



PHOTONICS Research

Accurate inverse design of Fabry–Perot-cavity-based color filters far beyond sRGB via a bidirectional artificial neural network

PENG DAI,^{1,†} YASI WANG,^{2,†} YUEQIANG HU,² C. H. DE GROOT,¹ OTTO MUSKENS,³ 
HUIGAO DUAN,^{2,4} AND RUOMENG HUANG^{1,5} 

¹School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK

²National Engineering Research Center for High Efficiency Grinding, College of Mechanical and Vehicle Engineering, Hunan University, Changsha 410082, China

³School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK

⁴e-mail: duanhg@hnu.edu.cn

⁵e-mail: r.huang@soton.ac.uk

Received 26 November 2020; revised 21 February 2021; accepted 1 March 2021; posted 1 March 2021 (Doc. ID 415141); published 30 April 2021

Structural color based on Fabry–Perot (F-P) cavity enables a wide color gamut with high resolution at submicroscopic scale by varying its geometrical parameters. The ability to design such parameters that can accurately display the desired color is therefore crucial to the manufacturing of F-P cavities for practical applications. This work reports the first inverse design of F-P cavity structure using deep learning through a bidirectional artificial neural network. It enables the production of a significantly wider coverage of color space that is over 215% of sRGB with extremely high accuracy, represented by an average ΔE_{2000} value below 1.2. The superior performance of this structural color-based neural network is directly ascribed to the definition of loss function in the uniform CIE 1976-Lab color space. Over 100,000 times improvement in the design efficiency has been demonstrated by comparing the neural network to the metaheuristic optimization technique using an evolutionary algorithm when designing the famous painting of “Haystacks, end of Summer” by Claude Monet. Our results demonstrate that, with the correct selection of loss function, deep learning can be very powerful to achieve extremely accurate design of nanostructured color filters with very high efficiency. © 2021 Chinese Laser Press

<https://doi.org/10.1364/PRJ.415141>

1. INTRODUCTION

Structural color filters can display various colors by selectively transmitting or reflecting a specific wavelength by varying structural parameters rather than material components in the visible region [1]. They have received enormous interest recently due to their potential applications in chromatic display [2,3], color printing [4,5], optical encryption [6,7], solar cells [8], and so on. They largely exceed the conventional colorant-pigment-based filters in multiple aspects including nontoxicity, great scalability and durability, high resolution, and easy tenability [9]. Generation of structural color normally employs resonances (e.g., plasmonics, Mie scattering, guided mode resonance) from subwavelength patterns [10–13]. However, all these require nanoscale patterning, which involves complicated fabrication steps and can be cost prohibitive for high-volume and large-area applications.

Planar thin-film structures based on Fabry–Perot (F-P) cavity resonances are an alternative structural color filter technology based on the thin-film interference [14]. A typical F-P

resonator consists of a lossless dielectric that is sandwiched between two reflective metal layers. The function of color filtering is achieved by multiple round-trip phase delays of electromagnetic waves in the F-P resonator [15]. The resonant peak location can therefore be controlled by varying the dielectric layer thickness to obtain different colors. The full width at half-maximum (FWHM) of the peak, which relates to the color purity and brightness, can be tuned through the two metal layer thicknesses. Compared with the pattern-based filters, F-P-cavity-based structural color offers a much lower-costing and higher-scalability way of structural color manufacturing while offering larger color gamut and high color purity and contrast [16]. The high-index materials of phase compensation overlayers can also achieve great angle insensitivity for both transverse-magnetic (TM) and transverse-electric (TE) polarizations [17]. There are several works published on the aspect of improving the performance of F-P cavity-based structural color in both transmissive and reflective modes [18–20]. In addition, the F-P-cavity-based color filter also demonstrates high lateral

resolution. Wang *et al.* reported that its color crosstalk is comparable to those of plasmonic filters and pigmented filters with the same sizes [4]. A minimum pixel dimension of 500 nm (*ca.* 50,800 dpi) has also been successfully demonstrated, showing a great potential for F-P-cavity-based color filters to be applied in high-resolution colorization.

An important aspect for structural color is the ability to design a structure that can accurately display the desired color [21]. This conventional design task is realized by a trial-and-error method in which an initial random design is converged to the desired design through iterative optimization. Identification of a reasonable design therefore often requires an experienced designer with prior knowledge of the problem and a significant amount of calculations or simulations, which could be prohibitively slow as the complexity of the structure increases. However, this paradigm has recently been changed by the unprecedented development of deep learning techniques [21–23].

Deep learning is an important subset of machine learning in which multilayered artificial neural networks are utilized to achieve high-accuracy predictions and classifications [24]. Several works have been reported recently in which neural networks were used to inversely design the nanostructure and material to achieve desired optical responses [25–31]. Before the networks can perform the intended functions, a training process needs to take place in which a dataset of structural parameters to color relations is required. This dataset normally involves a large amount of relations, and it needs to be generated by theoretical calculation or simulation. However, this is a one-time investment, and no more computational resources will be consumed once the network is properly trained. Pioneering works have also been reported on deep-learning-aided structural color design [32–36]. Hemmatyar *et al.* designed and optimized the hafnia array based all-dielectric metasurfaces via neural network to generate a wide color gamut [32]. Gao *et al.* reported a structural color inverse design by employing a bidirectional neural network [35], which was first proposed by Liu *et al.* [37]. The model consists of a fully connected inverse network that directly connects to a pretrained forward network. The inverse network can produce the design of the structural parameters for desired colors, while the forward network predicts the colors from the structural parameter inputs. The training of the inverse network is conducted by feeding the inversely designed structures directly into the pretrained forward network, and the network is optimized by minimizing the difference between the predicted color and the input color through backpropagation. This model has been widely used in different nanophotonic applications, as it overcomes the issue of nonuniqueness [31,38,39].

Although previous works have achieved remarkable progress in the structural color inverse design, one remaining challenge is the relatively low design accuracy (large color differences between design and target) even when extremely low validation loss is achieved. One key reason for that is the nonuniformity of the color space from which the loss function was defined. The loss function measures the difference between the network predicted value and the true value in the dataset. It is crucial for neural networks, as the training is a process of minimizing the

loss function [40]. An unsuitable selection of loss function may lead to a very low loss function but relatively higher prediction error [41]. Defining loss function in a nonuniform color space such as CIE 1931-XYZ means the same Euclidean distance among the XYZ vectors may signify different color differences, resulting in biased optimization against some colors [42].

Here we report, to the best of our knowledge, the first inverse design of F-P cavity structure using deep learning through a bidirectional artificial neural network. This structure enables the production of a significantly wider color gamut that is over 215% of sRGB color space, which is essential for real display applications. By defining the training loss function in the uniform CIE 1976-Lab color space where the same Euclidean distance indicates same color difference, our network is able to achieve a much higher design accuracy with average color difference ΔE_{2000} below 1.2. The high design efficiency of the network is also evaluated by comparison to the evolutionary algorithm, which shows over 100,000 times savings of computational resources. The demonstration of this work offers exciting prospects for deep learning techniques to be used to achieve accurate and efficient structural color designs for a wide range of different applications.

2. DATASET GENERATION

The schematic diagram of the transmissive F-P-cavity-based color filter employed in this work is illustrated in Fig. 1(a). A transmissive-type of color filter was chosen because of its wide application in spectrometers, CMOS image sensors, and liquid crystal displays [43–45]. Compared with reflective filters, transmissive filters enjoy the advantages of single-mode operation, which could lead to a higher color purity. However, it is worth mentioning that a similar approach can also be used on the reflective type. The system has a trilayer metal-insulator-metal (MIM) films stack on the quartz substrate in which a SiO₂ dielectric layer is sandwiched between two Ag metal layers. Optical interference occurs when white light enters the cavity which filters out the wavelengths that are not matching the resonant wavelength of this multilayer system. With the materials of the dielectric and metal layers fixed, the most critical parameters here affecting the resonant wavelengths are the thickness of each layer, which are represented as d_1 , d_2 , and d_3 as shown in Fig. 1(a). In this work, the ranges of d_1 , d_2 , and d_3 are set to be 0 to 50 nm, 0 to 1000 nm, and 0 to 50 nm to allow a large color gamut coverage. Only integer values are selected in this work to ensure the compatibility with fabrication techniques. A total of 101,000 parameter combinations are randomly generated, and the corresponding transmissive spectra from 380 to 780 nm are computed by the multiple beam interference formulas [16]. However, color is not a property of electromagnetic radiation but a subjective perception of an observer. Color-matching functions are then required to convert the transmissive spectra into corresponding color coordinates in CIE 1931-XYZ color space to generate a dataset containing 101,000 parameters \mathbf{D} (d_1, d_2, d_3) to color XYZ (X, Y, Z) relations. This dataset is divided into three groups for training (90,000), validation (10,000), and testing (1000) purposes. All colors generated in the training and validation dataset are plotted in the CIE 1931-xy chromaticity diagram in

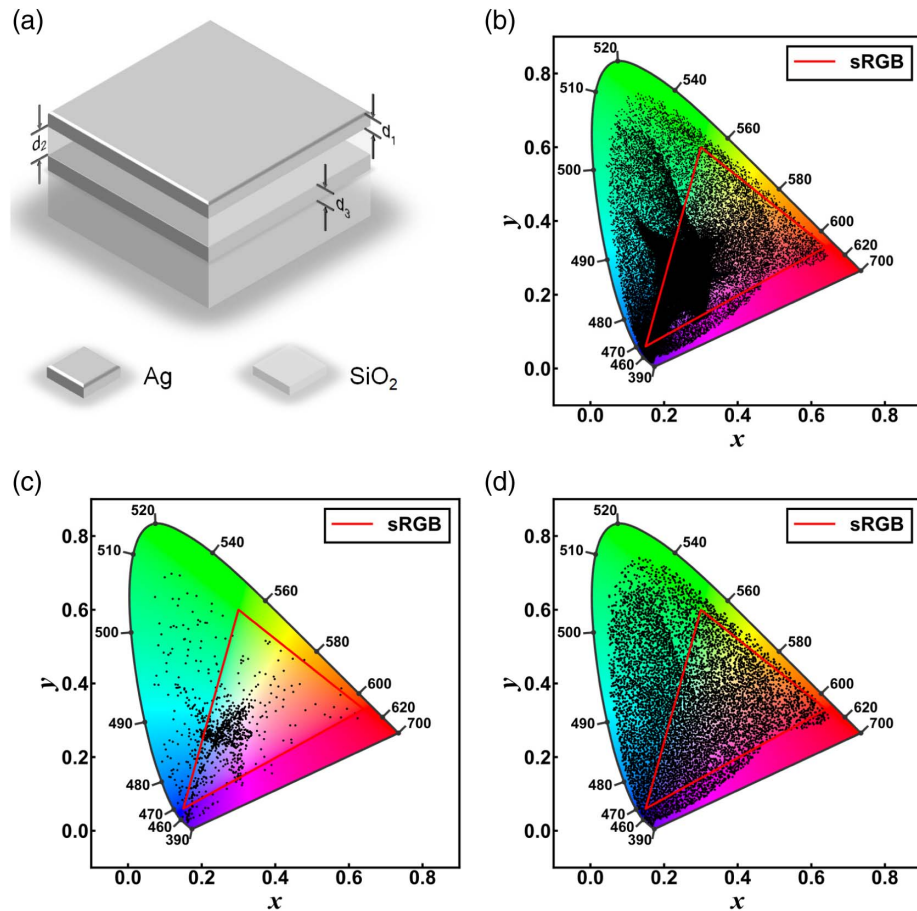


Fig. 1. (a) Schematic illustration of a transmissive F-P-cavity-based color filter with an MIM structure. (b) The training and validation dataset generated by the F-P cavity plotted in a CIE 1931- xy chromaticity diagram. The red triangle depicts the boundary of sRGB color space for a guide of the eye. (c) The original 1000 testing data and (d) the 7000 uniformly distributed testing data plotted in the CIE 1931- xy chromaticity diagram.

Fig. 1(b). It is clear that our Ag-SiO₂-Ag F-P-cavity-based color filter in this work can achieve a substantially larger (*ca.* 215%) color gamut than the sRGB color space (plotted by red lines for reference). The large gamut of coverage is one of the advantages of F-P-cavity-based structural color and is particularly beneficial for real applications such as display and full-color nanoprinting [4,19]. By varying the thickness of the dielectric layer (d_2), the transmission peak could be swept across the whole color in the visible light range. The modification of the Ag layer thicknesses (d_1 and d_3) serve to further tune the FWHM of the transmittance peak, resulting in a large gamut of coverage. The testing dataset is plotted in Fig. 1(c) (referred to as the testing set below). To further test the robustness of our networks, we also generated an additional testing set with 7000 colors that are uniformly distributed on the CIE 1931- xy chromaticity diagram [shown in Fig. 1(d) and referred to as the uniform testing set below].

3. FORWARD NEURAL NETWORK CONSTRUCTION

A forward neural network (FNN) was first trained to obtain accurate prediction of colors based on the layer thicknesses. Prior to training, a loss function needs to be established.

Although our original dataset converts the spectrum into a CIE 1931-XYZ tristimulus vector, it is not a suitable output for loss function definition due to its nonuniformity [42]. sRGB color space is also not ideal, as the conversion between XYZ and sRGB is not reversible when the color is outside of the sRGB color space. On the other hand, the CIE 1976-Lab color space has a one-to-one correspondence to the CIE 1931-XYZ but with much better uniformity, rendering it a more suitable color space for accurate color difference identification. In fact, the color difference function CIE ΔE_{1976} is defined by the Euclidean distance of two Lab vectors (L, a, b). This property is particularly beneficial in neural network training. By defining the loss function to be mean squared error (MSE) between the predicted and original Lab values, it can be directly converted to the actual color difference (ΔE_{1976}) and enable higher accuracy. Figure 2(a) shows the summary of the dataset preparation process for the FNN. After obtaining XYZ from the spectrum, it was converted to Lab, which was then used to construct the loss function and identify color difference. The architecture of the FNN is illustrated in Fig. 2(b) and is composed of a fully connected neural network (NN) including one input layer, one output layer, and several hidden layers. It takes the parameter **D** as input and outputs the Lab, which can be converted to other color vectors such as XYZ and sRGB for different applications.

During the deep learning process, the training data group is fed to the FNN to continuously adjust the weight and bias of each connection with every batch of data in epochs. This is achieved by backpropagation of the loss function. The selection of the hyperparameters (i.e., number of hidden layers and neurons per layer) is crucial to the performance of the network [21]. A systematic study was therefore conducted to investigate the impact of hyperparameters for this FNN. The CIE ΔE_{2000} color difference was chosen here to provide a better quantification of the FNN performance. Similar to the ΔE_{1976} , the ΔE_{2000} color difference is also a function of the two **Lab** values but corresponds better with the way in which human observers perceive small color differences and hence is used as the metric for design accuracy in this work [46]. The ΔE_{2000} can be classified into five groups: 1) $\Delta E_{2000} < 1$, it can be considered no color difference; 2) $1 < \Delta E_{2000} < 2$, the difference can be observed by experienced persons; 3) $2 < \Delta E_{2000} < 3.5$, the difference can be observed by unexperienced persons; 4) $3.5 < \Delta E_{2000} < 5$, a clear difference can be noticed; and 5) $5 < \Delta E_{2000}$, two different colors are observed [47]. The distribution of ΔE_{2000} from the testing set of each FNN is plotted in Fig. 3(a) as a function of layer number. It provides a clear indication on the color prediction performance of each FNN. The average ΔE_{2000} values (blue squares) can therefore be used to provide a quantified metric for the performance. It can be observed that the average ΔE_{2000} in the testing set decreases from 1.87 with two hidden layers to 0.44 with seven hidden

layers, indicating that the FNN with seven hidden layers has the best performance in our optimization range. The impact of number of neurons per layer was also investigated while the number of layers was fixed at seven.

Similar to the optimization of hidden layer numbers, the ΔE_{2000} distribution and average values are plotted in Fig. 3(b). The average ΔE_{2000} in the testing set plunged sharply from 3.05 to 0.44 with the increase of neuron number from 10 to 50 and subsequently bottomed out at 0.38 for 250 neurons. Therefore, the architecture of the FNN is optimized with seven hidden layers and 250 neurons in each hidden layer. The influence of the dataset size was also investigated as shown in Fig. 3(c). The results suggest that sufficient dataset size (over 50,000 in our case) is required for the network to achieve a good performance of an average ΔE value below 0.5 in our case.

The high accuracy of our FNN is further confirmed by evaluation using the uniform testing set, resulting in a ΔE_{2000} of 0.35. We believe this high accuracy can be ascribed to the use of uniform CIE 1976-Lab color space for the network training. An FNN with the same network architecture using **XYZ** as the output was also trained for comparison as shown in Figs. 3(d) and 3(e). Although a lower MSE (*ca.* 10^{-6}) was achieved by the FNN with **XYZ** output, the ΔE_{2000} distribution is significantly poorer than the one with **Lab** output with a decrease of ΔE_{2000} values observed from 1.03 to 0.38. To better explain the advantage of our approach, Fig. 3(f)

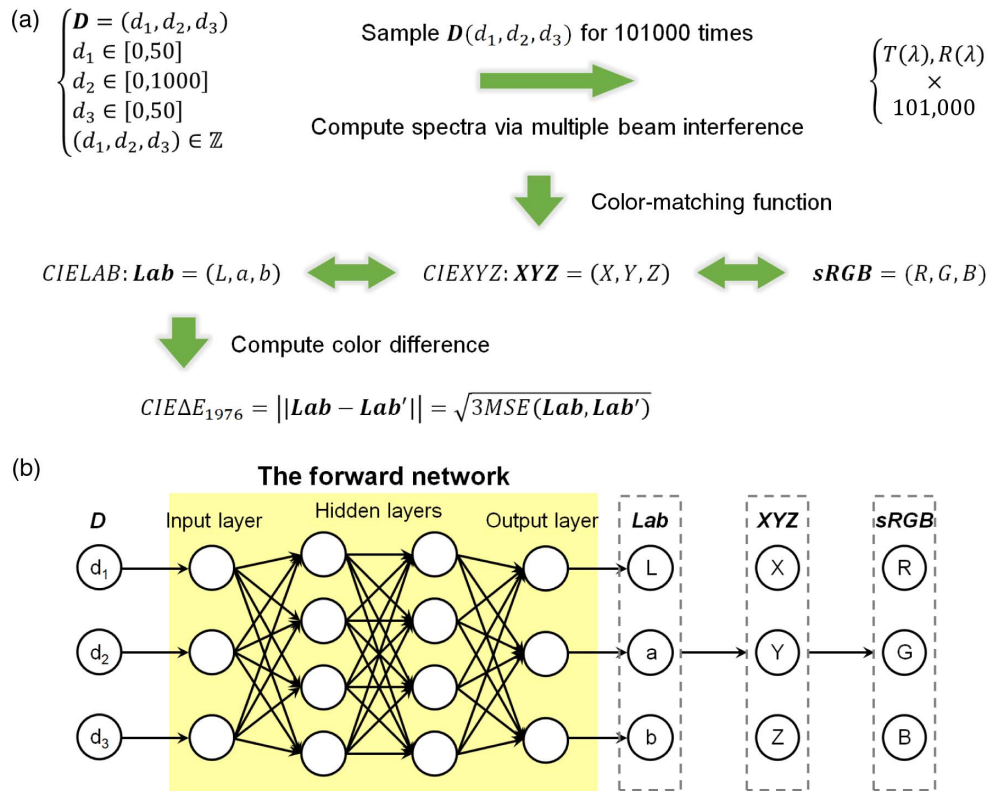


Fig. 2. Forward neural network for predicting F-P cavity structural colors. (a) The relationships of different parameters in the dataset. The single arrow means the transformation is unidirectional, the double arrow means the transformation is reversible. (b) The architecture of the forward neural network with input layer of geometric parameter **D**, hidden layers, and output layers of **Lab** color values.

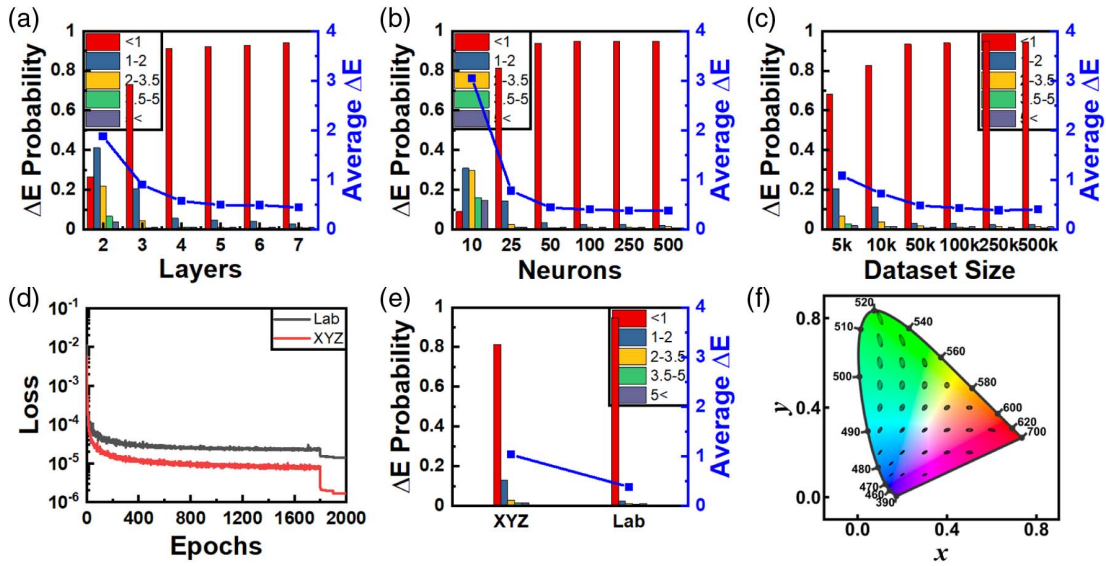


Fig. 3. Forward neural network training for predicting F-P cavity structural colors. The histogram of the probability and average values of ΔE_{2000} of the FNNs with different (a) hidden layer number, (b) neuron number per layer, and (c) the FNN with seven hidden layers and 250 neurons in each hidden layer for different dataset size. (d) The training loss curves for defining the loss function in CIE 1931-XYZ and CIE 1976-Lab color spaces. (e) The probability histogram and average values of ΔE_{2000} comparisons while the loss function is defined in CIE 1931-XYZ color space and CIE 1976-Lab color space. (f) The selected colors in the CIE 1931- xy chromaticity diagram with the boundary of each ellipse representing the colors that have a ΔE_{1976} of 6 to the selected color.

presents a selection of colors in the CIE 1931- xy chromaticity diagram with the boundary of each ellipse representing the colors that have a ΔE_{1976} of 6 to the selected color. It is clear that the ellipses in the green and red regions are larger than that in the blue region due to the nonuniformity of the CIE 1931-XYZ color space. One could obtain a very small ΔXYZ in the green region, but the actual reduction of color difference might be limited. This suggests that minimizing ΔXYZ values, especially in the green and red regions, will not be as effective in reducing color difference as the approach suggested in our work. This proves that the correct selection of loss function is critical to the performance of the neural network.

4. INVERSE NEURAL NETWORK CONSTRUCTION

The training of the inverse neural network (INN) is more challenging due to the nonuniqueness nature that one color can be formed by different F-P cavity structures. This multisolution property could lead to the adjustment of weight being pulled to different local or global minima during the training process, making the training difficult to converge. The bidirectional neural network architecture and tandem training strategy were therefore employed in this work in which the output parameter $\mathbf{D}'(d'_1, d'_2, d'_3)$ from the INN was directly fed into our pre-trained FNN to generate predicted $\mathbf{Lab}'(L', a', b')$ as shown in Fig. 4(a). The parameters of FNN are fixed during the INN training. The loss function can be defined as the MSE between the original and predicted Lab vectors instead of parameter \mathbf{D} . This means the INN will be optimized to match the desired color rather than the structure, avoiding the non-uniqueness problem in the training process. Here we adopted a

recently reported penalized tandem training strategy to further enhance the robustness of the network in the initial 100 epochs [39], which indicates that the inverse error (the MSE between the original \mathbf{D} and predicted \mathbf{D}') was included in the loss function for the first 100 epochs to avoid the predictions violating the ground truth of \mathbf{D} .

Another important factor in INN training is the selection of a random seed. The initial weights are obtained from random sampling using determined distributions (e.g., Xavier or Kaiming initialization). The training is therefore repeatable once the selection of a random seed was fixed. Different random seeds could place the INN at different starting positions in the loss plane [Fig. 4(b)]. In the case of multiple solutions, the loss plane is very complicated due to the interference of different global or local minima [Fig. 4(c), left]. Different starting positions could cause the INN to converge into a different local/global minimum, or in the case of a poor initialization, struggle to even converge. Selection of the random seed therefore plays a vital role in the INN training process. In this work, each INN underwent a process of random seed selection before hyperparameter optimization took place. The summary of the MSEs after 200 training epochs for INNs with varying hidden layers is presented in Fig. 4(d). It is obvious that the training progress differs significantly due to the selection of different random seeds, and different INN architectures prefer different random seed groups for optimized performance. This is particularly important for multisolution questions with the existence of a large number of local and global minima. It is less critical for single-solution questions (e.g., the FNN in this work), as the loss plane is relatively simple for the network to converge [Fig. 4(c), right] regardless of the initial starting position.

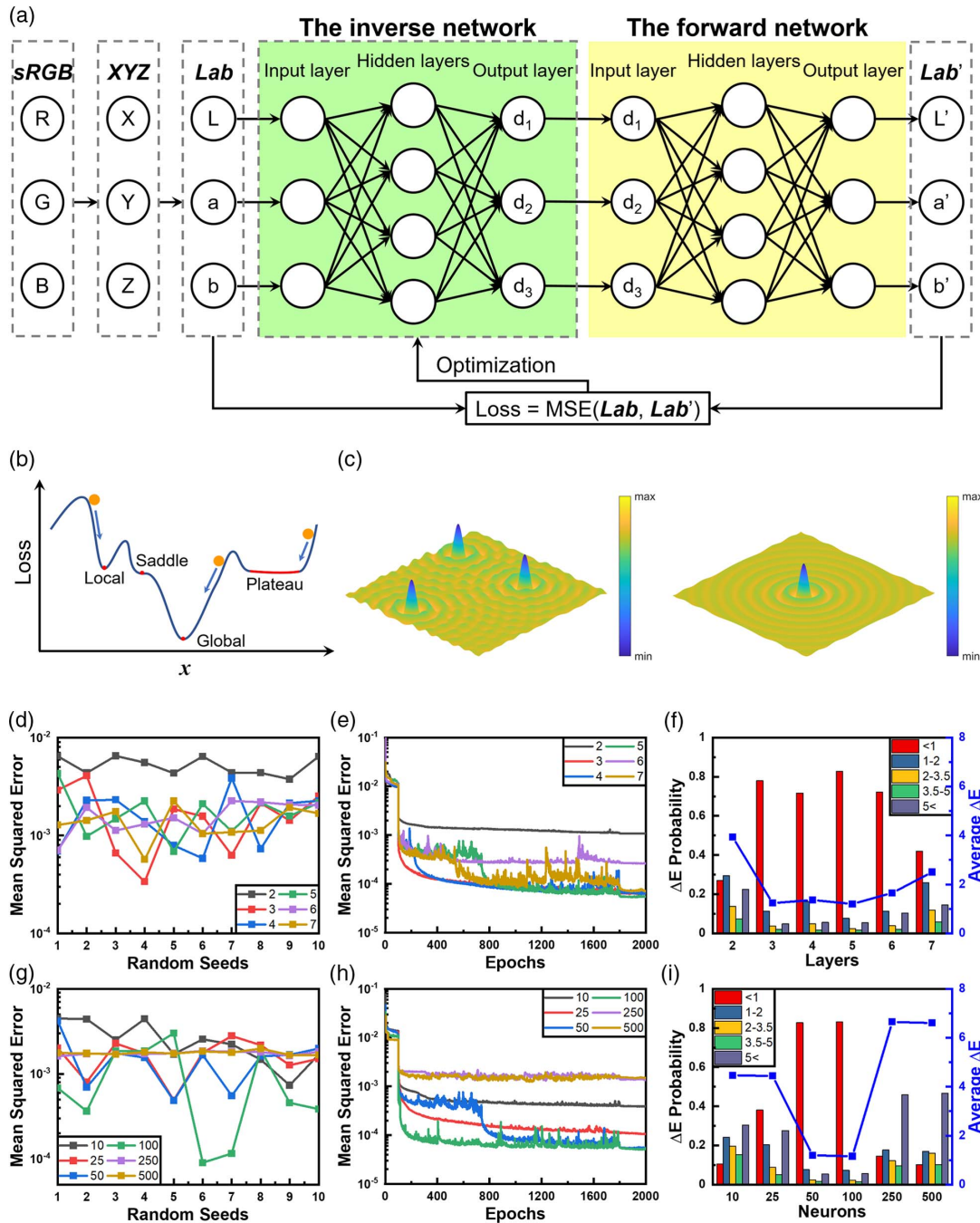


Fig. 4. Inverse neural network for predicting F-P cavity structural colors. (a) The bidirectional architecture with input layer of Lab values and output layer of geometric parameter \mathbf{D} and connected to the pretrained forward neural network. (b) The schematic of different weights' initialization positions. (c) The loss surface schematics of nonuniqueness (left) and uniqueness (right) problems, respectively. The MSEs after 200 epochs as a function of random seed, the training loss curves, the histogram of the distribution, and average values of ΔE_{2000} of the INNs with (d)–(f) different numbers of hidden layers and (g)–(i) different numbers of neurons per layer.

Figures 4(e) and 4(f) present the optimization of hidden layer numbers for INN. A drop of MSE at the 100th epoch in each training is due to the removal of the inverse error term in the loss function as mentioned above. Unlike the FNN where higher network complexity always leads to a better performance in the training set, a reduction of performance was observed for the INN with more than five layers. This inferior

performance for more complicated INNs may be caused by the increased dimensions in the loss plane and number of existing global minima, which causes the network convergence to be more difficult. Similar behavior was also observed in the process of neuron number optimization as shown in Figs. 4(g)–4(i). A final INN with five hidden layers with 100 neurons was identified to be the best INN architecture to be used in this work.

This network enables the average ΔE_{2000} of 1.16 in the non-uniform testing set and 1.18 in the uniform testing set, which are superior to most of the commercial display equipment. For example, the high-end Dell UltraSharp 32 PremierColor UltraHD 8K Monitor is factory calibrated at 100% sRGB coverage to an accuracy of ΔE_{2000} less than 2.

5. PERFORMANCE EVALUATION

To better evaluate the performance of our INN, we randomly selected 6 F-P cavity structures (**D** values) that the INN has never seen before, as listed in Table 1. These structures were converted to the corresponding spectra and subsequently **Lab** values as discussed previously. The target **Lab** values were then fed into our INN to obtain the designed structures **D'**. Further tests by converting the designed structures to designed colors through the theoretical calculation have resulted in very close matches with the targeted colors. The color differences, represented by the ΔE_{2000} values, are below 1 in all six cases, suggesting that human eyes are not able to distinguish their differences. More importantly, in five of the six cases, our

INN has produced designed structures that are significantly different from the original to the targeted structures while obtaining small color difference. This demonstrates the existence of nonuniqueness solutions for INN. More importantly, it highlights the ability of a neural network to discover solutions outside of the boundaries of the training data. Nanophotonic research has become more computation intensive due to the large spatial degrees of freedom and wide choice of materials [22]. Such ability to identify new design structures that can never be found through a conventional forward design technique would be extremely beneficial for discovering novel findings in nanophotonics.

The excellent performance of our network is further supported by designing the F-P cavity structures to reproduce the painting of “Haystacks, end of Summer” by Claude Monet. This was done by extracting the **sRGB** values of all 2000×1176 (2,352,000) color pixels from the original painting [shown in Fig. 5(a)] and inputting them into the bidirectional network. The INN outputs the designed geometric parameters for all 2,352,000 pixels, and the reconstructed paintings were subsequently generated and plotted in Fig. 5(b).

Table 1. Comparison of the Six Randomly Selected Target Structure **D, **Lab**, and sRGB Color Values with the Corresponding Values Inversely Designed by the Inverse Neural Network**

Target D (nm)	Target sRGB	Target Lab	Design D (nm)	Design Lab	Design sRGB	ΔE
31-365-31	142-27-80 142-27-80	L: 32.0 a: 50.2 b: -1.7	32-365-30	L: 31.8 a: 49.6 b: -1.4	141-28-79 141-28-79	0.27
44-152-11	95-67-59 95-67-59	L: 31.2 a: 10.7 b: 11.7	43-345-15	L: 31.5 a: 10.4 b: 12.4	96-68-55 96-68-55	0.70
29-141-12	163-142-97 163-142-97	L: 59.8 a: 1.7 b: 26.8	25-336-15	L: 60.0 a: 1.6 b: 27.5	164-142-96 164-142-96	0.37
29-121-26	(-600)-165-114 0-165-114	L: 56.4 a: -84.2 b: 11.0	23-293-20	L: 56.5 a: -84.8 b: 11.5	(-605)-165-113 0-165-113	0.25
20-385-44	89-(-44)-110 89-0-110	L: 16.9 a: 61.8 b: -46.0	41-387-23	L: 16.9 a: 62.0 b: -45.7	89-(-45)-109 89-0-109	0.18
19-839-18	44-140-166 44-140-166	L: 54.2 a: -19.9 b: -21.6	24-271-5	L: 54.3 a: -19.1 b: -21.2	44-141-166 44-141-166	0.29

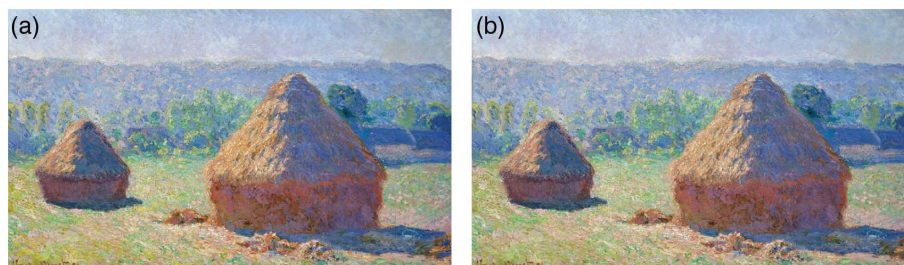


Fig. 5. Actual application of the INN-based structural color design. (a) The origin and (b) the reproduction via INN of the painting “Haystacks, end of Summer” by Claude Monet. Reproduction of “Haystacks, end of Summer” is permitted by the Musée d’Orsay, Paris (RF 1975 3).

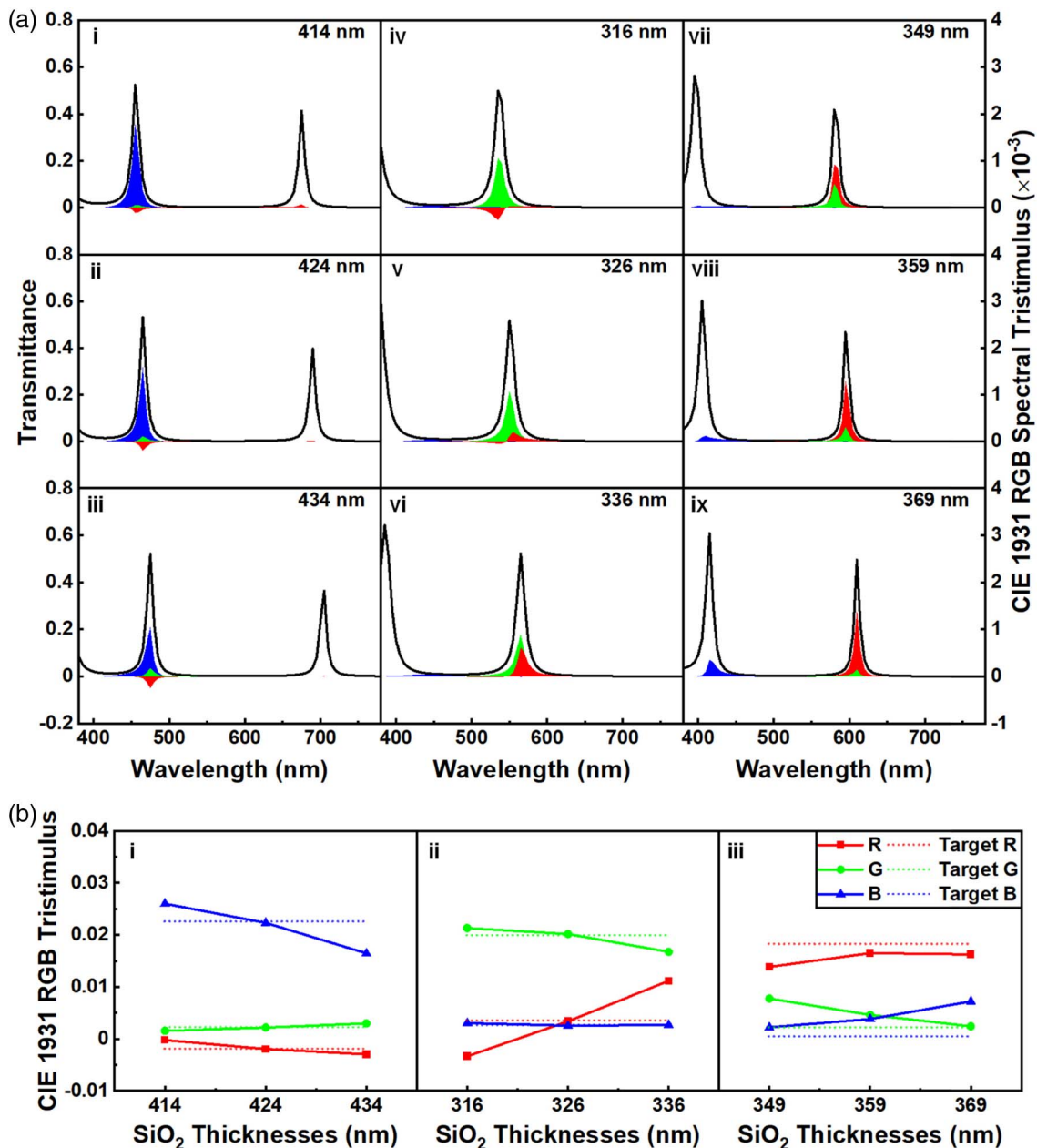


Fig. 6. Transmissive spectra and corresponding CIE 1931-RGB tristimulus values for the designed colors. The transmissive spectra (black line) and the contribution from the three stimuli (shades underneath the line) for the (i, ii, iii) blue, (iv, v, vi) green, and (vii, viii, ix) red color designed by the INN in this work. Within each figure, the middle row figure presents the original design, while the top and bottom figures represent the spectra from a 10 nm thinner layer and a 10 nm thicker dielectric layer. (b) The CIE 1931-RGB tristimulus values as a function of dielectric layer thickness for the blue, green, and red colors, respectively. The CIE 1931-RGB tristimulus values of the targeted colors are also included (dotted lines) for comparison.

An extremely high accuracy can be clearly observed, as the difference between the two images is almost undistinguishable. This proves the robustness and accuracy of our network in designing colors over a wide gamut.

We will now provide a detailed analysis of the network performance by evaluating the spectra of the blue (sRGB 0, 0, 102), green (sRGB 0, 102, 0), and red (sRGB 102, 0, 0) color filters designed by our INN. The designed F-P cavity for blue

color has a dielectric thickness (d_2) of 424 nm, and its transmissive spectrum is characterized by a main peak at 465 nm with a secondary peak at 690 nm as shown in Fig. 6(a-ii). The contribution from each CIE 1931-RGB spectral tristimulus is demonstrated by the size of the shades underneath the spectrum. It is obvious that the designed blue color consists of a majority of blue stimuli with a small proportion of green and red stimuli, the integrals of which are extremely close to the

definition of blue color in the sRGB color space [shown by the dashed line in Fig. 6(b-i), and the corresponding CIE 1931-RGB tristimulus values can be obtained by converting the CIE 1931-XYZ tristimulus]. Almost identical values are obtained with a ΔE_{2000} of 0.07. An unoptimized design with a dielectric layer 10 nm thinner will cause a blueshift of the spectrum [Fig. 6(a-i)]. This results in an increased contribution from the red stimuli but a reduced contribution from the green stimuli, driving the designed color away from the blue definition with a ΔE_{2000} of 6.16. Similarly, a 10 nm thicker dielectric layer induces a redshift with more contribution from green stimuli and less contribution from red stimuli, resulting in a ΔE_{2000} of 11.83. Similar behavior is also observed on the design of green color as shown in Fig. 6(a-iii). Our design with a dielectric thickness (d_2) of 326 nm results in a ΔE_{2000} of 0.35 [Figs. 6(a-v) and 6(b-ii)], whereas a ± 10 nm change of the thickness alters the contribution from the CIE 1931-RGB tristimulus and results in inferior color design with ΔE_{2000} of 7.72 ($d_2 - 10$ nm) or 11.83 ($d_2 + 10$ nm) as shown in Fig. 6(a-iv/vi). It is clear that our network is able to identify the optimized thickness that places the spectrum in the right position to achieve minimum difference from the desired color.

It is also worth pointing out the limitation of our network, which is manifested in the design of the red color. The network selects a design with d_2 of 369 nm, which uses the second-order peak at 595 nm, resulting in a ΔE_{2000} of 8.93 [Fig. 6(a-viii)]. Although it outperforms the other two scenarios with ± 10 nm of d_2 [ΔE_{2000} of 20.21 and 19.64, respectively, shown in Fig. 6(a-vii/ix)], it fails to select the first-order peak, which could realize a much better design by using a smaller thickness ($d_2 \sim 160$ nm). This suggests that the network is not flexible enough when a large change of thickness is required. This may be attributed to two factors. The first factor is the uneven color distribution of the training dataset caused by the nonlinear relation among thickness, spectrum, and color. The resonant cavity lengths in the F-P cavity are close to the positive integer times of $\lambda/4n$, where n is the refractive index of the dielectric layer and λ is the light wavelength in free space. The wavelengths of blue (~ 490 nm) and green (~ 550 nm) are shorter than that of the red (~ 610 nm). Higher-order blue and green peaks therefore appear more frequently than red within the same dielectric layer range. Hence, colors near the red region are significantly under-represented. This leads to the network being trained in favor towards optimization of colors in the blue and green regions. The second factor lies in the inherent limitation with tandem network architecture, which suffers from mode collapse [23]. The high-quality first-order red peak was abandoned by the network during the training process to achieve a high overall accuracy for all colors. This results in the red colors being predicted to the higher orders, leading to limited quality. These two factors are believed to contribute to the difficulty of designing colors in the red region, and further improvements in both the network design and training process are required to tackle this challenge.

Finally, we evaluate the computational efficiency of our network by comparing it with an evolutionary algorithm—a popular metaheuristic optimization technique [48]. Compared with a neural network, an evolutionary algorithm performs better in

the design of red color (details of EA can be found in Section 7). This forward design process was able to find the first-order red peak, achieving a ΔE_{2000} of 0.73. The designs of green and blue colors are not as good as our INN with ΔE_{2000} of 1.36 and 1.11, respectively. Moreover, this method demands much more computational resources than INN and is impractical in real applications. For example, the time required to design the painting of “Haystacks, end of Summer” (500×297 , 148,500 pixels) via our INN was 0.17 s. It took the evolutionary algorithm 4.8 h to design the same number of pixels under the same computational environment (see Section 7, Method). This translates to over 100,000 times of savings in time and computational resources for our network. In addition, the overall design accuracy for those 148,500 colors obtained from the network is significantly better (ΔE_{2000} of 0.78) than that designed by EA (ΔE_{2000} of 1.18), further proving the superior performance in both computational efficiency and design accuracy of our network.

6. CONCLUSION

In conclusion, we demonstrate the use of a bidirectional neural network to inversely design the geometric structures of F-P cavity color filters. This work leads to a gamut coverage that is 215% of the sRGB color space. By selecting the uniform CIE 1976-Lab color space over the conventional CIE 1931-XYZ color space as the representation of color, the bidirectional network has shown a superior accuracy for color design with an average ΔE_{2000} value below 1.2 in the testing set. This excellent performance is also verified by comparison with the gradient-free evolutionary algorithm in which our network demonstrates a 100,000 times design efficiency improvement with higher accuracy when designing the famous painting “Haystacks, end of Summer” by Claude Monet. The challenges in designing colors at longer wavelength due to uneven dataset distribution and the continuous gradient descent nature of artificial neural networks are also discussed. This proposed model will contribute to the establishment of standard procedure for future design of nanostructured color filters with deep learning technology.

7. METHOD

Data Processing. Before the neural networks are trained, the datasets are normalized from 0 to 1. By doing this, the effects of unit, magnitude, and dimension of the dataset are waived, which is able to improve the performance and probability of convergence. This process was done by the open-source machine learning library *Scikit-Learn*.

Color Conversion. The color conversions, which include the conversions between different color spaces (e.g., sRGB to CIE 1931-XYZ and CIE 1931-XYZ to CIE 1976-Lab) and color difference (ΔE_{2000}), were performed by the open-source library *Colour-Science*.

Deep Learning. All the deep learning models and training were developed and performed on the open-source deep learning framework *PyTorch*.

Training Hyperparameters. The training hyperparameters are listed as follows. Epochs: 2000; batch size: 64; activation function: ReLU; loss function: mean squared error (MSE); optimizer: Adam; learning rate: 0.001; learning rate scheduler:

MultiStepLR; milestones = [1800,1900] (forward training) and [1,1800,1900] (inverse training); gamma = 0.1. The Kaiming uniform initialization method was adopted in this work to investigate the impact of random seed selection [49].

Loss Functions. In the training of the forward neural network, the loss function is defined as $\text{Loss}_{\text{forward}} = \text{MSE}(\text{Lab}_{\text{predicted}}, \text{Lab}_{\text{truth}})$. In the training of the inverse neural network, a penalty term in the loss function has been introduced for the first 100 epochs to ensure $D_{\text{predicted}}$ is not too far away from the input dimension D_{truth} (i.e., reduce the chance of generating negative values). The penalized loss function is defined as $\text{Loss}_{\text{inverse}} = \text{MSE}(\text{Lab}_{\text{predicted}}, \text{Lab}_{\text{truth}}) + 0.2 \times \text{MSE}(D_{\text{predicted}}, D_{\text{truth}})$. The second term in the equation is a penalty term, which is used to control the inverse output error compared with the ground truth of the geometric parameters. The penalty coefficient λ is set to be 0.2 in our case. After the 100 epochs, the loss function is defined as $\text{Loss}_{\text{inverse}} = \text{MSE}(\text{Lab}_{\text{predicted}}, \text{Lab}_{\text{truth}})$.

Evolutionary Algorithm. The evolutionary algorithm (EA) is one type of gradient-free optimization method, and it is an appealing option for solving this optimization problem. It uses a refined iterative process in which an elite percentage of the individuals are retained through each iteration, allowing the samples to genetically evolve until the best option has been identified. The evolutionary algorithm in this work was realized by open-source EA library **DEAP**. The EA generation and population applied in this work are 100 and 50, respectively. The EA designed geometrical parameters (d_1, d_2, d_3) are (39 nm, 163 nm, 46 nm) for red (102, 0, 0); (33 nm, 138 nm, 50 nm) for green (0, 102, 0); and (21 nm, 417 nm, 49 nm) for blue (0, 0, 102). The ΔE_{2000} are 0.73, 1.36 and 1.11, respectively.

Computational Environment. CPU: Intel Core i9-9900K; GPU: nVIDIA RTX 2070; RAM: 48 GB; OS: Windows 10 Pro; Python version: 3.7. The same computational environment was used for the ANN and EA methods in this work.

Funding. International Exchange Scheme (IECNSFC \170193) between Royal Society (UK) and the National Natural Science Foundation of China (China).

Acknowledgment. All data supporting this study are openly available from the University of Southampton repository at DOI: <https://doi.org/10.5258/SOTON/D1686>.

Disclosures. The authors declare no conflicts of interest.

[†]These authors contributed equally to this paper.

REFERENCES

- N. Dean, "Colouring at the nanoscale," *Nat. Nanotechnol.* **10**, 15–16 (2015).
- J. Hong, E. Chan, T. Chang, T.-C. Fung, B. Hong, C. Kim, J. Ma, Y. Pan, R. Van Lier, S. Wang, B. Wen, and L. Zhou, "Continuous color reflective displays using interferometric absorption," *Optica* **2**, 589–597 (2015).
- L. Shao, X. Zhuo, and J. Wang, "Advanced plasmonic materials for dynamic color display," *Adv. Mater.* **30**, 1704338 (2018).
- Y. Wang, M. Zheng, Q. Ruan, Y. Zhou, Y. Chen, P. Dai, Z. Yang, Z. Lin, Y. Long, Y. Li, N. Liu, C.-W. Qiu, J. K. W. Yang, and H. Duan, "Stepwise-nanocavity-assisted transmissive color filter array microprints," *Research* **2018**, 8109054 (2018).
- Y. Hu, X. Luo, Y. Chen, Q. Liu, X. Li, Y. Wang, N. Liu, and H. Duan, "3D-integrated metasurfaces for full-colour holography," *Light Sci. Appl.* **8**, 86 (2019).
- X. Luo, Y. Hu, X. Li, Y. Jiang, Y. Wang, P. Dai, Q. Liu, Z. Shu, and H. Duan, "Integrated metasurfaces with microprints and helicity-multiplexed holograms for real-time optical encryption," *Adv. Opt. Mater.* **8**, 1902020 (2020).
- Y. Hu, L. Li, Y. Wang, M. Meng, L. Jin, X. Luo, Y. Chen, X. Li, S. Xiao, H. Wang, Y. Luo, C. W. Qiu, and H. Duan, "Trichromatic and tripolarization-channel holography with noninterleaved dielectric metasurface," *Nano Lett.* **20**, 994–1002 (2020).
- K.-T. Lee, J. Y. Lee, S. Seo, and L. J. Guo, "Colored ultrathin hybrid photovoltaics with high quantum efficiency," *Light Sci. Appl.* **3**, e215 (2014).
- M. Song, D. Wang, S. Peana, S. Choudhury, P. Nyga, Z. A. Kudyshev, H. Yu, A. Boltasseva, V. M. Shalaev, and A. V. Kildishev, "Colors with plasmonic nanostructures: a full-spectrum review," *Appl. Phys. Rev.* **6**, 041308 (2019).
- K. Kumar, H. Duan, R. S. Hegde, S. C. W. Koh, J. N. Wei, and J. K. W. Yang, "Printing colour at the optical diffraction limit," *Nat. Nanotechnol.* **7**, 557–561 (2012).
- M. L. Tseng, J. Yang, M. Semmlinger, C. Zhang, P. Nordlander, and N. J. Halas, "Two-dimensional active tuning of an aluminum plasmonic array for full-spectrum response," *Nano Lett.* **17**, 6034–6039 (2017).
- B. Yang, W. Liu, Z. Li, H. Cheng, S. Chen, and J. Tian, "Polarization-sensitive structural colors with hue-and-saturation tuning based on all-dielectric nanopixels," *Adv. Opt. Mater.* **6**, 1701009 (2018).
- V. Flauraud, M. Reyes, R. Paniagua-Domínguez, A. I. Kuznetsov, and J. Brugger, "Silicon nanostructures for bright field full color prints," *ACS Photon.* **4**, 1913–1919 (2017).
- H. Shin, M. F. Yanik, S. Fan, R. Zia, and M. L. Brongersma, "Omnidirectional resonance in a metal-dielectric-metal geometry," *Appl. Phys. Lett.* **84**, 4421–4423 (2004).
- D. Zhao, L. Meng, H. Gong, X. Chen, Y. Chen, M. Yan, Q. Li, and M. Qiu, "Ultra-narrow-band light dissipation by a stack of lamellar silver and alumina," *Appl. Phys. Lett.* **104**, 221107 (2014).
- M. A. Kats, R. Blanchard, P. Genevet, and F. Capasso, "Nanometre optical coatings based on strong interference effects in highly absorbing media," *Nat. Mater.* **12**, 20–24 (2013).
- C. Ji, C. Yang, W. Shen, K.-T. Lee, Y. Zhang, X. Liu, and L. J. Guo, "Decorative near-infrared transmission filters featuring high-efficiency and angular-insensitivity employing 1D photonic crystals," *Nano Res.* **12**, 543–548 (2019).
- Z. Li, S. Butun, and K. Aydin, "Large-area, lithography-free super absorbers and color filters at visible frequencies using ultrathin metallic films," *ACS Photon.* **2**, 183–188 (2015).
- Z. Yang, Y. Chen, Y. Zhou, Y. Wang, P. Dai, X. Zhu, and H. Duan, "Microscopic interference full-color printing using grayscale-patterned Fabry-Perot resonance cavities," *Adv. Opt. Mater.* **5**, 1700029 (2017).
- Z. Yang, C. Ji, D. Liu, and L. J. Guo, "Enhancing the purity of reflective structural colors with ultrathin bilayer media as effective ideal absorbers," *Adv. Opt. Mater.* **7**, 1900739 (2019).
- S. So, T. Badloe, J. Noh, J. Bravo-Abad, and J. Rho, "Deep learning enabled inverse design in nanophotonics," *Nanophotonics* **9**, 1041–1057 (2020).
- R. S. Hegde, "Deep learning: a new tool for photonic nanostructure design," *Nanoscale Adv.* **2**, 1007–1023 (2020).
- P. R. Wiecha, A. Arbouet, C. Girard, and O. L. Muskens, "Deep learning in nano-photonics: inverse design and beyond," *Photon. Res.* **9**, B182–B200 (2021).
- Q. Zhang, H. Yu, M. Barbiero, B. Wang, and M. Gu, "Artificial neural networks enabled by nanophotonics," *Light Sci. Appl.* **8**, 42 (2019).
- Y. Kiarashinejad, S. Abdollahramezani, and A. Adibi, "Deep learning approach based on dimensionality reduction for designing electromagnetic nanostructures," *npj Comput. Mater.* **6**, 12 (2020).
- S. So, J. Mun, and J. Rho, "Simultaneous inverse design of materials and structures via deep learning: demonstration of dipole resonance engineering using core-shell nanoparticles," *ACS Appl. Mater. Interfaces* **11**, 24264–24268 (2019).

27. J. Jiang and J. A. Fan, "Global optimization of dielectric metasurfaces using a physics-driven neural network," *Nano Lett.* **19**, 5366–5372 (2019).
28. Y. Chen, J. Zhu, Y. Xie, N. Feng, and Q. H. Liu, "Smart inverse design of graphene-based photonic metamaterials by an adaptive artificial neural network," *Nanoscale* **11**, 9749–9755 (2019).
29. T. Zhang, J. Wang, Q. Liu, J. Zhou, J. Dai, X. Han, Y. Zhou, and K. Xu, "Efficient spectrum prediction and inverse design for plasmonic waveguide systems based on artificial neural networks," *Photon. Res.* **7**, 368–380 (2019).
30. J. Peurifoy, Y. Shen, L. Jing, Y. Yang, F. Cano-Renteria, B. G. DeLacy, J. D. Joannopoulos, M. Tegmark, and M. Soljačić, "Nanophotonic particle simulation and inverse design using artificial neural networks," *Sci. Adv.* **4**, eaar4206 (2018).
31. I. Malkiel, M. Mrejen, A. Nagler, U. Arieli, L. Wolf, and H. Suchowski, "Plasmonic nanostructure design and characterization via deep learning," *Light Sci. Appl.* **7**, 60 (2018).
32. O. Hemmatyar, S. Abdollahramezani, Y. Kiarashinejad, M. Zandehshahvar, and A. Adibi, "Full color generation with Fano-type resonant HfO₂ nanopillars designed by a deep-learning approach," *Nanoscale* **11**, 21266–21274 (2019).
33. Z. Huang, X. Liu, and J. Zang, "The inverse design of structural color using machine learning," *Nanoscale* **11**, 21748–21758 (2019).
34. J. Baxter, A. Calà Lesina, J.-M. Guay, A. Weck, P. Berini, and L. Ramunno, "Plasmonic colours predicted by deep learning," *Sci. Rep.* **9**, 8074 (2019).
35. L. Gao, X. Li, D. Liu, L. Wang, and Z. Yu, "A bidirectional deep neural network for accurate silicon color design," *Adv. Mater.* **31**, 1905467 (2019).
36. P. R. Wiecha, A. Lecestre, N. Mallet, and G. Larrieu, "Pushing the limits of optical information storage using deep learning," *Nat. Nanotechnol.* **14**, 237–244 (2019).
37. D. Liu, Y. Tan, E. Khoram, and Z. Yu, "Training deep neural networks for the inverse design of nanophotonic structures," *ACS Photon.* **5**, 1365–1369 (2018).
38. L. Xu, M. Rahmani, Y. Ma, D. A. Smirnova, K. Z. Kamali, F. Deng, Y. K. Chiang, L. Huang, H. Zhang, S. Gould, D. N. Neshev, and A. E. Miroshnichenko, "Enhanced light-matter interactions in dielectric nanostructures via machine-learning approach," *Adv. Photon.* **2**, 026003 (2020).
39. W. Ma, F. Cheng, and Y. Liu, "Deep-learning-enabled on-demand design of chiral metamaterials," *ACS Nano* **12**, 6326–6334 (2018).
40. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* **521**, 436–444 (2015).
41. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016).
42. F. W. Billmeyer, *Color Science: Concepts and Methods, Quantitative Data and Formulae* (Wiley, 1983).
43. C. Williams, G. S. D. Gordon, T. D. Wilkinson, and S. E. Bohndiek, "Grayscale-to-color: scalable fabrication of custom multispectral filter arrays," *ACS Photon.* **6**, 3132–3141 (2019).
44. Y. Horie, S. Han, J.-Y. Lee, J. Kim, Y. Kim, A. Arbabi, C. Shin, L. Shi, E. Arbabi, S. M. Kamali, H.-S. Lee, S. Hwang, and A. Faraon, "Visible wavelength color filters using dielectric subwavelength gratings for backside-illuminated CMOS image sensor technologies," *Nano Lett.* **17**, 3159–3164 (2017).
45. Q. Chen, X. Hu, L. Wen, Y. Yu, and D. R. S. Cumming, "Nanophotonic image sensors," *Small* **12**, 4922–4935 (2016).
46. M. Habekost, "Which color differencing equation should be used?" *Int. Circ. Graph. Educ. Res.* **6**, 1–33 (2013).
47. W. S. Mokrzycki and M. Tatol, "Color difference delta E—a survey," *Mach. Graph. Vis.* **20**, 383–411 (2011).
48. A. Slowik and H. Kwasnicka, "Evolutionary algorithms and their applications to engineering problems," *Neural Comput. Appl.* **32**, 12363–12379 (2020).
49. K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: surpassing human-level performance on imagenet classification," in *IEEE International Conference on Computer Vision* (2015), Vol. **2015**, pp. 1026–1034.