and from theoretical calculations [5] to engineering applications [6]. In the field of nanophotonics [7], the DNNs have been employed to design plasmonics structures, metasurfaces [8], integrated photonic devices [9], and nanocavities [10]. The popularity of the DNNs stems from the fact that they can handle complex problems requiring high scientific rigor and precision. Instead of expressing and solving the physical equations in an explicit way, the DNNs can extract the physics hidden in the statistics of data and memorize them as network parameters through the learning process. In principle, the DNNs can approximate any physical phenomenon, provided that the capacity is high enough. However, the tuning of the DNNs is a nontrivial task. The performance of the DNNs is highly sensitive to the training data, DNN architecture, hyperparameter tuning, and regularization techniques. All these aspects have to be considered before applying the DNNs to a specific problem.

Localized surface plasmons (LSPs) of metals are immensely important for applications such as surface-enhanced spectroscopy [11,12], sensing [13,14], optoelectronic devices [15–17], and imaging [18,19]. The aggregation of plasmonic nanostructures gives rise to rich physics phenomena including Fano resonances [20,21] and complicated hybridization [22,23].

Harnessing the power of the coupling LSPs can enable designs of optical structures with an on-demand spectral response. Conventionally, such a task requires a deep understanding of the physics by solving Maxwell's equations using finite-difference time-domain (FDTD) and finite-element method (FEM). The design methods typically comprise the sweeping of parameters and trial and error. While this is not so difficult in a simple symmetric system with fewer parameters, the challenge grows exponentially as the problem becomes asymmetric and complex [24]. For instance, the unit cell considered here can be described by six parameters, which means searching for the optimal value in a six-dimensional space. Since the unit cell size is sub-wavelength, the coupling between the unit cells has to be considered, making the system design more complicated. Recently, there is a growing interest in tackling these problems using fully connected (FC) DNNs [9,24–27] and convolutional neural networks (CNN) [28]. However, a comparison of these two types of structures in the photonics inverse design problem is still lacking. Moreover, the numerical accuracy of neural networks in the multi-parameter inverse design problem has not been addressed adequately in the previous reports.

In this work, we demonstrate that convolutional neural networks (CNNs) are significantly better in the inverse design of plasmonic metasurfaces based on a target spectrum. The advantages are two-fold. First, since the starting point of such an inverse design problem is usually a target spectrum where the valleys and the peaks are the defining features, the CNNs can extract these features and perform analysis on them, a process which has been extensively demonstrated in image recognition [29]. In this sense, the CNNs yield more accountability for the inverse design process. Second, the connections of the CNNs are sparse [30]; namely, only a small patch of neurons are connected to the previous layer. It is a way to remove the redundant information, thus reducing the difficulty in training and improving both accuracy and generalization capabilities [31]. In the inverse design problem based on the spectral response, the peaks and valleys are the main important information. The CNNs can take advantage of this and allow for more accurate and robust inverse design.

We consider a plasmonic metasurface whose unit cell consists of three gold (Au) nanodisks as shown in Fig. 1(a). The fabrication process is as follows. First, a layer of Au (50 nm) is deposited on a double-side polished sapphire substrate by sputtering;

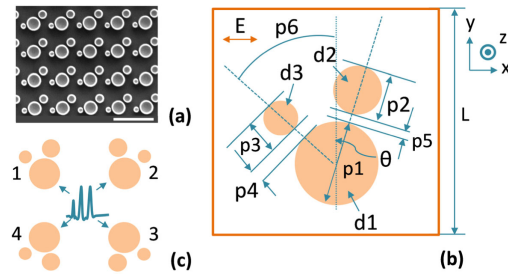


Fig. 1. (a) SEM image of the fabricated plasmonic metasurface, and the scale bar is 1 μm . (b) Schematic of geometrical parameters of the square unit cell. (c) Illustration of how the symmetry of the structure can cause data inconsistency.

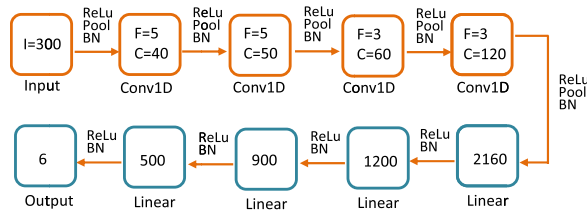


Fig. 2. Detailed parameters of the CNN.

subsequently, the photoresist is spin-coated on the substrate and electron-beam lithography is used to write the pattern. After developing, the Au is etched by inductively coupled plasma etching (ICP) with the photoresist as the etching mask. Six parameters can be defined to fully describe the geometry of a unit cell. They are the diameters of the three disks p_1 , p_2 , p_3 , the inter-disk distances p_4 , p_5 , and the angle p_6 as shown in Fig. 1(b). Previously, stand-alone plasmonic trimer systems have been shown to exhibit rich physics phenomena by virtue of the interactions between the plasmonic nanoparticles [22,32,33].

An illustration of the detailed network structure can be found in Fig. 2. It is comprised of convolutional layers (Conv1D) and linear layers. The input data of the neural network are the target spectra with an array size of 1×300 . In each Conv1D layer, F stands for the filter size, and C is the number of channels. Each Conv1D layer is followed by a rectified linear unit (ReLu) activation function, a max-pooling layer (Pool), and a batch normalization layer (BN). Each linear layer is followed by a ReLu and a BN layer. A total number of four convolutional layers and four linear layers are employed. After varying the number of the Conv1D layers and the linear layers, we find the current one is optimal. The output layer has six nodes representing the parameters that define a unit cell of the plasmonic metasurface.

We generate 25,000 sets of training data with randomly assigned parameters p_1 - p_6 using a commercial FDTD solver (FDTD solutions, Lumerical Inc). A broadband plane wave with wavelength ranging from 600 to 1600 nm is launched toward the structure from the $+z$ direction [Fig. 1(b)]. The boundary condition in the x - and y -directions is periodic, and those in the z -direction are perfectly matched layers (PML). The unit cell length L is 700 nm. The polarization of the source is in the x -direction. As a proof of concept, we set the thickness as a constant value of 50 nm; however, it can also be taken as a parameter for the network to learn. The absorption spectra are

collected as the training data and the parameters p_1 - p_6 are saved as the training labels for the CNN.

We note that the symmetry of the model could pose some challenges in the data generation. The conundrum is shown in Fig. 1(c). By mirroring the whole structure in the x and y -directions, the spectral response remains the same. But the structure is described by four different sets of parameters p_1 - p_6 . When the spectral data is fed to the CNN during the training, it does not know which set of parameters to converge to. While a few abnormal data points in the data set would not harm the performance of the CNNs, the symmetry problem presented here could give rise to a systematic data inconsistency with four subsets of conflicting data in the total set. The performance of the network will deteriorate greatly. Since such data inconsistency from symmetry can be a common occurrence, and it is not straightforward to find out, special attention has to be paid to eliminate these problems in any inverse design problem using the DNN techniques. Such a non-unique solution problem has been noted and solved by an extra network structure to pick out only one set of solutions [27], which increases the training difficulty and the size of the neural network. The origin of the problem is the symmetry in this work, so a more economical and robust solution is to just restrict the symmetry during the data generalization process. Therefore, we fix the center position of d_1 to be $(0, 0)$ and the angle θ to be 18° as shown in Fig. 2(a). Meanwhile, the rotation angle p_6 is set to be less than 90° , and the constraint $p_1 > p_2 > p_3$ is imposed on the diameter. Thus only one set of the configuration will be fed to the network, eliminating the non-unique solution problem. Another important point to consider is the normalization of the training data. The parameters p_1 - p_6 are on the order of 200, which is not in the best range for the DNNs. Thus each of them is scaled by 100 and shifted by -1 to be in the range of $[-1, 1]$. The spectrum data is scaled by 4 and shifted by -1 to be in the range of $[-1, 1]$.

The whole data set is divided into three subsets: 16,000 for training, 4000 for validation, and the rest of 5000 for testing. The mean squared error (MSE) is the performance indicator of the CNN as the loss function during training and as the error function during validation. The loss function is minimized by gradient descent using the Adam optimization algorithm [30] in the training process. The learning rate is set as 0.01 initially, and it is dropped by a factor of 1.2 every 50 epochs. The training sets are fed to the network with batch sizes of 100. The training and validation losses are obtained by a 5-fold cross-validation procedure and are shown in Figs. 3(a) and 3(b). To make a comparison between the CNNs and FC DNNs, another model that consists of only linear layers is built. In the FC DNN structures, the convolutional layers are replaced by three linear layers with the node numbers of 300, 900, and 2200, respectively. The ReLu and BN layers are maintained after each linear layer. The parameters are chosen based on the considerations to make the FC DNN network balanced. However, the total trainable parameters of these three layers alone are 2.25×10^6 , in contrast, the total trainable parameters of the Conv1D layers are 1260. We compare the CNN and FC DNN with and without batch normalization in Fig. 3. A few sharp drops in Fig. 3(a) can be discerned at epochs 100, 150, and 200 due to the learning rate modulation, which shows that the modulation of the learning rate is helping to accelerate the training process.

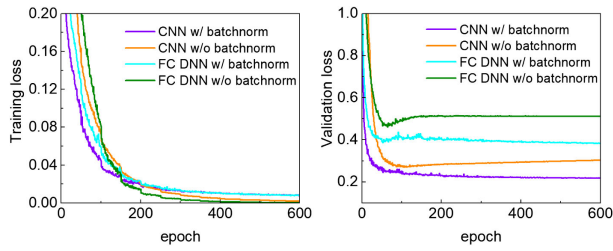


Fig. 3. (a) Training loss and (b) validation loss obtained during the training process for different neural network architectures.

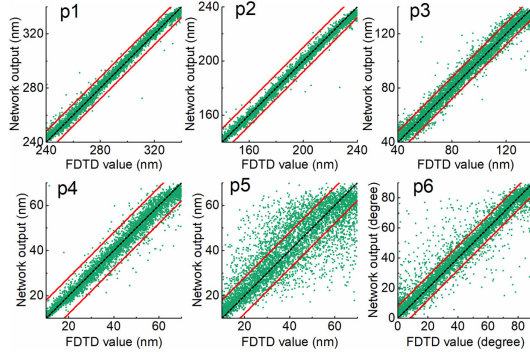


Fig. 4. Comparison of network prediction against FDTD values in the testing set. The black dashed lines represent $y = x$, and the red lines represent $y = x \pm 8$.

Without the batch normalization layers, the training loss decreases continuously to almost zero while the validation loss starts increasing after 100 epochs for both FC DNN and CNN, which is a clear sign of overfitting [34]. However, with the addition of the batch normalization layers, both the training and validation loss stabilized to a relatively low value, which indicates the batch normalization layers are helping to prevent overfitting in both CNN and FC DNN. The comparison of the FC DNN and the CNN is shown by the purple and cyan curves in Figs. 3(a) and 3(b). The training loss in Fig. 3(a) for both CNN and FC DNN converge to the same level, which indicates that both are well-trained and have the capability to fit training data well. However, the validation loss of the FC DNN is twice higher than that of CNN, indicating poor generalization capabilities, which can be attributed to the dense connections of FC DNN [31,35,36].

After the CNN is trained, the spectra in the testing set are fed to it and the predicted parameters p1-p6 are compared with the initial values used to generate the spectra. Figure 4 shows the comparison. The black and red dashed lines representing $y = x$ and $y = x \pm 8$, respectively, are plotted as guidelines. We observe there is a hierarchy in the accuracy. The parameters p1, p2, and p3 have the highest accuracy while the accuracy of p4, p5, and p6 is lower. The parameters p1, p2, and p3 decide the diameter of the three nanodisks. Hence, they are the most important parameters defining the overall shapes of the spectra and approximate peak and valley positions. The predicted p1, p2, and p3 follow the trend of $y = x$ closely and with 90% of the points fall within the boundaries of $y = x \pm 8$ nm. The parameters p4, p5, and p6 are related to the coupling strength between disks, and therefore they are parameters to fine-tune

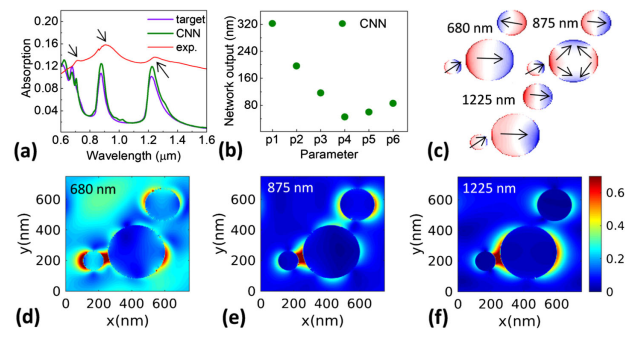


Fig. 5. (a) Comparison of the target spectrum, the spectral response of the plasmonic metasurface predicted by the CNN, and the measured spectrum of the fabricated metasurface. (b) Geometrical parameters p1-p6 predicted by the CNN. (c) Charge distribution at the plasmonic resonance wavelengths where black arrows indicate the direction of the electric field. (d)–(f) Normalized electric field intensity distribution at the resonance peaks of (d) 680 nm, (e) 875 nm, and (f) 1225 nm. The scale bar applies to (d), (e), and (f).

the spectrum. The parameter p5 has the lowest accuracy with the maximum deviation as high as 40 nm. This is because the parameter p5 controls the coupling between d1 and d2, which is weak in this work since the polarization direction is largely perpendicular to the coupling axis. On the contrary, the parameter p4 controls the coupling between d1 and d3, and the polarization is not always perpendicular to the coupling axis due to the rotation of p6. Thus the accuracy of p4 is higher. This point will be addressed in detail in the following paragraphs. The hierarchy of the accuracy is an indication that the network is learning the importance of each parameter p1-p6.

To further evaluate the performance of the proposed approach, an arbitrary spectrum modified from the testing data set with three plasmonic resonance peaks [the purple curve in Fig. 5(a)] is fed to the CNN. Afterward, an FDTD simulation is carried out using the predicted parameters p1-p6 to get the absorption spectrum. The comparison between the target spectrum and the one obtained from the network geometry (the green curve) is shown in Fig. 5(a). There are three peaks in the target spectrum at 680, 875, and 1225 nm, which are all well fitted by the network predicted parameters. Figure 5(b) shows the parameters p1-p6 predicted by the CNN. The metasurface is fabricated based on these parameters [Fig. 1(a)], and the absorption spectrum [Fig. 5(a)] is measured by a VIS/NIR spectrophotometer with a polarizer. There are some shifts and broadening of the resonant peaks, which are mainly due to the errors of the fabrication of small feature sizes down to 50 nm. However, the overall resonant peak positions marked by black arrows agree well with the prediction by the CNN. To investigate the origin of each plasmonic peak, we plot the charge distribution and normalized electric field intensity distribution at the peak wavelengths in Figs. 5(c)–5(f). The electric field direction is generally horizontal due to the polarization of the incoming light [Fig. 5(c)]. Due to the hybridization of each plasmonic mode, we observe the antibonding mode at 680 nm and bonding mode at 1225 nm. At 875 nm, d1 shows quadrupole modes due to the coupling with d3. We observe that the electric field is much weaker between d1 and d2 compared to d1 and d3. This is because the incident light is almost perpendicular to the axis of d1 and d2, which results in reduced coupling strength

[37,38]. Consequently, the parameter p_5 has a higher error with deviations up to 40 nm.

On the contrary, the interaction between d_1 and d_3 is more prominent, as observed from the enhanced field between d_1 and d_3 in Figs. 5(c)–5(f). This could be attributed to the smaller size of d_3 and the fact that the polarization is parallel to the axis of d_1 and d_3 . However, the rotation of d_3 is controlled by p_6 , and there are situations where the axis between d_1 and d_3 are almost perpendicular to the polarization direction, such as the case shown in Fig. 2(b). In these situations, the accuracy of p_4 should deteriorate, however, we observe the accuracy of p_4 is consistently high. This is because the coupling between p_2 and p_3 become prominent when p_6 is small, thus p_4 is still a critical parameter. Despite the deviations of p_5 and p_6 , the overall accuracy is adequate, considering the deviation of ± 8 nm of the critical geometrical parameters p_1 , p_2 , and p_3 is merely 1.5% of the shortest wavelength in this study (600 nm) and similar to the resolution limit of state-of-the-art nanofabrication techniques.

In conclusion, we demonstrate the CNN as an excellent tool to achieve high accuracy inverse design of plasmonic metasurfaces. The superiority stems from the fact that it can recognize the peaks and valleys of a spectrum and it is computationally less costly. We also show that batch normalization can improve the performance of the CNN. A high design accuracy of ± 8 nm is achieved for the critical geometrical parameters. Analysis of the results suggests the critical geometrical parameters possess significantly smaller errors than the less important ones. The technique and methodology can be applied to other inverse designs of nanophotonics involving target spectra.

Funding. King Abdullah University of Science and Technology (Baseline Fund BAS/1/1664-01-01, Competitive Research Grants URF/1/3437-01-01, Competitive Research Grants URF/1/3771-01-01, GCC Research Council Grant REP/1/3189-01-01).

Acknowledgment. The authors would like to thank Prof. Xiangliang Zhang from KAUST for her advice.

Disclosures. The authors declare no conflicts of interest.

REFERENCES

- Q. Liu, L. Fang, G. L. Yu, D. P. Wang, C. L. Xiao, and K. Wang, *Nat. Commun.* **10**, 2449 (2019).
- K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, *Nature* **559**, 547 (2018).
- P. Baldi, P. Sadowski, and D. Whiteson, *Nat. Commun.* **5**, 4308 (2014).
- P. M. R. DeVries, F. Viegas, M. Wattenberg, and B. J. Meade, *Nature* **560**, 632 (2018).
- L. F. Arsenault, A. Lopez-Bezaniilla, O. A. von Lilienfeld, and A. J. Millis, *Phys. Rev. B* **90**, 155136 (2014).
- H. Pinkard, Z. Phillips, A. Babakhani, D. A. Fletcher, and L. Waller, *Optica* **6**, 794 (2019).

- K. Yao, R. Unni, and Y. B. Zheng, *Nanophotonics* **8**, 339 (2019).
- Z. C. Liu, D. Y. Zhu, S. P. Rodrigues, K. T. Lee, and W. S. Cai, *Nano Lett.* **18**, 6570 (2018).
- M. H. Tahersima, K. Kojima, T. Koike-Akino, D. Jha, B. N. Wang, C. W. Lin, and K. Parsons, *Sci. Rep.* **9**, 1368 (2019).
- T. Asano and S. Noda, *Opt. Express* **26**, 32704 (2018).
- C. Matricardi, C. Hanske, J. L. Garcia-Pomar, J. Langer, A. Mihi, and L. M. Liz-Marzan, *ACS Nano* **12**, 8531 (2018).
- W. Lee, S. Y. Lee, R. M. Briber, and O. Rabin, *Adv. Funct. Mater.* **21**, 3424 (2011).
- N. Liu, M. Mesch, T. Weiss, M. Hentschel, and H. Giessen, *Nano Lett.* **10**, 2342 (2010).
- M. Dowran, A. Kumar, B. J. Lawrie, R. C. Pooser, and A. M. Marino, *Optica* **5**, 628 (2018).
- N. N. Jiang, X. L. Zhuo, and J. F. Wang, *Chem. Rev.* **118**, 3054 (2018).
- X. Yan, B. Li, Y. Wu, X. Zhang, and X. M. Ren, *Appl. Phys. Lett.* **109**, 053109 (2016).
- R. F. Service, *Science* **339**, 263 (2013).
- S. Kawata, Y. Inouye, and P. Verma, *Nat. Photonics* **3**, 388 (2009).
- T. L. Vasconcelos, B. S. Archanjo, B. Fragneaud, B. S. Oliveira, J. Riikonen, C. F. Li, D. S. Ribeiro, C. Rabelo, W. N. Rodrigues, A. Jorio, C. A. Achete, and L. G. Cancado, *ACS Nano* **9**, 6297 (2015).
- J. Ye, F. F. Wen, H. Sobhani, J. B. Lassiter, P. Van Dorpe, P. Nordlander, and N. J. Halas, *Nano Lett.* **12**, 1660 (2012).
- J. A. Fan, C. H. Wu, K. Bao, J. M. Bao, R. Bardhan, N. J. Halas, V. N. Manoharan, P. Nordlander, G. Shvets, and F. Capasso, *Science* **328**, 1135 (2010).
- L. Chuntunov and G. Haran, *Nano Lett.* **11**, 4503 (2011).
- E. Calandrini, A. Cerea, F. De Angelis, R. P. Zaccaria, and A. Toma, *Nanophotonics* **8**, 45 (2019).
- J. Peurifoy, Y. C. Shen, L. Jing, Y. Yang, F. Cano-Renteria, B. G. DeLacy, J. D. Joannopoulos, M. Tegmark, and M. Soljacic, *Sci. Adv.* **4**, eaar4206 (2018).
- I. Malkiel, M. Mrejen, A. Nagler, U. Arieli, L. Wolf, and H. Suchowski, *IEEE International Conference on Computational Photography (ICCP)* (2018).
- I. Malkiel, M. Mrejen, A. Nagler, U. Arieli, L. Wolf, and H. Suchowski, *Light Sci. Appl.* **7**, 60 (2018).
- D. J. Liu, Y. X. Tan, E. Khoram, and Z. F. Yu, *ACS Photon.* **5**, 1365 (2018).
- W. Ma, F. Cheng, and Y. M. Liu, *ACS Nano* **12**, 6326 (2018).
- A. Krizhevsky, I. Sutskever, and G. E. Hinton, *Adv. Neural Inf. Process. Syst.* **25**, 1097 (2012).
- D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," arXiv:1412.6980 (2014).
- D. C. Mocanu, E. Mocanu, P. Stone, P. H. Nguyen, M. Gibescu, and A. Liotta, *Nat. Commun.* **9**, 2383 (2018).
- T. Shegai, Z. P. Li, T. Dadoosh, Z. Y. Zhang, H. X. Xu, and G. Haran, *Proc. Natl. Acad. Sci. USA* **105**, 16448 (2008).
- E.-M. Roller, L. V. Besteiro, C. Pupp, L. K. Khorashad, A. O. Govorov, and T. Liedl, *Nat. Phys.* **13**, 761 (2017).
- N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, *J. Mach. Learn. Res.* **15**, 1929 (2014).
- J. Patterson and A. Gibson, *Deep Learning: A Practitioner's Approach* (O'Reilly Media, 2017).
- M. Thom and G. Palm, *J. Mach. Learn. Res.* **14**, 1091 (2013).
- P. K. Jain, W. Y. Huang, and M. A. El-Sayed, *Nano Lett.* **7**, 2080 (2007).
- D. Punj, R. Regmi, A. Devilez, R. Plauchu, S. B. Moparthy, B. Stout, N. Bonod, H. Rigneault, and J. Wenger, *ACS Photon.* **2**, 1099 (2015).