Data-Driven Design of Thin-Film Optical Systems using Deep Active Learning

YOUNGJOON HONG[1] AND DAVID P. NICHOLLS[2,*]

[1]*Department of Mathematics, Sungkyunkwan University, Suwon 440-746, Republic of Korea*
[2]*Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, IL 60607, USA*
[*]*davidn@uic.edu*

Abstract: A deep learning aided optimization algorithm for the design of flat thin-film multilayer optical systems is developed. The authors introduce a deep generative neural network, based on a variational autoencoder, to perform the optimization of photonic devices. This algorithm allows one to find a near-optimal solution to the inverse design problem of creating an anti-reflective grating, a fundamental problem in material science. As a proof of concept, the authors demonstrate the method's capabilities for designing an anti-reflective flat thin-film stack consisting of multiple material types. We designed and constructed a dielectric stack on silicon that exhibits an average reflection of 1.52 %, which is lower than other recently published experiments in the engineering and physics literature. In addition to its superior performance, the computational cost of our algorithm based on the deep generative model is much lower than traditional nonlinear optimization algorithms. These results demonstrate that advanced concepts in deep learning can drive the capabilities of inverse design algorithms for photonics. In addition, the authors develop an accurate regression model using deep active learning to predict the total reflectivity for a given optical system. The surrogate model of the governing partial differential equations can then be broadly used in the design of optical systems and to rapidly evaluate their behavior.

1. Introduction

Multilayered flat thin-film diffraction gratings are essential optical elements in nanoplasmonic and photonic devices as they modulate light intensity and spectral composition in such systems [1]. In practice it is often important to determine the best multilayer design among a wide choice of dielectrics and metals of varying thicknesses to achieve a desired reflection and transmission spectrum. This problem has a long and distinguished history, beginning at least with Baumeister's characterization of optimal coating design as an optimization problem in 1958 [2]. Numerous inverse algorithms have been constructed to find such optimal designs for use in efficient photonic devices [3]. For instance, a multitude of conventional optimization methods perform well at this particular task including genetic algorithms [4, 5], topology optimization [6, 7], and the needle optimization technique [8, 9] which, e.g., have delivered the remarkable results on design of anti–reflective coatings found in [10, 11]. However, it is well known that this problem is highly nonlinear and non-convex, featuring numerous suboptimal local minima which make it difficult to find the global optimum [3].

In the past decade the artificial intelligence technique of machine learning, in particular deep learning, has revolutionized many fields of computational science, and nano-phototonics is no exception. There are already a multitude of survey articles on the topic, e.g., [12], and a staggering number of techniques have been brought to bear on this task. Of particular note, we point out the work of Peurifoy *et al* [13] who trained an artificial neural network (ANN) and then optimized this using conventional techniques, and Liu *et al* [14] who considered a tandem approach which begins the same, but then trains a second ANN to produce desired reflectivity values. We also

mention the work of Unni *et al* [15] on convolutional mixture density networks and the paper of Barry *et al* [16] demonstrating how nature itself identifies optimal structures for its own purposes. Recently, deep generative models including (conditional) variational autoencoders have been introduced to optimize the shape of two-dimensional unit cells giving a two-dimensional binary image [17–20]. As deep neural networks deliver such superior performance in computer vision, one expects that shape optimization algorithms can also be improved. We close by mentioning the deep learning approach that has been considered to find optimal designs using a ResNet generative model [21].

In this paper, we develop an alternative approach based on deep learning strategies which is both fast and efficient, and which can readily be extended to more general thin-film structures featuring, e.g., corrugated layer interfaces. In this way we view our new algorithm as particularly promising in the design of new metamaterials [22]. More specifically, in this paper we develop a novel and effective inverse design algorithm with the aid of deep learning to identify m-layer flat thin-film stacks composed of materials with varying refractive indices and thicknesses. In this scheme we begin by constructing a structure/response database using a rapid and accurate classical Fresnel solver [23]. Then, we make use of a generative deep neural network (DNN), a conditional variational autoencoder (CVAE), to obtain a nearly optimal design of the optical system. The goal of this CVAE is to minimize the average reflection of the structure over a range of incident illumination angles ($0 \le \theta \le \pi/3$) and wavelengths (400 nm $\le \lambda \le$ 1600 nm), which we denote

$$O(\mathbf{p}) = \frac{3}{\pi}\frac{1}{1200}\int_0^{\pi/3}\int_{400}^{1600} \mathcal{R}(\lambda,\theta,\mathbf{p})\, d\lambda\, d\theta, \tag{1}$$

where $\mathbf{p}$ is the design vector and $\mathcal{R}$ is specified in (4). We also propose a deep active learning algorithm to effectively search for the optimal solution. While our method effectively and efficiently delivers a configuration with quite a small reflectivity, we cannot guarantee that it is the "best" design with the smallest possible reflectivity.

2. Methods

In this section we present the governing equations for a thin-film multilayer optical system and an accurate method for numerically approximating its solution based upon the classical Fresnel equations [23]. We then present a novel approach to coupling this to not only a Deep Learning (DL) algorithm (we consider a CVAE), but also an active learning methodology.

2.1. Governing equations

We consider a thin-film multilayer optical system consisting of flat layers of varying materials with rather arbitrary thicknesses. The variety of dielectrics and thicknesses leads to a considerable practical design challenge of finding a "best" grating structure. To be more precise, dielectrics occupy each of the m-many domains

$$S^1 := \{y > g_1\}; \quad S^m := \{y < g_{m-1}\}; \quad \text{and} \quad S^j := \{g_j < y < g_{j-1}\}, \quad 2 \le j \le m-1;$$

where the (flat) interface locations are given by $\{y = g_j\}_{j=1}^m$. This structure is illuminated by incident radiation of frequency ω and angle θ in the uppermost layer, S^1, of the form

$$\underline{v}^{inc} = e^{-i\omega t + i\alpha x - i\beta y}, \quad \alpha = k_0 \sin(\theta), \quad \beta = k_0 \cos(\theta), \quad k_0 = \omega/c_0,$$

and c_0 is the speed of light in the vacuum. Factoring out time-dependence of the form $\exp(-i\omega t)$ we define the (reduced) scattered fields

$$v_j = v_j(x,y) \quad \text{in } S^j \text{ for } 1 \le j \le m,$$

and seek α-quasiperiodic, outgoing (upward/downward propagating) solutions of the following system of Helmholtz equations [23]

$$\Delta v_j + k_j^2 v_j = 0, \qquad \text{in } S^j,\ 1 \le j \le m, \tag{2a}$$
$$v_1 - v_2 = -e^{i\beta g_1} e^{i\alpha x}, \qquad \text{at } y = g_1, \tag{2b}$$
$$\partial_y v_1 - \tau_1^2 \partial_y v_2 = (i\beta) e^{i\beta g_1} e^{i\alpha x}, \qquad \text{at } y = g_1, \tag{2c}$$
$$v_j - v_{j+1} = 0, \qquad \text{at } y = g_j,\ 2 \le j \le m, \tag{2d}$$
$$\partial_y v_j - \tau_j^2 \partial_y v_{j+1} = 0, \qquad \text{at } y = g_j,\ 2 \le j \le m, \tag{2e}$$

where $k_j = n_j k_0 = n_j(\omega/c_0)$ is the wavenumber in layer j (with refractive index n_j) and

$$\tau_j = \begin{cases} 1, & \text{for Transverse Electric (TE) polarization,} \\ (k_j/k_{j+1})^2, & \text{for Transverse Magnetic (TM) polarization,} \end{cases}$$

for $1 \le j \le m-1$.

2.2. Numerical method: The Fresnel solver

It is well known [23] that the most general solutions of (2a) are

$$v_j(x,y) = \left(U_j e^{i\beta^j y} + D_j e^{-i\beta^j y} \right) e^{i\alpha x}, \quad D_1 = U_m = 0,$$

where $\beta^j := \sqrt{(k^j)^2 - \alpha^2}$, $\mathrm{Im}(\beta^j) \ge 0$, and the upward/downward propagating wave conditions enforce $D_1 = U_m = 0$, respectively. The remaining constants $\{U_1, D_2, U_2, \ldots, D_{m-1}, U_{m-1}, D_m\}$ are determined from the boundary conditions (2b)–(2e). This results in a linear system of equations to be solved

$$A\vec{v} = \vec{r}, \quad \vec{v} := (U_1, D_2, U_2, \ldots, D_{m-1}, U_{m-1}, D_m)^T, \quad \vec{r} := e^{i\beta g_1}(-1, (i\beta), 0, \ldots, 0)^T, \tag{3}$$

and A is pentadiagonal with readily derived entries [23]. We denote the direct solution of the Fresnel equations, $A\vec{v} = \vec{r}$, as the *Fresnel Solver* which, we point out, can be accomplished in *linear* (in m) time via the Thomas Algorithm [24]. Of the many quantities that one can compute from this solution, the one of paramount importance to our current study is the reflectivity

$$\mathcal{R} = |U_1|^2 . \tag{4}$$

We point out that the solution of these equations is the order-zero approximation produced by our recently developed High-Order Perturbation of Surfaces (HOPS) algorithm [25] implemented with a Transformed Field Expansions approach. As this methodology is designed for structures with *corrugated* interfaces the full power of this algorithm is not necessary in the current context, however, in a forthcoming publication we will describe the extension of our algorithm to the case of corrugated interfaces which will require the full HOPS methodology.

2.3. A variational autoencoder

Generative models in combination with neural networks, such as variational autoencoders (VAEs), are used to learn complex distributions underlying datasets, see e.g. [26, 27]. Described simply, VAEs consist of an encoder, a latent space, and a decoder (see Figure 1). VAEs presume that a given training sample is generated from a latent representation, which is then sampled by a decoder (with a prior Gaussian distribution). As a generative neural network, VAEs have been successfully utilized in various domains from image generation and natural language processing to anomaly detection and clustering tasks (see, e.g., [26] and references therein).

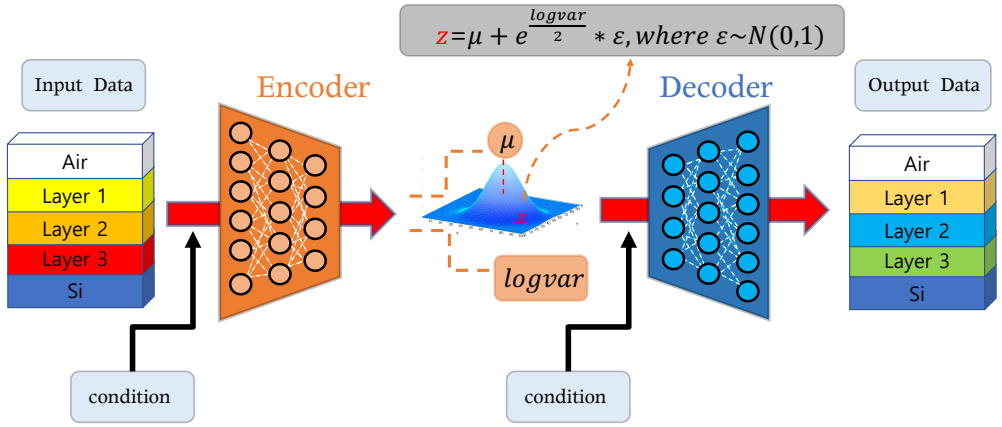


Fig. 1. Depiction of a varitational autoencoder.

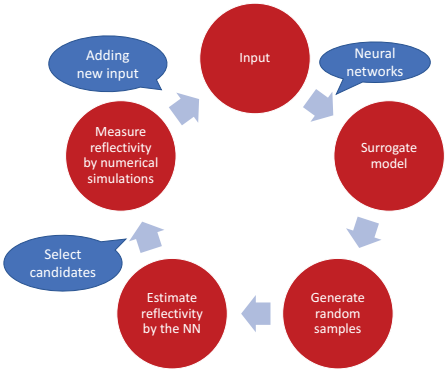


Fig. 2. Depiction of an active learning implementation.

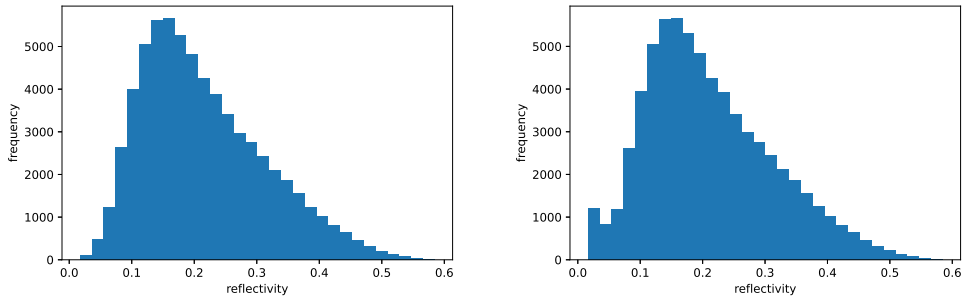
VAEs are regularized autoencoders which also feature an encoder and decoder. The encoder maps high-dimensional data to low-dimensional latent vectors that capture principal features, then the decoder maps the latent vector back to the high-dimensional space. While there are many applications of autoencoders (such as dimensionality reduction, anomaly detection, and noise removal) they are not generally adequate as generative models [28]. Indeed, once the autoencoder is trained there is no opportunity to produce any new content with both encoder and decoder. By contrast, a VAE regularizes the encoding distribution to ensure that its latent space has good properties to generate a new dataset. More precisely, the encoder in the VAE maps input data points not to the latent space but to the *distribution* of the latent space. Then, the encoder produces the mean and covariance matrix values that are a function of the input data. The decoder exploits the latent space distribution as an input to generate distributions of data. In the VAE, the loss function consists of reconstruction loss and regularization loss. The reconstruction loss is identical to that used by autoencoders, while the regularization loss is the Kullback-Leibler (KL) divergence between the Gaussian distribution from the encoder and a

standard Gaussian distribution [29].

For the optical system we consider here the input vector x consists of a collection of refractive indices and layer thicknesses. The VAE architecture aims to learn a stochastic mapping between the observed data space x and a latent space z which can be interpreted as a directed model with a joint distribution $p_\theta(x, z)$ such that $p_\theta(x, z) = p_\theta(x|z)p_\theta(z)$, where θ is a learnable parameter and $p_\theta(z)$ is the prior distribution of the latent variable. The conditioned distribution $p_\theta(x|z)$ can be parameterized by a decoder but the distribution is generally intractable. To resolve this issue a VAE introduces another deep neural network (encoder) to map x back to the latent vector z by approximating the posterior distribution (see Figure 1). With the encoder and decoder networks, the likelihood function for the training has a tractable representation and can be derived by the evidence lower bound (ELBO):

$$\text{Loss} = \mathbb{E}_{q(z|x)}[\log(p(x|z))] - \mathbb{D}_{KL}[q(z|x)||p(z)],$$

where $\mathbb{D}_{KL}$ stands for the KL divergence. In this paper, we employ a conditional variational autoencoder (CVAE), which is an extension of the VAE suitable for incorporating a control on a specified condition [27]. The CVAE is believed to insert label information in the latent space to force a deterministic, constrained, representation of the learned data. In contrast to a VAE, a CVAE has control on the data generation process so, by changing the conditional variable (which refers to the reflectivity in our model), inputs of an optical system for a specified reflectivity can be generated.



(a) Histogram of the original dataset.

(b) Histogram of the CVAE augmented dataset.

Fig. 3. Histogram of the (a.) original and (b.) CVAE augmented data sets.

2.4. Active learning

Active learning is a machine learning strategy which interactively queries a user to generate new data points with desirable properties. These queries are usually in the form of unlabeled data and, to improve the underlying machine learning model, it is crucial for the system to propose records for interactive labelling effectively. Active learning is often called optimal experimental design (OED) in the engineering literature (see, e.g., [30, 31]) as it finds the most uncertain points and adds them into the training set in an iterative way [32]. In this paper we develop surrogate models using deep learning techniques (implemented by a multilayer perceptron (MLP)) to identify efficient device designs. However, in order to improve these designs, additional training of the surrogate model becomes increasingly expensive.

We aim to obtain a better model by adding minimal additional training data, and, for this purpose, we adopt an active learning strategy that selects only training points which enhance the accuracy of the MLP surrogate model. To produce such training samples, the CVAE is used to

generate an optimal system design which hopefully gives small reflectivity (see Figure 2). To be more precise, let L be the number of training set data points initially generated by the Fresnel solver, and, with these L points, we train our CVAE model to generate K many optical systems (which hopefully possess small reflectivity) by imposing the smallness of the reflectivity as the condition on the CVAE. Few (say P many) of these K optical systems generated by the CVAE produce the reflectivity expected by the condition enforced since the randomly generated training set does not contain sufficient data close to our target. However, they are sufficiently close that they merit further investigation with our (relatively expensive) Fresnel solver and we add them to our training set; this is then repeated M times. This algorithm is depicted in Figure 2. The initial reflectivity histogram of randomly generated data (the "initial training set") is plotted in Figure 3a while the augmented dataset generated by our CVAE and active learning process is shown in Figure 3b. The second figure is, naturally, nearly identical to the first figure as all of the original data is retained while only "desirable" (i.e., small reflectivity) configurations are added.

3. Numerical experiments

As we described above, the CVAE architecture consists of an encoder and decoder network. The encoder network can be described as

$$\text{INPUT} \to \text{FC} \to \text{sigmoid} \to \text{FC} \to \text{BN} \to \text{sigmoid} \to \text{FC} \to \text{sigmoid} \to \text{FC} \to \text{OUTPUT},$$

while the decoder architecture is given by

$$\text{INPUT} \to \text{FC} \to \text{RELU} \to \text{FC} \to \text{BN} \to \text{RELU} \to \text{FC} \to \text{RELU} \to \text{FC} \to \text{OUTPUT},$$

where "FC" stands for a fully connected layer and "BN" denotes batch normalization. In more detail:

- The input of the encoder consists of refractive indices and thickness for each layer with reflectivity as a condition.
- The dimension of the output of the encoder is 4×2, which is the same as the dimension of the latent space.
- In the encoder the output sizes of the FCs are $384, 384, 128$, and 4, respectively.
- The input of the decoder consists of samples of the four latent variables (for which we have four means and four variances).
- In the decoder the output sizes of the FCs are $128, 384, 384$, and 6, respectively.
- The BN has momentum = 0.1 and stability constant $\varepsilon = 10^{-5}$.

We applied our algorithm to the design of an anti-reflection (AR) coating for a silicon solar cell consisting of three layers of dielectrics [3]. This thin-film stack was designed to minimize the average reflection at an air-silicon interface over the incident illumination angle range $[0, \pi/3]$ and wavelength range $[400, 1100]$ nm in TM polarization as in (1). As a benchmark, we compared our results with those from [3] which provides a guaranteed global optimum solution using a parallel branch-and-bound method. Their algorithm required extensive searching through the full design space and utilized more than two weeks of CPU time to solve for the global optimum.

To be consistent with [3] we generated an $L = 50,000$ member training set from our Fresnel solver whose refractive indices and layer thicknesses were randomly selected from the intervals $[1.09, 2.60]$ and $[5, 200]$ nm, respectively. Here we supplemented with active learning using $K = 5000$ at each of $M = 3$ iterations resulting in $P = 2000$ additional datapoints. A histogram

of the resulting reflectivities are given in Figure 4 which shows that the optimized devices generated from our CVAE and deep active learning algorithms have average reflectivities from approximately 1.5% to 3%, which is quite a small range of values compared to a randomly generated set. A fraction of the suggested devices were near the global optimum, and the best device had an efficiency of 1.52 %. This best device had layer thicknesses

$$151.722\text{nm}, 91.258\text{nm}, 61.637\text{nm},$$

and refractive indices

$$1.187, 1.771, 2.717,$$

from top layer to bottom. The total computing time that our CVAE and deep active learning

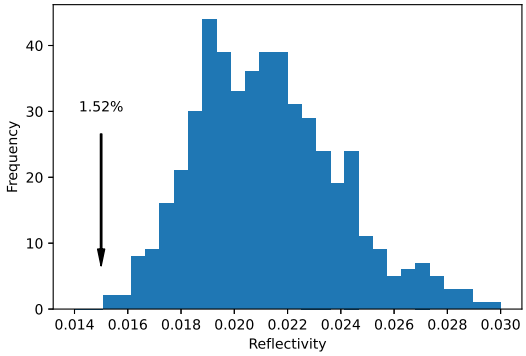


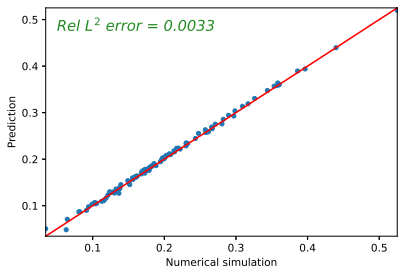
Fig. 4. Histogram of reflectivities generated by our CVAE and deep active learning algorithm.

algorithm required for training was less than 30 minutes with a single GPU (NVIDA RTX-3090). All of the devices sampled from our algorithm were near the global optimum, showing the ability of the generative network to produce a narrow distribution of devices. Another feature of such biased data generation is the construction an accurate regression model using the MLP. Figures 5a–5d show that the MLP model trained by the augmented data set (generated by the active learning algorithm) gives a better result than the model trained by the original (normally distributed) dataset. The reflectivity map of the best device (with efficiency 1.52 %) is depicted in Figure 6a with the full set of incidence angles, and in Figure 6b for three choices of the angle. We remark that the data-driven design under consideration can be generalized to an arbitrary number of layers, and the machine learning procedure becomes more effective since the computational cost is very expensive in this case.

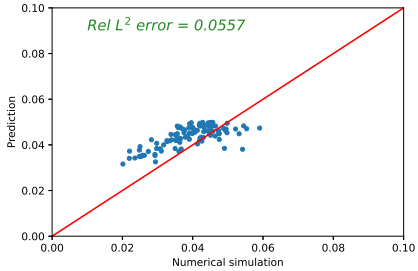
4. Conclusions

In this paper we introduced a deep learning aided optimization algorithm for thin-film multilayer optical systems. We constructed a deep generative neural network, based on a variational autoencoder, to perform optimization of photonic devices. The incorporation of the variational autoencoder helps to improve our search for the optimal grating design. Benchmark calculations of our algorithm for the problem of designing anti-reflection coatings show that the generative model is effective in searching for global optima, is computationally efficient, and outperforms a number of alternative design algorithms.

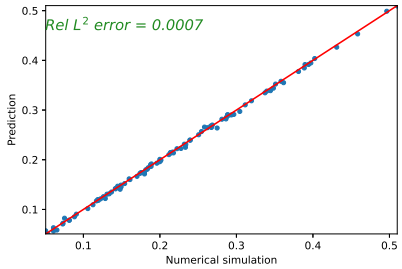
Funding. The work of Y. H. was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2021R1A2C1093579). D.P.N. gratefully acknowledges support from the National Science Foundation through grant No. DMS–2111283.



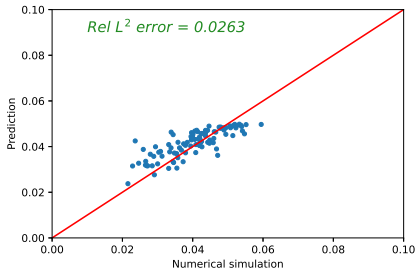
(a) Regression using deep learning with the original dataset. This plot shows reflectivities which range from 2 % to 50 %.



(b) Regression using deep learning with the original dataset. This plot shows only reflectivities which range from 2 % to 7 %.

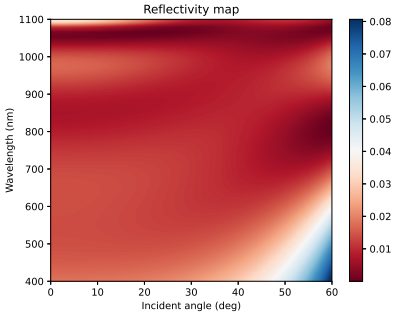


(c) Regression using *deep active learning* with the augmented dataset. This plot shows reflectivities which range from 2 % to 50 %.

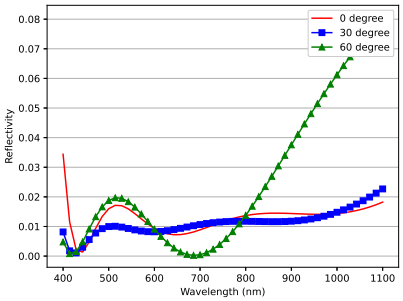


(d) Regression using *deep active learning* with the augmented dataset. This plot shows only reflectivities which range from 2 % to 7 %.

Fig. 5. Regression model comparison with the original and augmented datasets. Clearly, the relative L^2-errors of the regression model with the augmented datasets are lower than the original regression model.



(a) Optimal grating reflectivity versus incident angle (degrees) and wavelength (nanometers).



(b) Optimal grating reflectivity versus three choices of incident angle (degrees) and wavelength (nanometers).

Fig. 6. Optimal grating reflectivity versus incident angle and wavelength.

Disclosures. The authors declare no conflicts of interest.

Data Availability Statement. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. S. A. Maier, *Plasmonics: Fundamentals and Applications* (Springer, New York, 2007).
2. P. Baumeister, "Design of multilayer filters by successive approximations," J. Opt. Soc. Am. **48**, 955–958 (1958).
3. P. Azunre, J. Jean, C. Rotschild, V. Bulovic, S. G. Johnson, and M. A. Baldo, "Guaranteed global optimization of thin-film optical systems," New J. Phys. **21**, 073050 (2019).
4. J. Dobrowolski and R. Kemp, "Refinement of optical multilayer systems with different optimization procedures," Appl. Opt. **29**, 2876–93 (1990).
5. S. Martin, J. Rivory, and M. Schoenauer, "Synthesis of optical multilayer systems using genetic algorithms," Appl. Opt. **34**, 2247–54 (1995).
6. J. Jensen and O. Sigmund, "Topology optimization for nano-photonics," Laser Photonics Rev. **5**, 308 (2011).
7. A. M. Hammond, A. Oskooi, M. Chen, Z. Lin, S. G. Johnson, and S. E. Ralph, "High-performance hybrid time/frequency-domain topology optimization for large-scale photonics inverse design," Opt. Express **30**, 4467–4491 (2022).
8. A. V. Tikhonravov, M. K. Trubetskov, and G. W. DeBell, "Application of the needle optimization technique to the design of optical coatings," Appl. Opt. **35**, 5493–5508 (1996).
9. A. V. Tikhonravov and M. K. Trubetskov, "Modern design tools and a new paradigm in optical coating design," Appl. Opt. **51**, 7319–7332 (2012).
10. J. Dobrowolski and B. Sullivan, "Universal antireflection coatings for substrates for the visible spectral region," Appl. Opt. **35**, 4993–4997 (1996).
11. F. Lemarquis, T. Begou, A. Moreau, and J. Lumeau, "Broadband antireflection coatings for visible and infrared ranges," CEAS Space J. **11**, 567–578 (2019).
12. P. R. Wiecha, A. Arbouet, C. Girard, and O. L. Muskens, "Deep learning in nano-photonics: inverse design and beyond," Photon. Res., PRJ **9**, B182–B200 (2021).
13. J. Peurifoy, Y. Shen, L. Jing, Y. Yang, F. Cano-Renteria, B. G. DeLacy, J. D. Joannopoulos, M. Tegmark, and M. Soljacic, "Nanophotonic particle simulation and inverse design using artificial neural networks," Sci. Adv. **4**, eaar4206 (2018).
14. D. Liu, Y. Tan, E. Khoram, and Z. Yu, "Training deep neural networks for the inverse design of nanophotonic structures," ACS Photonics **5**, 1365–1369 (2018).
15. R. Unni, K. Yao, and Y. Zheng, "Deep convolutional mixture density network for inverse design of layered photonic structures," ACS Photonics **7**, 2703–2712 (2020).
16. M. A. Barry, V. Berthier, B. D. Wilts, M.-C. Cambourieux, P. Bennet, R. Pollès, O. Teytaud, E. Centeno, N. Biais, and A. Moreau, "Evolutionary algorithms converge towards evolved biological photonic structures," Sci. Rep. **10**, 12024 (2020).
17. W. Ma, F. Cheng, Y. Xu, Q. Wen, and Y. Liu, "Probabilistic representation and inverse design of metamaterials based on a deep generative model with semi-supervised learning strategy," Adv. Mater. **31**, 1901111 (2019).
18. Z. Liu, L. Raju, D. Zhu, and W. Cai, "A hybrid strategy for the discovery and design of photonic structures," IEEE J. on Emerg. Sel. Top. Circuits Syst. **10**, 126–135 (2020).
19. W. Ma and Y. Liu, "A data-efficient self-supervised deep learning model for design and characterization of nanophotonic structures," Sci. China Physics, Mech. Astron. **63** (2020).
20. Z. A. Kudyshev, A. V. Kildishev, V. M. Shalaev, and A. Boltasseva, "Machine learning–assisted global optimization of photonic devices," Nanophotonics **10**, 371–383 (2021).
21. J. Jiang and J. A. Fan, "Multiobjective and categorical global optimization of photonic structures based on ResNet generative neural networks," Nanophotonics **10**, 361–369 (2021).
22. L. Wang, Y.-C. Chan, F. Ahmed, Z. Liu, P. Zhu, and W. Chen, "Deep generative modeling for mechanistic-based learning and design of metamaterial systems," Comput. Methods Appl. Mech. Eng. **372**, 113377 (2020).
23. P. Yeh, *Optical waves in layered media*, vol. 61 (Wiley-Interscience, 2005).
24. R. J. LeVeque, *Finite difference methods for ordinary and partial differential equations* (Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2007). Steady-state and time-dependent problems.
25. Y. Hong and D. P. Nicholls, "A high–order perturbation of surfaces method for scattering of linear waves by periodic multiply layered gratings in two and three dimensions," J. Comput. Phys. **345**, 162–188 (2017).
26. D. P. Kingma and M. Welling, "An introduction to variational autoencoders," Foundations Trends Mach. Learn. **12**, 307–392 (2019).
27. K. Sohn, H. Lee, and X. Yan, "Learning structured output representation using deep conditional generative models," in *Advances in Neural Information Processing Systems,* vol. 28 C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, eds. (Curran Associates, Inc., 2015), pp. 1–9.
28. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016). http://www.deeplearningbook.org.
29. D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," J. Am. Stat. Assoc. **112**, 859–877 (2017).
30. H. Ranganathan, H. Venkateswara, S. Chakraborty, and S. Panchanathan, "Deep active learning for image classification," in *2017 IEEE International Conference on Image Processing (ICIP),* (2017), pp. 3934–3938.
31. B. Settles, "Active learning," Synth. Lect. on Artif. Intell. Mach. Learn. **6**, 1–114 (2012).
32. R. Pestourie, Y. Mroueh, T. V. Nguyen, P. Das, and S. G. Johnson, "Active learning of deep surrogates for pdes:

application to metasurface design," npj Comput. Mater. **6**, 164 (2020).