IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 15 June 2022, accepted 8 July 2022, date of publication 12 July 2022, date of current version 18 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3190419

RESEARCH ARTICLE

Symmetry Enhanced Network Architecture Search for Complex Metasurface Design

TIANNING ZHANG[1], CHUN YUN KEE[1], YEE SIN ANG[1], (Member, IEEE), ERPING LI[2], (Fellow, IEEE), AND LAY KEE ANG[1], (Fellow, IEEE)

[1]Science, Mathematics and Technology, Singapore University of Technology and Design, Singapore 487372
[2]College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China

Corresponding authors: Tianning Zhang (tianning_zhang@mymail.sutd.edu.sg) and Lay Kee Ang (ricky_ang@sutd.edu.sg)

This work was supported in part by the USA Office of Naval Research Global under Grant N62909-19-1-2047, and in part by the Singapore University of Technology and Design (SUTD)—Zhejiang University (ZJU) Visiting Professor (VP) under Grant 201303. The work of Tianning Zhang was supported by the Singapore Ministry of Education Ph.D. Research Scholarship. The work of Yee Sin Ang was supported by SUTD Start-Up Research Grant SRT3CI21163.

ABSTRACT Using machine learning (ML) techniques like deep learning (DL) for accelerated design (forward and inverse) of metasurfaces has attracted great interest. However, most studies are focused on using relatively regular and less complex patterns for specific photonics applications. In this paper, we report significant improvements of our prior developed DL model tested on complex and random metasurfaces by combining the the Network Architecture Search (NAS) method and the spatial symmetry information of the complex metasurfaces. It is found that a shallow and wide neural network will provide better performance for the complex and physics based metasurfaces problem, which is in contrast to the deep trend in existing DL models. Our method can now accurately identify the EM response locations from arbitrary random and complex metasurfaces while the conventional models fail to accomplish. It can also accurately predict the EM response curve by injecting correct symmetry information into the architecture design step. Thus this paper offers a platform to distill the influence of different fundamental operations for complex metasurface design problems. In future, it may play an essential role in determining the most suitable neural network for the complex metasurface problems. Finally, we are sharing this home-generated physics-based dataset [SUTD polarized reflection of complex metasurfaces (SUTD-PRCM)] for future testings from the research community that we believe the best DL model is yet to be found.

INDEX TERMS Artificial intelligence, machine learning (ML), deep learning (DL), neural networks, metasurfaces, electromagnetics (EMs).

I. INTRODUCTION

Due to the electromagnetic (EM) wave or light-matter interaction, specified designed metasurfaces in some unique patterns can exhibit remarkable EM responses or light outputs for many applications [1]–[15]. Some examples are heat transforming [16], cloaking [17], [18], hologram [19], conversion [20], absorption [21], [22], scattering [23], polarization [24]–[26], transmission [27], different colors [28], [29], meta-lense [25], [26], programmable metasurfaces [30]–[32], and many others [33]–[36]. These applications are made possible by the rapid advancement in micro- and even nano- fabrication technologies and computational modeling over the past decades.

In the design of these metasurfaces in specified geometrical arrangements, machine learning (ML) methods like deep learning (DL) techniques have demonstrated unprecedented performance in providing rapid and accurate prediction [37] if a well-defined EM response is given. Particularly, the DL technique has been mainly applied for forward modeling and inverse design generation [38]–[51]. For forward modeling, instead of explicitly solving the governing Maxwell equations, DL models can learn the complex and non-linear mapping between input parameters to output EM response with a sufficiently large and high-quality dataset. Compared to computationally expensive EM numerical solvers, efficient

The associate editor coordinating the review of this manuscript and approving it for publication was Su Yan.

and accurate evaluation of the DL model can lead to faster computational time and larger design search dimensions. The inverse design problem aims to return the potential pattern for the desirable EM response. The high-fidelity surrogated model can replace the costly numerical solvers in the traditional design methodology based on evolutionary algorithms such as genetic algorithm (GA) [52], particle swarm optimization (PSO) [53], and ant colonization optimization (ACO) [54]. Usually, a generative adversarial network (GAN) system will involve a forward model that can be numerical solvers or DL models.

The involvement of DL typically started from the exploration with the most basic neural architecture, a fully connected network (FCN) [50], [55], [56] for supervised learning. With this approach, the electromagnetic scattering behavior of alternating dielectric thin films parameterized on thicknesses and dielectric constants of the films were successfully predicted [57]. In dealing with the instability and inconsistency problem, a bidirectional encoder-decoder model (Tandem) is proposed [57]. A recent paper [58] studied this problem when the output is only a scalar parameter. ML algorithms on densely sampled spectral work such as reflection and transmission were also tested [59], [60]. For inverse design, deep generative models are employed for generating new meta-atom designs to achieve the desired EM response. Various groups [59], [61]–[63] have used the GAN system to quantify a differential mapping from desired EM response to the discrete 2D pattern. Another paper [64] further enhances the expression capability of the DL model by appending to CNN a recurrent neural network (RNN) which is more often seen in sequence modeling.

However, we observe that majority of the DL-related works in metasurface design are restricted to canonical shapes or connected polygons, which belong to a relatively limited and straightforward design dimension. The findings reported in using such a limited and regular dataset are also qualitative in comparing different neural architectures. It is less intuitive to make meaningful comparison across different models without a common and more complex dataset. Easy access to such standard dataset will allow a more quantitative and fair comparison between different neural architectures and training strategies in the research community. Inspired by how the standard datasets in CV community has advanced the state-of-the-art in their field, we are ready to share our dataset (SUTD-PRCM), which was first generated in a previous work and tested on different DL models [62]. This dataset is essentially a collection of numerical simulated results of EM wave reflection of randomly created metasurfaces.We used one floquet port with two modes in the unit cell simulation within the frequency domain solver. The reflection spectra are results from S-parameter from CST simulation. We will use the polarization to interpret the result under a more general setting of wave propagation. Each sample in the dataset consists of an input metasurface of 16×16 binary image, and its associated output EM reflection as a function of frequency from 2 to 10 GHz. The randomly generated samples are sufficiently complex that are suitable for forward prediction and inverse design in testing different DL models in a simple GPU setting.

In our recent work [62], we have demonstrated that in using the SUTD-PRCM dataset tested with some existing DCNN based neural architectures that might not be the most optimal neural architecture yet that we will report the further improvements in this paper. Thus, we are motivated to find the best AI architecture for complex metasurface. An important paper [65] reports a very novel way to automatically identify the best architecture and parameter selection for modeling Minkowski Fractal-Based FSS problem. The geometry of their metasurface is controlled by width, height, thickness, and offset length. They use the Bayesian Optimization method in network architecture search and optimize the architecture of a fully-connected regression model (FCRM). It has been proved that their method can reduce sensitivity for training and considerably improve prediction performance. This paper will focus on using 2D patterns to implement a metasurface rather than shape parameters. The deep CNN module is the main machine learning operation in this method. Moreover, we will use a gradient-based method named Differentiable Architecture Search (DARTS) to optimize architecture. We will propose a framework that can help us to understand how to design the potential 'best' DL model for our complex metasurface datasets. We regard a deep learning model as two parts, the basic operations, and the architecture to assemble those operations. We would demonstrate the general approach to find suitable architecture and design suitable fundamental operations.

The focus of this paper is to use a physics-generated dataset to test the performance of some deep learning models, and we conclude the well-known models will not work well in this dataset. Note the model is not aimed to obtain any specific design or performance of metasurfaces. The paper can be viewed as a random-shape-level metasurface machine learning task that tries to deal with complex binary metasurface that has not been studied before. Most literature on metasurface designs will limit the pattern to reach some performance. In this paper, we use a full random metasurface, which has no constrain in its pattern and to test the limit of AI in this complicated dataset by using Deep Convolutional Neural network technology, network architecture search technology, and symmetry model technology. Our experiments aim to identify the best architecture of AI, and also to share this home-made dataset with the research community of further improvement.

The rest of the paper is organized as follows. Firstly, we introduce the main methods used in this paper like Network Architecture Search, geometrical machine learning concept, and symmetric neural network in Section.II. Secondly, we briefly introduce the main experiment setup like the Maximum-peak-binary classification (MPBC) task and Magnitude Regression (MR) task in Section.III. Thirdly, we present the improved performance of these methods compared to the traditional DCNN model in Section.IV. Finally,

we conclude the paper with a summary and possible future works in Section.V. The details of the public SUTD-PCRM dataset are presented in the Supplementary material. Its symmetry property is also shown in Appendix.

II. METHODS

A. NETWORK ARCHITECTURE SEARCH (NAS)

Neural architecture search (NAS) [66] is a technique for automatically designing the structure or layout of the neural network (NN) in the field of machine learning (ML). The objective of NAS is to discover the best architecture for a neural network tailored for a specific requirement based on a given dataset. It represents a set of tools that will test and evaluate many potential architectures in order to select the most suitable one for a given problem by maximizing a fitting function. The most well-known NAS method is Google's NASNet [67], but this method requires thousands of TPU/GPU resources that are not affordable for common research groups based in universities. There are various NAS methods like Reinforcement Learning (RL) Methods [68], Gradient-based (GB) Methods [69], Evolutionary Algorithms (EA) [70] and Bayesian Optimization (BO) [71].

In this paper, we adopt the Differentiable Architecture Search (DARTS) [69], [72], [73], as it requires significantly less computational resources as compared to other NAS methods. DARTS introduces a continuous relaxation scheme that enables differentiable learning objectives, unlike the RL or EA approaches. This differentiability is the key to the computational feasibility following a gradient-based approach. It allows the researchers to apply NAS only using one GPU. Many seek to further improve DARTS and leads to many variants such as PC-DARTS [73], SharpDARTS [74], RC-DARTS [75] and Fair DARTS [76]. In our experiment, we adopt the DARTS [69] with Geometry-Awared gradient algorithm [72] and PC-DARTS setup [73] which can converge quickly and escape from local minima. The PC-DARTS is a memory-efficient and faster variant of DARTS that reduces the redundancy in navigating the search space by performing the search in a subset of channels. The Geometry-Awarded gradient algorithm proposes a better optimization and regularization strategy to achieve fast and high-quality convergence. We modify the branches and blocks in the original DARTS structure [69], [72], [73] and make it feasible for the SUTD-PRCM dataset. Tracking the evolution of the searched architect during the training process of DARTS can help us to summarize the correct structure-property required by the best model. The full framework is shown in Fig. 1. We use 3000 samples from the 108,000 training dataset as the validation set for updating architecture parameters. The remaining 105,000 samples construct the new training dataset and are used for updating model parameters. An algorithm flowchart of DARTS is added in the supplementary materials.

B. GEOMETRIC DEEP LEARNING

ML aims to construct algorithms that can learn the latent rule of a dataset by approximating the underlying nonlinear relationship with the typical objective of accurately predicting a set of input parameters. So the learning algorithm would deal with millions of training examples that demonstrate the intended relation between input and output. After training, the model should approximate correct outputs, even for not appearing samples. Such an ability is called the generalization of the ML model. It is well-known that the success of modern ML algorithms will depend heavily on the correct assumptions about the nature of the target function when they are established. These assumptions are subsumed in the phrase inductive bias [77], which is becoming a popular concept in the ML community after Google put forward this paper [78].

Recently, other researchers also put forward a new concept called Geometric deep learning [79], which is a common mathematical framework to study the ML neural network architectures with incorporate prior physical knowledge. It provides a consistent way to build future architectures by considering geometry and symmetry. Exploiting these known symmetries of an extensive system is a powerful method that enables deep learning systems to exploit the low-dimensional geometry arising from basic physics principles. For example, reveal the space-translation symmetry for grid images, time-translation symmetry in the time-series sequences, and rotation symmetry for molecules. In short, this theory can distill into two concepts: the $\mathcal{G}$-invariant function and $\mathcal{G}$-equivariant property. Generally, a $\mathcal{G}$-invariant function is required. But it is more flexible to use $\mathcal{G}$-equivariant when the Signal has deformation stability. For example, Recurrent Neural Network (RNN) is well-known to conserve the time translation since it can be viewed as a step as a Markov dependence updating processing. Convolution Neural Network (CNN) could hold the spatial translation because of the convolutional operation. Graph Neural Network (GNN) focuses on more interrelation between two meta-unit, with a good affinity for permutations symmetry.

Our Metasurface ML dataset (SUTD-PRCM) has some symmetrical information as shown in Appendix.A. Different branches and parts will have different symmetries. For example, the x(co)-polarization branch owns the $Z2$ symmetry. The y(cross)-polarization complex number branch owns the $P4Z2_{\pi}$ symmetry. Since the traditional DCNN models don't contain the symmetry information, it is straightforward to embed these corresponding symmetrical properties or information into the network and to explore its improvement on DL models and NAS search.

C. SYMMETRIC CNN MODULE

There are a lot of works about how to include symmetry information from the data sets into ML. For example, [80] takes a detailed study about how to integrate the group techniques into kernel learning. The main technique is averaging all

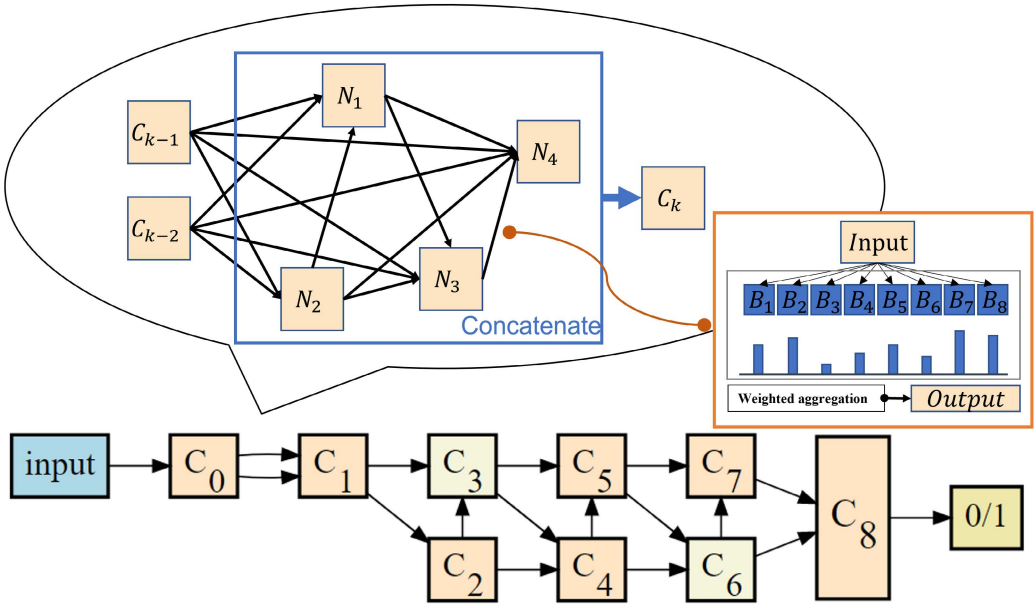


FIGURE 1. The main structure of the network architecture search (NAS) module. The orange and yellow square represent the two architectural types: normal and reduction, respectively. There are nine cells from C_0 to C_8. Each cell is the interaction of the last two cells. The interaction contains 4 intermediate units (N_1 to $N4$). Each unit is the sum aggregation of several mapping (black arrow). There are 14 mappings in one interaction. Each mapping is a weighted aggregation of candidate operations. In this paper, we use eight symmetry operations S_i. A final linear layer is applied to convert C_8 to the binary output of positive 1 or negative 0.

possible nonlinear functions over the symmetry group, which has been used in [81] for biomedical image analysis, [82] for shape matching, and [83]. Another example is the Scattering Convolution Neural Networks (SCNN) [84], which uses wavelet instead of conventional convolution operation to capture transition invariance. SCNN is later extended [85] to deal with translations, rotations, and scaling groups in object recognition. Other research groups have addressed this problem by learning equivariant representations, like equivariant Boltzmann machines [86], [87], and equivariant descriptors [88]. Nevertheless, concepts from locally compact topological groups are used [89] to develop a theory of statistically efficient learning in the sensory cortex, which has achieved great success in speech recognition [90]. In 2015, researchers proved that convolutional networks could exploit the rotation symmetry for galaxy morphology prediction by rotating feature maps [91]. The technique was further extended to solve various computer vision tasks with cyclic symmetry [92].

In this paper, we will use $\mathcal{G}$-average [79] way to build an $\mathcal{G}$- equivariant convolutional layer

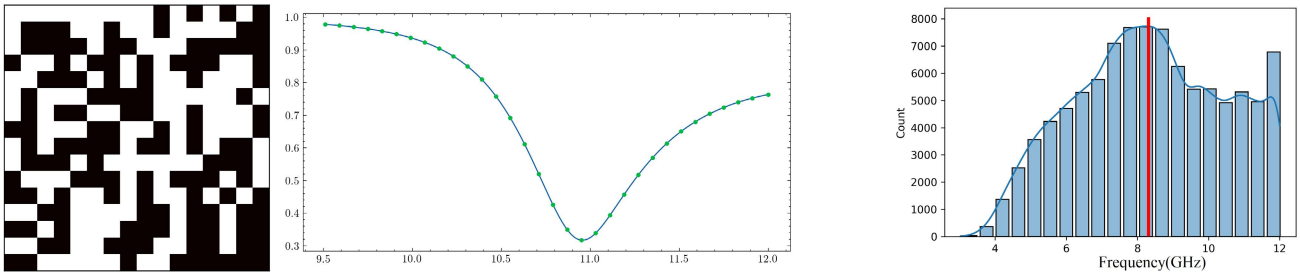
$$\begin{aligned}\tilde{f}(K;x) &= \frac{1}{\mu(\mathcal{G})}\int_{\mathcal{G}} f(g(K;x))d\mu(g) \\ &= \frac{1}{Card(\mathcal{G})}\sum_{g\in\mathcal{G}} f(g(K);x), \end{aligned} \tag{1}$$

where the $\mu(\mathcal{G})$ is the Haar measure of group $\mathcal{G}$, the $\tilde{f}$ is the symmetric convolutional operation and f is normal convolutional operation. When the group $\mathcal{G}$ is a discrete group, the $\mu(\mathcal{G})$ becomes the element counting measure, that is, the Card of the group set $Card(\mathcal{G})$. The $\mathcal{G}$-equivariant operation for the discrete group is the average of all possible results under the symmetry group.

For example,there exists $\mathcal{Z}_2$ symmetry in co-polarized reflection $\mathcal{T}$ in our dataset (see Appendix.A). We convolute the input 4 times with the different kernels under different $\mathcal{Z}_2$ operations. Note that such converting only works on CNN operation with stride $s = 1$. For $s > 1$, CNN operation with a certain kernel config would fail to maintain a complete symmetry property. For example, if we take kernel $k = 3$, stride $s = 2$ and padding $p = 1$, the symmetric CNN module would produce half size $(w \times h)$ feature map for even size input $(2w \times 2h)$ but fail to maintain $Z2$ equivariant property since there is a extra line wouldn't get counted. We would call this CNN type "unfilled." Some operations in our NAS search space are unfilled, so the naive symmetric processing is partially symmetry injecting in strict. To overcome this problem, we would carefully modify CNN's hyper config and result in a completely different search space. For example, the $(k = 3, s = 2, p = 1)$ CNN layer would be replaced by $(k = 2, s = 2, p = 0)$ CNN layer for even size input. The cross convoluting operation named factorized-reduce layer will also be prohibited. It may alter our model's learning ability and result in the searched architect being harder to compare with the previous results. By default, we would demonstrate the performance after naive symmetric converting. We will note if we are using the completely symmetric model.

III. PERFORMANCE EVALUATION TASK

To test the effectiveness of the NAS method, we apply the method to two ML problems: the Maximum-peak-



(a) (b) (c)

FIGURE 2. **(a) One sample pattern from RDN class of metasurface. (b) MR task: The corresponding magnitude spectrum, $|\mathcal{R}|$, for the pattern in (a). (c) MPBC task: The histogram of location of maximum peak in the magnitude spectra, $|\mathcal{R}|$, of RDN class of metasurface. The binary label represents that the peak locates at the left or right side of the middle number (red line) .**

binary classification (MPBC) task and the Magnitude Regression (MR) task.

MPBC task aims to predict whether the maximum peak in a spectrum is located at a frequency larger than a threshold frequency. If the maximum peak is at a frequency larger than the threshold, the binary image is assigned a positive label; otherwise, a negative label. We will focus on the RDN dataset in this Section and set the threshold as $f_0 = 8.31$GHz, which is the median peak frequency of the RDN dataset. It would provide balanced positive and negative labels and avoid unbalanced classification (see Fig.2c). The RDN dataset has 110,000 samples, 108,000 are used as training samples, and the remaining 3000 are test samples. There are 55,364 positive labels and 52,636 negative labels in the training dataset, and 1508 positive labels and 1492 negative labels in the test dataset. The non-trained classifier produces a baseline accuracy of around $1508/3000 = 50.27\%$ according to the roughly equal positive and negative classes in the training dataset.The reason why we choose MPBC is because it will provide the least level for distilling physics knowledge only from images. In this benchmark, we can see the Resnet model and other deep learning models may face overfitting issues when the testing accuracy is poor. For more detail, please refer the Supplementary Material. On the other hand, most DARTS models can converge into a relatively higher accuracy. Meanwhile, the classification task is validated by accuracy rather than Mean Square Error (MSE) loss, where the MSE is a soft metric. For example, when the MSE is small, we are not able to evaluate properly the ''matching of curves'' via MSE. A MSE $= 0.005$ model may perform worse of a MSE $= 0.006$ model when later only perform badly on some large number. By contrast, the accuracy of classification is a very solid metric that directly reflect the performance of predictor.

MR task uses the same dataset in our previews work [62] which pair the input random pattern images and their corresponding EM responses as shown in Fig.2a and Fig.2b. From the prior work, [62], we achieve a preliminary performance for the accurate forward prediction, and we want to figure out whether the NAS model and extra symmetry information would help to improve the forward prediction.

Following a similar methodology in [62], we perform a cross comparison study to evaluate the generalization performance of the new proposed model across different type of patterns. For more details about the dataset, please see the Supplementary Material.

IV. RESULTS

A. NAS

In this section, we benchmark the performance of the DARTS framework to existing ML models using the MPBC task. With the help of the NAS method, we can achieve success on the MPBC task in contrast to failed results by using the conventional machine learning and deep learning method. Nevertheless, we can figure out some key features that a model should have for the physics-embedding dataset.

Firstly, we apply traditional machine learning models such as random forest classifier (RFC) and linear/log support vector machine (SVM) classifier and off-the-shelf neural network architectures, such as deep multi-layer perception (MLP), Resnet18 (RS18), Resnet34 (RS34), ResSimple(RSSM), and SqueezeNet (SQN1) to tackle the MPBC task. ResSimple is a tiny model consisting of 3 layers of Resnet block, which has a similar parameter number to the DARTS model. To our surprise, the results are marginally better than the non-train accuracy (50%). For example, the RFC and SVM achieve 53%, and 56%, respectively. None of them can identify the location of the maximum peak. While deep learning architectures are well known to perform excellent with CV-related tasks, to our surprise, all these neural network models do not score above 60% accuracy, as shown in Fig.3. Note in our prior work [62], a modified version of Resnet18 was reported to achieve excellent performance in a related regression problem formulated to predict the EM response of the RDN class in the SUTD-PRCM dataset. With the initial findings from [62], we speculate that the architecture of the existing neural network used for the SUTD-PRCM dataset might not be optimal yet. To exclude the influence of the size and data normalization, we also test the same combination of ML/DL models on a scaled binarized handwritten digits (MNIST) dataset with the same characteristics of the image as the SUTD-PCRM dataset (binary colored 16×16 image). It shows that those ML models can perform very

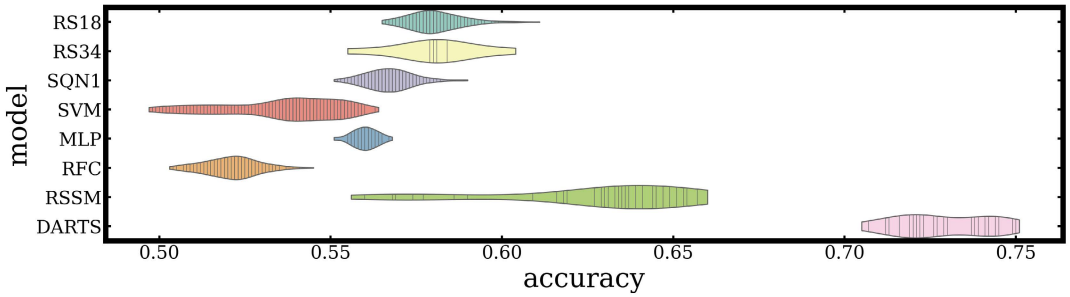


FIGURE 3. Every scattering point is representing one trial with different hyper parameters. Different colors represent different machine learning models. Each model is trained to convergence. The models (on the left with lower accuracy) are the traditional machine learning models listed from top to bottom: RS18 (Resnet18s), RS34 (Resnet34S), SQN1 (SqueezeNet1S), SVM (support vector machine),MLP(Multilayer perceptron),RSSM(ResSimple), RFC (random forest classification machine). The one on the right with much higher accuracy is the best NAS based model (DARTS), which represents the state-of-the-art performance in MPBC task.

TABLE 1. The accuracy of selected ML models on MNIST and MPBC dataset. The red accuracy is the best test accuracy during training; its corresponding training accuracy at that epoch is shown in blue. It means that post-training after that epoch is overfitting, and we will take an early stop.

	SVM	RFC	MLP	SQN1	RSSM	RS18	RS34	DARTS
MNIST	.91/.91	.96/.99	.96/.97	.95/.95	.87/.88	.96/.95	.98/.98	.97/.97
MPBC	.56/.56	.53/.70	.56/.56	.57/.57	.66/.70	.61/.67	.60/.64	.75/.81

well: SVM (91%),RFC (96%),MLP(96%),ResSimple(87%) and SqueezeNet1S (95%), Resnet18S (96%), DARTS (97%), and Resnet34S (98%) as shown in Table 1. Notice, we did not try for finetuning the model's performance on MNIST. So, the affinity between dataset and model is the only reason to explain the failures of traditional models. The optimal architecture searched out by the DARTS framework is able to achieve accuracy over 75% as shown in the last row in Fig.3, which surpasses all the ML models.

The failure of those traditional classification methods is due to the significant differences between the random metasurfaces (RDN) patterns and the MNIST. For example, each image of MNIST is a centralized and continuous image, and each digit (0-9) can smoothly deform to each other. This property restricts the possible patterns to a small subset of a 16×16 binary image domain. On the other hand, the pattern in RDN can be any random binary 16×16 image. Apparently, the complexity of the MPBC problem studied here in our dataset is higher than that of the MNIST digit recognition and thus explains the less accuracy as summarized in Table.1. Notice that if we use an MNIST-like dataset like the PLG dataset of SUTD-PRCM, the deep learning-based model can also do excellent classification as well as DARTS model. A new table is now added in the Supplementary Material. This result implies that the DCNN model is not suitable for dealing with the complex metasurface dataset. Thus it requires new architects and new fundamental components to help AI understand latent physics.

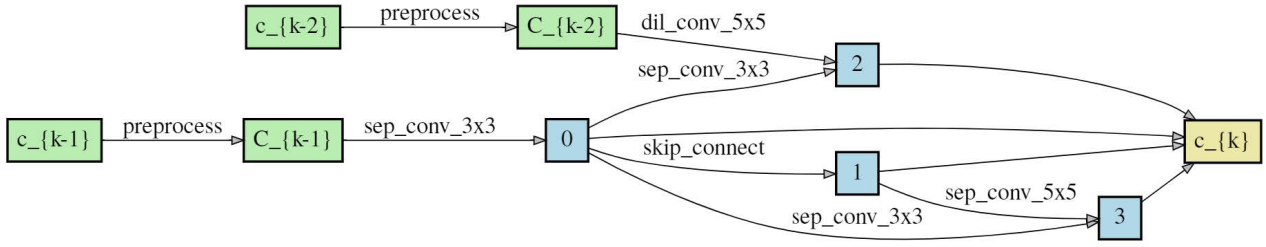
Furthermore, the optimal network architecture suggested by NAS-DARTS converges into a relatively simple structure shown in Fig.4a and Fig.4b which requires only $233,682$ parameters. It is only about $1/5$ and $1/10$ of the requirements from Resnet18S and Resnet34S, respectively. On the other hand, the operation numbers of DARTS, like the number of adding and producing, are much larger than others.

TABLE 2. The comparison between different neural architectures and NAS based neural architecture. The parameters are referring to the total free parameters in all operations. For example, convolution with kernel size (C_2, C_1, w, h) is one operation with $w \times h \times C_1 \times C_2$ parameters.

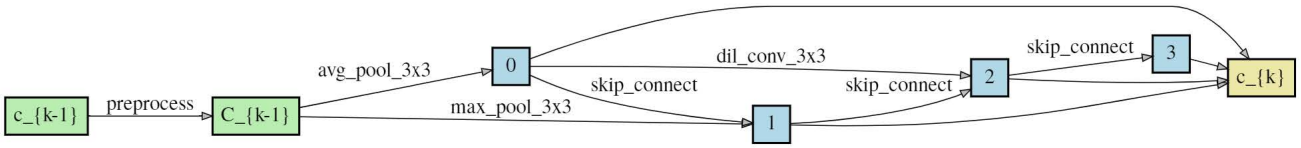
Model	No. Operations	No. Parameters
ResSimple	62	242434
SqueezeNet1S	51	742306
MLP	30	5659074
Resnet18S	62	11174338
Resnet34S	110	21282498
DARTS	194	232034

Table 2 shows the number of parameters and operations of different models used in the SUTD-PRCM (RDN Class) dataset.

Intuitively, more parameters accompanying a bigger model will provide a better capability to capture the inherent relationship between inputs and outputs. Surprisingly, the NAS-DARTS approach suggests an alternative shallow and smaller architecture will perform better than the other models. ResSimple is a tiny model consisting of 3 layers of Resnet block, which has a similar parameter number to the DARTS model.. Such a phenomenon highlights the importance of suitable neural network architecture. Designing a suitable meta-operation may be more effective than building a large and deep neural network architecture. The detailed structure of our NAS-based DL model (see Fig.4) has shown that convolution stacking is not the dominant element of the architecture anymore. It preferred low-level features over deep hierarchical high-level features, which is common for traditional CNN. Thus, CNN-based models like Resnet and SqueezeNet are not performing well in the RDN-SUTD-PRCM dataset. By investigating different search results for the different performances, we can conclude that a wide network with a fruitful operation combination is more suitable for a deep network with simple stacked operations. More searched out structures can be found in the Supplementary.



(a) Normal branch.



(b) Reduce branch.

FIGURE 4. **The final truncated architecture produced by NAS. (a) is the normal branch. (b) is the reduce branch. Only the first and the second important path are kept for each node.**

The successful implementation of the MPBC task gives us confidence that the NAS method is a valid technology for investigating the influence of architecture on the machine learning of our random complex metasurfaces dataset. One concern is that the performance of DARTS cannot reach the same level of accuracy demanded by the traditional dataset like MNIST. One possible reason is the fundamental operation used in our NAS presented above is still based on the traditional CNN network, which may not fully fit the inductive bias of the Metasurface dataset. We then would investigate the influence of fundamental operation by changing the search space to improve the search strategy of DARTS.

B. EFFECT OF SYMMETRY

To reduce the reliance on large data and improve the interpret ability of ML models, there is growing research that effort to incorporate prior knowledge into training. The symmetry, which is a recurring theme in understanding a physics-based problem, is of our interest to enforce physical consistency in DL models. After performing an exploratory study on the SUTD-PRCM dataset, we identified a few symmetry groups, $\mathcal{P}_4\mathcal{Z}_2$ and $\mathcal{Z}_2$, that are associated with the cross-polarized ($y-$) and co-polarized ($x-$) reflection magnitude spectrum, $|\mathcal{R}|$ and $|\mathcal{T}|$, respectively.

We test the effectiveness of symmetry by using the MPBC problem mentioned above which theoretically holds the Z2 symmetry since it bases on the $|\mathcal{T}|$ branch. The DARTS model in the previous section is, hereinafter, referred to as DartsB (without symmetry information). The convolutional layers in DartsB are replaced by their $\mathcal{P}_4\mathcal{Z}_2$ and $\mathcal{Z}_2$ symmetric version with the resulting models denoted as DartsB_P4Z2 and DartsB_Z2, respectively. Next, we will search a new architecture with the symmetric DARTS search space to see whether it can reach a structure better than Darts_Z2. We name this architecture DartsB_Z2R. Notice, the symmetry information in above structure is not necessarily be retained throughout all the layers. Notice, symmetry information in above structure is not necessarily be retained throughout all the layers.Firstly, there are conventional convolutional layers in the pre-processing steps as shown in Fig.4. These layers do not preserve symmetry. Secondly, it is impossible to preserve symmetry under some kernel configurations. Under these settings, only partial symmetry is enforced. The network will be biased towards preserving symmetry but not fully compliant. Furthermore, we create a fully symmetry compliant version that allows only kernel configuration capable of preserving symmetry as shown in Section.II-C, denoted as DartsB_Z2P. For comparison, Resnet18S is adapted to a fully symmetric version denoted as SymResnet. All the models are initialized randomly and trained to convergence. The performances are summarized in Table 3. By simply replacing the operation with the symmetric version, we observe 0.87% improvement from DartsB_NORMAL to DartsB_Z2. Replacing the operation with an inappropriate symmetry version will degrade the performance as demonstrated from DartsB_NORMAL to DartsB_P4Z2 (a drop of 7.2%). Surprisingly, the DartsB_Z2R architecture searched from a sketch will perform poorer than the modified model. It is because of the inconsistent symmetry information in passing its influence during the searching processing, which will cause a structure deeper than the DartsB_NORMAL. However, the MPBC gains more benefits from a wide and shallow network. Thus, the partial symmetry injection may make the searching and training unpredictable. On the other hand, if full symmetry is preserved, a 1.77% improvement is observed from DartsB_NORMAL to DartsB_Z2P. For comparison, SymResnet gains 8% accuracy after conversion from Resnet18S to the full symmetric version.

TABLE 3. The accuracy of best performing DARTS binary classifier (DartsB) with different symmetry configuration tested on MPBC problem. The DartsB_Z2 and DartsB_P4Z2 are the symmetric version of the DartsB model (DartsB_NORMAL) - without prior knowledge of symmetry. The DartsB_Z2R is the new architecture searched out from the Z2 symmetry operation space directly. The DartsB_Z2P is the new architecture searched out not only from the Z2 symmetry operation space but also modified all the operations carefully that make the whole model symmetry equivalent. The SymResnet is the symmetry equivalent version of Resnet structure.The accuracy is tested on the same MPBC dataset, which contains the $\mathcal{Z}_2$ symmetry. The results suggests that directly adding proper symmetry would slightly improve the performance of the MPBC task. The DartsB_Z2P shows the symmetric searched model would perform about 2% better than the non-symmetry model. The DartsB_Z2R and DartsB_Z2 results show the comparison between the influence of architecture and operation.

dataset/model	DartsB_Z2	DartsB_P4Z2	DartsB_NORMAL	DartsB_Z2R	DartsB_Z2P	SymResnet
MPBC($\mathcal{Z}_2$)	75.97%	67.90%	75.10%	73.14%	76.87%	70.01%

To study the impact of injecting symmetry on the spectrum prediction problem, we look into the co-polarized and cross-polarized magnitude spectrum $|\mathcal{T}|$ and $|\mathcal{R}|$ in the SUTD-PRCM dataset. Symmetry group $\mathcal{P}_4\mathcal{Z}_2$ and $\mathcal{Z}_2$ are identified for $|\mathcal{R}|$ and $|\mathcal{T}|$, respectively. For each set of data corresponding to different polarization, the NAS-GAEA-DARTS method (see Sec.IV-A) will search for the best-performing architecture within DARTS_Z2, DARTS_P4Z2, and DARTS_NORMAL search space. DARTS_NORMAL search space does not have symmetric operations. DARTS_Z2 contains operation with $\mathcal{Z}\in$ symmetry, which is the most suitable symmetry for $|\mathcal{T}|$ but not the most ideal for $|\mathcal{R}|$. DARTS_P4Z2 includes the operation with $\mathcal{P}_4\mathcal{Z}_2$ symmetry, which is the most suitable symmetry for $|\mathcal{R}|$ but not appropriate for $|\mathcal{T}|$. The DARTS_NORMAL search space consists of 7 block candidates, including 1 zero operation block, 2 pooling blocks, and 4 convolutional blocks as shown in Fig.1. The DARTS_P4Z2 and DARTS_Z2 search space will convert 4 CNN blocks into corresponding symmetry versions. Although the symmetry group $\mathcal{Z}_2$ is subsets $\mathcal{P}_4\mathcal{Z}_2$, there is no subset relationship between the two search spaces, DARTS_Z2 and DARTS_P4Z2, since an average is involved as illustrated in Eq. (1).

Table 4 summarizes the best performing model resulting from each search space of different symmetric properties, $\mathcal{Z}_2$, $\mathcal{P}_4\mathcal{Z}_2$ and without-symmetry. The scores under the same column are the performances of the best architecture obtained from the NAS method searching in the same search space but with different datasets. The scores in the same rows are the performance of the best performing model resulting from the NAS method searching in different search spaces for the same dataset. For example, the performance of DARTS_Z2 on $|\mathcal{T}|$ is obtain from two step: Firstly, we apply DARTS_Z2 model(whose search space are symmetric modules) on the RDN-$|\mathcal{T}|$ dataset and search out the best architecture; Secondly, we train this architecture on RDN-$|\mathcal{T}|$ dataset and produce the score in Table 4 It is evident that injecting appropriate symmetry has led to better performance (lower error shown in the table) as observed from DARTS_NORMAL to DARTS_Z2 for $|\mathcal{T}|$ and from DARTS_NORMAL to DARTS_P4Z2 for $|\mathcal{R}|$. On the hand, injecting inappropriate symmetry may degrade the performance. Note if the model gets the wrong symmetry information, it must overcome the inconsistency, which will produce poorer performance. Note the Z2 symmetry contain the P4Z2 symmetry which means that if a system has P4Z2 symmetry it much have Z2 symmetry; By contrast, if the system has only Z2 symmetry, it may not have P4Z2 symmetry. Thus the Z_2 model can still performance well on P4Z2 dataset. Overall, injecting appropriate symmetry into an existing model is beneficial. If fully symmetric architecture is searched from scratch, further improvements can be achieved.

C. GENERALIZATION VERSUS PERFORMANCE

In our previous work [62], we find the RDN108000 model can perform well on its RDN dataset but lose its generalization on its subset dataset: PLG and PTN. In Table.5, we can see that the models searched by DARTS under proper symmetry can achieve better performance than the Deep Neural Network as well as the test score on PLG and PTN. This comparison directly shows that the symmetry NAS model holds more generalization than the DNN model. If the DARTS_NORMAL model is searched under unsymmetric candidate spacing, we find that its performance score can also reach a good value close to the DCNN model. However, the smaller scores on PLG and PTN patterns suggested it has better generalization than DCNN. Symmetric realization of RDN108000 is impossible because most CNN operations are unfilled, and the tail-layer of RDN108000 will flatten the feature map, which destroys the symmetry passing. Partially symmetric converting would make the model hard to train. From our results, only the pure symmetric version of RDN108000, which replaces all CNN kernels to a filled hyper configuration and applies an all-in-one pooling layer at the tail, can achieve successful training. We note this model as Z2_RDN. However, such a model is quite different from the original model proposed [62]. Our findings also show that the RDN108000 and Z2_RDN can achieve similar performance but take less generalization than the DARTS model. The DARTS_Z2 model so far is the best architecture and model for our regression task. It shows better performance in both the native score (in RDN) and cross scores (in PLG and PTN). We notice that the DARTS searched architecture on regression usually has more trainable operation than the architecture search from MBPC to enhance the ability to parse curve information. Thus it can gain benefit from partial symmetry injecting. We also searched for the DARTS model named DARTS_Z2P from a pure symmetric search space and modified its pre-processing branch following the same spirit as DartsB_Z2P. The pure-symmetric requirement will change some CNN kernels and remove some operations as introduced in Sec.II-C. Such modification will decrease the

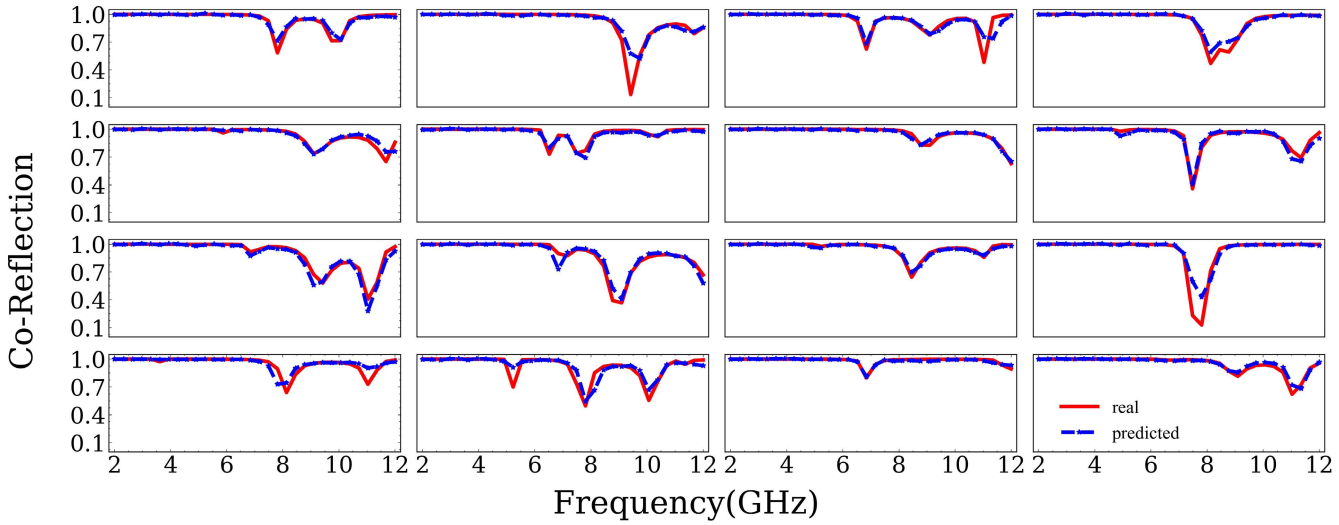


FIGURE 5. 16 randomly selected prediction results (reflection coefficient as a function of frequency) from test dataset. Differing from Resnet18S [62], there is no mode collapse issue in the DARTS model. The overall shapes of the spectra are captured with location of dips correctly identified.

TABLE 4. The best performance that DARTS_Z2, DARTS_P4Z2 and DARTS_NORMAL can achieve for each dataset. Notice the models in same colume, for example DARTS_Z2 for $|\mathcal{T}|$ and $|\mathcal{R}|$ are totally different architecture but only with same architecture search spacing. The result shows with property symmetry adding, the performance can indeed get improvement. Lower score is better performance.

Dataset/Model	DARTS_Z2	DARTS_P4Z2	DARTS_NORMAL
$\lvert\mathcal{T}\rvert(\mathcal{Z}_2)$	0.00322	0.00522	0.00521
$\lvert\mathcal{R}\rvert(\mathcal{P}_4\mathcal{Z}_2)$	0.00549	0.00498	0.00736

ability of the DARTS model, resulting in a slightly worse performance than the DARTS_Z2. However, it still performs better than the Sequential model RDN1080000, as shown in Table 5. In Table.5, one can find that RDN10800 outperforms the DARTS_NORMAL on the RDN dataset, which is quite surprising since the DARTS architecture is supposed to be suitable for the task. From the view of minimizing loss, the DARTS_NORMAL does perform poorer than the RDN108000. We speculate there are several reasons. Unlike the MPBC task, now the model can receive more information for the curve side. Some parts of the artificial architecture models may naturally be good at dealing with the curve information. This explains why the Resnet model can train well on the Regression task but fail in the classification task—for example, the tail layer. In RDN108000, it is a dual-layer Linear block but only has one plain Linear operation at the end of DARTS. Those settings in DARTS are fixed and do not join the architecture optimization, which may limit the ability of NAS. We believe a better score is possible if we use the pre-trained DARTS model and vary the tail layer structure. Such a problem will be studied in the future. Moreover, the MSE score is a soft metric. A small difference in such a low score may not well reflect the difference between ''performance of curve matching''. Our previous study shows that the RDN108000 has become an extremely statistical performance fitting. Thus, it loses part of the ability to be aware of the fruitful peak-curve behavior among datasets (the physical findings we want to achieve eventually). This can be proven by looking at its cross score, where the Z2_RDN and DARTS_NORMAL perform better than RDN108000, while their native score is worse than RDN108000. For a normal model, if it eventually acts as a fitting machine, we think it has a tradeoff behavior between better generalization and accuracy since the statistic behaviors between RDN, PLG, and PTN are totally different: a better accuracy means better statistic fitting on the native dataset and worse performance on other datasets. Under this scenario, our goal is to find a way to help the model to escape from the statistic-fitting trap and to handle both accuracy and generalization. The first step is using the Symmetry information as shown in the DARTS_Z2 model, which will improve both naïve and cross scores. However, the bad cross score suggests that more improvements are required in the future studies. These results prove the feasibility of NAS method studied in this paper.

In Fig.5, we show 16 predicted results that are randomly selected from the test dataset. Notice we train on the fully random complex dataset and completely isolate the training set and test dataset. Thus it is quite difficult to realize perfect prediction like DL model on transitional task. The improvements we made in this paper, like using the NAS approach and symmetry information, have greatly improved our prior works [62]. Our improved model can now correctly predict the peak resonance location with a small error between the predicted values and the target ones. Although the model will sometimes make ''less estimation'' for the peak value, it can accurately predict the trend of the curve and the location of peaks which is excellent in the complex metasurface regression task so far. There is still room for improvement in our NAS framework, like we still register the CNN block as the main equivariant layer. The RDN dataset of SUTDPRCM shows much ''weaker'' transition symmetry than our PLG dataset or other traditional computer vision tasks (see Appendix.A and Supplementary Materials). Thus we may design other equivariant operations in future works

TABLE 5. The comparison of generalization performance using mean square error. RDN108000, PLG27000 and PTN27000 are the best performing Resnet18S models obtained previously [62]. The DARTS_Z2 and DARTS_NORMAL model are consistent with Table 4. The Z2_RDN model is the fully symmetric version of Resnet18S. However, its kernel configuration is adapted to preserve full symmetry as discussed in Sec.II-C. The models are tested on SUTD-PRCM dataset for co-polarized ($x-$ reflection) reflection of PLG, PTN, and RDN classes of metasurfaces.

dataset	DARTS_Z2	DARTS_Z2P	RDN108000	Z2_RDN	DARTS_NORMAL	PLG27000	PTN27000
PLG $\lvert\mathcal{T}\rvert$	0.010433	0.012654	0.014787	0.012210	0.011804	0.000408	0.016790
PTN $\lvert\mathcal{T}\rvert$	0.011838	0.010720	0.013552	0.012102	0.010357	0.007517	0.004382
RDN $\lvert\mathcal{T}\rvert$	0.003216	0.003891	0.004209	0.004522	0.004713	0.017448	0.012390

to solve this ''weak'' transition and connect symmetry of the SUTD-PCRM dataset to achieve better performance.

V. CONCLUSION

In this paper, we use the Differentiable Architecture Search (DARTS) based Neural Architecture Search (NAS) method to achieve better performance than our prior paper [62]. We show the advantages over the conventional ML methods such as SVM and Resnet for the dataset that is a physics-based problem such as our complex random metasurfaces (SUTD-PRCM dataset). The NAS-DARTS results could achieve high accuracy on the Maximum-peak-binary classification (MPBC) task reported above, while the conventional deep learning models fail to predict. Apart from providing a straightforward method to investigate the influence of the model's architecture, it can also help us to understand the fundamental operation's effect. We show this advantage by introducing proper symmetry into the search space. The NAS-DARTS framework would provide a better architecture that performs better than our previews work [62]. From the results searched by NAS-DARTS, we find the SUTD-PRCM dataset prefers shallow and wide neural networks, which is a surprizing finding as compared to the normal experience in computer vision. To realize a feasible and general Deep learning model for the complex metasurface problem, it is essential to implement proper fundamental operations with the correct symmetry information from the dataset. With the modern approach of inverse design, which often features a fast surrogate model in terms of DNN, this finding has a profound impact on the AI design applications of complex metasurfaces.

In future work, one can probably improve by fine-tuning the NAS method, which is beyond the scope of this paper. In particular, making a tradeoff for search space of neural architecture can be tedious and frustrating. The inverse design framework for the complex metasurface problem was introduced in our previous work [62], so it is also an interesting question on how to use the NAS method or symmetry effect in such an inverse design problem. Through performing exploratory analysis with this physics-based SUTD-PRCM dataset, we also hope to alleviate some of the difficulties, and this will be investigated in our future works, including in using complex number-based DL models. It is clear that the CNN block is not the best choice to match the inductive bias of this physics based dataset (SUTD-PCRM). We may also design other equivariant operation that fits the ''weak'' transition and connect symmetry condition (see Appendix). By posting the details of our dataset in the Supplementary and public domain, we welcome the AI community to use our physics-based dataset (SUTD-PCRM) to find the most suitable DL model.

APPENDIX A SYMMETRY OF DATASET

The SUTD polarized reflection of complex metasurfaces (SUTD-PRCM) dataset has global transitional symmetry, local transitional symmetry and connecting symmetry as shown in Fig.6. Firstly, the 1st image I_1 and 2nd image I_2 hold the same pattern - a binary dog - at different corners in the 16×16 design panel. Figure 7 shows $|\mathcal{R}_1|$ and $|\mathcal{R}_2|$, which are their corresponding EM- response simulated curve. These two curves (blue and green) are overlapping as the same under the machine precision limitation, which means the dataset holds the global transition symmetry, which implies that by moving the whole pattern in any direction, and any distance, the response curve will not be affected. Such property is similar to the traditional computer vision classification task - no matter where the dog is, the label is always set 1. A quick explanation is that we use the periodic boundary condition in simulation, so the symmetry spontaneously hold global transition property.

In Fig.6, the 3rd image I_3 and 4th image I_4 break this symmetry by adding fixed pixels at three corners. In such a case, translation invariance would not conserve anymore. However, their response curves $|\mathcal{R}_3|$ and $|\mathcal{R}_4|$ (in Fig.7) show that a transition operation in the pattern side only induces a negligible influence. It implies that the deformation stability could hold under transition symmetry. From the geometric view points [79], the dataset holds the ''approximate'' local transition symmetry, which is noted as $\mathcal{T}$-equivariant. Such a property implies a good affinity between convolutional operation and the dataset.

If we continuously move the 'dog' pattern from the right-down corner to the left-down corner (5th image I_5) and check the variation in the response side. We find those variations would remain very small until this 'dog' encounters another metasurface unit. In Fig.7, it shows that the response $|\mathcal{R}_5|$ becomes extremely large and induces a metasurface resonance peak at a certain frequency. A possible physics explanation is that the induced current suddenly changes when two isolated blocks meet, so the EM response appears to be a phase transition. This scenario shows that the SUTD-PRCM dataset (a physics based dataset) differs from the traditional Computer Vision (CV) task, which is usually interpreted as a divisible pattern segment [93]. In this scenario, the pattern in I_5 should not be regarded as 'dog at the corner'; Instead, it should be viewed as a new image that integrates two blocks.

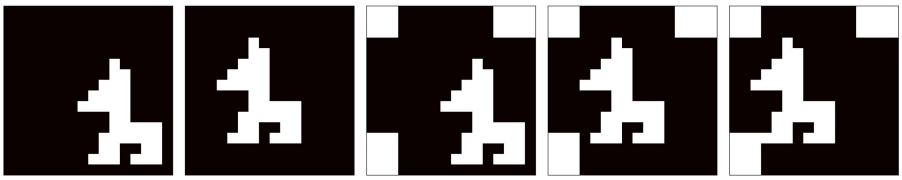


FIGURE 6. Five dog patterns demonstrate the "transition symmetry" in the SUTD-PRCM dataset. Patterns 1 and 2 reveal the global transition symmetry. Patterns 3,4 and 5 have three extra corner patches which break the global transition symmetry but hold'weak' $\mathcal{T}$-equivariant. Pattern 5 reveals the contact/connect property of the SUTD-PRCM dataset which does not hold overlapping invariant.

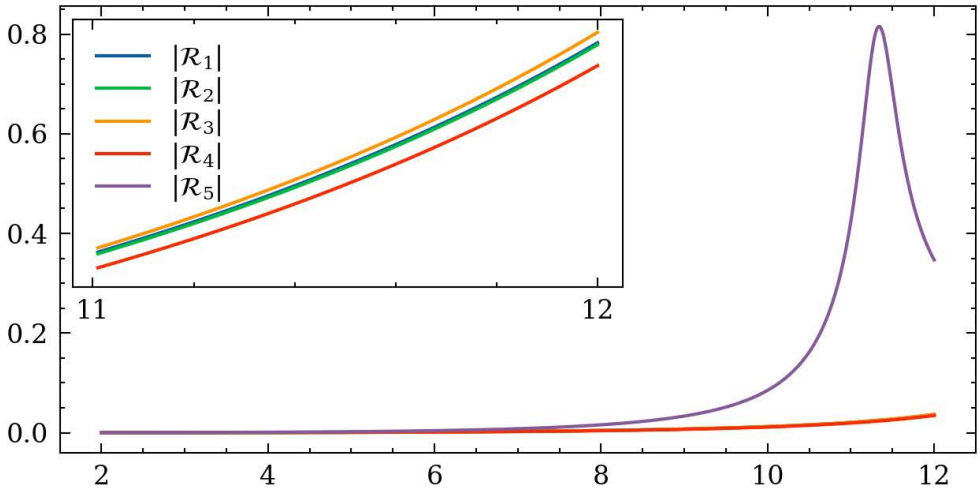


FIGURE 7. The target Response for Fig.6. The Response $|\mathcal{R}_1|$ to $|\mathcal{R}_4|$ is very close to each other which proves the $\mathcal{T}$-equivariant property of the dataset. The Response $|\mathcal{R}_5|$ has a sharp peak induced by the encounter of 'dog' and 'corner'.

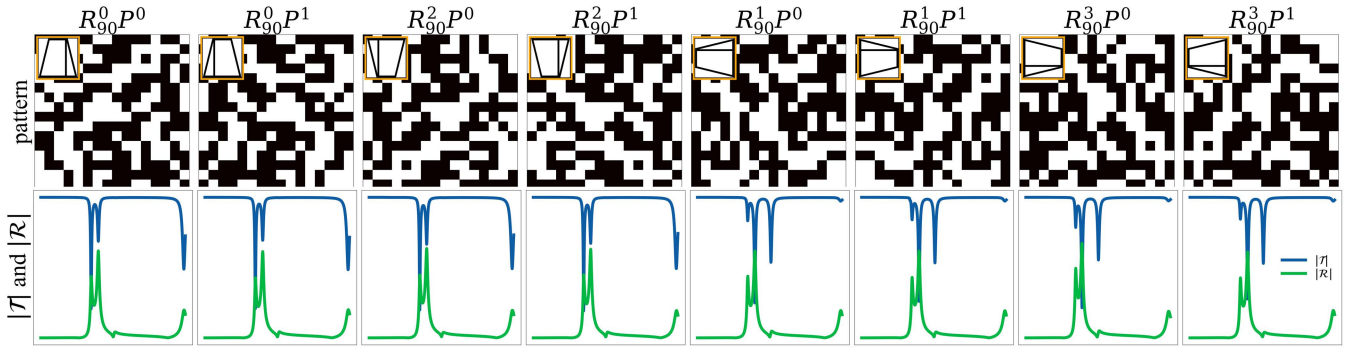


FIGURE 8. The global rotation reflection symmetry test. The 1st row is 1 random pattern and its 7 symmetry partners. From left to right, they are identity, horizontal flip, 180 degree rotation, vertical flip, 90 degree rotation, flip along the lower-left upper right diagonal, 270 degree rotation and flip along the lower-right upper left diagonal. We use an asymmetry ladder to clearly view the direction situation for each pattern at the left-upper corner. The ladder is not a part of pattern. The colour varying from blue to green represents the frequency varying from 2 to 12 GHz. The 2th row shows the magnitude value of two branch $|\mathcal{R}|$ and $|\mathcal{T}|$. It shows the indensity of cross-polarized reflection shows complete rotation reflection $\mathcal{P}_4\mathcal{Z}_2$ symmetry and indensity of co-polarized reflection only has $\mathcal{Z}_2$ symmetry.

In comparison, for most CV tasks, the active responses (label or segment) are required to distinguish from the overlapping object. In our metasurface dataset, latent physics is not divisible by pattern, so the Deep learning model is required to handle these type of scenario. On the other hand, the metasurface fabrication industry may focus more on those resonance peaks induced by the irregular block combination. Thus, generating these kind of examples, which break the local transition symmetry will be important for the inverse design using AI deep learning. We may call this property in the SUTD-PRCM dataset "connect character or connect symmetry". Since it only breaks the $\mathcal{T}$-equivariant when overlapping, we will name such symmetry as 'weak' $\mathcal{T}$-equivariant.

The SUTD-PRCM dataset has global rotation $\mathcal{P}_4$ symmetry and reflection $\mathcal{Z}_2$ symmetry as shown in Fig.8. More specific, different branches in SUTD-PRCM dataset has different rotation/reflection
symmetry:

- The co-polarized reflection $|\mathcal{T}|$ has flip symmetry $\mathcal{Z}_2$.
- The cross-polarized reflection $|\mathcal{R}|$ has full rotation and flip symmetry $\mathcal{P}_4 + \mathcal{Z}_2$.

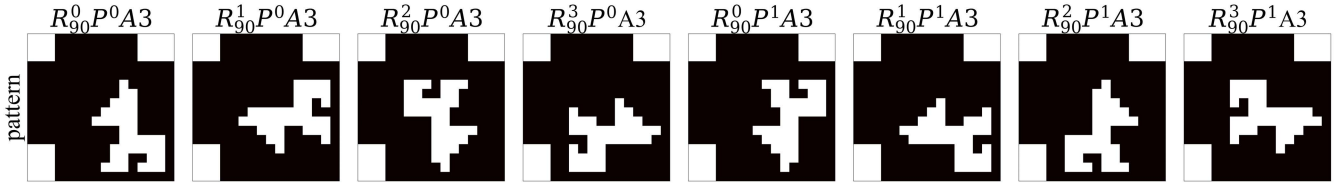


FIGURE 9. The local rotation reflection symmetry test on 'dog' image. The 1st row is 1 dog pattern and its 7 symmetry partners. The 'A3' suffix indicate there are 3 corner pixels to break the global transition symmetry.

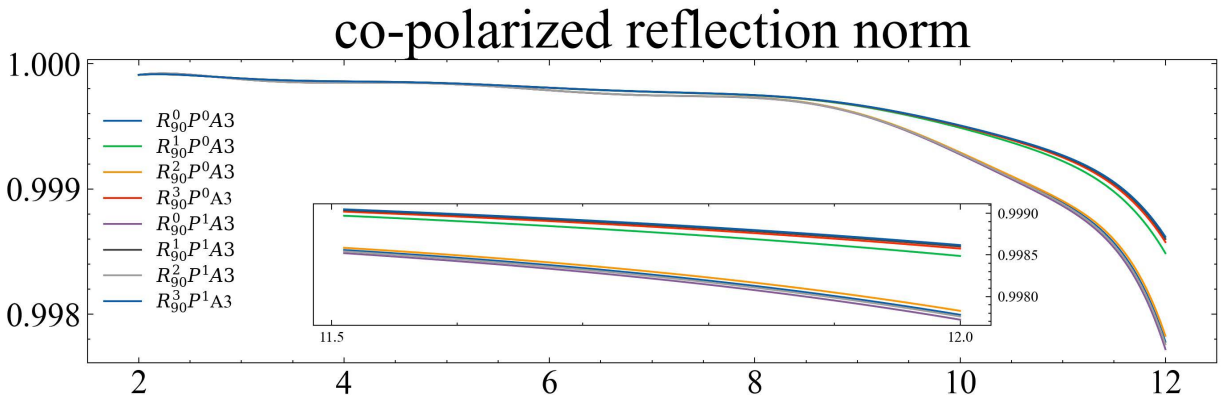


FIGURE 10. The norm of co-polarized reflection from 2 to 12 GHz for 8 different local rotation reflection objects. They will be clearly divided into two classes by a R_90 difference, which implies the $\mathcal{Z}_2$ symmetry of co-polarized reflection. The high frequency curve will get slight variation. This variation is much smaller than the difference caused by structure difference, and it followed the deformation stability condition, so the dataset holds the local $\mathcal{Z}_2$-equivariant.

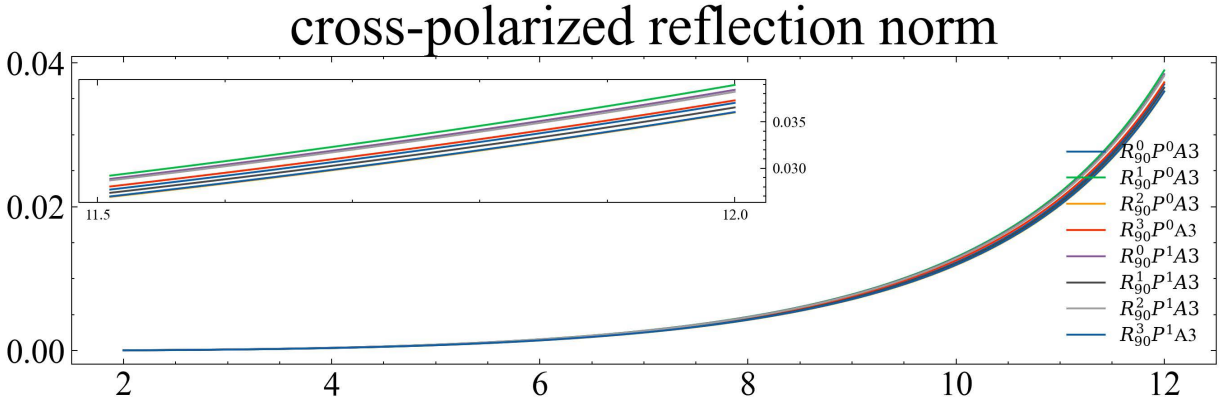


FIGURE 11. The norm of cross-polarized reflection from 2GHz to 12GHz for 8 different local rotation reflection objects. They are almostly same at low frequency and get slight diff at high frequency region. This shows consistent $\mathcal{P}_4\mathcal{Z}_2$ symmetry of global cross-polarized reflection. The difference is much smaller than the difference caused by structure and it follows the deformation stability condition, so the dataset holds the local $\mathcal{P}_4$-equivariant + $\mathcal{Z}_2$-equivariant.

- The totally reflection $P = \sqrt{|\mathcal{T}|^2 + |\mathcal{R}|^2}$ only has flip symmetry $\mathcal{Z}_2$.

Fig.8 shows the full behavior of a fully random metasurface under global rotation-reflection($\mathcal{P}_4\mathcal{Z}_2$) symmetry. The first row gives the view of a random pattern and its seven symmetry partners under $\mathcal{P}_4\mathcal{Z}_2$ symmetry. At the left-upper corner, there is an anisotropy ladder to point out the rotation and reflection direction. It is not part of the pattern. It shows the intensity of cross-polarized reflection shows complete rotation reflection $\mathcal{P}_4\mathcal{Z}_2$ symmetry and intensity of co-polarized reflection only has $\mathcal{Z}_2$ symmetry.

The local rotation and reflection symmetry will be shown via the 'dog' image too. In Fig.9, we show all the 'dog' images under $\mathcal{P}_4\mathcal{Z}_2$ symmetry. Notice there is no exact centre of the 'dog', so we apply the symmetry operation on the 10×10 block at the lower-right corner. There are only slight differences between the eight symmetry curves. Fig.10 and Fig.11 show the norm values ($|\mathcal{T}|$ and $|\mathcal{R}|$) of the 8 curves. The Fig.10 shows the local $\mathcal{Z}_2$ symmetry of the co-polarized reflection. The eight curves are divided into two groups (see the sub figure), and each group holds only small variation within. Compared with the inter-groups gap and intra-groups variation, we know the variation caused by the symmetry is much smaller than the difference due to the structure. Thus for local rotation symmetry and reflection, it is better to consider $\mathcal{P}_4\mathcal{Z}_2$ or $\mathcal{Z}_2$ symmetry under deformation stability conditions.

APPENDIX B DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available at the following URL/DOI:

- https://github.com/veya2ztn/SUTD_PRCM_dataset
- https://github.com/veya2ztn/MetaSurface

REFERENCES

[1] O. Quevedo-Teruel, H. Chen, A. Díaz-Rubio, G. Gok, A. Grbic, G. Minatti, E. Martini, S. Maci, G. V. Eleftheriades, and M. Chen, "Roadmap on metasurfaces," *J. Opt.*, vol. 21, no. 7, 2019, Art. no. 073002.

[2] R. Li, Z. Wang, and H. Chen, *Metamaterials and Negative Refraction*. Cambridge, U.K.: Cambridge Univ. Press, 2020.

[3] A. E. Cardin, S. R. Silva, S. R. Vardeny, W. J. Padilla, A. Saxena, A. J. Taylor, W. J. M. Kort-Kamp, H.-T. Chen, D. A. R. Dalvit, and A. K. Azad, "Surface-wave-assisted nonreciprocity in spatio-temporally modulated metasurfaces," *Nature Commun.*, vol. 11, no. 1, pp. 1–9, Dec. 2020.

[4] Z. Li, X. Tian, C.-W. Qiu, and J. S. Ho, "Metasurfaces for bioelectronics and healthcare," *Nature Electron.*, vol. 4, no. 6, pp. 1–10, 2021.

[5] R. Zhu, T. Qiu, J. Wang, S. Sui, C. Hao, T. Liu, Y. Li, M. Feng, A. Zhang, C.-W. Qiu, and S. Qu, "Phase-to-pattern inverse design paradigm for fast realization of functional metasurfaces via transfer learning," *Nature Commun.*, vol. 12, no. 1, pp. 1–10, Dec. 2021.

[6] X. Z. Chen, L. L. Huang, H. Mühlenbernd, G. X. Li, and B. F. Bai, "Dual-polarity plasmonic metalens for visible light," *Nature Commun.*, vol. 3, no. 1, pp. 1–6, 2012.

[7] Q. Ma, G. D. Bai, H. B. Jing, C. Yang, L. Li, and T. J. Cui, "Smart metasurface with self-adaptively reprogrammable functions," *Light, Sci. Appl.*, vol. 8, no. 1, pp. 1–12, 2019.

[8] K. Wang, J. Zhao, Q. Cheng, D. S. Dong, and T. J. Cui, "Broadband and broad-angle low-scattering metasurface based on hybrid optimization algorithm," *Sci. Rep.*, vol. 4, no. 1, pp. 1–6, May 2015.

[9] J. Ding, N. Xu, H. Ren, Y. Lin, W. Zhang, and H. Zhang, "Dual-wavelength terahertz metasurfaces with independent phase and amplitude control at each wavelength," *Sci. Rep.*, vol. 6, no. 1, pp. 1–9, Dec. 2016.

[10] J. Ding, S. An, B. Zheng, and H. Zhang, "Multiwavelength metasurfaces based on single-layer dual-wavelength meta-atoms: Toward complete phase and amplitude modulations at two wavelengths," *Adv. Opt. Mater.*, vol. 5, no. 10, May 2017, Art. no. 1700079.

[11] X. Wang, J. Ding, B. Zheng, S. An, G. Zhai, and H. Zhang, "Simultaneous realization of anomalous reflection and transmission at two frequencies using bi-functional metasurfaces," *Sci. Rep.*, vol. 8, no. 1, pp. 1–8, Dec. 2018.

[12] T. J. Cui, S. Liu, and L. Zhang, "Information metamaterials and metasurfaces," *J. Mater. Chem. C*, vol. 5, no. 15, pp. 3644–3668, 2017.

[13] Y. Zhang, C. Fowler, J. Liang, B. Azhar, M. Y. Shalaginov, S. Deckoff-Jones, S. An, J. B. Chou, C. M. Roberts, V. Liberman, M. Kang, C. Ríos, K. A. Richardson, C. Rivero-Baleine, T. Gu, H. Zhang, and J. Hu, "Electrically reconfigurable non-volatile metasurface using low-loss optical phase-change material," *Nature Nanotechnol.*, vol. 16, no. 6, pp. 661–666, Jun. 2021.

[14] S. Abdollahramezani, O. Hemmatyar, M. Taghinejad, H. Taghinejad, A. Krasnok, A. A. Eftekhar, C. Teichrib, S. Deshmukh, M. El-Sayed, E. Pop, M. Wuttig, A. Alu, W. Cai, and A. Adibi, "Electrically driven programmable phase-change meta-switch reaching 80% efficiency," 2021, *arXiv:2104.10381*.

[15] X. Liu, I. V. Shadrivov, K. Fan, and W. J. Padilla, "Experimental realization of a terahertz all-dielectric metasurface absorber," *Opt. Exp.*, vol. 25, no. 1, pp. 191–201, Jan. 2017.

[16] Y. Li, W. Li, T. Han, X. Zheng, J. Li, B. Li, S. Fan, and C.-W. Qiu, "Transforming heat transfer with thermal metamaterials and devices," *Nature Rev. Mater.*, vol. 6, no. 6, pp. 488–507, Jun. 2021.

[17] Y. Yang, L. Jing, B. Zheng, R. Hao, W. Yin, E. Li, C. M. Soukoulis, and H. Chen, "Full-polarization 3D metasurface cloak with preserved amplitude and phase," *Adv. Mater.*, vol. 28, no. 32, pp. 6866–6871, Aug. 2016.

[18] C. Qian and H. Chen, "A perspective on the next generation of invisibility cloaks—Intelligent cloaks," *Appl. Phys. Lett.*, vol. 118, no. 18, May 2021, Art. no. 180501.

[19] C. Liu, Q. Ma, L. Li, and T. J. Cui, "Work in progress: Intelligent metasurface holograms," in *Proc. 1st ACM Int. Workshop Nanosc. Comput., Commun., Appl.*, Nov. 2020, pp. 45–48.

[20] C. Liu, Y. Bai, Q. Zhao, Y. Yang, H. Chen, J. Zhou, and L. Qiao, "Fully controllable pancharatnam-berry metasurface array with high conversion efficiency and broad bandwidth," *Sci. Rep.*, vol. 6, no. 1, pp. 1–7, Dec. 2016.

[21] O. Mitrofanov, T. Siday, R. J. Thompson, T. S. Luk, I. Brener, and J. L. Reno, "Efficient photoconductive terahertz detector with all-dielectric optical metasurface," *APL Photon.*, vol. 3, no. 5, May 2018, Art. no. 051703.

[22] O. Mitrofanov, L. L. Hale, P. P. Vabishchevich, T. S. Luk, S. J. Addamane, J. L. Reno, and I. Brener, "Perfectly absorbing dielectric metasurfaces for photodetection," *APL Photon.*, vol. 5, no. 10, Oct. 2020, Art. no. 101304.

[23] H. Zhang, Q. Cheng, H. Chu, O. Christogeorgos, W. Wu, and Y. Hao, "Hyperuniform disordered distribution metasurface for scattering reduction," *Appl. Phys. Lett.*, vol. 118, no. 10, Mar. 2021, Art. no. 101601.

[24] M. I. Khan, Q. Fraz, and F. A. Tahir, "Ultra-wideband cross polarization conversion metasurface insensitive to incidence angle," *J. Appl. Phys.*, vol. 121, no. 4, Jan. 2017, Art. no. 045103.

[25] M. Khorasaninejad, W. T. Chen, R. C. Devlin, J. Oh, A. Y. Zhu, and F. Capasso, "Metalenses at visible wavelengths: Diffraction-limited focusing and subwavelength resolution imaging," *Science*, vol. 352, no. 6390, pp. 1190–1194, Jun. 2016.

[26] M. Khorasaninejad and F. Capasso, "Metalenses: Versatile multifunctional photonic components," *Science*, vol. 358, no. 6367, Dec. 2017, Art. no. eaam8100.

[27] X. You, R. T. Ako, W. S. L. Lee, M. Bhaskaran, S. Sriram, C. Fumeaux, and W. Withayachumnankul, "Broadband terahertz transmissive quarter-wave metasurface," *APL Photon.*, vol. 5, no. 9, Sep. 2020, Art. no. 096108.

[28] F. Cheng, J. Gao, T. S. Luk, and X. Yang, "Structural color printing based on plasmonic metasurfaces of perfect light absorption," *Sci. Rep.*, vol. 5, no. 1, p. 11045, Sep. 2015.

[29] J. Proust, F. Bedu, B. Gallas, I. Ozerov, and N. Bonod, "All-dielectric colored metasurfaces with silicon Mie resonators," *ACS Nano*, vol. 10, no. 8, pp. 7761–7767, 2016.

[30] L. Bao, Q. Ma, R. Y. Wu, X. Fu, J. Wu, and T. J. Cui, "Programmable reflection–transmission shared-aperture metasurface for real-time control of electromagnetic waves in full space," *Adv. Sci.*, vol. 8, no. 15, Aug. 2021, Art. no. 2100149.

[31] X. G. Zhang, Q. Yu, W. X. Jiang, Y. L. Sun, L. Bai, Q. Wang, C.-W. Qiu, and T. J. Cui, "Polarization-controlled dual-programmable metasurfaces," *Adv. Sci.*, vol. 7, no. 11, 2020, Art. no. 1903382.

[32] X. G. Zhang, W. X. Jiang, H. L. Jiang, Q. Wang, H. W. Tian, L. Bai, Z. J. Luo, S. Sun, Y. Luo, C. W. Qiu, and T. J. Cui, "An optically driven digital metasurface for programming electromagnetic functions," *Nature Electron.*, vol. 3, no. 3, pp. 165–171, 2020.

[33] B. Vasić and G. Isić, "Refractive index sensing with hollow metal–insulator–metal metasurfaces," *J. Phys. D, Appl. Phys.*, vol. 54, no. 28, Jul. 2021, Art. no. 285106.

[34] R. Lin, Y. Zhai, C. Xiong, and X. Li, "Inverse design of plasmonic metasurfaces by convolutional neural network," *Opt. Lett.*, vol. 45, no. 6, pp. 1362–1365, Mar. 2020.

[35] L. Y. Niu, H. C. Zhang, P. H. He, M. Tang, M. Wang, L. P. Zhang, G. D. Bai, J. Mao, and T. J. Cui, "Dual-band and dual-polarized programmable metasurface unit with independent channels," *J. Phys. D, Appl. Phys.*, vol. 54, no. 14, Apr. 2021, Art. no. 145109.

[36] M. K. Taher Al-Nuaimi, W. Hong, and W. G. Whittow, "Nature-inspired orbital angular momentum beam generator using aperiodic metasurface," *J. Phys. D, Appl. Phys.*, vol. 54, no. 27, Jul. 2021, Art. no. 275106.

[37] O. Khatib, S. Ren, J. Malof, and W. J. Padilla, "Deep learning the electromagnetic properties of metamaterials—A comprehensive review," *Adv. Funct. Mater.*, vol. 31, no. 31, Aug. 2021, Art. no. 2101748.

[38] X. Shi, T. Qiu, J. Wang, X. Zhao, and S. Qu, "Metasurface inverse design using machine learning approaches," *J. Phys. D, Appl. Phys.*, vol. 53, no. 27, Jul. 2020, Art. no. 275105.

[39] A. Mall, A. Patil, D. Tamboli, A. Sethi, and A. Kumar, "Fast design of plasmonic metasurfaces enabled by deep learning," *J. Phys. D, Appl. Phys.*, vol. 53, no. 49, Dec. 2020, Art. no. 49LT01.

[40] G. Ding, W. Xiong, P. Wang, Z. Huang, Y. He, J. Liu, Y. Li, D. Fan, and S. Chen, "Spatial phase retrieval of vortex beam using convolutional neural network," *J. Opt.*, vol. 24, no. 2, Feb. 2022, Art. no. 025701.

[41] O. Khatib, S. Ren, J. Malof, and W. J. Padilla, "Deep learning the electromagnetic properties of metamaterials-a comprehensive review," in *Proc. Adv. Funct. Mater.*, 2021, Art. no. 2101748.

[42] C. C. Nadell, B. Huang, J. M. Malof, and W. J. Padilla, "Deep learning for accelerated all-dielectric metasurface design," *Opt. Exp.*, vol. 27, no. 20, pp. 27523–27535, Sep. 2019.

[43] Y. Deng, S. Ren, K. Fan, J. M. Malof, and W. J. Padilla, "Neural-adjoint method for the inverse design of all-dielectric metasurfaces," *Opt. Exp.*, vol. 29, no. 5, pp. 7526–7534, 2021.
[44] Z. Zhen, C. Qian, Y. Jia, Z. Fan, R. Hao, T. Cai, B. Zheng, H. Chen, and E. Li, "Realizing transmitted metasurface cloak by a tandem neural network," *Photon. Res.*, vol. 9, no. 5, pp. B229–B235, 2021.
[45] C. Qian, B. Zheng, Y. Shen, L. Jing, E. Li, L. Shen, and H. Chen, "Deep-learning-enabled self-adaptive microwave cloak without human intervention," *Nature Photon.*, vol. 14, no. 6, pp. 383–390, Jun. 2020.
[46] X. Zhou, Q. Xiao, and H. Wang, "Metamaterials design method based on deep learning database," *J. Phys., Conf.*, vol. 2185, no. 1, Jan. 2022, Art. no. 012023.
[47] W. Ma, F. Cheng, and Y. Liu, "Deep-learning-enabled on-demand design of chiral metamaterials," *ACS Nano*, vol. 12, no. 6, pp. 6326–6334, 2018.
[48] W. Ma, Z. Liu, Z. A. Kudyshev, A. Boltasseva, W. Cai, and Y. Liu, "Deep learning for the design of photonic structures," *Nature Photon.*, vol. 15, no. 2, pp. 77–90, Feb. 2021.
[49] J. Jiang, M. Chen, and J. A. Fan, "Deep neural networks for the evaluation and design of photonic devices," *Nature Rev. Mater.*, vol. 6, no. 8, pp. 1–22, 2020.
[50] S. An, C. Fowler, B. Zheng, M. Y. Shalaginov, H. Tang, H. Li, L. Zhou, J. Ding, A. M. Agarwal, C. Rivero-Baleine, K. A. Richardson, T. Gu, J. Hu, and H. Zhang, "A deep learning approach for objective-driven all-dielectric metasurface design," *ACS Photon.*, vol. 6, no. 12, pp. 3196–3207, 2019.
[51] L. Li, H. Ruan, C. Liu, Y. Li, Y. Shuang, A. Alù, C.-W. Qiu, and T. J. Cui, "Machine-learning reprogrammable metasurface imager," *Nature Commu.*, vol. 10, no. 1, pp. 1–8, 2019.
[52] S. Jafar-Zanjani, S. Inampudi, and H. Mosallaei, "Adaptive genetic algorithm for optical metasurfaces design," *Sci. Rep.*, vol. 8, no. 1, pp. 1–16, Dec. 2018.
[53] Q. Zhang, X. Wan, S. Liu, J. Y. Yin, L. Zhang, and T. J. Cui, "Shaping electromagnetic waves using software-automatically-designed metasurfaces," *Sci. Rep.*, vol. 7, no. 1, pp. 1–11, Jun. 2017.
[54] D. Z. Zhu, E. B. Whiting, S. D. Campbell, D. B. Burckel, and D. H. Werner, "Optimal high efficiency 3D plasmonic metasurface elements revealed by lazy ants," *ACS Photon.*, vol. 6, no. 11, pp. 2741–2748, Nov. 2019.
[55] I. Malkiel, M. Mrejen, A. Nagler, U. Arieli, L. Wolf, and H. Suchowski, "Plasmonic nanostructure design and characterization via deep learning," *Light: Sci. Appl.*, vol. 7, no. 1, pp. 1–8, Dec. 2018.
[56] J. Peurifoy, Y. Shen, L. Jing, Y. Yang, F. Cano-Renteria, B. G. DeLacy, J. D. Joannopoulos, M. Tegmark, and M. Soljačić, "Nanophotonic particle simulation and inverse design using artificial neural networks," *Sci. Adv.*, vol. 4, no. 6, Jun. 2018, Art. no. eaar4206.
[57] D. Liu, Y. Tan, E. Khoram, and Z. Yu, "Training deep neural networks for the inverse design of nanophotonic structures," *ACS Photon.*, vol. 5, no. 4, pp. 1365–1369, Apr. 2018.
[58] T. Asano and S. Noda, "Optimization of photonic crystal nanocavities based on deep learning," *Opt. Exp.*, vol. 26, no. 25, pp. 32704–32717, 2018.
[59] J. Jiang, D. Sell, S. Hoyer, J. Hickey, J. Yang, and J. A. Fan, "Free-form diffractive metagrating design based on generative adversarial networks," *ACS Nano*, vol. 13, no. 8, pp. 8872–8878, Aug. 2019.
[60] J. Jiang and J. A. Fan, "Global optimization of dielectric metasurfaces using a physics-driven neural network," *Nano Lett.*, vol. 19, no. 8, pp. 5366–5372, Aug. 2019.
[61] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[62] T. Zhang, C. Y. Kee, Y. S. Ang, and L. K. Ang, "Deep learning-based design of broadband GHz complex and random metasurfaces," *APL Photon.*, vol. 6, no. 10, Oct. 2021, Art. no. 106101.
[63] X. Han, Z. Fan, Z. Liu, C. Li, and L. J. Guo, "Inverse design of metasurface optical filters using deep neural network with high degrees of freedom," *InfoMat*, vol. 3, no. 4, pp. 432–442, 2021.
[64] I. Sajedian, J. Kim, and J. Rho, "Finding the optical properties of plasmonic structures by image processing using a combination of convolutional neural networks and recurrent neural networks," *Microsystems Nanoeng.*, vol. 5, no. 1, pp. 1–8, Dec. 2019.
[65] N. Calik, M. A. Belen, P. Mahouti, and S. Koziel, "Accurate modeling of frequency selective surfaces using fully-connected regression model with automated architecture determination and parameter selection based on Bayesian optimization," *IEEE Access*, vol. 9, pp. 38396–38410, 2021.
[66] T. Elsken, J. H. Metzen, and F. Hutter, "Neural architecture search: A survey," *J. Mach. Learn. Res.*, vol. 20, no. 55, pp. 1–21, 2019.
[67] X. Qin and Z. Wang, "NASNet: A neuron attention stage-by-stage net for single image deraining," 2019, *arXiv:1912.03151*.
[68] B. Zoph and Q. V. Le, "Neural architecture search with reinforcement learning," 2016, *arXiv:1611.01578*.
[69] H. Liu, K. Simonyan, and Y. Yang, "DARTS: Differentiable architecture search," 2018, *arXiv:1806.09055*.
[70] E. Real, S. Moore, A. Selle, S. Saxena, Y. L. Suematsu, J. Tan, Q. V. Le, and A. Kurakin, "Large-scale evolution of image classifiers," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 2902–2911.
[71] K. Kandasamy, W. Neiswanger, J. Schneider, B. Poczos, and E. Xing, "Neural architecture search with Bayesian optimisation and optimal transport," 2018, *arXiv:1802.07191*.
[72] L. Li, M. Khodak, M.-F. Balcan, and A. Talwalkar, "Geometry-aware gradient algorithms for neural architecture search," 2020, *arXiv:2004.07802*.
[73] Y. Xu, L. Xie, X. Zhang, X. Chen, G.-J. Qi, Q. Tian, and H. Xiong, "PC-DARTS: Partial channel connections for memory-efficient architecture search," 2019, *arXiv:1907.05737*.
[74] A. Hundt, V. Jain, and G. D. Hager, "SharpDARTS: Faster and more accurate differentiable architecture search," 2019, *arXiv:1903.09900*.
[75] X. Jin, J. Wang, J. Slocum, M.-H. Yang, S. Dai, S. Yan, and J. Feng, "RC-DARTS: Resource constrained differentiable architecture search," 2019, *arXiv:1912.12814*.
[76] X. Chu, T. Zhou, B. Zhang, and J. Li, "Fair darts: Eliminating unfair advantages in differentiable architecture search," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 465–480.
[77] D. F. Gordon and M. Desjardins, "Evaluation and selection of biases in machine learning," *Mach. Learn.*, vol. 20, nos. 1–2, pp. 5–22, 1995.
[78] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner, and C. Gulcehre, "Relational inductive biases, deep learning, and graph networks," 2018, *arXiv:1806.01261*.
[79] M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković, "Geometric deep learning: Grids, groups, graphs, geodesics, and gauges," 2021, *arXiv:2104.13478*.
[80] M. Reisert, "Group integration techniques in pattern analysis—A kernel view," Univ. Med. Center Freiburg, Germany, Tech. Rep., 2008.
[81] H. Skibbe, "Spherical tensor algebra for biomedical image analysis," Ph.D. dissertation, Dept. Comput. Sci., Albert-Ludwigs-Universität Freiburg, Germany, 2013.
[82] S. Manay, D. Cremers, B.-W. Hong, A. J. Yezzi, and S. Soatto, "Integral invariants for shape matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1602–1618, Oct. 2006.
[83] R. Kondor, "A novel set of rotationally and translationally invariant features for images based on the non-commutative bispectrum," 2007, *arXiv:cs/0701127*.
[84] J. Bruna and S. Mallat, "Invariant scattering convolution networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1872–1886, Aug. 2013.
[85] E. Oyallon and S. Mallat, "Deep roto-translation scattering for object classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2865–2873.
[86] J. J. Kivinen and C. K. Williams, "Transformation equivariant Boltzmann machines," in *Proc. Int. Conf. Artif. Neural Netw.* Cham, Switzerland: Springer, 2011, pp. 1–9.
[87] K. Sohn and H. Lee, "Learning invariant representations with local transformations," 2012, *arXiv:1206.6418*.
[88] U. Schmidt and S. Roth, "Learning rotation-aware features: From invariant priors to equivariant descriptors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2050–2057.
[89] F. Anselmi, J. Z. Leibo, L. Rosasco, J. Mutch, A. Tacchetti, and T. Poggio, "Unsupervised learning of invariant representations," *Theor. Comput. Sci.*, vol. 633, pp. 112–121, Jun. 2016.
[90] C. Zhang, S. Voinea, G. Evangelopoulos, L. Rosasco, and T. Poggio, "Discriminative template learning in group-convolutional networks for invariant speech representations," in *Proc. 16th Annu. Conf. Int. Speech Commun. Assoc.*, Sep. 2015, pp. 1–5.
[91] S. Dieleman, K. W. Willett, and J. Dambre, "Rotation-invariant convolutional neural networks for galaxy morphology prediction," *Monthly Notices Roy. Astronom. Soc.*, vol. 450, no. 2, pp. 1441–1459, 2015.
[92] S. Dieleman, J. D. Fauw, and K. Kavukcuoglu, "Exploiting cyclic symmetry in convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1889–1898.
[93] C. Molnar, *Interpretable Machine Learning*. Morrisville, NC, USA: Lulu, 2020.

TIANNING ZHANG received the B.S. degree in physics from Zhejiang University, Hangzhou, China, in 2017. He is currently pursuing the Ph.D. degree in science, mathematics, and technology with the Singapore University of Technology and Design. His current research interests include machine learning techniques for arbitrary meta-material design and novel machine learning tool in AI for science.

CHUN YUN KEE received the B.E. degree in engineering science and the Ph.D. degree in electrical and computer engineering from the National University of Singapore, Singapore, in 2010 and 2019, respectively.

He is currently a Research Fellow with the Science, Mathematics and Technology (SMT) Cluster. His current research interests include numerical modeling in electromagnetic problem and semiconductor devices. He is also interested in the application of fractional calculus in capturing non-local dynamics in physical systems.

YEE SIN ANG (Member, IEEE) received the Ph.D. degree in condensed matter physics from the University of Wollongong, Australia, in 2014, under the supervision of Professor Chao Zhang. He is currently an Assistant Professor with the Science, Mathematics and Technology (SMT), Singapore University of Technology and Design (SUTD). His current research interests include the device physics and modeling of high-performance, low-energy electronic, optoeletronic, photonic and valleytronic devices using 2D, topological and quantum materials, and the design of novel electron devices for beyond-CMOS. He is a Lindau Alumni of the 69th Lindau Nobel Laureate Meeting.

ERPING LI (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Sheffield Hallam University, Sheffield, U.K., in 1992.

He is currently a Changjiang-Qianren Distinguished Professor with the Department of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China, and the Dean of the Joint Institute of Zhejiang University—University of Illinois at Urbana–Champaign, Champaign, IL, USA. Since 1989, he has been a Research Fellow, a Principal Research Engineer, an Associate Professor, and the Technical Director of the Singapore Research Institute and University, Singapore. In 2000, he joined the A*STAR Research Institute of High-Performance Computing, Singapore, as a Principal Scientist and the Director. His research interests include electrical modeling and design of microscale/nanoscale integrated circuits, 3-D electronic package integration, and nanoplasmonic technology.

Dr. Li is also a fellow of the MIT Electromagnetics Academy, USA. He is also a Founding Member of the IEEE MTT-RF Nanotechnology Committee. He was a recipient of the 2015 IEEE Richard Stoddard Award on Electromagnetic Compatibility (EMC), the IEEE EMC Technical Achievement Award, the Singapore IES Prestigious Engineering Achievement Award, the Changjiang Chair Professorship Award from the Ministry of Education in China, and number of best paper awards. He has been the general chair and the technical chair for many international conferences. He was elected to the IEEE EMC Distinguished Lecturer, in 2007. He served as an Associate Editor for the IEEE Microwave and Wireless Components Letters, from 2006 to 2008, and a Guest Editor for 2006 and 2010 IEEE Transactions on Electromagnetic Compatibility Special Issues and the 2010 IEEE Transactions on Microwave Theory and Techniques APMC Special Issue. He is also an Associate Editor of the IEEE Transactions on Electromagnetic Compatibility and IEEE Transactions on Components, Packaging and Manufacturing Technology. He was the President of the 2006 International Zurich Symposium on EMC, the Founding General Chair for Asia-Pacific EMC (APEMC) Symposium, and the General Chair of 2008, 2010, 2012, 2016 APEMC, and the 2010 IEEE Symposium on Electrical Design for Advanced Packaging Systems. He has been invited to give numerous invited talks and plenary speeches at various international conferences and forums.

LAY KEE ANG (Fellow, IEEE) received the B.S. degree from the National Tsing Hua University, Hsinchu, Taiwan, in 1994, and the M.S. and Ph.D. degrees in nuclear engineering and plasma physics from the University of Michigan, Ann Arbor, USA, in 1996 and 1999, respectively.

He is currently the Head of the Science and Mathematics and Technology (SMT) Cluster, and the Ng Teng Fong Chair Professor with the Singapore University of Technology and Design, Singapore. His research interests include the development of theoretical scaling laws and models that can capture the essential physics in any interesting problems with a strong focus on the physics and applications of charge particles (electrons) and photons (laser and electromagnetic wave) in various medium and structures. He is also interested in using fractional calculus in modeling complexity in physical sciences and engineering.