Opto-Electronic Science

Original Article

2022, Vol. 1, No. 1

DOI: 10.29026/oes.2022.210012

Benchmarking deep learning-based models on nanophotonic inverse design problems

Taigao Ma[1], Mustafa Tobah[2], Haozhu Wang[3]* and L. Jay Guo[3]*

Photonic inverse design concerns the problem of finding photonic structures with target optical properties. However, traditional methods based on optimization algorithms are time-consuming and computationally expensive. Recently, deep learning-based approaches have been developed to tackle the problem of inverse design efficiently. Although most of these neural network models have demonstrated high accuracy in different inverse design problems, no previous study has examined the potential effects under given constraints in nanomanufacturing. Additionally, the relative strength of different deep learning-based inverse design approaches has not been fully investigated. Here, we benchmark three commonly used deep learning models in inverse design: Tandem networks, Variational Auto-Encoders, and Generative Adversarial Networks. We provide detailed comparisons in terms of their accuracy, diversity, and robustness. We find that tandem networks and Variational Auto-Encoders give the best accuracy, while Generative Adversarial Networks lead to the most diverse predictions. Our findings could serve as a guideline for researchers to select the model that can best suit their design criteria and fabrication considerations. In addition, our code and data are publicly available, which could be used for future inverse design model development and benchmarking.

Keywords: inverse design; photonics; machine learning; neural networks; generative models

Ma TG, Tobah M, Wang HZ, Guo LJ. Benchmarking deep learning-based models on nanophotonic inverse design problems. *Opto-Electron Sci* **1**, 210012 (2022).

Introduction

Nanophotonics has become an important platform for exploring the light-matter interaction[1–5] and wavefront manipulation[6–12], and is critical for realizing the advanced photonic-electronic integrated circuits[13]. Most nanophotonic devices are based on carefully designed nanostructured plasmonic[14] and dielectric[15] materials. These emerging devices have surpassed conventional photonic devices for many applications, such as on-chip coherent light sources[16–17], communication[18–19], information processing[20], and sensing[21–22], to name an important few.

Nanophotonic devices usually have different structures, which can uniquely determine their optical responses and functionality. Researchers can simulate or measure the optical response of a nanophotonic device through the electromagnetic (EM) simulation or experiment. However, it is nontrivial to inverse design the nanostructures from desired optical responses and features. One of the challenges is that different structures can have similar responses, which leads to the one (optical response) -to-many (structures) mapping issue. Usually, inverse design problems are solved by human experts through a time-consuming iterative trial-and-

[1]Department of Physics, The University of Michigan, Ann Arbor, Michigan 48109, USA; [2]Department of Materials Science and Engineering, The University of Michigan, Ann Arbor, Michigan 48109, USA; [3]Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, Michigan 48109, USA.

*Correspondence: HZ Wang, E-mail: hzwang@umich.edu; LJ Guo, E-mail: guo@umich.edu

Received: 29 October 2021; Accepted: 10 December 2021; Published online: 7 January 2022

error approach, which is guided by domain knowledge. For example, to realize a polarization-insensitive optical response, the symmetric structure should be typically considered[23–24]. However, because of the one-to-many mapping issue, we still do not know if the intuitive symmetric structure gives the best performance. Additionally, relying on human expertise alone to design complicated structures with a large number of degrees of freedom (DOF) could result in a slow design process.

On the other hand, optimization-based methods have been widely used in inverse design for a long time[25]. The optimization-based methods are a combination of forward simulations and automatic iterative searches, where in each iteration, the optimization starts from a set of structures and requires the EM simulations to obtain the corresponding optical responses. The difference between the simulated and the target optical response is later used to update the structures with the objective of minimizing this difference. After sufficient iterations, a structure with desired optical responses can be found. Different optimization methods differ from each other in terms of the mechanism for updating the structure, including the local optimization, e.g., Newton's methods[26] and gradient descent[27], and the global optimization, such as simulated annealing[28], adjoint variable algorithms[29], evolutionary algorithms[30], particle-swarm algorithms[31], and Bayesian optimization[32]. A summary and benchmark of optimization methods in the photonic inverse design can be found in ref.[33].

Though proven powerful for a wide range of nanophotonic inverse design problems, optimization-based methods are often target-specific, i.e., the optimization process needs to start from scratch for each new inverse design target. Because EM simulation is performed each iteration during the optimization-based inverse design process, applying such methods for many different inverse design targets is time-consuming or even intractable. For example, when designing photonic nanostructures to reconstruct all colors in a painting[34], one needs to perform the optimization process for potentially thousands of different inverse design problems, which can take an extremely long time.

Recently, deep learning models have been demonstrated as an efficient alternative to the optimization-based methods for nanophotonic inverse design. Rather than target-specific as in optimization-based methods, deep learning models have a strong generalization ability and can learn the mapping between the structural parameters and the optical responses. After being trained on a dataset containing pairs of structural parameters and the corresponding optical responses, deep learning models can accurately predict the structural parameters given a design target within milliseconds, which greatly improves the efficiency of the inverse design process. For example, Liu et al.[35] trained the tandem networks for the inverse design of optical multilayer thin films. Ma et al.[36] applied Variational Auto-Encoders (VAEs) for the inverse design of metamaterial elements, including cross, split-ring, and H-shape nanostructures. Liu et al.[37] used the Generative Adversarial Networks (GANs) to inverse design the nanostructures for customer-defined optical spectra. There have been several excellent reviews on deep learning-based inverse design published recently, including these three commonly used models for inverse designs[38–41].

Although deep learning-based methods have been shown to give accurate predictions efficiently on different nanophotonic inverse design problems, existing works mostly overlook other requirements that are also important for real applications. For example, due to the constraint of existing nanofabrication techniques, structures with high-aspect-ratio or sharp corners can be difficult or even impossible to realize. Therefore, if a diverse set of designs with optical responses close to the target responses can be identified, researchers can choose designs with lower aspect-ratios or smoother shapes that are more amenable to nanofabrication. Thus, whether and to what extent an inverse design method can learn the one-to-many mapping, i.e., come up with a diverse set of designs for a single design target, is a critical feature of practical inverse design methods. Apart from diversity, robustness also plays an important role when considering the real fabrication. If the predicted structures from certain inverse design models violate physical constraints, e.g., the dimension of the designed nanostructure for a metasurface exceeds the size of a unit cell, those structures should never be considered because their optical responses will not be reasonable. In addition, optical responses of fabricated structures may deviate from the desired responses because of the variations in the fabrication process, i.e., the fabrication tolerance. The optical responses corresponding to the predicted structures given by different models may have different sensitivity to such fabrication variations. We believe that in addition to accuracy, both the diversity of the predicted structures and the robustness of their optical

response against predicted structures are important considerations when applying deep learning models to inverse design. Unfortunately, no existing work has systematically considered and compared these two properties.

To bridge the gap of the diversity and robustness for deep learning-based inverse design models, and provide a direct comparison for design accuracy, we benchmark three widely used deep learning models: Tandem networks, VAEs, and GANs, on two representative nanophotonic inverse design problems. Performance metrics including accuracy, diversity, and robustness are quantitatively evaluated on both problems with held-out test datasets. Based on our comparisons, we provide recommendations on how to select from these inverse design models based on different requirements and highlight the important future research directions for developing inverse design models that can be adopted more widely for practical nanophotonic inverse design applications.

Methods

Neural networks (NNs) contain multiple layers of neurons that are connected in series. Each neuron takes in one or multiple inputs from the previous layer, sums up all the inputs based on learnable weights, and passes the outputs through a nonlinear activation as the inputs to the next layer. By stacking multiple layers of neurons together, complex information can be processed by these interconnected neurons, enabling NNs to learn the mapping between inputs and outputs. In terms of the inverse design, the inputs are the optical responses, and the outputs are the designs of structures (i.e., structural parameters). However, using the conventional NNs to inverse design directly will give inaccurate results[42]. Because of the one-to-many mapping issue, there are multiple possible structures for a given target optical response. Minimizing the loss during training (i.e., the difference between the target structure and designed structures, which are usually represented by the Mean Square Error (MSE)) will make it hard for the conventional NNs to converge and lead NNs to output the averaged structures, which usually will not have the desired optical responses. Therefore, special constructions of NNs are required to deal with this one-to-many mapping issue. The following three models are widely known to solve this issue properly and are commonly used in inverse design and, therefore, are examined in this benchmark work. Specifically, tandem networks[43] can learn a one-to-one mapping that accurately maps the given optical response to one of the potential structures. Generative models, including VAEs[44] and GANs[45], leverage the stochastic generation process to directly capture the one-to-many mapping. We classify these three models into two categories based on whether their outputs are deterministic or stochastic (generative models). We use x, y to denote optical responses and structures, respectively, and z to denote the latent variables or random variables used in VAEs and GANs, respectively. Detailed network constructions and training can be found in the supplementary information.

Deterministic models

Tandem networks are the combinations of the Forward Neural Networks (FNNs) and the Inverse Neural Networks (INNs), which are shown in Fig. 1(a). The FNN takes in the structure parameters and outputs the predictions of their corresponding optical responses and is used to approximate the solution of Maxwell's equations. We use the MSE loss to train the FNN:

$$MSE_x^{\mathrm{FNN}} = \frac{1}{N}\sum_i \left(x_i - f_\theta\left(y_i\right)\right)^2 , \tag{1}$$

where y and x are the structures and corresponding optical responses, respectively, $f_\theta(y)$ are the predicted responses of FNN based on the network parameters θ, and N is the number of training samples. Once trained, we can use the FNN to predict the optical responses accurately for given structures. The INN takes in the target optical responses and outputs inverse predictions of possible structures. The idea of tandem networks is to train the FNN first, and then connect the output of INN to this pre-trained FNN and use the forward prediction loss to supervise the learning of INN:

$$MSE_x^{\mathrm{INN}} = \frac{1}{N}\sum_i \left(x_i - f_\theta\left(g_\phi(x_i)\right)\right)^2 , \tag{2}$$

where $g_\phi(x)$ are the predicted structures for given optical responses x based on the INN's parameters ϕ, and $f_\theta\left(g_\phi(x_i)\right)$ is the predicted optical responses given by the pre-trained FNN, which correspond to the predicted structures. Using this two-step training, tandem networks circumvent the one-to-many mapping issue by enforcing the INN to converge to only one possible solution suggested by FNN. Tandem networks have been widely used in a variety of inverse design problems, such as multi-layer transmission spectra[35], silicon structure colors[46], and chiral metamaterials[47].

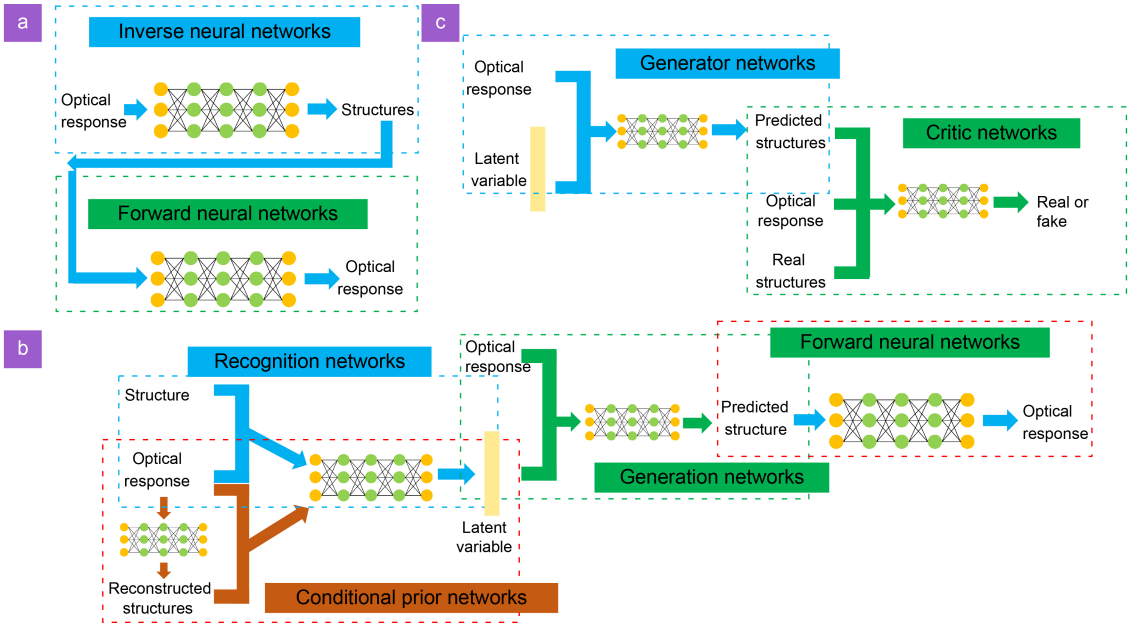

Fig. 1 | The structure of three considered models: (**a**) Tandem networks, (**b**) VAEs, (**c**) GANs. The detailed descriptions of building and training each neural networks can be found in the supplementary information.

Stochastic models

Unlike tandem neural networks that can only map the input to a single output deterministically, both VAEs and GANs are generative models that can stochastically output multiple different predictions given the same input. For VAEs, we consider a specific variant conditional-VAEs (c-VAEs)[44] to inverse predict structures given specific optical responses. There are three networks in VAEs (for the remaining parts, we will use VAEs to refer to c-VAEs for simplicity): the recognition networks, the generation networks, and the conditional prior networks. During training, the recognition networks learn to encode the structures and the optical responses together into the latent variables z, and the generation networks learn to decode the structures from the latent variables z based on the conditional optical responses[48]. The latent variables z follow the normal distribution. Because of the introduction of latent variables z, VAEs can give multiple predictions when decoding from different latent variables. The conditional prior networks provide reconstructions of structures and are useful during the inverse prediction. We find that connecting the pre-trained FNNs to the output of VAEs can improve the accuracy. The overall network structures for VAEs are shown in Fig. 1(b). The loss for training VAEs is:

$$L_{\mathrm{VAE}} = -\frac{1}{N}\sum_{i} KL\left(q_{\varphi}(z_i|x_i,y_i)\,||\,p_{\theta}(z_i|x_i)\right) + MSE_{\mathrm{pred}} + MSE_{\mathrm{recon}} + \alpha * MSE_x\ , \tag{3}$$

where the $KL\left(q_{\phi}(z|x,y)\,||\,p_{\theta}(z|x)\right)$ is the Kullback-Leibler (KL) divergence between the latent distribution $q_{\phi}(z|x,y)$ and prior distribution $p_{\theta}(z|x)$, MSE_{pred} is the prediction loss between the target structures y and the inverse designed structures $\hat{y}$ predicted by VAEs, MSE_{recon} is the reconstruction loss between the target structures and the reconstructed structures y_{recon} given by the conditional prior networks, MSE_x is the forward prediction loss between the target responses x and the predicted responses given by the FNNs, which correspond to the inverse designed structures $\hat{y}$. The α is the weight factor for forward prediction loss. More details can be found in supplementary information.

GANs are another type of generative models. We consider the conditional-GANs (c-GANs)[45] to inverse predict structures given specific optical responses. There are two networks in GANs (for remaining parts, we will use GANs to refer to c-GANs for simplicity): the generator networks that generate structures based on the random variables z and the optical responses, and the critic networks that attempt to distinguish if a structure is from the dataset or from the generator networks. The idea of the GAN is based on the game theory, where the generator networks always learn to generate structures that are distributed as close to the test dataset as possible in order to fool the critic networks, while the critic networks always learn to distinguish the generated structures from real structures. The loss for training GANs is:

$$L_{\mathrm{GAN}} = \frac{1}{N}\sum_i [\log D_\theta(y_i|x_i) + \log(1 - D_\theta(G_\phi(z_i|x_i)))], \tag{4}$$

where $G_\phi(z|x)$ are the predicted structures from the generator networks, $D_\theta(y|x)$ are the scores given by critic networks for the structures from the training dataset, and $D_\theta(G_\phi(z|x))$ are the scores given by critic networks for structures predicted by the generator networks. The ϕ and θ are the parameters of the generator and critic networks, respectively. The random variables z are sampled from the normal distribution. We minimize this loss function when training the generator networks, while maximizing this loss function when training the critic networks.

Both VAEs and GANs are widely used in inverse designing free-form structures, including metamaterials[36], diffractive metagratings[49], and nano-antennas[50]. Detailed descriptions for constructing and training each model can be found in the supplementary information.

Experiments

We formally introduce two inverse design problems as the benchmarking problems to evaluate inverse design models. A set of evaluation metrics regarding the design accuracy, diversity, and robustness to fabrication variations are later described in detail. We report the benchmarking results and summarize the relative performance of tandem networks, VAEs, and GANs at the end of this section. All data and code are publicly available[51].

Inverse design problems

Nanophotonic inverse design problems can be grouped into two categories[40] based on the number of DOF associated with the structure design. On the one hand, when the number of DOF is small, a structure template based on simple building block elements, such as nanodisks and nanobricks, can be used to form the design. A few structural parameters, including height, width, and radius, can be carefully designed to describe the structural elements. Thus, a 1D vector containing the structural parameters is used as the representation for the design. On the other hand, when the number of DOF is large, the nanostructures have free-form geometries and cannot be represented by a small set of structural parameters. Instead, 2D binarized images are used to represent these free-form structures. In terms of the construction of neural networks, we use Multilayer Perceptron[52] (MLP) and Convolutional Neural Networks[53] (CNN) for the vector representation and the image representation, respectively.

To ensure conclusions are generalizable on most nanophotonic inverse design problems, we consider two different inverse design problems from the template design and free-form design categories, respectively.

Template structures: Silicon structure color inverse design

For the inverse design based on template structures, we choose a design task that has been investigated in ref.[46]. As shown in Fig. 2(a), the template structure is a unit cell arranged periodically and consists of four identical and uniformly spaced silicon nanorods. A layer of 70 nm Si_3N_4 is located between the nanorod structures and the bottom silicon substrate layer. This periodic template structure is represented by a vector with four structural parameters (D, H, G, P), where D and H refer to the diameter and height of each nanorod, respectively, G refers to the gap between two nearby nanorods, and P refers to the period of the unit cell. The inverse design target of optical responses is the reflective structural color, which can be described by three-dimension CIE 1931 coordinates (x, y, Y).

For data collection, we use the Rigorous Coupled Wave Analysis (RCWA)[54] to simulate 8411 samples. Structure parameters of (D, H, G, P) are uniformly and randomly sampled in the ranges of (80, 160) nm, (30, 200) nm, (160, 320) nm, and (300, 700) nm, respectively. A physical constraint $G + D < P$ is used during the sampling process to make sure all four nanorods are within one unit cell. The reflection spectrum is computed between the (380, 780) nm wavelength range with a 5 nm step size, which is then converted to CIE 1931 coordinates (x, y, Y). Detailed information can be found in ref.[46]. In all three models, we use 6,000 samples for training, 1,000 samples for validation, and the rest 1,411 samples for testing. The obtained structural colors in the training dataset are plotted in the CIE chromatic diagram in Fig. 2(b).

Free-form structures: Silicon transmission filter inverse design

For the inverse design based on free-form structures, we choose a design task that we investigated before, where we used NNs to inverse design metasurface filters[55]. As shown in Fig. 2(c), the structure is a 2D periodic pattern on the silicon substrate. The pattern is made of

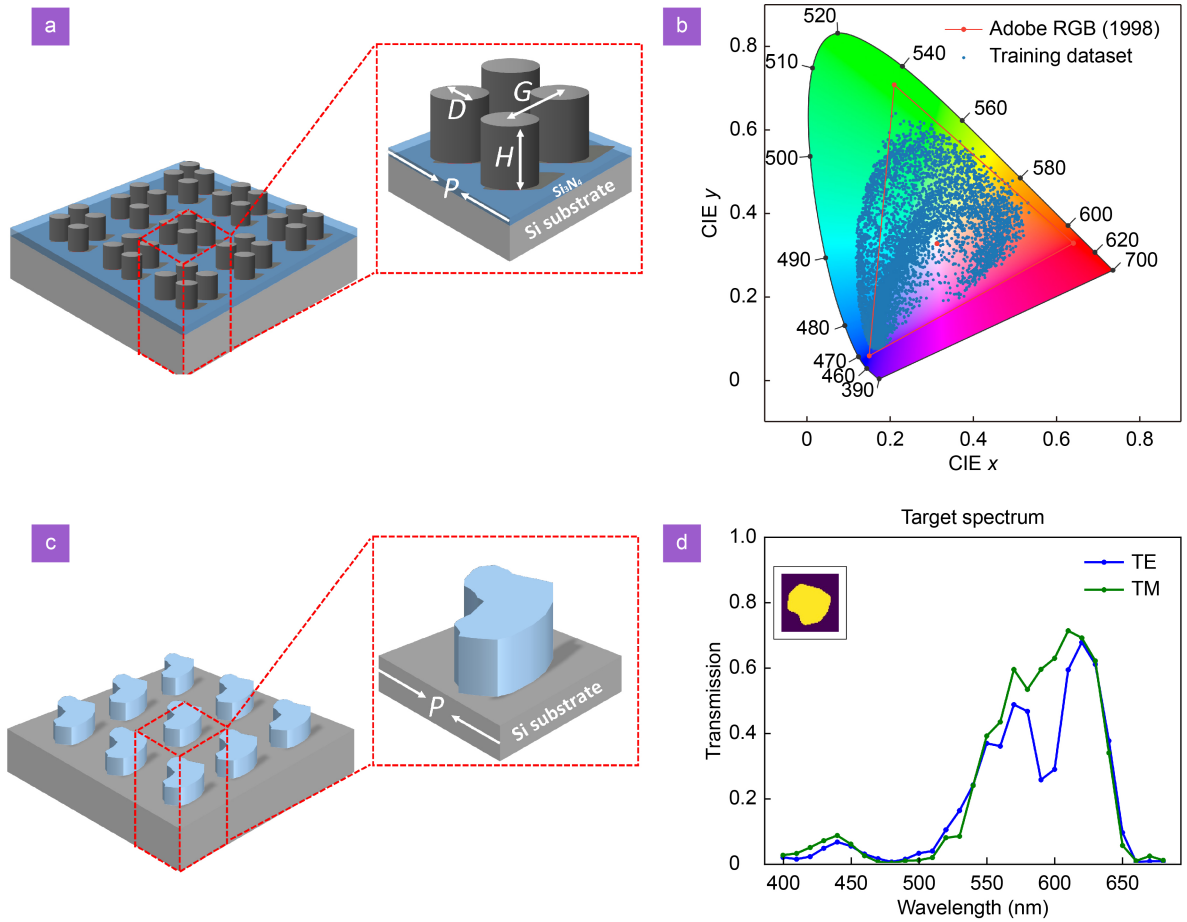

Fig. 2 | (**a**) The template structure for silicon structural color inverse design. Four structural parameters (*D*, *H*, *G*, *P*) are shown in the inset figure. (**b**) The obtained structural colors in the training dataset embedded in the CIE 1931 chromatic diagram, which cover a wide color gamut. (**c**) The free-form structure for silicon transmission filter inverse design. The inset is the period of the structure and the 2D pattern treated as an image. (**d**) One example of the TE/TM transmission spectra in the training dataset. The inset shows the 2D free-form structure with period 283 nm, where the yellow and black regions are the dielectric material and air, respectively.

polycrystalline silicon (Poly-Si) with a fixed thickness of 500 nm and is represented by a 2D 64 × 64 pixeled binarized image. We also include a scalar parameter ranging from 200 nm to 400 nm as the period of the unit cell. For the inverse design target of optical responses, we consider the transmission spectra for both TE and TM polarized normal incident light. The spectrum target is within the visible band and ranges from 400 nm to 680 nm, with a 10 nm step size.

Again, we use RCWA to simulate 63,757 samples. The free-form 2D patterns are randomly generated, and the period is uniformly and randomly sampled between (200, 400) nm. During the image pattern generation, to make sure the corresponding structures satisfy the fabrication limitation, all sharp features are smoothed to fulfil the minimum curvature with a 20 nm radius. Detailed descriptions can be found in the supplementary information. In all three models, we use 53,750 samples for training, 5000 samples for validation, and 5007 samples for testing. Figure 2(d) gives one example of the free-form structure as well as the corresponding transmission spectra.

Evaluation metrics

As stated earlier, practical inverse design problems often involve considerations beyond accuracy. However, no previous research work has systematically studied the properties of deep learning-based inverse design models for practical applications. To bridge this gap, we propose a set of evaluation metrics based on practical considerations that are generalizable for extensive inverse design problems:

Accuracy: The design accuracy is most widely considered in previous research works, and it quantifies how close we can design a structure that achieves the target response. We use both MAE and Root Mean Square

Error (RMSE) to measure the accuracy. The MAE is expressed as:

$$MAE = \frac{1}{N}\sum_i |x_i - \hat{x}_i| \,, \tag{5}$$

where x and $\hat{x}$ refer to the target optical responses and inverse designed optical response, respectively. The RMSE is expressed as:

$$RMSE = \sqrt{\frac{1}{N}\sum_i (x_i - \hat{x}_i)^2} \,. \tag{6}$$

RMSE is more sensitive to large difference than MAE because of the squared error. Thus, if a design method can output accurate designs on average but predicts poor designs occasionally, its MAE will be low while the RMSE could be high. Therefore, including MAE and RMSE for the accuracy evaluation allows us to investigate if an inverse design model exhibits such behavior.

Diversity: This evaluates if the examined model can give multiple predictions for one specific task; and if so, how diverse these predicted structures are distributed in the structure space. As mentioned above, an inverse design model that can output a diverse set of structure designs given a design target is highly desired. This is because the diverse designs could facilitate the fabrication process by providing more candidate designs for researchers to choose from, which can be beneficial, especially when the designs involve shapes that are challenging for nanofabrication. In addition, an inverse design model that can capture the one-to-many mapping may provide physical insights for the inverse design problems.

Robustness: Two different aspects of robustness are considered. First, we examine the robustness of neural network models by checking if the predicted structures satisfy the constraint of the physical system. In addition, we also examine the optical performance drop caused by fabrication variations as the second type of robustness. This is because during nanofabrication, the fabricated structures may slightly deviate from the expected designs due to variations in the fabrication process, leading to different optical responses.

To compare the performance, we report each model's best performance on these two inverse design problems found through an extensive hyperparameter search (details in the supplementary information).

Performance comparisons

Template structure: Silicon structure color inverse design

For the aspect of accuracy, we compare the predicted colors given by the inverse designed structures with the target colors. To calculate the predicted colors, we first use RCWA to simulate the reflection spectra for the inverse-designed structures, then covert the spectra into the CIE color coordinates. When measuring the accuracy, in addition to the MAE and RMSE, we also calculate the R^2 scores for each CIE coordinate (x, y, Y). In order to improve the statistics confidence, we train each model five times with five different random seeds. All five models use the same hyperparameters, which are found through the hyperparameters search. We report the average accuracy results of all five models in Table 1, where the standard deviations are also included. More details can be found in the supplementary information. To visualize the difference between the designed color and the target color, we randomly select and show five examples of structural color inverse design given by tandem networks, VAEs, and GANs in Fig. 3(a). Additional results on the color pixel generation to reproduce a painting with the inverse designed structures are also included in Fig. 3(c–h). Based on the high R^2 scores and small MAE and RMSE values, as well as the accurate color inverse predictions, we can see that all these three models can give accurate results of color inverse design, although tandem networks give slightly more accurate results than the VAE and the GAN.

However, tandem networks can only give one prediction for a specific color task, which could lead to a potential negative impact on fabrication (will be discussed later). Since the VAEs and GANs introduce extra latent variables or random variables, every time they will give different structure predictions[44,45]. By inverse predicting the same color task multiple times and choosing the best

Table 1 | Table of performance comparisons for the silicon structure color inverse design problem. Best results are given by bold type. All R^2 scores, MAE, RMSE, and robustness are averaged in five models, which are trained from different random seeds. Their standard deviations are also given.

Models	$R^2(x)$	$R^2(y)$	$R^2(Y)$	MAE	RMSE	Fault rate	Robustness
Tandem networks	**0.998±0.0004**	**0.997±0.0004**	**0.996±0.009**	**0.0043±0.0002**	**0.0070±0.0004**	19/1411 (1.35%)	0.0611±0.0004
VAEs	0.992±0.001	0.990±0.001	0.992±0.0004	0.0074±0.0002	0.0112±0.0003	**0/1411 (0.00%)**	**0.0520±0.0011**
GANs	0.991±0.003	0.986±0.004	0.982±0.008	0.0069±0.0014	0.0138±0.0024	3/1411 (0.21%)	0.0613±0.0027

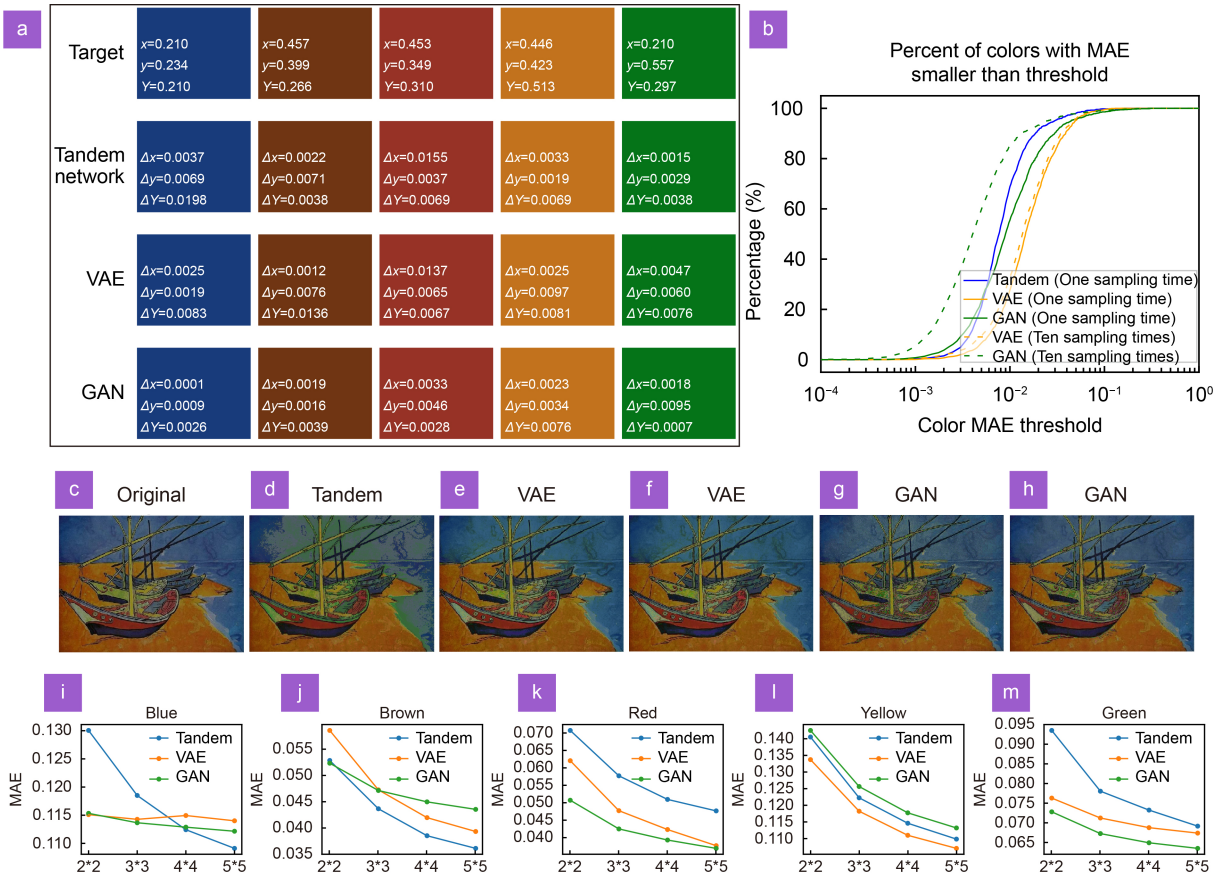

Fig. 3 | (**a**) Five randomly selected examples of color inverse design (blue, brown, red, yellow, and green). The first row is the target color, where the inset numbers are the target CIE (*x*, *y*, *Y*) coordinates. The second, third, and forth row corresponds to the predicted structural color by tandem networks, VAE and GAN, respectively, where the inset numbers are the absolute difference of each CIE coordinate. (**b**) The percent of predicted color tasks that have MAE is smaller than a given threshold. The solid lines show the results when only sampling once for each model, while the dashed lines show results when sampling ten times and picking the most accurate one. As we increase the predicting times, the accuracy of generative models (VAEs and GANs) improved. (**c**–**h**) Comparison of one specific application of structural color inverse design: reproducing a paint. (c) The original image of the Vincent van Gogh's painting: Fishing Boats on the Beach at Saintes Maries-de-la-Mer. (d) The image reconstructed by the predictions of Tandem networks. (e) The image reconstructed by the predictions of the VAE. (f) The image reconstructed by the predictions of VAE when sampling ten times. (g) The image reconstructed by the predictions of GAN. (h) The image reconstructed by the predictions of GAN when sampling ten times. (**i**–**m**) The comparison of three models' robustness with respect to the size of the array for five colors: (i) Blue, (j) Brown, (k) Red, (l) Yellow, (m) Green. To calculate the color, we are not considering the structure to be infinitely periodic anymore. Instead, we are simulating the color within a limited region that only contains the 2 by 2, 3 by 3, 4 by 4, and 5 by 5 array of unit cells, respectively. Fishing Boats on the Beach at Saintes-Maries-de-la-Mer" is reproduced with the permission of the Van Gogh Museum, Amsterdam (Vincent van Gogh Foundation).

structure that gives the most accurate color, the accuracy of inverse design for VAEs and GANs can be further improved. In Fig. 3(b), we calculate how many color inverse design tasks have MAE smaller than a specific MAE threshold and show the tendency as the sampling times of inverse prediction change from once to ten times. We can see that when prediction is carried out for ten sampling times, the accuracy of the GAN and VAE is improved, giving a higher percent of tasks for a specific MAE threshold. The accuracy of GAN is improved more than the improvement of VAE, which is related to the diversity of each neural network and will be discussed later. In addition, we do want to mention that inverse predicting multiple times costs extra time since it requires more simulations for validating the predicted colors and picking up the best structure.

For the diversity of the generated candidate structures, we compare how diverse the distributions of the predicted structures are for each neural network model. Specifically, we start from the inverse design of the green color with the CIE coordinates $(x, y, Y) = (0.2917, 0.5711, 0.4720)$. The original green color is shown in

Fig. 4(c). For each model, we inverse predict 1000 times using this specific color target as the input, which gives 1000 predicted structures. To visualize and compare the distribution, we first calculate the frequency histogram of each predicted structure parameter, then divide each histogram by its respective maximum value to ensure that the peak density for each histogram is one. We show this normalized density of each structure parameter (D, H, G, P) in Fig. 4(a). Since the tandem networks are deterministic models, these 1,000 predicted structures are the same, i.e., there is no diversity at all. Both VAEs and GANs can give diverse structure distributions, but with different levels of diversity. For the VAEs, since the recognition networks learn to map structures into a single-modal normal distribution, they can only output a narrow and single-peaked structural distribution. In comparison, there is no such limitation for GANs, therefore, they can capture a multi-modal distribution caused by the intrinsic one-to-many mapping, where the learned structure distributions of height, period, and diameter exhibit multiple peaks. The overlapped distributions of structural parameters show that both the tandem networks and the VAEs methods only learn one specific mode in the multi-modal distribution learned by GANs. This diverse distribution in structure space clearly reveals the intrinsic one-to-many mapping feature, which is expected in physics.

To validate the color accuracy of 1000 predicted structures in such diverse distributions, we simulate the colors of these structures using RCWA and show all predicted colors in the 3d (x, y, Y) space in Fig. 4(b). For further comparison, we randomly show 40 colors predicted by the VAE and GAN in Fig. 4(e, f). The color predicted by the tandem networks is also shown in Fig. 4(d). Although the predicted structures are broadly distributed, their corresponding colors are close to the target color (an illustration of the one-to-many mapping), and their color differences cannot be distinguished by human eyes. This diverse distribution is highly desirable in practice since more broadly distributed structure spaces provide more choices during fabrication. Specifically, structures with greater gaps or greater diameters are easier to fabricate, allowing researchers to pick the structures that are more suitable for fabrication from these 1000 predicted structures. Therefore, when diversity is of high design

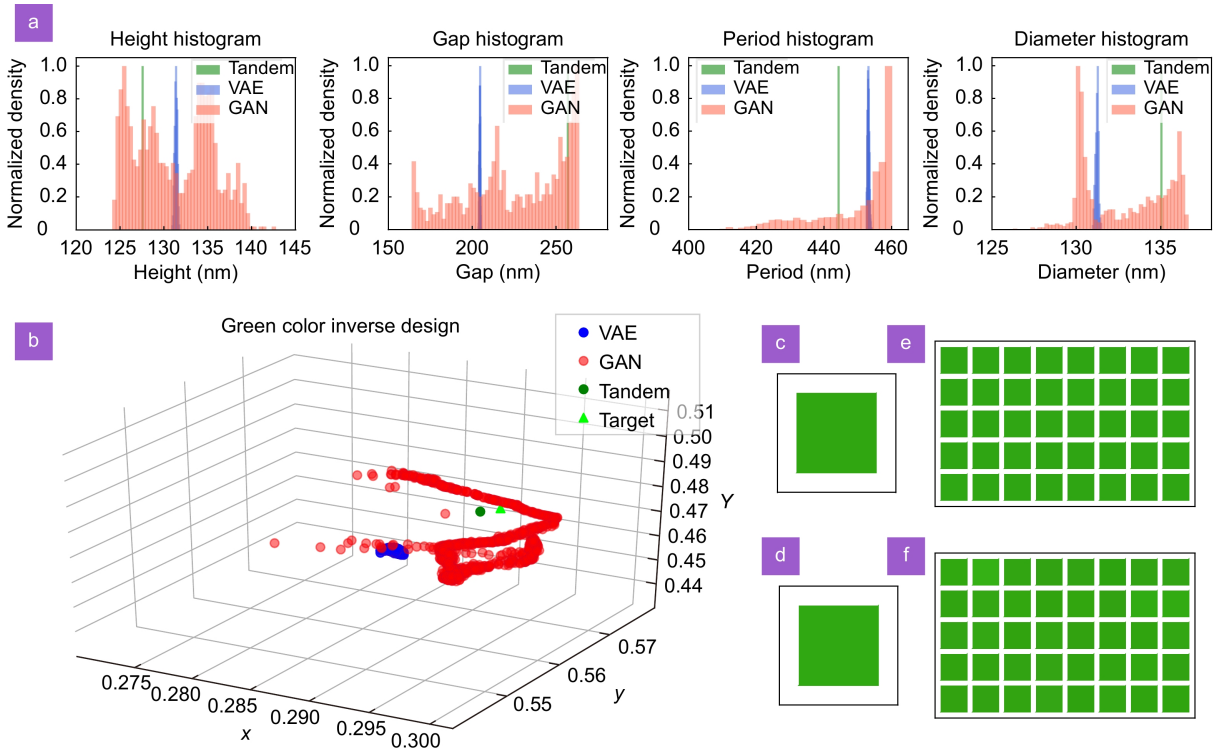

Fig. 4 | (**a**) The normalized density distribution of 1000 inverse designed structure parameters for the green color (**c**) with the CIE coordinates (x, y, Y) = (0.2917, 0.5711, 0.4720). (**b**) The 3-dimensional color distribution related to 1,000 inverse designed structures. We can see all these predicted structures give a fairly accurate green color. (c) The target green color with coordinates (x, y, Y) = (0.2917, 0.5711, 0.4720). (**d**) The color corresponding to the structure predicted by tandem networks. (**e**) The randomly selected 40 different colors corresponding to the structures predicted by the VAE. (**f**) The randomly selected 40 different colors corresponding to the structures predicted by the GAN.

priority, the GAN will be more preferred in inverse design. More examples with yellow and brown colors can be found in supplementary information, and both give similar conclusions.

We want to emphasize the relationship between accuracy and diversity. In Fig. 4(b), each neural network model exhibits different distribution behaviors in the color space, which originates from the different distributions behaviors in the structure spaces. Tandem networks only give one predicted color that is close to the target color. The colors given by the VAE are located within a narrow color space close to the target color, while the GAN's are more diversely spread out in the color space, surrounding the target color. Therefore, if we only predict once for the GAN, it is possible that the inverse designed structure gives the color with a large color difference from the target color. By predicting multiple times and picking the most accurate one, we can minimize this randomness and improve the accuracy of the GAN. Similar procedures are also applicable to the VAE, but its accuracy may not be improved too much because the structure distributions are localized, leading to the localized color distribution.

In terms of robustness, there are three aspects to examine. First, we examine if the generated structures satisfy the constraints of physical systems. We need to make sure that all predicted structure parameters are positive and satisfy another physical constraint of $G + D < P$, meaning that the sum of the gap and the diameter of nanorods should be smaller than the period of the unit cell. Any structure that does not satisfy these two constraints is treated as a *faulty design*. For a given color design target, we run each model ten times, which gives ten predicted structures. When all ten structures are *faulty designs*, this design task is treated as a *faulted task*. We calculate the number of *fault tasks* in the test dataset and summarize the *fault rate* for each model in Table 1. Another example of analyzing the robustness of the image reconstruction in Fig. 3(c–h) is shown in Fig. S10. We can see that the chance that tandem networks fail to give a prediction is higher than VAEs and GANs, which could limit its applications when these failed tasks are necessary. Secondly, we examine how generated structures are susceptible to fabrication variations. This is done by adding a +5/ –5 nm perturbation to structure parameters and measuring the shifts of CIE coordinates. We randomly select and inverse predict 100 color targets in the test dataset and calculate the color related to the perturbated structures. We use the MAE between this perturbated color and the target color to represent the fabrication robustness, i.e., smaller MAE corresponds to higher robustness. Again, we average the robustness from five different models and show the results in the last column in Table 1. We can see VAE gives slightly higher robustness than the other two models. But overall, all these three models give similar robustness in terms of the fabrication variation. This result aligns with our expectation because the loss functions of all three inverse design models do not include components that promote robustness with respect to fabrication variation.

All of the colors in the training dataset are obtained based on the infinite periodic array of unit cells, which cannot be used in many actual applications, e.g., reproduce a paint. Therefore, we evaluate the third robustness, which is to examine how accurate the predicted colors are when only a finite size of the array of the unit cell are used for one color pixel[34]. Here we consider that a color pixel is made up of an array with a finite number of unit cells, with array size to be 2 by 2, 3 by 3, 4 by 4, and 5 by 5. Specifically, we calculate and compare the robustness of these five colors in Fig. 3(a) as an example. For each color task, we inverse predict twenty times and choose the best structure that gives the most accurate color. In order to calculate the predicted color related to different sizes of the array of unit cells, because the considered simulation region is no longer periodic, we change the periodic boundary conditions to perfect matching layers and use the Finite-Difference Time-Domain (FDTD) to simulate the reflection spectrum. Detailed descriptions can be found in the supplementary information. We calculate the MAE between the predicted color and target color and show the relationship with respect to the size of the unit cell array in Fig. 3(i–m). As we expected, when we increase the size of the unit cells array, the color difference with respect to the target color decreases. Overall, again generative models, including both VAEs and GANs, are more robust than tandem networks when using a finite array size to reconstruct one color pixel. This is because generative models can give multiple structure predictions, which is possible to provide more robust structures.

Free-form structure: Silicon transmission filter inverse design

In terms of accuracy, we compare the MAE and RMSE between the simulated spectra related to the inverse designed structures and the target spectra. Because the pixel values of predicted 2D image patterns are not exactly zero or one, we binarize the predicted image patterns by setting the binarization threshold to be 0.5. The

corresponding transmission spectra are simulated using RCWA based on the binarized 2D image pattens. Again, in order to improve the statistics confidence, we train each model three times with three different random seeds. All three models use the same hyperparameters, which are found through the hyperparameters search. We report the average accuracy results over all three models in Table 2, where the standard deviations are also included. Figure 5 also gives two examples for transmission spectrum inverse design, where the inset (lower) shows the inverse designed 2D structure pattern. By comparison, we can see that if researchers care more about the accuracy, they can refer to tandem networks or VAEs.

In terms of the diversity, we compare how diverse the distributions of the predicted 2D patterns are for each model. To quantify the diversity of free-form structures, we introduce a quantity to describe the irregularity of 2D patterns, which is defined as:

$$irr = \frac{(\max(Dis) - \min(Dis))}{\text{mean}(Dis)}, \tag{7}$$

where Dis is the distance between the extracted edges and the center of the 2D pattern. We give several examples of the 2D image patterns with different irr in the supplementary information. $irr = 0$ means a perfect circle pattern and a greater irr means a more nonuniform pattern. By examining the distribution of irr, we can reveal the distribution of the predicted structures. Some other evaluation methods for irregularity can also be used.

As an example, we start from the inverse design of a randomly chosen target spectrum in Fig. 6. Similarly, for each model, we inverse predict 1000 times using this specific spectrum target as the input, which gives 1,000 predicted structures. We show the normalized density of irregularity distribution in Fig. 6(a–c). Again, we observe that tandem networks only give one structure prediction, while the VAE tends to give a single-peak distribution,

Table 2 | Table of performance comparison for the silicon free-form transmission spectrum inverse design. The best results are given by bold type. All MAE, RMSE, and robustness are averaged in three models, which are trained from different random seeds. Their standard deviations are also given.

Model	MAE	RMSE	Robustness
Tandem networks	0.0322±0.0027	0.0517±0.0046	0.0656±0.0015
VAEs	**0.0277±0.0002**	**0.0444±0.0003**	**0.0617±0.0004**
GANs	0.0508±0.0049	0.0829±0.0074	0.0705±0.0037

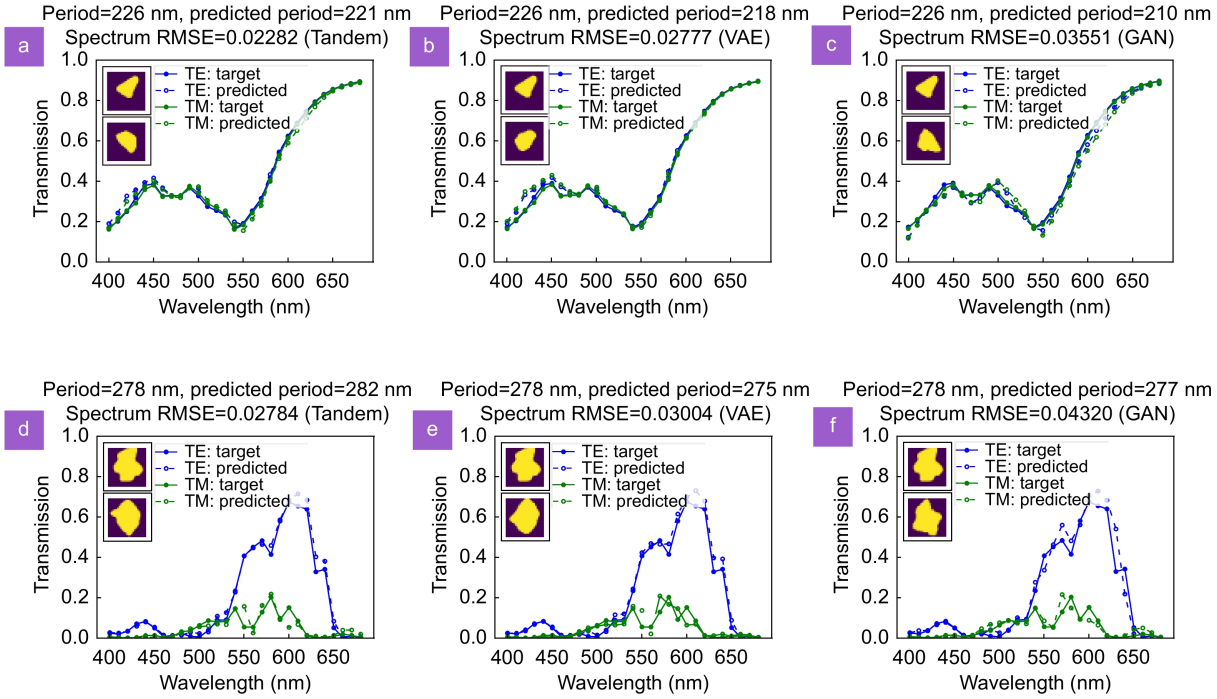

Fig. 5 | Two randomly selected examples of transmission spectrum inverse design for the tandem networks (**a**)(**d**), VAEs (**b**)(**e**), and GANs (**c**)(**f**). The inset shows the original structure (upper) in the test dataset and the inverse predicted structure (lower) by each model.

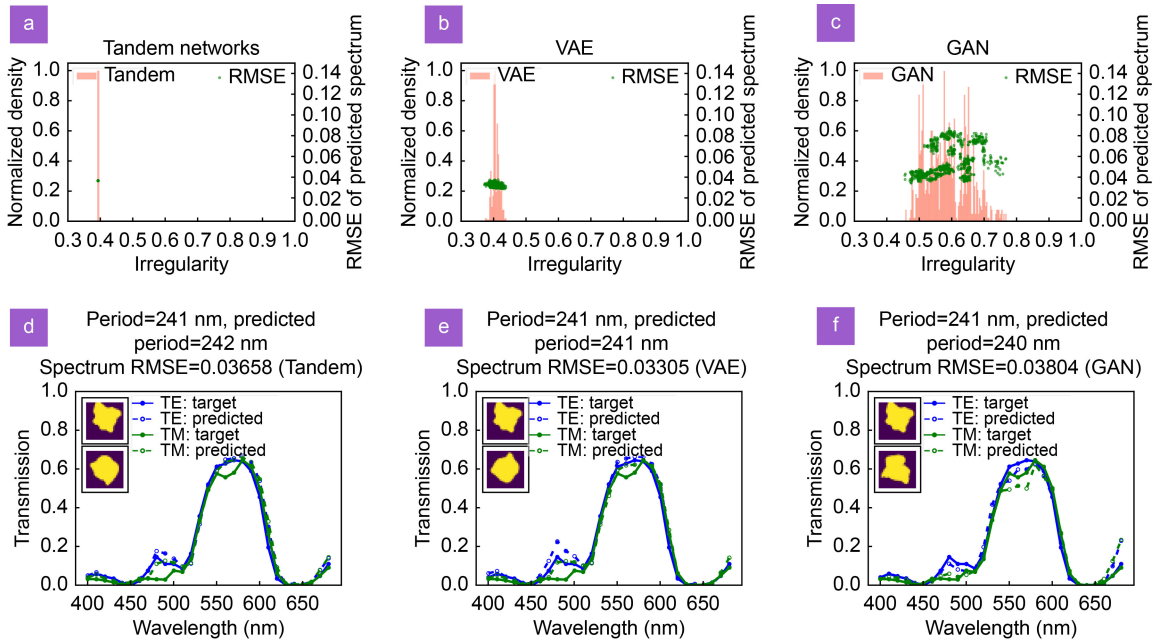

Fig. 6 | Comparisons of the diversity for three models. (**a**–**c**) The red bar shows the distribution of the normalized density of irregularity for 1000 inverse predicted structures by tandem networks, VAEs and GANs, while the green points are the scatter plot of spectrum RMSE V.S. the irregularity. According to the distribution of irregularity, we can see that tandem networks only give one structure prediction, where the VAE gives limited diversity, and the GAN gives a multi-modal structure distribution that covers a wide region. (**d**–**f**) A randomly chosen inverse designed structure predicted by tandem networks, VAEs and GANs, as well as its corresponding transmission spectra. The inset shows the original structure (upper) in test dataset and the predicted structure (lower) given by each model.

and the GAN gives a multi-modal distribution. To better demonstrate that these predicted structures give accurate spectra, we simulate their transmission spectra using the RCWA. In Fig. 6(a–c), we show the MAE of the 1,000 spectra v.s. the irregularity of 1,000 predicted structures. One specific example of the inverse designed structure as well as its transmission spectra is also shown in Fig. 6(d–f). We can see that although the predicted structures are distributed in a wide irregularity range, their spectra are close to the target spectrum. In this case, a more diverse set of structure predictions can benefit the fabrication since a smaller irregularity means a more uniform shape, which leads to easier fabrication. Therefore, researchers can always pick the best structure that can facilitate the fabrication while still giving an accurate spectrum. In this case, the GAN would be more preferred. Another example of diversity in the spectrum inverse design can be found in the supplementary information, which gives similar conclusions.

In terms of robustness, we only investigate the robustness against fabrication variation, since all generated patterns are images, and they do not need to satisfy any physical constraint similar to the color inverse design task (any negative image pixel can be attributed as 0 during binarization). We measure the fabrication robustness by testing how much the spectrum will shift under a small perturbation of the inverse designed structures, mimicking the fabrication variations induced by nanofabrication tolerance. This is done by shrinking, expanding, or smoothing the shapes of predicted structures by a small factor. More details can be found in the supplementary information. We randomly pick 100 spectrum targets from the test dataset and use the MAE of the perturbated spectra with respect to the target spectra to represent the fabrication robustness. Again, we average the robustness from three different models and show the results in the last column in Table 2. We can see that VAE gives slightly higher robustness than the other two models. But overall, all these three models give similar robustness measurements, which aligns with the observation in the template structure inverse design task.

Results and discussion

For all evaluating metrics including accuracy, diversity, and robustness, we give a qualitative comparison of all three models in Table 3, where a greater number of stars correspond to a better performance in each evaluation metric. We find that tandem networks and VAEs give higher accuracy than GANs. However, tandem networks can fail when predicting some tasks, which can be

problematic if this specific task is important because there is no way to find another structure to replace this failed structure. Generative models like VAEs and GANs can solve this problem and give multiple predictions by introducing random variables. GANs give a better diversity than the tandem networks and VAE as well as demonstrate multi-modal outputs for the inverse-designed structures, providing designers with more options to select the best structure. By running predictions multiple times, it is possible to find structures that give more accurate results, thus increasing the accuracy of the GAN. We do not observe a significant difference in robustness among the three studied models, but VAEs give a slightly better result. We need to emphasize that during training, there is no loss function terms or training data that incorporates the fabrication variations. Therefore, the observation that all three models perform similarly in terms of robustness is not surprising.

Table 3 | Conclusion of performance measure for all three neural networks. The number of stars is proportional to the performance.

Models	Accuracy	Diversity	Robustness
Tandem	☆☆☆	☆	☆☆
VAEs	☆☆☆	☆☆	☆☆☆
GANs	☆☆	☆☆☆	☆☆

Although we only consider two specific inverse design problems, these neural networks models and introduced evaluation metrics are applicable for many other nanophotonic inverse design problems with different structures and materials, including the multilayer thin films[35], plasmonic nanostructures[56], and metasurfaces[37], etc, where their structures can be described either by a vector or an image when processed by appropriate neural networks. Therefore, our conclusions are generalizable to a wide range of nanophotonic inverse design problems.

Conclusions

In conclusion, we benchmark the performance of three deep learning-based methods that are commonly used in the current deep learning-based inverse design: tandem networks, VAEs, and GANs. To compare their performance and give guidance to researchers and engineers, each model is evaluated in terms of accuracy, diversity, and robustness, where the last two aspects are seldomly explored in the current domain of deep learning-based inverse design. Detailed comparisons and discussions are included. We hope our work can provide insights for researchers and engineers to correctly select their target model that best fits their specific needs. For example, if researchers want the predicted structures to give the most accurate optical responses, then they can choose tandem networks or VAEs. If they want to have multiple structures for easier fabrication, GANs or VAEs will be preferred.

All three models show similar performance on robustness, although VAEs give slightly better performance. Fabrication robustness is very important for real application and should be considered when dealing with nanofabrications. Additional model development beyond these studied models is necessary to incorporate fabrication robustness as a learning objective. For example, by re-parametrizing the structures[57], or building suitable datasets and incorporating the fabrication variation into loss functions[58], it is possible for neural networks to learn these properties and output predicted structures that are robust to fabrication variations.

We also want to mention that the current machine learning models can only work well for in-distribution inverse design, where the target optical responses should follow a similar distribution of the training dataset. Otherwise, the NNs may give erroneous predictions. This is because NNs can only accurately interpolate within the training dataset, while the extrapolation capability beyond the training distribution is limited. For inverse design problems that may require a high degree of extrapolation, forward search approaches based on reinforcement learning[59] or conventional optimization-based methods should be used. Hybrid methods that combine neural networks with physics-driven solvers can also be used for solving the extrapolation issue[60,61].

References

1. Shen YZ, Friend CS, Jiang Y, Jakubczyk D, Swiatkiewicz J et al. Nanophotonics: interactions, materials, and applications. *J Phys Chem B* **104**, 7577–7587 (2000).
2. Pu MB, Guo YH, Li X, Ma XL, Luo XG. Revisitation of extraordinary young's interference: from catenary optical fields to spin –orbit interaction in metasurfaces. *ACS Photonics* **5**, 3198–3204 (2018).
3. Gan XT, Mak KF, Gao YD, You YM, Hatami F et al. Strong enhancement of light–matter interaction in graphene coupled to a photonic crystal nanocavity. *Nano Lett* **12**, 5626–5631 (2012).
4. de Leon NP, Shields BJ, Yu CL, Englund DE, Akimov AV et al. Tailoring light-matter interaction with a nanoscale Plasmon resonator. *Phys Rev Lett* **108**, 226803 (2012).
5. Baranov DG, Wersäll M, Cuadra J, Antosiewicz TJ, Shegai T. Novel nanostructures and materials for strong light–matter interactions. *Acs Photonics* **5**, 24–42 (2018).

6. Yu NF, Capasso F. Flat optics with designer metasurfaces. *Nat Mater* **13**, 139–150 (2014).
7. Yu NF, Genevet P, Kats MA, Aieta F, Tetienne JP et al. Light propagation with phase discontinuities: generalized laws of reflection and refraction. *Science* **334**, 333–337 (2011).
8. Huang YJ, Luo J, Pu MB, Guo YH, Zhao ZY et al. Catenary electromagnetics for ultra‐broadband lightweight absorbers and large‐scale flat antennas. *Adv Sci* **6**, 1801691 (2019).
9. Li X, Chen LW, Li Y, Zhang XH, Pu MB et al. Multicolor 3D meta-holography by broadband plasmonic modulation. *Sci Adv* **2**, e1601102 (2016).
10. Zheng GX, Mühlenbernd H, Kenney M, Li GX, Zentgraf T et al. Metasurface holograms reaching 80% efficiency. *Nat Nanotechnol* **10**, 308–312 (2015).
11. Staude I, Miroshnichenko AE, Decker M, Fofang NT, Liu S et al. Tailoring directional scattering through magnetic and electric resonances in subwavelength silicon nanodisks. *ACS Nano* **7**, 7824–7832 (2013).
12. Lin DM, Fan PY, Hasman E, Brongersma ML. Dielectric gradient metasurface optical elements. *Science* **345**, 298–302 (2014).
13. Nagarajan R, Joyner CH, Schneider RP, Bostak JS, Butrie T et al. Large-scale photonic integrated circuits. *IEEE J Sel Top Quant Electron* **11**, 50–65 (2005).
14. Maier SA. Metamaterials and imaging with surface Plasmon polaritons. In Maier SA. Plasmonics: Fundamentals and Applications. 193–200 (Springer, 2007); http://doi.org/10.1007/0-387-37825-1_11.
15. Decker M, Staude I, Falkner M, Dominguez J, Neshev DN et al. High‐efficiency dielectric Huygens' surfaces. *Adv Opt Mater* **3**, 813–820 (2015).
16. Stern B, Ji XC, Okawachi Y, Gaeta AL, Lipson M. Battery-operated integrated frequency comb generator. *Nature* **562**, 401–405 (2018).
17. Sun J, Timurdogan E, Yaacobi A, Hosseini ES, Watts MR. Large-scale nanophotonic phased array. *Nature* **493**, 195–199 (2013).
18. Cheng QX, Bahadori M, Glick M, Rumley S, Bergman K. Recent advances in optical technologies for data centers: a review. *Optica* **5**, 1354–1370 (2018).
19. Thomson D, Zilkie A, Bowers JE, Komljenovic T, Reed GT et al. Roadmap on silicon photonics. *J Opt* **18**, 073003 (2016).
20. Walmsley I. Photonic quantum technologies. *Proceedings of SPIE* **11844**, 11844OF (2021).
21. Tittl A, Leitis A, Liu MK, Yesilkoy F, Choi DY et al. Imaging-based molecular barcoding with pixelated dielectric metasurfaces. *Science* **360**, 1105–1109 (2018).
22. Chen LW, Yin YM, Li Y, Hong MH. Multifunctional inverse sensing by spatial distribution characterization of scattering photons. *Opto-Electron Adv* **2**, 190019 (2019).
23. Nguyen TT, Lim S. Wide incidence angle-insensitive metamaterial absorber for both TE and TM polarization using eight-circular-sector. *Sci Rep* **7**, 3204 (2017).
24. Kim I, So S, Rana AS, Mehmood MQ, Rho J. Thermally robust ring-shaped chromium perfect absorber of visible light. *Nanophotonics* **7**, 1827–1833 (2018).
25. Campbell SD, Sell D, Jenkins RP, Whiting EB, Fan JA et al. Review of numerical optimization techniques for meta-device design [Invited]. *Opt Mater Express* **9**, 1842–1863 (2019).
26. Hansen E. Interval forms of Newtons method. *Computing* **20**, 153–163 (1978).
27. Ruder S. An overview of gradient descent optimization algorithms. arXiv: 1609.04747 (2017).
28. Kim WJ, O'Brien J. Optimization of a two-dimensional photonic-crystal waveguide branch by simulated annealing and the finite-element method. *J Opt Soc Am B* **21**, 289–295 (2004).
29. Lalau-Keraly CM, Bhargava S, Miller OD, Yablonovitch E. Adjoint shape optimization applied to electromagnetic design. *Opt Express* **21**, 21693–21701 (2013).
30. Storn R, Price K. Differential evolution – A simple and efficient heuristic for global optimization over continuous spaces. *J Glob Optim* **11**, 341–359 (1997).
31. Poli R, Kennedy J, Blackwell T. Particle swarm optimization. *Swarm Intell* **1**, 33–57 (2007).
32. Snoek J, Larochelle H, Adams RP. Practical Bayesian optimization of machine learning algorithms. In *Proceedings of the 25th International Conference on Neural Information Processing Systems* 2951–2959 (Curran Associates Inc. , 2012).
33. Schneider PI, Santiago XG, Soltwisch V, Hammerschmidt M, Burger S et al. Benchmarking five global optimization approaches for nano-optical shape optimization and parameter reconstruction. *ACS Photonics* **6**, 2726–2733 (2019).
34. Yang WH, Xiao SM, Song QH, Liu YL, Wu YK et al. All-dielectric metasurface for high-performance structural color. *Nat Commun* **11**, 1864 (2020).
35. Liu DJ, Tan YX, Khoram E, Yu ZF. Training deep neural networks for the inverse design of nanophotonic structures. *ACS Photonics* **5**, 1365–1369 (2018).
36. Ma W, Cheng F, Xu YH, Wen QL, Liu YM. Probabilistic representation and inverse design of metamaterials based on a deep generative model with semi‐supervised learning strategy. *Adv Mater* **31**, 1901111 (2019).
37. Liu ZC, Zhu DY, Rodrigues SP, Lee KT, Cai WS. Generative model for the inverse design of metasurfaces. *Nano Lett* **18**, 6570–6576 (2018).
38. Wiecha PR, Arbouet A, Girard C, Muskens OL. Deep learning in nano-photonics: inverse design and beyond. *Photonics Res* **9**, B182–B200 (2021).
39. Khatib O, Ren SM, Malof J, Padilla WJ. Deep learning the electromagnetic properties of metamaterials—a comprehensive review. *Adv Funct Mater* **31**, 2101748 (2021).
40. Jiang JQ, Chen MK, Fan JA. Deep neural networks for the evaluation and design of photonic devices. *Nat Rev Mater* **6**, 679–700 (2021).
41. Ma W, Liu ZC, Kudyshev ZA, Boltasseva A, Cai WS et al. Deep learning for the design of photonic structures. *Nat Photonics* **15**, 77–90 (2021).
42. Jordan MI. Constrained supervised learning. *J Math Psychol* **36**, 396–425 (1992).
43. Jordan MI, Rumelhart DE. Forward models: supervised learning with a distal teacher. *Cogn Sci* **16**, 307–354 (1992).
44. Sohn K, Yan XC, Lee H. Learning structured output representation using deep conditional generative models. In *Proceedings of the 28th International Conference on Neural Information Processing Systems* 3483–3491 (MIT Press, 2015).
45. Mirza M, Osindero S. Conditional generative adversarial nets. arXiv: 1411.1784 (2014).
46. Gao L, Li XZ, Liu DJ, Wang LH, Yu ZF. A bidirectional deep neural network for accurate silicon color design. *Adv Mater* **31**, 1905467 (2019).

47. Ma W, Cheng F, Liu YM. Deep-learning-enabled on-demand design of chiral metamaterials. *ACS Nano* **12**, 6326–6334 (2018).
48. Kingma DP, Welling M. Auto-encoding variational Bayes. arXiv: 1312.6114 (2014).
49. Jiang JQ, Sell D, Hoyer S, Hickey J, Yang JJ et al. Free-form diffractive metagrating design based on generative adversarial networks. *ACS Nano* **13**, 8872–8878 (2019).
50. So S, Rho J. Designing nanophotonic structures using conditional deep convolutional generative adversarial networks. *Nanophotonics* **8**, 1255–1261 (2019).
51. https://github.com/taigaoma1997/benchmark_in_de.git.
52. Pal SK, Mitra S. Multilayer perceptron, fuzzy sets, and classification. *IEEE Trans Neural Netw* **3**, 683–697 (1992).
53. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* **60**, 84–90 (2017).
54. Hugonin JP, Lalanne P. RETICOLO software for grating analysis. arXiv: 2101.00901 (2021).
55. Han X, Fan ZY, Liu ZY, Li C, Guo LJ. Inverse design of metasurface optical filters using deep neural network with high degrees of freedom. *InfoMat* **3**, 432–442 (2021).
56. Malkiel I, Mrejen M, Nagler A, Arieli U, Wolf L et al. Plasmonic nanostructure design and characterization via Deep Learning. *Light:Sci Appl* **7**, 60 (2018).
57. Chen MK, Jiang JQ, Fan JA. Design space reparameterization enforces hard geometric constraints in inverse-designed nanophotonic devices. *ACS Photonics* **7**, 3141–3151 (2020).
58. Sell D, Yang JJ, Doshay S, Yang R, Fan JA. Large-angle, multifunctional metagratings based on freeform multimode geometries. *Nano Lett* **17**, 3752–3757 (2017).
59. Wang HZ, Zheng ZY, Ji CG, Guo LJ. Automated multi-layer optical design via deep reinforcement learning. *Mach Learn:Sci Technol* **2**, 025013 (2021).
60. Jiang JQ, Fan JA. Global optimization of dielectric metasurfaces using a physics-driven neural network. *Nano Lett* **19**, 5366–5372 (2019).
61. Jiang JQ, Fan JA. Simulator-based training of generative neural networks for the inverse design of metasurfaces. *Nanophotonics* **9**, 1059–1069 (2019).

Acknowledgements

We are grateful for financial support from NSF Data Science Supplement of SNM program (CMMI-1635636)

Competing interests

The authors declare no competing financial interests.

Supplementary information

Supplementary information for this paper is available at https://doi.org/10.29026/oes.2022.210012