Prediction of Plasmonic Metasurface Nanofeatures Using a Modified VAE Regressor

Prajith Pillai, Parama Pal
TCS Research
Bangalore 560066
p.prajith@tcs.com, parama.pal@tcs.com

Abstract

We apply a modified variational autoencoder (VAE) regressor for inversely retrieving topological parameters of plasmonic composites for generating structural colors 'at will'. We demonstrate results of a comparison study between inverse models based on generative VAEs as well as conventional tandem networks. We describe our strategy for filtering a simulated dataset for training both the models. The VAE- based inverse model links the electromagnetic response to the geometrical dimensions from the latent space using a multilayer perceptron regressor and shows better accuracy over the conventional tandem model.

Introduction

Plasmonic 'metasurfaces' are artificially engineered, two-dimensional nanocomposites that have the ability to demonstrate structural color effects. Plasmonic metamaterials can provide selective coloration and have been proposed as an ideal material for applications such as intelligent displays for augmented reality/virtual reality devices and high-volume consumer electronics. The use of plasmonic effects to achieve structural color as opposed to mixing together different colorants has been shown to be a robust, cost-effective scheme to produce colors at will. Metasurfaces comprise of periodic, sub-wavelength unit cells or 'meta-units' that interact with incoming radiation in unconventional ways thereby enabling manipulation of physical phenomena (such as reflection, absorption, diffraction) by varying the unit cell dimensions [1]. One example is that of localized surface plasmon resonances (LSPRs) which occur when light is trapped between conductive, nanometer-sized features leading to selective resonant absorption and hybridized reflectance modes resulting in colored metasurface films. Selective coloration can therefore be achieved by engineering unit cell dimensions appropriately [1,2]. In this work, we describe a model for inversely obtaining meta-unit geometries to achieve structural color produced 'at will'. We have used the spectral response data (provided in [2]) corresponding to a metasurface made up of a regularly spaced array of polymeric PDMS (polydimethylsiloxane) nanopillars uniformly coated with an aluminum layer (depicted in Figure 1).We use a prefiltered uniformly spaced dataset (10% of the original dataset) for our study. The spectral data can be interpreted via the CIE 1931 color coordinate system wherein any color on the chromaticity chart can be expressed as x and y coordinate pairs based on the three CIE primaries (Figure 1 inset). The human eye has three types of color sensors that respond to different wavelengths and can be visually depicted as a three-dimensional plot. In the CIE 1931 colour space, this 3D plot can be reduced to a 2D plot by using two parameters, namely brightness and chromaticity . This has been explained in detail in the supplementary information. Our aim is to inversely retrieve the dimensions of the nanoscale topological features from the observed (perceived) structural color via the corresponding spectral response.In recent years, an overwhelming quantum of work has been dedicated to demonstrating the applicability of data-driven frameworks for designing metamaterials 'by specification', i.e., inversely mapping the desired functionality to the structural design. Deep learning models, owing to their inherent capability for strong generalization, renders them as an ideal choice for mapping morphological and topological features to electromagnetic responses. Early work in this domain has provided only anecdotal evidence for the translatability of inverse models into actual fabrication scenarios whereas reality demands that the design models be robust to variations within fabrication tolerances.Conventional tandem networks with pre trained decoder is used for some cases of metasurface design [3]. Convolutional Neural Network based autoencoders are also employed for inverse parameter retrieval [4]. Generative models such as variational autoencoders (VAEs) and GANs, which define the latent space via the mean (μ) and standard deviation () of a normal distribution and use latent variables to encode the model input(s) to the output(s), have not been sufficiently explored. VAEs and GAN have high diversity compared to tandem networks because of the modal distribution of latent variables [5] providing good reconstruction. VAEs are an unsupervised learning method and hence quantifying the latent space (for predicting geometrical dimensions) becomes challenging [6,7].We therefore use a modified VAE regressor (Figure 2) which provides diversity as well as prediction accuracy [8,9].Modified VAEs have a probabilistic encoder that maps the chromaticity coordinates to the latent space and a probabilistic decoder that maps this latent space back to the coordinates. The addition of a multi-layer perceptron (MLP) from the latent space to

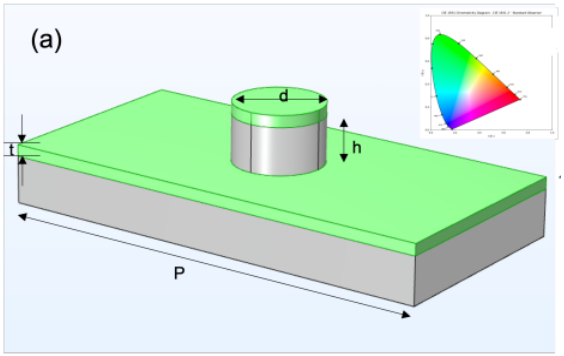

Figure 1: Annotated schematic of Al-coated PDMS meta unit (period P, pillar diameter d, pillar height h, and aluminium thickness t); inset depicts the CIE 1931 chromaticity chart.

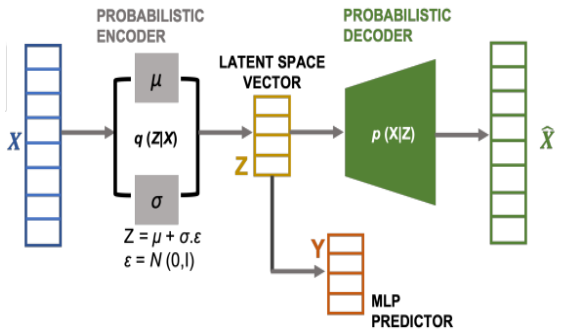

Figure 2: Representative diagram of a modified variational-autoencoder with input X (CIE chromaticity coordinates x, y) and output Y (d, h, and t)

the vector space defined by the meta unit's structural parameters, makes the training versatile by taking into account the reconstruction as well as dimensional prediction. The model is thus simultaneously trained with a combined loss function having three components: the mean squared error in the reconstruction of the coordinate space, Kullback-Leibler (KL) divergence (which captures the disparity between the learned distribution and the distribution of the prior network) and the mean squared error in the prediction of dimension through the dense layer added to latent space.

Related Work

Data-driven inverse methods for designing nanostructures have gained significant traction over the last decade. However, inverse models for plasmonic structures remain a challenge due to the large simulation times required for generating the datasets. Conventional design methods require using first principles in an iterative manner wherein parameters influencing the device response are tweaked till the desired performance is achieved. For plasmonic metasurfaces, there can be instances where multiple meta-units may yield near-identical colors implying that tandem networks will be a reasonable choice as they force the model to converge to a single solution. However, this typically requires large, well-distributed datasets. The distribution of the training data pairs plays an important role during inverse design as uneven spreads and clusters hinders learning significantly. Tandem autoencoders have been used for inversely obtaining plasmonic meta unit dimensions for given structural color(s) [2]. This disadvantages of this approach is that uses a considerably large dataset for inverse prediction. In addition, the colour space prediction suffers from low accuracy in regions corresponding to x and y coordinates proximal to the 45 degree axis in the CIE 1931 colour space (close to the achromatic 'white' point) [2]. In this work, we clearly demonstrate the benefits of our approach over tandem autoencoders via a methodical comparison study, as described in the following sections.

Methods

Dataset Refinement

The nanopillar feature under consideration here generates colors concentrated within an oblong region around the achromatic or 'white' point of the CIE chart resulting in uneven clustering and hence insufficient learning owing to the possibility of multiple outputs for similar inputs. To avoid skewing towards this cluster, we evaluated the Euclidean distance between all the points in the coordinate space and rejected all points whose Euclidean distance lay below a threshold set by us. A qualitative study was done wherein 5 threshold values were randomly chosen and the reduction in the clustering was evaluated qualitatively. We chose a threshold value that reduced the dataset to 299 points (shown in Figure 3).We evaluated the losses for original and reduced datasets as well (supplementary information). Although this meant a reduction of the dataset size by a factor of nearly 8, it resulted in more evenly spaced points with no clustering and significantly improved the losses and learning for the modified VAE. This dataset was applied to both the tandem autoencoder model as well as the modified VAE model.

Model Evaluation

Our modified VAE network takes in the CIE x and CIE y chromaticity coordinates as inputs for the encoder. The output B of the MLP (linked to the latent space) yields the meta-unit (nanopillar) dimensions. We have a distribution z for the latent space and we try to generate a sample x containing CIE coordinates from this distribution which is represented by $p(x|z)$. From Bayes theorem we have

$$p(x|z) = \frac{p(z|x) * p(x)}{p(z)} \tag{1}$$

We approximate $p(z|x)$ using a variational posterior represented as $q(z|x)$ and the difference in the distribution is evaluated using the Kullback-Leibler (KL) divergence loss. Therefore, our objective becomes a minimisation problem wherein;

$$min\ KL(q(z|x)||p(z|x)) \tag{2}$$

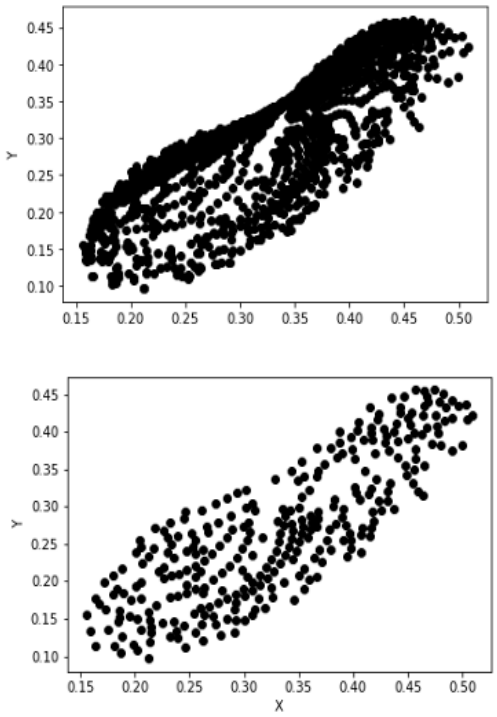

Figure 3: *x* and *y* chromaticity coordinates : original dataset (top), filtered dataset (bottom)

The above minimisation problem is equivalent to the following maximisation problem

$$E_{q(z|x)}logp(x|z) - KL(q(z|x)||p(z)) \quad (3)$$

where the first term is the reconstruction likelihood for the coordinate space and the second term represents the learned distribution's similarity to the prior distribution. The combined loss for the variational autoencoder modifies to,

$$Loss = L(x, \hat{x}) + \Sigma_j KL(q_j(z|x)||p(z)) \quad (4)$$

Now, we add an MLP regressor to the latent space which tries to recreate the dimensions corresponding to the colour coordinate space. This loss gets added to the combined loss giving the total loss equation as shown below

$$Loss_{total} = L(y, \hat{y}) + L(x, \hat{x}) + \Sigma_j KL(q_j(z|x)||p(z)) \quad (5)$$

where first term is the dimension loss, second term reconstruction loss and the third term KL divergence loss The training dataset, which is a collection of tensors encoding the nanopillar diameter *'d'*, height *'h'* and thickness *'t'* of the Al coating with the corresponding structural color coordinates. We performed a feature selection exercise to feed the most relevant features into the model thereby avoiding overfitting from using highly correlated predictors. A correlation matrix (shown in Table 1) was computed for the structural parameters (*d*, *h*, and, *t*) which showed minor correlation between the pillar diameter d and aluminium thickness t but almost zero correlation for the other parameter pairs indicating that all three parameters could be used for our

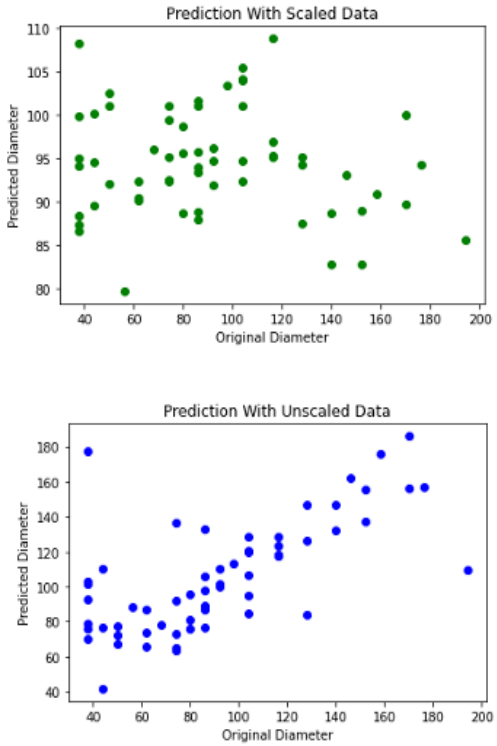

Figure 4: Prediction of pillar diameter *'d'* for scaled (top) and unscaled (bottom) data

model. We also conducted a study on the effect of scaling on the reduced dataset. The data for the geometrical variables was scaled using 'MinMaxScaler' from Python scikit-learn 1.1.2. The scaled data does not work well compared to the unscaled data for predicting the meta unit's geometrical parameters. This can be seen in figure 4(b) where the actual and predicted diameters show a linear relationship with the initial hyperparameters. Hence, we proceeded with unscaled data for our model. The total dataset was split in the ratio 80:10:10. 10 percentage data was used as unseen test set for performance evaluation. The model is trained for 2500 epochs with an early stopping criteria by monitoring the validation loss.

Table 1: Correlation matrix for the geometric dimensions

	Diameter	Height	Thickness
Diameter	1.0000	-0.0324	-0.1977
Height	-0.0324	1.0000	-0.0486
Thickness	-0.1977	-0.0486	1.0000

Hyperparameter Tuning

To optimize the number of network layers and neurons (in each layer), the reconstruction loss, dimensional loss, reconstruction validation loss and dimensional validation loss were calculated first by varying the number of layers from

1 to 5 (for 512 neurons). Subsequently, the neurons in each layer were varied as 256, 512, and 1024. We observed that the lowest validation losses corresponded to 2 dense layers having 512 neurons each. We also tuned model hyperparameters including the learning rate (varied as 0.1, 0.001, 0.001, 0.0001), activation functions (Sigmoid, Tanh, ReLU, SoftPlus), optimizers (SGD, Adam, RMSprop), latent space dimensions (varied as 3, 5, 10, 100) and regularization (L1, L2). The best performance was obtained with an Adam optimizer (learning rate of 0.01), and ReLU activation function for the encoder, decoder, and predictor. The final model had a latent space dimension equal to 10 and L2 regularization with lambda (regularization parameter) of 0.0001. The loss tables for the hyper-parameter tuning is provided in the supplementary information. Post hyper-parameter tuning, the model reconstruction and dimensional validation losses were 345.1534 and 329.7767 respectively.

Results and Discussion

We have added a predictor to the latent space of the variational autoencoder and trained the network in a unconventional manner to inversely retrieve structural parameters of Al-coated PDMS nanopillars for structural color generation. We train the model using a total loss function that tries to optimize the model using regeneration accuracy for the decoder and prediction accuracy for the regressor. We used a Euclidean filter to trim the original dataset to reduce clustering and further optimized our model by tuning critical hyperparameters. Our model offers significant improvement over tandem network-based frameworks thereby bringing us closer to a 'realizable' approach offering accuracy as well as diversity. Figure 5 depicts color prediction results from randomly selected points from the unseen test data close to the achromatic point in the CIE chart. A comparison of colour predicted using tandem network is added in the supplementary information.Our model yields high accuracies for reconstruction (as shown in Figure 5) as well as for prediction. Our model outperforms inverse models based on conventional tandem autoencoders in terms of reconstruction accuracy as well as dimensional prediction (Supplementary data) and performs well with only 300 training data points thereby obviating the need to generate large datasets. The mean absolute errors while predicting the nanopillar diameter, height and thickness (test set) were 15.97 nm, 14.84 nm, and 4.51 nm respectively for the test set. The inversely predicted dimensions done using the trained network is shown in Table 2 for few test cases. A benchmark study of the losses using both tandem network and modified generative VAE regressor is also done for the reduced dataset and the results are shown in Figure 6.

Conclusion

In this work we explore a modified VAE regressor for inversely retrieving plasmonic metasurface dimensions from colour coordinates in CIE 1931 colour space. The highly clustered dataset is initially refined by using a eucledian filter. The evenly spaced dataset is then used for training the model. A comparative study is also done with non generative tandem autoencoder inverse model with the reduced dataset. The modified VAE regressor model is observed to outperform the traditional tandem networks without much drop in the accuracy. Thus we can conclude that generative models such as variational auto encoders (VAEs) and generative adversarial networks (GANs) are the 'go-to' choice for inverse problems in the area of metamaterials when dataset availability is low and we do not want to compromise on prediction accuracy. We anticipate that the advantages associated with these architectures will play a key role in accelerating the 'discovery' of versatile devices based on high-performance functional materials.

Actual colour	Predicted colour	Actual colour Coordinate (x)	Actual colour Coordinate (y)	Predicted colour Coordinate (x)	Predicted colour Coordinate (y)
		0.205311901	0.198157787	0.23447141	0.20010978
		0.47936778	0.441238057	0.4949526	0.46358418
		0.279568389	0.211715401	0.2592351	0.21884245
		0.490952676	0.418480077	0.47314385	0.42702973
		0.20564836	0.245515287	0.21156514	0.24471009
		0.343576543	0.34846719	0.3511328	0.34889418
		0.39805409	0.255730626	0.40551338	0.25606203

Figure 5: Comparison of randomly selected actual color and model-predicted color and their corresponding CIE 1931 coordinate values

Table 2: Comparison of actual and predicted dimensions (3 cases from test data)

Case	Diameter	Height	Thickness
Actual Case1	170	44	92
Predicted Case1	166.44	41.48	92.53
Actual Case2	55	35	50
Predicted Case2	56.47	42.13	50.81

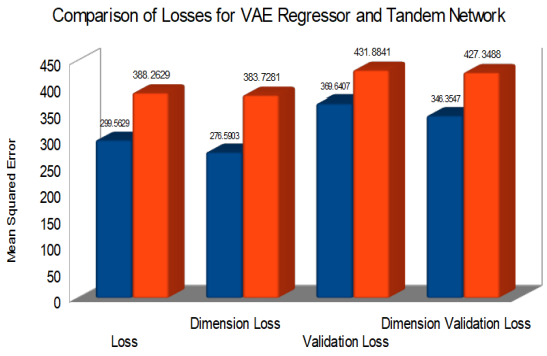

Figure 6: Losses comparison VAE Regressor (blue) and Tandem Autoencoder (red)

References

Chen, K., Wen., Y., and Lin, Y-S (2022) Plasmonic enhancement of metamaterial nanograting with high color purity and wide color gamut characteristics, Optics Communications, 506,127575.

Roberts, N.B., Hedayati, M.K. (2021) A deep learning approach to the forward prediction and inverse design of plasmonic metasurface structural color, Appl. Phys. Lett., 119 (061101).

Liu, D., Tan, Y., Khoram, E., Yu, Z. (2018). Training Deep Neural Networks for the Inverse Design of Nanophotonic Structures. ACS Photonics, 5(4), 1365–1369.

Rahmani, B., Loterie, D., Kakkava, E., Borhani, N., Teğin, U., Psaltis, D., Moser, C. (2020). Actor neural networks for the robust control of partially measured nonlinear systems showcased for image propagation through diffuse media. Nature Machine Intelligence, 2(7), 403–410

Ma, T., Tobah, M., Wang, H., and Guo, J.L (2022) Benchmarking deep learning-based models on nanophotonic inverse design problems, Opto -Electronic Science, 1 (210012).

Palsson, S., Cerri, S., Dittadi, A., and Limput, K.V. (2019) Semi-Supervised Variational Autoencoder for Survival Prediction, arXiv:1910.04488.

Almaeen, M., Alanazi, Y., Sato, N., Melnitchouk, W., Kuchera, M.P., and Li, Y. (2021), Variational Autoencoder Inverse Mapper: An End-to-End Deep Learning Framework for Inverse Problems, International Joint Conference on Neural Networks (IJCNN) Proceedings, 1-8.

Mao, S., Cheng, L., Zhao, C., Khan, F.N., Li, Q., and Fu, H.Y. (2021) Inverse Design for SiliconPhotonics: From Iterative Optimization Algorithms to Deep Neural Networks, Applied Science, 11, 10.3390.

Ma, W., Cheng, F., Xu, Y., Wen, Q., and Liu, Y. (2019) Probabilistic representation and inverse design of metamaterials based on a deep generative model with semi-supervised learning strategy, Adv. Mater., 31, 1901111.