

# Inverse design of two-dimensional freeform metagrating using an adversarial conditional variational autoencoder

Kojima, Keisuke; Koike-Akino, Toshiaki; Wang, Ye; Jung Minwoo; Brand, Matthew

TR2023-004 January 31, 2023

## Abstract

Inverse design of two-dimensional freeform metagrating using an adversarial conditional variational autoencoder Keisuke Kojimaa, Toshiaki Koike-Akinob, Ye Wangb, Minwoo Jungb, c, and Matthew Brandb aBoston Quantum Photonics LLC, 588 Bost Post Rd #315, Weston, MA 02493, USA bMitsubishi Electric Research Laboratories, 201 Broadway, Cambridge, MA 02139, USA cDepartment of Physics, Cornell University, Ithaca, NY 14853, USA. ABSTRACT For the inverse design of metagratings and metasurfaces, generative deep learning has been widely explored. Most of the works are based on a conditional generative adversarial network (CGAN) and its variants, however, selecting proper hyper parameters for efficient training is challenging. An alternative approach, an adversarial conditional variational autoencoder (A-CVAE) has not been explored yet for the inverse design of metagrat- ings and metasurfaces, even though it has shown great promise for the inverse design of planar nanophotonic waveguide power/wavelength splitters recently. In this paper, we discuss how A-CVAE can be applied for two-dimensional freeform metagratings, including the training dataset preparation, construction of the network, training techniques, and the performance of the inverse-designed metagratings.

*SPIE Photonics West 2023*



# Inverse design of two-dimensional freeform metagrating using an adversarial conditional variational autoencoder

Keisuke Kojima<sup>a</sup>, Toshiaki Koike-Akino<sup>b</sup>, Ye Wang<sup>b</sup>, Minwoo Jung<sup>b, c</sup>, and Matthew Brand<sup>b</sup>

<sup>a</sup>Boston Quantum Photonics LLC, 588 Bost Post Rd #315, Weston, MA 02493, USA

<sup>b</sup>Mitsubishi Electric Research Laboratories, 201 Broadway, Cambridge, MA 02139, USA

<sup>c</sup>Department of Physics, Cornell University, Ithaca, NY 14853, USA.

## ABSTRACT

For the inverse design of metagratings and metasurfaces, generative deep learning has been widely explored. Most of the works are based on a conditional generative adversarial network (CGAN) and its variants, however, selecting proper hyper parameters for efficient training is challenging. An alternative approach, an adversarial conditional variational autoencoder (A-CVAE) has not been explored yet for the inverse design of metagratings and metasurfaces, even though it has shown great promise for the inverse design of planar nanophotonic waveguide power/wavelength splitters recently. In this paper, we discuss how A-CVAE can be applied for two-dimensional freeform metagratings, including the training dataset preparation, construction of the network, training techniques, and the performance of the inverse-designed metagratings.

**Keywords:** Metasurface, metagratings, deep learning, variational autoencoder, CVAE, A-CVAE

## 1. INTRODUCTION

Metasurfaces<sup>1</sup> are expected to play a major role in ultracompact and multi-functional lenses and various other applications. Possible applications include quarter-wave plates, vortex plates to create orbital angular momentum (OAM), plasmonic collimators,<sup>2</sup> ultra wideband lenses,<sup>3</sup> and thermal emitters.<sup>4</sup> There are also researches on the application of metasurfaces for quantum photon sources<sup>5-7</sup> and quantum state engineering.<sup>8</sup> However, designing metasurfaces also has the same issue of vast design space and time consuming electromagnetic simulations.

Metasurfaces are in principle ultrathin layer of metamaterials which control the wavefront of optical beams.<sup>1,2</sup> There is so much freedom in the design of the nanostructures, and the wavefront can be made to be dependent on the wavelength, polarization, or the combination of both. These are expected to create ultra-thin, high-performance, and multi-functionality lenses. However, there are so many design parameters, so it can be very challenging to design metasurfaces.

The most common way to inverse-design metasurfaces is to use a gradient-based adjoint method.<sup>9-11</sup> This often creates very good designs, however, the end results depend on the initial conditions, and multiple runs need to be carried, taking a long time to generate near-optimal results. In order to complement the inverse-design method, various types of deep learning has been explored.<sup>12-17, 17-22</sup>

Many of the works are based on a generative adversarial network (GAN) and its variants,<sup>12,23</sup> however, selecting proper hyper parameters for efficient training is usually challenging. An alternative approach, based on an autoencoder (AE), tries to condense the training data geometry information into a small number of latent variables,<sup>4,15</sup> analogous to dimensionality reduction, and optimization in the latent space becomes effective.<sup>4,24</sup>

As an extension of the well-established CVAE,<sup>25</sup> we have developed an adversarial conditional variational autoencoder (A-CVAE),<sup>26</sup> wherein the latent variable distribution becomes closer to normal distribution. It has been successfully demonstrated to have excellent performance for classification of brain waves by learning subject-invariant representations.<sup>27</sup> We also applied A-CVAE for the inverse-design planar nanophotonic devices,<sup>28-30</sup> but its application for the inverse design of metasurfaces have not been explored so far.

In this paper, we discuss how A-CVAE can be applied for two-dimensional freeform metagratings, which is a special type of metasurfaces and comprising of periodic arbitrary-shape unit cells, including the training dataset preparation, construction of the network, and the performance of the inverse-designed metagratings.

---

Further author information: (Send correspondence to Keisuke Kojima: E-mail: kkojima@bostonqp.com)

## 2. DEVICES

### 2.1 Device Structure

The metagratings are made of 325 nm-thick polycrystalline silicon on an SiO<sub>2</sub> substrate. A unit-cell comprises of a two-dimensional freeform silicon pattern and deflect normal-incident light to the +1 diffraction order. As shown in Fig. 1, the unit cell is subdivided into 256 × 128 pixels wherein "1" represents silicon and "0" represents air.<sup>31</sup>

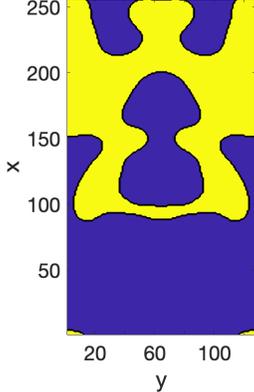


Figure 1: Schematic of the unit cell pattern of the metagrating, where the yellow part depicts the region with poly Si layer, while the blue region shows the area where the poly Si is etched.

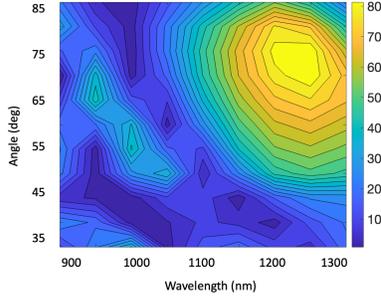


Figure 2: An example of a deflection efficiency map evaluated at 9 wavelengths and 11 angles.

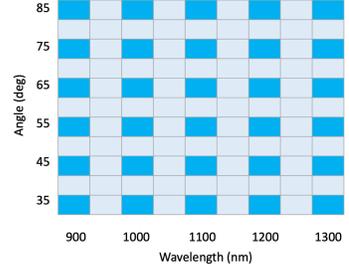


Figure 3: The group of patterns used for the training of the network, indicated by the dark blue boxes.

### 2.2 Training Data Preparation

From the Metanet database developed by the Stanford team,<sup>32</sup> we select two-dimensional freeform metagratings targeting wavelength of 900, 1000, 1100, 1200, and 1300 nm, and deflection angle in the x-direction of 35, 45, 55, 65, 75, and 85 degrees<sup>3</sup>, comprising of 25 groups (cells) and the total of 518 patterns. The polarization is chosen to be in the transverse electric (TE) mode, i.e., the electric field is in the direction of y-axis. The deflection angle in the y-direction is zero. Each device pattern is evaluated for the deflection efficiency at the wavelength of 900, 950, ..., 1250, and 1300 nm, and at the deflection angle of 35, 40, ..., 85 degrees, totaling 99 conditions. The open source code RETICOLO, based on the rigorous coupled-wave analysis (RCWA) method, is used for the calculation of the deflection efficiency.<sup>33</sup> An example of the efficiency map is shown in Fig. 2.

Among these 518 patterns, 93 patterns showed sharp changes in the efficiency map, and are excluded these from the training data. We also add 8 patterns enhancing the consistency among the training data. Here, at the angle of 75 deg, using the adjoint method<sup>34</sup> we optimize a device pattern at 1000 nm using the best device pattern at 900 nm as an initial condition, and another device pattern using the best device at 1100 nm as an initial condition. The same procedure is repeated at other wavelengths at the angle of 75 deg. Their benefits will be discussed in a later section 4.1. So the total number of patterns used in the training is 433.

The unit cell inherently has ambiguity in the phase (position) both in x- and y-directions. If we neglect the optical phase of the deflected light, the same pattern with the different translational phase gives the same efficiency. For CVAE, this becomes an issue since for the same target optical response, there are multiple possibilities. In order to minimize this ambiguity issue, we standardize the training data patterns, such that we maximize the difference in average weight of the pixel values in the x-direction (vertical) by translation. Also, the choice of the y-axis was made such that the weight is larger near the edge of the y-axis. We plot 12 device patterns in Fig. 4, before (left) and after (right) the standardization procedure.

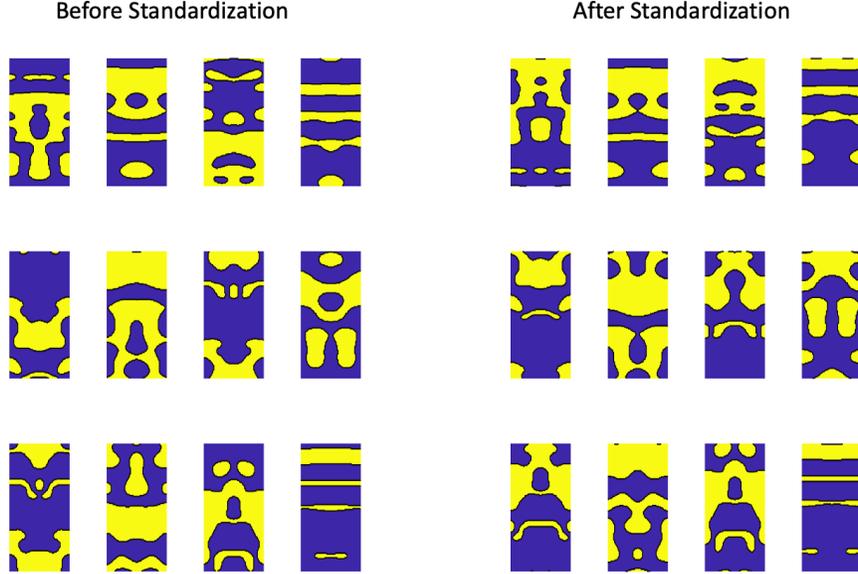


Figure 4: Device patterns before and after standardization. In the vertical direction, "1" (yellows) pixels are collected more in the upper half. In the horizontal direction, the "1" pixels are collected towards the edge.

### 3. DEEP LEARNING PROCEDURES

#### 3.1 A-CVAE Overview

The goal of our inverse design is to generate a set of highly efficient metagrating patterns, given a target wavelength and deflection angle. These are validated by the actual RCWA simulations and the best device patterns are selected.

We use an A-CVAE<sup>28</sup> as shown in Fig. 8, where a separate branch to the adversary block is used for isolating the latent variable  $\tilde{s}$  from the conditional variations  $s$  (the target optical response). Since we treat the input pattern as an image (i.e. pixels), we use two-dimensional convolutional neural networks (CNNs) for the encoder and the decoder. The encoder consists of three layers of two-dimensional convolutional layers, each followed by a tanh activation functions and two-dimensional maxpool layer.

The optical response (99-element vector), i.e., the deflection efficiency over 9 wavelengths and 11 angles is fed through a multilayer perceptron (MLP), used as the target, and is added to the latent variables. Each of the two MLPs consist of a single linear layer followed by a tanh activation function. The same optical response is fed through another MLP, and is added to the second channel of the encoder. The adversarial block consists of a linear layer, followed by a relu activation function, a linear layer, and a sigmoid function. The number of latent variables is ten.

Further, we implement loop consistency to further improve the training of the network.<sup>35,36</sup> The benefit of loop consistency along with the A-CVAE model is that extra loop for the calculation of the additional loss makes the loss more sensitive to the reconstruction loss, making the training more effective.

Since it is difficult to identify the optimal training epochs, we stored weight matrices at multiple epochs, and tried to find the optimal epochs using the validation results.<sup>28</sup>

#### 3.2 Latent space optimization

Using the trained decoder and MLP, we explore the optimization in the latent space.

As the target optical response, we pick eight device patterns from the existing dataset which give the highest efficiency at the target wavelength and angle. If the peak position is off, we synthesize target optical response by shifting the efficiency map in the wavelength direction. Since we do not know which gives the best outcome,

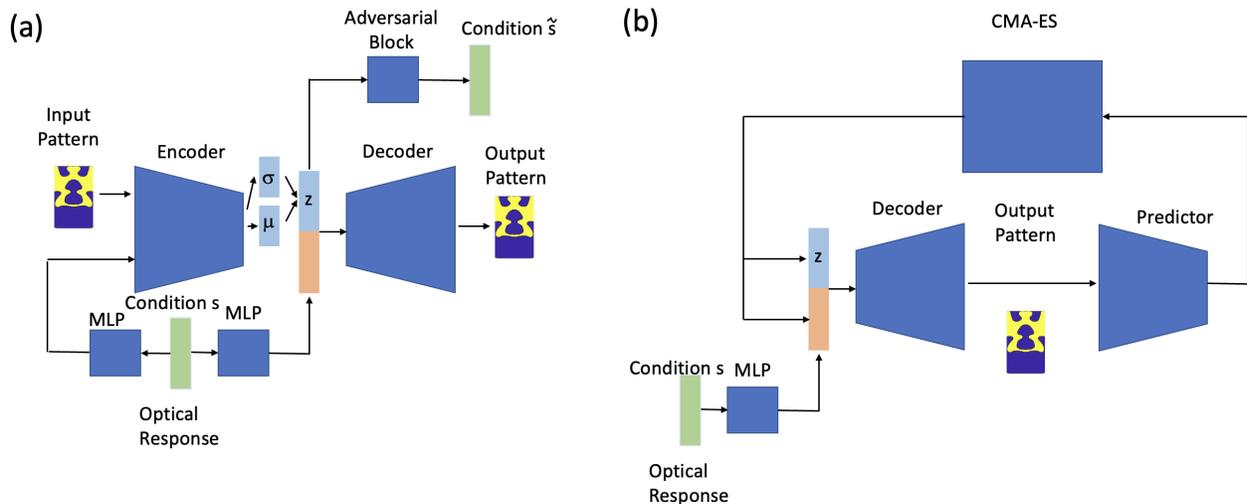


Figure 5: (a) Schematic of the A-CVAE in the training phase, and (b) schematic of the decoder and the CMA-ES optimization in the latent space. The decoder and the MLP are the same ones in the trained process. MLP: Multilayer Perceptron.

use variable weight factors to generate the optimal optical response. The output device patterns of the decoder go through a gaussian filter and a threshold to smooth the edges.

We create a predictor, which is a forward model for the optical responses. The input is the device patterns ( $256 \times 128$  matrices), and the output is the efficiency map (99 element). The network layer structure is similar to that of the encoder, except that the former uses four convolutional layers, and have only one input channel. For the training of the predictor network, we use the same dataset as that used in the A-CVAE training, but is augmented with  $\pm 4$  and  $\pm 8$  shifts in the x-direction, making the dataset 5 times larger than the original dataset.

For the optimization algorithm, we use covariance matrix adaptation evolution strategy (CMA-ES).<sup>37</sup> The CMA-ES algorithm tries to maximize the deflection efficiency at the target wavelength and angle, by optimizing ten latent variables and eight weighting factors. For each generation (iteration) of the CMA-ES process, we generate 1000 device patterns using the decode-predictor pair, and 100 generations are used for optimizing the patterns. Since we are using neural networks to generate and predict the performances, this process is very fast. Finally, we validate the device patterns by using another gaussian filter and using multiple threshold values to vary the thickness of the patterns, followed by the RCWA method for calculating the deflection efficiency.

## 4. RESULTS

### 4.1 Data consistency

First, we compare the generated patterns with and without the training data for consistency. In Fig. 6 shows the results evaluated at 75 deg. There are no training data specifically targeted at 950, 1050, 1150, and 1250 nm as shown in Fig. 3. Therefore, the efficiency of the existing training data, as indicated by the blue bars in Fig. 6 are compromised at these wavelengths. The red bars indicate the maximum efficiency of the 50 pattern data generated by A-CVAE and with the training data for consistency, demonstrating that "interpolation" for high efficiency can be done. The orange bars in Fig. 6 show the maximum efficiency of the 50 pattern data generated by A-CVAE without the training data for consistency, indicating the importance of the extra training data.

To visualize the effect of data consistency, we compare the device patterns. Fig. 7 (a) shows the original device pattern in the training dataset which gives the highest deflection efficiency at 75 deg and 900 nm. Fig. 7 (c) shows the optimized device pattern at 75 deg and 900 nm using (a) as the initial condition, and is added in the training dataset for consistency. Fig. 7 (b) is the best pattern generated by the A-CVAE at 75 deg and

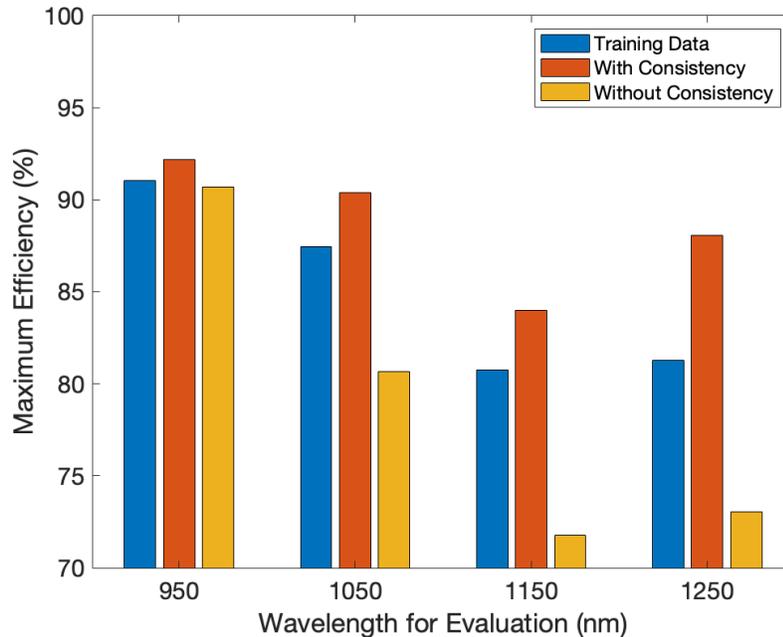


Figure 6: Maximum efficiency of the training data evaluated at each wavelength at 75 deg, device patterns generated by A-CVAE without consistency patterns, and patterns generated by A-CVAE with consistency patterns. Note that training data exist at 900, 1000, 1100, 1200, and 1300 nm.

950 nm. These indicate that the smooth interpolation of the patterns can be made with this A-CVAE method, giving high efficiency. This consistency concept is similar to the cascaded dataset as described in Fig.<sup>38</sup>

## 4.2 Comparison with the training data

Next, we evaluate the generated and validated data in the groups where there is already training dataset at 65 and 75 degrees and at wavelengths of 900, 1000, ... 1300 nm. Figure 8 shows the comparison, where blue bars are for the maximum efficiency of the training data, while red bars show the maximum efficiency of the generated devices. In seven out of ten groups, the generated patterns gave better results, with an average of 2.2%. This demonstrates that A-CVAE can improve the overall of the whole database, once these new datasets are added.

## 4.3 Active learning

Including the newly generated devices into the training data further improves the design capabilities proven very effective.<sup>4, 23, 28, 39, 40</sup> Even though A-CVAE shows the capabilities of generating overall better patterns than the training data, we have not performed this active learning. This iterative process is also expected to further improve the overall performance of the whole network.

## 4.4 Computational Time

We use a PC workstation with an Intel i9-12900K processor and NVIDIA RTX3090 graphics board. Python and Pytorch are used for the machine learning part, and Matlab and RCWA (RETICOLO) are used for the efficiency calculation and the validation of the device patterns. The A-CVAE training takes 7.1 hours for 4000 epochs. The training of the predictor network takes 4.4 hours for 4000 epochs. These are one-time investment to build the networks. Once the networks are trained, generating 50 device patterns takes 25 minutes, and validating them takes 10 minutes, for a total of 35 minutes. If we are optimizing device patterns from scratch for each group of wavelength-angle,  $\sim 25$  device patterns need to be generated to replicate the Metanet database. Since each adjoint method optimization takes about 4 hours, the total time of 100 hours is required for each group. Therefore, there is more than two orders of magnitude speed advantage for this deep learning method.

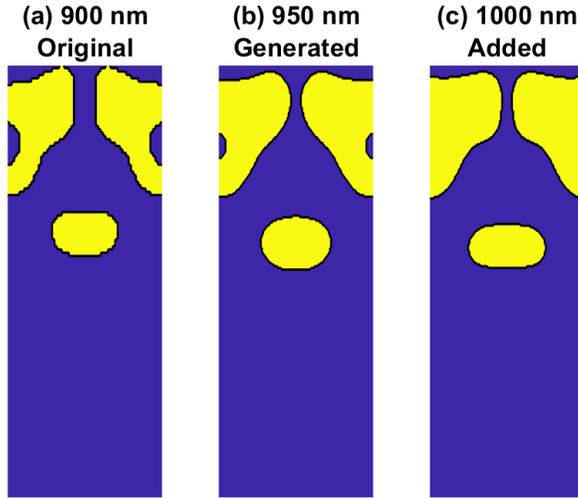


Figure 7: (a) Best device pattern at 75 deg and 900 nm, (b) best generated device pattern at 75 deg and 950 nm, and (c) additional training device pattern, using the pattern(a) as an initial condition and optimized at 1000 nm.

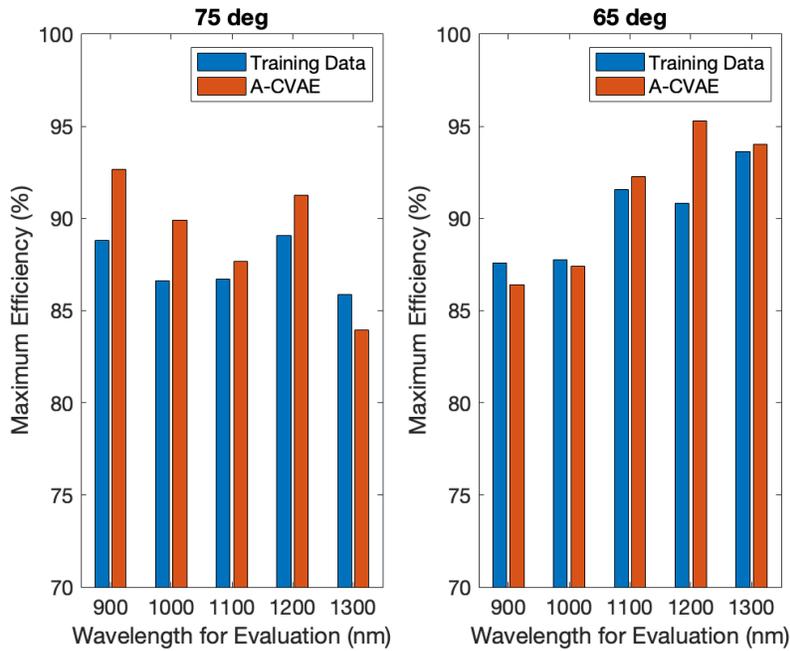


Figure 8: Maximum of the training data and that of patterns generated by A-CVAE evaluated at each wavelength at 75 deg and 65 deg.

## 5. FUTURE DIRECTIONS

As for the case of planar waveguides, we successfully demonstrated the transfer learning concept when the core Si layer thickness is changed and good inverse-design capabilities are demonstrated with minimal additional g data with new Si thickness.<sup>30</sup> Similar results are expected for metasurfaces.

We have also investigated different Bayesian models and found that nested A-CVAE is promising for planar wavelength splitters.<sup>38</sup> This is also expected to improve the performance for this metagrating inverse-design.

## 6. CONCLUSION

We applied the deep learning method based on A-CVAE to the inverse-design of freeform metagratings, and demonstrated that patterns with high efficiency can be generated from the training data. In particular, the interpolated patterns show higher efficiencies than the existing data. Once the network is constructed, we can generate high efficiency data patterns at arbitrary conditions very quickly. This inverse-design process is decoupled from any specific differential equations such as Maxwell equations, so it is expected to be generic and applicable to broader problems.

## REFERENCES

- [1] Yu, N. and Capasso, F., “Flat optics with designer metasurfaces,” *Nature materials* **13**(2), 139–150 (2014).
- [2] Yu, N. and Capasso, F., “Optical metasurfaces and prospect of their applications including fiber optics,” *J. Lightwave Technol.* **33**, 2344–2358 (Jun 2015).
- [3] Ndao, A., Hsu, L., Ha, J., Park, J.-H., Chang-Hasnain, C., and Kanté, B., “Octave bandwidth photonic fishnet-achromatic-metalens,” *Nature communications* **11**(1), 1–6 (2020).
- [4] Kudyshev, Z. A., Kildishev, A. V., Shalaev, V. M., and Boltasseva, A., “Machine-learning-assisted metasurface design for high-efficiency thermal emitter optimization,” *Applied Physics Reviews* **7**(2), 021407 (2020).
- [5] Huang, T.-Y., Grote, R. R., Mann, S. A., Hopper, D. A., Exarhos, A. L., Lopez, G. G., Klein, A. R., Garnett, E. C., and Bassett, L. C., “A monolithic immersion metalens for imaging solid-state quantum emitters,” *Nature communications* **10**(2392) (2019).
- [6] Li, L., Liu, Z., Ren, X., Wang, S., Su, V.-C., Chen, M.-K., Chu, C. H., Kuo, H. Y., Liu, B., Zang, W., et al., “Metalens-array-based high-dimensional and multiphoton quantum source,” *Science* **368**(6498), 1487–1490 (2020).
- [7] Santiago-Cruz, T., Gennaro, S. D., Mitrofanov, O., Addamane, S., Reno, J., Brener, I., and Chekhova, M. V., “Resonant metasurfaces for generating complex quantum states,” *Science* **377**(6609), 991–995 (2022).
- [8] Stav, T., Faerman, A., Maguid, E., Oren, D., Kleiner, V., Hasman, E., and Segev, M., “Quantum entanglement of the spin and orbital angular momentum of photons using metamaterials,” *Science* **361**(6407), 1101–1104 (2018).
- [9] Lalau-Keraly, C. M., Bhargava, S., Miller, O. D., and Yablonovitch, E., “Adjoint shape optimization applied to electromagnetic design,” *Optics express* **21**(18), 21693–21701 (2013).
- [10] Piggott, A. Y., Lu, J., Lagoudakis, K. G., Petykiewicz, J., Babinec, T. M., and Vučković, J., “Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer,” *Nature Photonics* **9**(6), 374–377 (2015).
- [11] Christiansen, R. E. and Sigmund, O., “Inverse design in photonics by topology optimization: tutorial,” *J. Opt. Soc. Am. B* **38**, 496–509 (Feb 2021).
- [12] Liu, Z., Zhu, D., Rodrigues, S. P., Lee, K.-T., and Cai, W., “Generative model for the inverse design of metasurfaces,” *Nano letters* **18**(10), 6570–6576 (2018).
- [13] Nadell, C. C., Huang, B., Malof, J. M., and Padilla, W. J., “Deep learning for accelerated all-dielectric metasurface design,” *Optics express* **27**(20), 27523–27535 (2019).
- [14] Ma, W., Cheng, F., and Liu, Y., “Deep-learning enabled on-demand design of chiral metamaterials,” *ACS Nano* **12**(6), 6326–6334 (2018).
- [15] Ma, W., Cheng, F., Xu, Y., Wen, Q., and Liu, Y., “Probabilistic representation and inverse design of metamaterials based on a deep generative model with semi-supervised learning strategy,” *Advanced Materials* **31**(35), 1901111 (2019).
- [16] An, S., Fowler, C., Zheng, B., Shalaginov, M. Y., Tang, H., Li, H., Zhou, L., Ding, J., Agarwal, A. M., Rivero-Baleine, C., et al., “A deep learning approach for objective-driven all-dielectric metasurface design,” *ACS Photonics* **6**(12), 3196–3207 (2019).
- [17] An, S., Zheng, B., Tang, H., Shalaginov, M. Y., Zhou, L., Li, H., Kang, M., Richardson, K. A., Gu, T., Hu, J., et al., “Multifunctional metasurface design with a generative adversarial network,” *Advanced Optical Materials* **9**(5), 2001433 (2021).

- [18] Wiecha, P. R. and Muskens, O. L., “Deep learning meets nanophotonics: A generalized accurate predictor for near fields and far fields of arbitrary 3D nanostructures,” *Nano Letters* **20**(1), 329–338 (2019).
- [19] Gao, L., Li, X., Liu, D., Wang, L., and Yu, Z., “A bidirectional deep neural network for accurate silicon color design,” *Advanced Materials* **31**(51), 1905467 (2019).
- [20] Liu, Z., Raju, L., Zhu, D., and Cai, W., “A hybrid strategy for the discovery and design of photonic structures,” *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* **10**(1), 126–135 (2020).
- [21] Jiang, J., Chen, M., and Fan, J. A., “Deep neural networks for the evaluation and design of photonic devices,” *Nature Reviews Materials* **6**, 679–700 (2021).
- [22] Tang, R., Lim, S. W. D., Yin, X., and Capasso, F., “Minimal memory differentiable ftd for inverse design,” in [*Conference on Lasers and Electro-Optics*], *Conference on Lasers and Electro-Optics*, FM5H.4, Optica Publishing Group (2022).
- [23] Wen, F., Jiang, J., and Fan, J. A., “Robust freeform metasurface design based on progressively growing generative networks,” *ACS Photonics* **7**(8), 2098–2104 (2020).
- [24] Tripp, A., Daxberger, E., and Hernández-Lobato, J. M., “Sample-efficient optimization in the latent space of deep generative models via weighted retraining,” in [*Advances in Neural Information Processing Systems*], Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., eds., **33**, 11259–11272, Curran Associates, Inc. (2020).
- [25] Sohn, K., Lee, H., and Yan, X., “Learning structured output representation using deep conditional generative models,” in [*Advances in neural information processing systems*], 3483–3491 (2015).
- [26] Wang, Y., Koike-Akino, T., and Erdogmus, D., “Invariant representations from adversarially censored autoencoders,” *arXiv preprint arXiv:1805.08097* (2018).
- [27] Özdenizci, O., Wang, Y., Koike-Akino, T., and Erdogmus, D., “Transfer learning in brain-computer interfaces with adversarial variational autoencoders,” in [*2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*], 207–210, IEEE (2019).
- [28] Tang, Y., Kojima, K., Koike-Akino, T., Wang, Y., Wu, P., Xie, Y., Tahersima, M. H., Jha, D. K., Parsons, K., and Qi, M., “Generative deep learning model for inverse design of integrated nanophotonic devices,” *Laser & Photonics Reviews* **14**, 2000287 (2020).
- [29] Kojima, K., Koike-Akino, T., Tang, Y., and Wang, Y., “Inverse design for integrated photonics using deep neural network,” in [*Integrated Photonics Research, Silicon and Nanophotonics*], IF3A–6, Optica Publishing Group (2021).
- [30] Kojima, K., Jung, M., Koike-Akino, T., Wang, Y., Brand, M., and Parsons, K., “Deep transfer learning for nanophotonic device design,” in [*Conference on Lasers and Electro-Optics Pacific Rim (CLEO-PR)*], CFA12E–05 (<https://www.merl.com/publications/docs/TR2022-107.pdf>) (2022).
- [31] Jiang, J. and Fan, J. A., “Global optimization of dielectric metasurfaces using a physics-driven neural network,” *Nano letters* **19**(8), 5366–5372 (2019).
- [32] Jiang, J., Lupoiu, R., Wang, E. W., Sell, D., Hugonin, J. P., Lalanne, P., and Fan, J. A., “Metanet: a new paradigm for data sharing in photonics research,” *Optics Express* **28**(9), 13670–13681 (2020).
- [33] Hugonin, J. P. and Lalanne, P., “Reticolo software for grating analysis,” (2021).
- [34] Sell, D., Yang, J., Doshay, S., Yang, R., and Fan, J. A., “Large-angle, multifunctional metagratings based on freeform multimode geometries,” *Nano letters* **17**(6), 3752–3757 (2017).
- [35] Tang, Y., Kojima, K., Koike-Akino, T., Wang, Y., Jha, D. K., Parsons, K., and Qi, M., “Nano-optic broadband power splitter design via cycle-consistent adversarial deep learning,” in [*CLEO: Science and Innovations*], SW4E–1, Optica Publishing Group (2021).
- [36] Koike-Akino, T., “Evolution of machine learning for photonic research.” Boston Chapter of IEEE Photonics Society (Dec. 2021). [http://www.bostonphotonics.org/files/seminars/TKoike\\_2021.pdf](http://www.bostonphotonics.org/files/seminars/TKoike_2021.pdf).
- [37] “The cma evolution strategy, <https://cma-es.github.io/>,” (2016).
- [38] Jung, M., Kojima, K., Koike-Akino, T., Wang, Y., Zhu, D., and Brand, M., “Finding the right deep neural network model for efficient design of tunable nanophotonic devices,” in [*CLEO: Science and Innovations*], SW5E–6, Optica Publishing Group (2022).
- [39] Asano, T. and Noda, S., “Iterative optimization of photonic crystal nanocavity designs by using deep neural networks,” *Nanophotonics* **8**(12), 2243–2256 (2019).

- [40] Kojima, K., Tahersima, M. H., Koike-Akino, T., Jha, D. K., Tang, Y., Wang, Y., and Parsons, K., “Deep neural networks for inverse design of nanophotonic devices,” *Journal of Lightwave Technology* **39**(4), 1010–1019 (2021).