Machine-Learning-Assisted Optimization for Antenna Geometry Design

Qi Wu, *Member, IEEE*, Weiqi Chen, *Graduate Student Member, IEEE*, Chen Yu, *Member, IEEE*, Haiming Wang, *Member, IEEE*, and Wei Hong, *Fellow, IEEE*

***Abstract*— A machine-learning-assisted optimization (MLAO) method for antenna geometry design (AGD) (MLAO-AGD) is proposed. By combining machine learning (ML) methods, including a convolutional neural network (CNN) and Gaussian process regression (GPR), MLAO-AGD achieves great efficient improvement compared with conventional evolutionary-algorithm-assisted AGD methods. The ML methods are introduced to build surrogate models between the antenna geometry and the antenna performance and then provide predictions of potential designs during optimization. The ML-based surrogate model is iteratively updated by verified optimization results using full-wave simulations. Three antenna design examples, including multiband and broadband antenna element design tasks and a mutual coupling reduction design task, are presented to show the advantages of the proposed MLAO-AGD, which include the convergence speed and antenna performance.**

***Index Terms*— Antennas, convolutional neural network (CNN), geometry, machine learning (ML), mutual coupling.**

I. Introduction

THE fast development of modern wireless communication and radar applications has raised interest in the design of antennas with complex structures. Classical antenna design methodologies start with choosing an existing geometry based on prior knowledge, followed by parameter studies and optimizations focused on certain design objectives such as the reflection coefficient and the directivity. Notably, the performance of the final antenna design greatly depends on the predefined antenna geometry. In other words, the quality of the optimized antenna design is constrained by the physical antenna geometry mechanism chosen in the first step. If the optimized result fails to meet the predefined design goals, then geometry modification is implemented, and the above-mentioned design process is repeated.

The introduction of geometry optimization can help alleviate the above-mentioned time-consuming iterative selection–optimization process by considering shapes and the connectivity of individual antenna regions as being subject to design objectives. Compared with the traditional antenna design methodologies, geometry-optimization-assisted methods enable flexible and powerful exploration of much larger degrees of design freedom and better antenna performance, especially for applications with limited prior knowledge and design guidelines. Geometry optimization always has the potential to result in an unexpected antenna shape that fulfills designers' predefined goals, thereby overcoming biases and creating novel antenna topologies.

From the perspective of the used optimization methods, antenna geometry design (AGD) methods can be divided into two algorithmic categories: gradient-algorithm-based topology optimization methods [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14] and evolutionary-algorithm-assisted geometry optimization methods [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29]. Gradient-algorithm-based topology optimization is based on adjoint calculation-directed gradient-based optimization [30], [31], [32]. In most cases, the antenna structure is defined using a fixed or moving mesh, and the gradient is calculated through the adjoint formulation to inform how an individual voxel/pixel/mesh element should change for the next iteration to improve the cost function. Such methods are quite desirable for hyperdimensional problems. Algorithms including the scalar isotropic material with penalization (SIMP) method [33] and level set method (LSM) [34] have been introduced to achieve antenna topology optimization for applications including linearly and circularly polarized radiation [1], conical-beam radiation [14], subwavelength antennas [13], near-field coupling [12], and mutual coupling reduction [2]. In [2], an effective hybrid topology optimization method combining the merits of both SIMP and LSM is introduced to design isolation structures in multi-input multi-output (MIMO) antenna systems. In [30], the adjoint method is applied to the design of gradient index lenses to rapidly optimize their performance using only a few simulations. One of the main limitations of gradient-algorithm-based topology optimization

Manuscript received 14 March 2023; revised 8 October 2023; accepted 6 November 2023. Date of publication 1 January 2024; date of current version 7 March 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62371120 and Grant 62271133, in part by the ZTE Industry-University-Institute Cooperation Funds under Grant IA20221230038, in part by the Fundamental Research Funds for the Central Universities under Grant 2242023R40039. *(Corresponding author: Qi Wu.)*

Qi Wu, Haiming Wang, and Wei Hong are with the School of Information Science and Engineering and the State Key Laboratory of Millimeter Waves, Southeast University, Nanjing 210096, China, and also with the Pervasive Communication Research Center, Purple Mountain Laboratories, Nanjing 211111, China (e-mail: qiwu@seu.edu.cn; hmwang@seu.edu.cn; weihong@seu.edu.cn).

Weiqi Chen and Chen Yu are with the School of Information Science and Engineering and the State Key Laboratory of Millimeter Waves, Southeast University, Nanjing 210096, China (e-mail: wq_chen@seu.edu.cn; chenyu@seu.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TAP.2023.3346493.

Digital Object Identifier 10.1109/TAP.2023.3346493

methods is that they require information about the sensitivity of the objectives with respect to many design variables, which limits their cooperation with commercial full-wave simulators. Another main limitation is that the applicability of gradient-based algorithms relies on the objective function being single-peaked with respect to the design variables, rather than nonconvex or nonsmooth. Therefore, gradient-algorithm-based topology optimization methods are more likely to become stuck in local optima compared with the evolutionary-algorithm-assisted methods. This disadvantage can be alleviated by introducing multiple initial guesses or conservative convex separable approximations, but these techniques will inevitably introduce trial-and-error processes and lead to inefficiency when applied to complex antenna designs.

Compared with gradient-based optimization algorithms, evolutionary algorithms offer an improved global-search ability. Moreover, evolutionary algorithms require no gradient information, therefore facilitating the implementation of commercial full-wave simulators within the design process, especially for complicated electromagnetic (EM) problems or applications under complex EM environments. However, the main difficulty of evolutionary algorithms lies in their large number of evaluation calls. Evolutionary-algorithm-assisted antenna geometry optimization methods typically view the antenna shape as a matrix in which the elements "1" and "0" indicate metal and nonmetal, respectively, and introduce binary evolutionary algorithms (BEAs) such as binary particle swarm optimization (BPSO) [15], [16], [29], binary differential evolutionary algorithm (BDEA) [17], and binary brain storm optimization (BBSO) [35] to solve the binary optimization problem. In these evolutionary-algorithm-based methods, the number of fitness function evaluations exponentially grows with increasing number of design parameters, therefore limiting the application of these methods in practical scenarios. Some advanced evolutionary-algorithm-based methods have been proposed to enhance the convergence speed and the optimization performance in antenna geometry optimization applications. In [29], an excellent improved BPSO is proposed for the design of high-dimensional multifunctional antennas. By introducing orthogonal array (OA)-based initialization and a new transfer function (TF) with a time-variant transfer factor, the algorithm is greatly improved over the basic BPSO. In [16], a systematic geometry optimization approach for mutual coupling reduction application is proposed based on BPSO. While these proposed evolutionary-algorithm-assisted methods achieve great performance in various kinds of AGD applications, most of them still need thousands of fitness function evaluations using full-wave simulations to converge to states of acceptable antenna performance.

The machine learning (ML) methods have been recently introduced into EM research, particularly antenna design, to accelerate design and optimization procedures [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46]. To build computationally efficient surrogate models, various kinds of ML methods have been introduced, including artificial neural networks (ANNs) [47], Gaussian process regression (GPR) [36], [48], and support vector machines [49]. The ML-assisted optimization (MLAO) methods can achieve efficient antenna design and explore greater antenna performance capabilities than conventional methods. However, most of these MLAO methods focus only on designing the parameters of a predefined antenna geometry instead of exploring the potential antenna geometry.

While some investigations have implemented ML methods in AGD, most of these demonstrations are offline methods, which means that thousands or even tens of thousands of full-wave simulations are needed to construct the database and ensure prediction accuracy at the very beginning of the algorithms. In [50], a convolutional neural network (CNN) is introduced to model the resonant frequencies of a dual-band pixelated microstrip antenna. The numbers of data points introduced for the training, validation, and test geometries are 578, 378, and 984, respectively, for combined design domains with only 2, 3, and 4 pixels added to the base geometry. In [51], a deep CNN is introduced to estimate the EM scalar field of a patch antenna, in which a total of 996 instances in the design parameter range are introduced for training and validation of the model. In [52], a CNN is introduced to model the vector-fitting-represented reflection coefficient performance of an ultrawideband antenna, in which 64 training samples are introduced to model four design parameters with a limited design range. In [53], a CNN is introduced for cooperation with PSO to model a fragment-type patch antenna using 500 antenna samples. Another issue introduced by the proposed ML-method-based AGD is the limited feature representation strategies for the antenna geometry. On one hand, most of the above-mentioned ML-based AGD methods search for pixelated antenna geometries, which may cause difficulties during manufacturing [54]. On the other hand, in many practical cases, aside from the limited areas for geometry optimization, many critical antenna design parameters that define structures with a fixed geometry also need to be optimized. However, most of these methods fail to optimize the antenna geometry and parameters simultaneously.

In this article, the MLAO scheme is introduced into AGD tasks to construct an online ML-assisted design methodology. Compared with the classic evolutionary-algorithm-assisted AGD methods, the proposed MLAO for AGD (MLAO-AGD) method achieves great enhancement in both the efficiency and performance. The main contributions of this work are listed as follows.

1) With the introduction of ML methods, including a CNN and GPR, the antenna geometry optimization process is greatly accelerated based on the proposed MLAO-AGD scheme. The proposed method requires no gradient information in the optimization process and therefore offers great generalization capability for extension to either commercial or in-house simulators. The correlations between the antenna geometry and the performance are built using ML methods, which are then used in cooperation with evolutionary algorithms to find potential antenna topologies that fulfill design goals. The MLAO-AGD scheme also offers opportunities to iteratively implement model updating, which facilitates

the algorithm starting with a small data size. The method scheme is very flexible and can be introduced for cooperation with various kinds of modeling techniques and optimization algorithms.

2) Different from the conventional MLAO methods, in which only antenna dimensions under a fixed geometry are optimized and designed, here, pixelated and normalized Gaussian network (NGnet) techniques are introduced to achieve AGD. The proposed MLAO-AGD can achieve topologies with either a checkerboard or smooth border required by designers. Moreover, the algorithm can achieve hybrid modeling and optimization considering both the antenna geometry and design parameters, which greatly extends the boundaries of the capability of the MLAO methods in antenna design.
3) Three typical antenna design applications are introduced to validate the proposed MLAO-AGD algorithm: a multiband monopole antenna, a wideband coaxial-fed microstrip antenna, and a double-layer decoupling structure (DS) for a two-element MIMO system. MLAO-AGD is compared with the classic evolutionary optimization algorithms, including BPSO, BBSO, GA, etc. MLAO-AGD achieves better performance considering both the convergence speed and final antenna performance. The low evaluation cost offered by the surrogate models largely accelerates the evolutionary optimization process, therefore helping to efficiently direct the evolution of the antenna geometry toward the preassigned design goals. In the decoupling case, both the design parameters of the microstrip antennas and the geometry of the DSs are simultaneously optimized, achieving a great decoupling performance in a fast manner.

The rest of this article is organized as follows. The problem formulation is presented in Section II. The basic techniques used in the proposed algorithms are given in Section III. The proposed MLAO-AGD is presented in Section IV. In Section V, with the application of the proposed algorithm, several typical antenna design cases are studied, and the performance is compared with that of the conventional evolutionary algorithms. Section VI concludes the work.

II. Problem Formulation

The geometry optimization task in antenna design can be formulated as

$$\begin{cases} \min\limits_{\rho} \ R = R(\boldsymbol{t}(\rho)) \\ \text{s.t. } g_1(\rho) \le 0 \\ \quad\ g_2(\rho) = 0 \end{cases} \tag{1}$$

in which ρ describes the geometry of the designated antenna, $\boldsymbol{t}(\rho)$ is the evaluated antenna performance obtained using full-wave simulations, such as the reflection performance at frequency points of concern and radiation performance in directions and at frequency points of concern, $R(\boldsymbol{t}(\rho))$ represents the fitness function calculated using $\boldsymbol{t}(\rho)$, and $g_1(\rho)$ and $g_2(\rho)$ represent the unequal and equal restrictions predefined before the optimization process, respectively. The structure of the designated antenna can be further defined as follows:

$$\rho = \left\{\rho_{\text{topo}}(\boldsymbol{x}), \rho_{\text{para}}(\boldsymbol{l})\right\} \tag{2}$$

in which $\rho_{\text{topo}}(\boldsymbol{x})$ describes the antenna geometry inside the design area for geometry optimization at position vector $\boldsymbol{x}$, and $\rho_{\text{para}}(\boldsymbol{l})$ describes the structure of the antenna affected by the variable design parameters $\boldsymbol{l}$. For planar AGD, the antenna geometry can be expressed as

$$\begin{cases} \rho_{\text{topo}}(\boldsymbol{x}) = 1, & \text{solid material} \\ \rho_{\text{topo}}(\boldsymbol{x}) = 0, & \text{void material} \end{cases} \tag{3}$$

in which the solid and void material are usually replaced with metal and air.

The most commonly used strategy is to discretize the design domain into smaller rectangular element cells named "pixels" and fill them with conductors or air. In this case, $\rho_{\text{topo}}(\boldsymbol{x})$ is mapped into a matrix $\boldsymbol{X} = (x_{ab}) \in \mathbb{R}^{m\times n}$, in which $x_{ab} \in \{0, 1\}$ represents the value at the a-row and b-column. With larger values of m and n, the design domain becomes larger, as does the difficulty of modeling and optimization. The pixelated antenna geometry mapping strategy facilitates implementation of BEAs such as BPSO, BBSO, BDEA, and binary genetic algorithm (BGA) to search for the global optimum, cooperate with commercial full-wave simulators, and simultaneously calculate the results for multiple samples on parallel computers. In these methods, the matrix $\boldsymbol{X}$ is commonly transformed into an $n = ab$-number-long vector to implement optimization processes such as crossover, mutation, and other update strategies. One of the most important limitations of this flattening operation is the loss of allocation information. In other words, the information given by the simulation results is not fully used in classic BEA methods. Moreover, the evaluations of the objective functions in classic BEA methods involve time- and resource-intensive numerical full-wave simulations. In contrast, the NGnet strategy uses vectors of real numbers to generate the antenna geometry. Therefore, classic EAs such as PSO, BSO, DEA, and GA can be easily introduced to address optimization. However, the direct fitness function calculation based on full-wave simulations still requires a great amount of calculation resources and time, which makes them less applicable in practical AGD.

III. Basic Techniques

In this section, the basic techniques used in the proposed MLAO-AGD, including NGnet, a CNN, GPR, and the modified BPSO, are discussed. These techniques are introduced to cooperate with the MLAO scheme in a flexible manner.

A. Normalized Gaussian Network

Two antenna geometry generation strategies, including the pixelated method and NGnet [19], are introduced, as shown in Fig. 1. The pixelated method, as illustrated in Fig. 1(a), maps the design domain into a matrix $\boldsymbol{X} = (x_{ab}) \in \mathbb{R}^{m\times n}$. Another typical feature representation strategy is NGnet, in which the

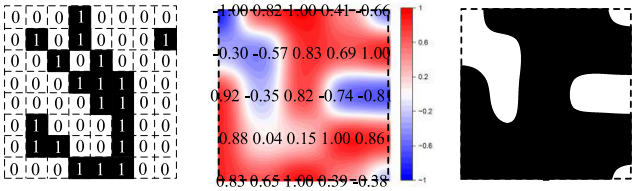

Fig. 1. Geometry generation strategies. (a) Pixelated geometry and its mapped matrix. (b) NGnet-generated geometry and its mapped matrix.

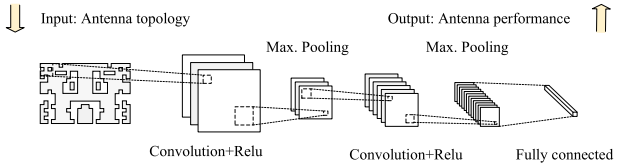

Fig. 2. CNN structure used in the ML modeling process.

antenna shape is expressed in terms of the linear combination of normalized Gaussian functions, as shown in Fig. 1(b). The antenna geometry can be expressed as

$$\begin{cases} \sum_{k=1}^{N} w_{i'} b_{i'}(\boldsymbol{x}) \geq 0, & \rho_{\text{topo}}(\boldsymbol{x}) = 1 \\ \sum_{k=1}^{N} w_{i'} b_{i'}(\boldsymbol{x}) < 0, & \rho_{\text{topo}}(\boldsymbol{x}) = 0 \end{cases} \tag{4}$$

in which $b_{i'}(\boldsymbol{x})$ represents the normalized Gaussian function, $w_{i'}$ is the weighting coefficient, and N is the number of Gaussian functions. $b_{i'}(\boldsymbol{x})$ is calculated using

$$b_{i'}(\boldsymbol{x}) = G_{i'}(\boldsymbol{x}) / \sum_{k=1}^{N} G_k(\boldsymbol{x}) \tag{5}$$

in which $G_k(\boldsymbol{x})$ represents the Gaussian function with center μ_k and covariance matrix Σ_k, expressed as follows:

$$G_k(\boldsymbol{x}) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-1/2(\boldsymbol{x}-\boldsymbol{\mu}_k)^T \Sigma_k^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_k)} \tag{6}$$

in which D signifies the dimension of input x. In this case, $\rho_{\text{topo}}(\boldsymbol{x})$ is mapped into a vector of hyperparameters $\boldsymbol{w} = (w_n) \in \mathbb{R}^N$. After considering both the antenna geometry and structure parameters, $\boldsymbol{t}(\rho)$ is then expressed as $\boldsymbol{t}(\boldsymbol{X}, \boldsymbol{l})$ using the pixelated strategy or as $\boldsymbol{t}(\boldsymbol{w}, \boldsymbol{l})$ using the NGnet strategy. The design task is then transformed as

$$\min_{\boldsymbol{X},\boldsymbol{l}} \; R = R(\boldsymbol{t}(\boldsymbol{X}, \boldsymbol{l})), \quad \text{pixelated method} \tag{7}$$

or

$$\min_{\boldsymbol{w},\boldsymbol{l}} \; R = R(\boldsymbol{t}(\boldsymbol{w}, \boldsymbol{l})), \quad \text{NGnet method} \tag{8}$$

under the given governing equations. By introducing the above-mentioned geometry generation strategies, MLAO-AGD transforms the AGD task into an optimization task, which can be solved under the classic MLAO scheme.

B. Convolutional Neural Network

A CNN is a mathematical construct that is typically composed of three types of layers: convolution, pooling, and fully connected layers. The used CNN structure is illustrated in Fig. 2. The first two layers play roles in feature extraction, and the third layer maps the extracted features into the final output, such as the antenna performance at frequency points of concern. The convolution layer is the fundamental component of the CNN, which typically consists of linear and nonlinear operations, i.e., convolution operation and activation operation. The convolution operation introduces a small array of numbers called a kernel to be applied to the input tensor, which is an array of numbers. The elementwise product between each element of the kernel and the input tensor is calculated at each location of the tensor and then summed to obtain the output value at the corresponding position of the output tensor, which is called a feature map. This procedure is then repeated by applying multiple kernels to form multiple feature maps to represent the characteristics of the input tensors. Two main hyperparameters define the convolution operation: the kernel size $p \times q$ and kernel number n_{kernel}. We suggest using a 3×3 size kernel in the proposed algorithm.

For overlap of the outermost element of the input tensor, zero padding is essential before the convolution operation. The kernels are shared across the entire image positions, which largely increases the modeling efficiency by reducing the number of parameters to learn. The outputs of the convolution are then passed through a nonlinear activation function, which is the rectified linear unit (ReLU) in the presented algorithm. Then, the output is passed through the pooling layer to implement the downsampling operation, which reduces the in-plane dimensionality of the feature maps. Here, max pooling is implemented, in which the maximum values in each patch extracted from the input feature maps are outputted. The used CNN consists of a stack of several convolution layers and pooling layers, followed by one fully connected layer and one regression layer. To achieve a balance between the computational cost and modeling ability, a shallow CNN with a layer number of 2 is recommended. The output feature maps of the final pooling layer are flattened and connected to the fully connected layer, in which every input is connected to every output by a learnable weight.

C. Gaussian Process Regression

GPR is a generalization of the Gaussian probability distribution. A typical single-objective GPR (SOGPR) approximates the target output $y = \hat{t}(\mathbf{x})$ by interpreting it as a probability distribution in function space as

$$\hat{t} \sim \text{GP}\big(m_{\text{GP}}(\mathbf{x}), k_{\text{GP}}(\mathbf{x}, \mathbf{x}')\big) \tag{9}$$

where $\mathbf{x}$ and $\mathbf{x}'$ are the positions in the $\mathbb{R}^N$ design space, with N being the sample number, $m_{\text{GP}}(\mathbf{x})$ is the mean, which can be taken as zero without loss of generality, and $k_{\text{GP}}(\mathbf{x}, \mathbf{x}')$ is the covariance function. Different covariance functions can be applied to describe the correlation between two candidate designs, such as the well-known squared exponential (SE) covariance function [55]

$$k_{\text{SE}}(\mathbf{x}, \mathbf{x}') = \sigma_{\text{f}}^2 \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{x}')^{\text{T}} \mathbf{P}^{-1} (\mathbf{x}-\mathbf{x}')\right) \tag{10}$$

where σ_{f} represents the output scale amplitude, and the diagonal matrix $\mathbf{P} = \text{diag}([l_{\text{GP},1}^2, l_{\text{GP},2}^2, \ldots, l_{\text{GP},N_{\text{GP}}}^2]^{\text{T}})$ contains

the characteristic length scales representing the oscillation frequencies along different directions. In the training process, these hyperparameters are determined using an optimization process based on the given dataset.

D. Modified BPSO

In this article, the modified BPSO is introduced to update the positions of the population, in which each particle e is initialized with binary data γ_e and randomized velocity v_e. At the t_{PSO}th iteration, each particle e is updated as follows [29]:

$$v_e(t_{\text{PSO}}) = wv_i(t_{\text{PSO}}-1) + c_1 r_1(P_{\text{best}} - \gamma_e(t_{\text{PSO}}-1)) + c_2 r_2(G_{\text{best}} - \gamma_e(t_{\text{PSO}}-1)) \tag{11}$$

$$\gamma_e(t_{\text{PSO}}) = \begin{cases} 1, & \text{if rand}() < S(v_e(t_{\text{PSO}})) \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

$$S(v_e(t_{\text{PSO}})) = \begin{cases} 1 - \dfrac{2}{1+e^{-\alpha(t)v_e(t_{\text{PSO}})}}, & v_e(t_{\text{PSO}}) \le 0 \\ \dfrac{2}{1+e^{-\alpha(t)v_e(t_{\text{PSO}})}} - 1, & v_e(t_{\text{PSO}}) > 0 \end{cases} \tag{13}$$

where $v_e(t_{\text{PSO}})$ and $v_e(t_{\text{PSO}}-1)$ are the current and previous velocities of particle e, respectively; $\gamma_e(t_{\text{PSO}})$ and $\gamma_e(t_{\text{PSO}}-1)$ are the current and previous values of the binary data of particle e, respectively; P_{best} and G_{best} are the best individual solution and the best global solution thus far, respectively; ω is the inertia weight controlling exploration of the particle; c_1 and c_2 are the learning factors; r_1 and r_2 are the random numbers within [0, 1]; $S(v_e(t_{\text{PSO}})))$ is the TF used to transform all the real values of velocities to probability values within [0, 1]; rand() is a random number with a uniform distribution within [0, 1]; and $\alpha(t_{\text{PSO}})$ is the time-variant transfer factor given by

$$\alpha(t_{\text{PSO}}) = \alpha_{\max} - \frac{\alpha_{\max} - \alpha_{\min}}{t_{\text{PSO}}} \tag{14}$$

where $\alpha_{\max}$ and $\alpha_{\max}$ are the maximum and minimum transfer factors, respectively. The CNN cannot provide confidence level information for potential design points, which may limit the exploration ability of the algorithm in some cases with large design domains. Therein, the relative shortest distance σ_s between the solution of concern and the training data can be introduced as an exploration characteristic to estimate the uncertainty in the fitness estimation [56].

IV. MLAO-AGD

In this section, the MLAO scheme is introduced into AGD. The key motivation of MLAO is to use efficient surrogate models to implement population-based metaheuristic optimization methods, thereby accelerating optimization. Various ML methods have been introduced to build data-driven surrogate models of EM components. The MLAO algorithm exploits the negligible calculation times of surrogate models to address the large number of function calls required by optimization algorithms at the possible cost of added iterations. The flow diagram of the proposed MLAO-AGD is shown in Fig. 3, in which methods and strategies, including

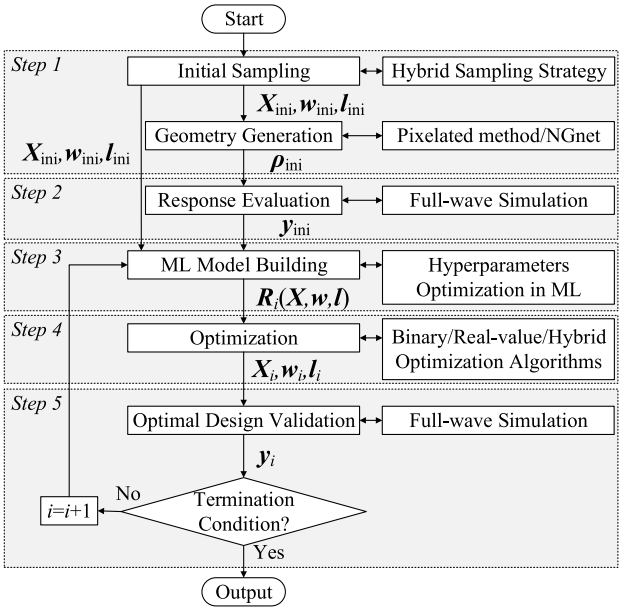

Fig. 3. Flow diagram of the MLAO-AGD algorithm.

a hybrid sampling strategy, geometry generation strategies, ML methods, evolutionary optimization algorithms, and online surrogate model updating, are introduced to ensure the convergence of the algorithm and the final performance of the designed antenna. The detailed steps are summarized as follows.

Step 1: Initial Sampling and Geometry Generation: One of the most critical aspects in MLAO-AGD is to develop generation strategies for the antenna geometry to ease the implementation of both ML modeling and optimization. Two antenna geometry generation strategies, including the pixelated method and NGnet, are considered. In this case, $\rho_{\text{topo}}(\boldsymbol{x})$ is mapped into a vector of hyperparameters $\boldsymbol{w} = (w_n) \in \mathbb{R}^N$. After considering both the antenna geometry and structure parameters, $\boldsymbol{t}(\rho)$ is then expressed as $\boldsymbol{t}(\boldsymbol{X}, \boldsymbol{l})$ using the pixelated strategy or as $\boldsymbol{t}(\boldsymbol{w}, \boldsymbol{l})$ using the NGnet strategy. The design task is then transformed as

$$\min_{\boldsymbol{X},\boldsymbol{l}} \; R = R(\boldsymbol{t}(\boldsymbol{X}, \boldsymbol{l})), \quad \text{pixelated method} \tag{15}$$

or

$$\min_{\boldsymbol{w},\boldsymbol{l}} \; R = R(\boldsymbol{t}(\boldsymbol{w}, \boldsymbol{l})), \quad \text{NGnet method} \tag{16}$$

under the given governing equations. By introducing the above-mentioned geometry generation strategies, MLAO-AGD transforms the AGD task to an optimization task, which can be solved under the classic MLAO scheme.

Another important characteristic of MLAO-AGD is that the proposed method tends to start with a very limited size of the initial database, with a population number of N_{ini}, which is different from other offline ML-assisted AGD methods [50], [51], [52], [53]. As seen in the verification examples in Section V, only 50 sampling topologies are used as the initial database to start the algorithm. On one hand, each AGD task has its specialty, which makes transfer of the dataset for one antenna geometry to another design task difficult. Thus, the benefit of building an initial dataset with a large data size

does not overcome the loss in computational expenses in most cases. On the other hand, the online surrogate model updating scheme of MLAO-AGD ensures the continuous updating of the prediction ability of the ML surrogate model, which reduces the size requirements of the initial dataset. However, due to the limited size of the initial dataset and the fact that no a priori information about the location of the global optimum is known, building an initial dataset with design parameters evenly distributed over the feasible solution space that covers quasioptimal topologies as special sampling points is favorable. These topologies should be those given by a priori EM knowledge and design guidelines and extreme topologies, such as all-metal or all-air cases. In pixelated-based strategies, initialization strategies, including randomized initialization and OA-based population initialization [29], can be introduced to form fairly and uniformly distributed candidates. Here, the ηth geometry solution in the initial database is expressed as

$$X_{\text{ini},\eta} = \begin{cases} \text{RA}(0,1), & 1 \le \eta \le N_{\text{ini}} - 2 \\ O_{a,b}, & \eta = N_{\text{ini}} - 1 \\ J_{a,b}, & \eta = N_{\text{ini}} \end{cases} \tag{17}$$

in which RA$(0, 1)$ represents a randomly generated $a \times b$ matrix of integers 0 or 1, and $O_{a,b}$ and $J_{a,b}$ represent an $a \times b$ zero matrix and a matrix of ones, respectively. In the NGnet-based strategies, the initial sampling points are built using the Latin hypercube sampling (LHS) strategy.

Step 2: Response Evaluation and Database Construction: The initial database is then constructed by simulating the sampled antenna geometry with a full-wave simulator as

$$\mathbb{D}_{\text{ini}} = \left\{ \left(X_{\text{ini},n}, t^k_{\text{ini},n} \right) | n = 1, \ldots, N_{\text{ini}}; k = 1, \ldots, K \right\}, \quad \text{pixelated method} \tag{18}$$

or

$$\mathbb{D}_{\text{ini}} = \left\{ \left(w_{\text{ini},n}, l_{\text{ini},n}, t^k_{\text{ini},n} \right) | n = 1, \ldots, N_{\text{ini}}; k = 1, \ldots, K \right\}, \quad \text{NGnet method} \tag{19}$$

where $t^k_{\text{ini},n}$ is the kth design objective, such as $|S_{11}|$ at the frequency point of concern or the realized gain in the direction and at the frequency point of concern. The database $\mathbb{D}_{i_{\text{iter}}}$ is constructed in the i_{iter}th iteration and is iteratively updated.

Step 3: Building of the ML Model Using the Obtained Database: Once the database $\mathbb{D}_{i_{\text{iter}}}$ is constructed, ML is then introduced to build surrogate models $\hat{\boldsymbol{t}}_{i_{\text{iter}}}$ between the design parameters and the antenna performance. In this step, the designers need to balance the tradeoff between the computational cost and modeling accuracy. Under a fixed amount of data, growth of the complexity of the modeling methods results in improved modeling capability, with a tradeoff in the modeling speed and a possible loss of generalizability, which are important in the later stages of optimization. To build ML models under the pixelated-method-based geometry generation strategy, a CNN is a straightforward solution. If NGnet is used to generate the antenna geometry, then GPR is recommended to build a surrogate model with input parameters with real values [55]. The CNN is trained to find the parameters of the kernels in convolution layers and their weights in fully connected layers that minimize the differences between output predictions and true values using the database $\mathbb{D}$ in every iteration. In the first iteration, the database is $\mathbb{D}_{\text{initial}}$. With either the CNN or GPR, the training step should be accomplished within a much shorter period than in the full-wave simulation process, especially at the later iterations of the MLAO-AGD process. In such cases, the size of the training dataset becomes much larger than that in the starting period of the algorithm, which increases the computational cost. Therefore, limiting the complexity of the used ML models is crucial.

Step 4: Optimization: Once the surrogate model $\hat{\boldsymbol{t}}_{i_{\text{iter}}}$ between the antenna geometry and the antenna performance is built using ML methods, it is integrated with optimization methods to search for the antenna geometry with the best performance. In the case of pixelated-based geometry generation, BEAs are used to create initial populations based on initialization strategies such as randomized initialization and OA-based population initialization [29], and then, the fitness functions are calculated using the results predicted by the CNN model. The calculated fitness function directs the evolution of the population, which forms an iterative process between the BEA and CNN.

In the case of NGnet-based geometry generation, real-valued EAs are introduced to create initial populations. The real-valued EAs are then combined with GPR to calculate the fitness functions based on the predicted values, make evaluations, and perform evolutions. The prediction mean $\tilde{y}$ and prediction variance $\sigma_g^2(\tilde{\mathbf{x}})$ are then given as

$$\tilde{y} = \hat{t}(\tilde{\mathbf{x}}) = K^{\text{T}}(\mathbf{X}, \tilde{\mathbf{x}}) \left[K(\mathbf{X}, \mathbf{X}) + \sigma_s^2 \mathbf{I} \right]^{-1} \mathbf{y} \tag{20}$$

and

$$\sigma_g^2(\tilde{\mathbf{x}}) = k(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}) - K^{\text{T}}(\mathbf{X}, \tilde{\mathbf{x}}) \left[K(\mathbf{X}, \mathbf{X}) + \sigma_s^2 \mathbf{I} \right]^{-1} K(\mathbf{X}, \tilde{\mathbf{x}}) \tag{21}$$

where $K(\mathbf{X}, \tilde{\mathbf{x}})$ denotes the covariance matrix between the training points and the new point. In this case, the exploration ability of the algorithm can be modulated by introducing the lower confidence bound (LCB) prescreening strategy [57]. Both GPR and the CNN can calculate the predictions of the entire population in a parallel manner, which is essential to accelerate the optimization process. The optimization task is then transformed as

$$\min_{\boldsymbol{X},\boldsymbol{l}} \ \hat{R} = \hat{R}(\boldsymbol{t}(\boldsymbol{X}, \boldsymbol{l})) - w_{\text{ps}}(t_{\text{MLAO}})\sigma_{\text{s}}, \quad \text{pixelated method} \tag{22}$$

or

$$\min_{\boldsymbol{w},\boldsymbol{l}} \ \hat{R} = \hat{R}(\boldsymbol{t}(\boldsymbol{w}, \boldsymbol{l})) - w_{\text{ps}}(t_{\text{MLAO}})\sigma_{\text{g}}, \quad \text{NGnet method} \tag{23}$$

under the given governing equations, in which $w_{\text{ps}}(t_{\text{MLAO}})$ normally offers a time-variant prescreening factor and decreases within the iteration time t_{MLAO} in MLAO-AGD.

Step 5: Optimal Design Validation and Termination Condition Check: In this step, the optimized design parameters such as $\mathbf{X}_i$, $\mathbf{w}_i$, and $\mathbf{l}_i$ given by Step 4 are transformed into the antenna geometry using geometry generation strategies such as

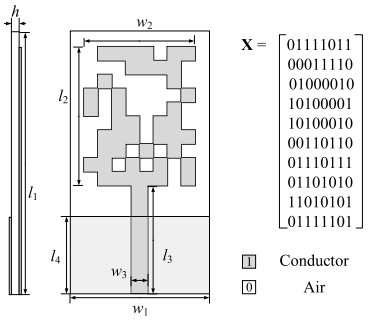

Fig. 4. Configuration of the triband monopole antenna and its mapped design matrix. $w_1 = 18, l_1 = 34, w_2 = 14.4, l_2 = 18, w_3 = 2.2, l_3 = 14, l_4 = 10$, and $h = 1.2$ (units: mm).

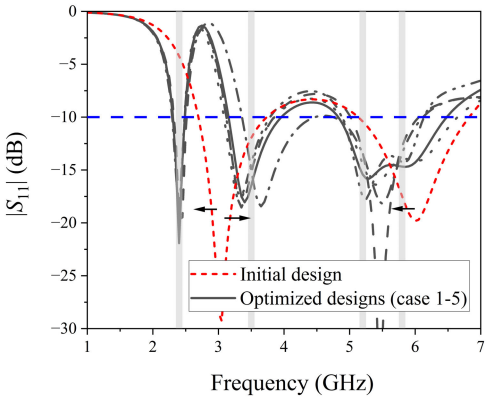

Fig. 5. Performance of the triband monopole antenna: initial design and optimized designs.

pixelated- and NGnet-based methods and then simulated using the full-wave simulation method. The termination conditions typically consist of the evaluation of the validated antenna performance and the maximum number of iterations or time limitation. If the termination conditions are not fulfilled, then the verified results $\mathbf{y}_i$ in this iteration i and the corresponding design parameters are added into the training dataset. The algorithm returns to Step 3 and trains the ML models using the updated training dataset. Typically, in most iterations, the validated results are different from the predicted results given by the ML methods in Step 4 and fail to achieve better results than the existing dataset. However, the iterative process of the algorithm provides a continual exploration of the design domain and an understanding of the design task. The following examples show that MLAO-AGD outperforms the conventional optimization algorithms in convergence speed due to the effective feature abstraction from the dataset.

V. Verification Examples and Discussion

In this section, several AGDs are introduced to verify the effectiveness of our proposed MLAO-AGD algorithm. The performance of our algorithm is compared with that of some representative evolutionary-algorithm-assisted antenna geometry methods, such as the improved BPSO [29], BBSO [35], and GA. We validate that the proposed MLAO-AGD can achieve a much faster convergence speed than classical algorithms considering applications including multiband and broadband radiation and mutual coupling reduction.

A. Triband Monopole Antenna Design

With growing requirements for various kinds of communication applications, multiband antenna design has become a vital task in many wireless communication systems. However, compared with the single-band antenna design, multiband antenna design lacks clear design guidelines. Geometry optimization is a very powerful and flexible method when facing this design task. A triband monopole antenna design problem considering the 2.4/5.2/5.8-GHz WLAN bands and the 3.5-GHz WiMAX band is presented here as the first example, as shown in Fig. 4. This classic design has been introduced to verify the effectiveness of many algorithms, including the improved BPSO [29] and the hybrid Taguchi BPSO [15]. The side length of the metallic-filled pixels is enlarged by 0.1 mm to ensure electrical connectivity. The objective function is described as

$$\min_{\mathbf{X}} t_1(\mathbf{X}) = \max\left(\left|\hat{S}_{11}(t_c, \mathbf{X})\right|\right) \tag{24}$$

in which $|\hat{S}_{11}(f_c, \mathbf{X})|$ represents the predicted reflection coefficient of the antenna with the geometry mapped as $\mathbf{X}$ under the frequency points f_c within the operating band. The used fitness function is different from the commonly used function [15], [29]

$$\begin{aligned} &\min_{\mathbf{X}} t_2(\mathbf{X}) = \frac{1}{n_{\text{fre}}} \sum_{i_{\text{fre}}=1}^{\text{fre}} Q(f_i) \\ &\text{s.t. } Q\left(f_{i_{\text{fre}}}\right) = \begin{cases} -10, & \left|S_{11}\left(f_{i_{\text{fre}}}\right)\right| \le -10 \\ \left|S_{11}\left(f_{i_{\text{fre}}}\right)\right|, & \text{otherwise} \end{cases} \end{aligned} \tag{25}$$

where $f_{i_{\text{fre}}}$ denotes the i_{fre}th sampling frequency within f_c, and n_{fre} is the number of sampling frequencies. The major reason why $|S_{11}(f_{i_{\text{fre}}})|$ values are used to calculate the fitness functions is that they are *true* results from classical evolutionary-assisted algorithms, indicating that these values are reliable for representing the merits of the corresponding antenna geometry. However, in MLAO-AGD, $|\hat{S}_{11}(f_{i_{\text{fre}}})|$ is predicted based on the existing CNN-based surrogate models, which have not been verified through a full-wave simulator. By applying the fitness function t_1 in the MLAO-AGD process, the optimization algorithm continues the search process for the potential design geometry after passing the design goals, which enhances the success possibility of the antenna geometry after verification using full-wave simulation. To achieve a balance between exploration and exploitation, the prescreening factor w_{ps} gradually decreases from 1 to 0 during the optimization process.

In this MLAO-AGD demonstration, the size of the initial dataset is $N_{\text{ini}} = 50$, and the maximum iteration number is $i_{\max} = 500$. The frequency point $f_{i_{\text{fre}}}$ is selected as [2.35, 2.4, 2.45, 3.45, 3.5, 3.55, 5.15, 5.2, 5.25, 5.75, 5.8, 5.85] GHz. The performance of the initial design and the optimized designs in five test cases is shown in Fig. 5. In the initial full-metal design, the antenna resonates at 3.05 and 6 GHz. After optimization, the antennas all completely fulfill the predefined 2.4/5.2/5.8-GHz WLAN band and 3.5-GHz WiMAX band requirements.

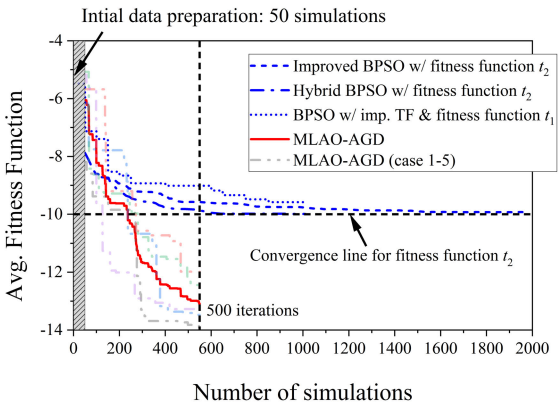

Fig. 6. Convergence rates for the triband monopole AGD using different methods and different fitness functions.

TABLE I
STATISTICS OF DIFFERENT ALGORITHM RESULTS FOR THE TRIBAND MONOPOLE DESIGN

Algorithm	Worst iter.	Best iter.	Mean iter.	Time*
Improved BPSO	>2000	630	> 1524	>134.11 h
Hybrid BPSO	1004	328	629	55.35 h
BPSO w/ imp. TF	>1000	850	>970	>85.36 h
MLAO-AGD	281	127	237	26.60 h

* The times for the improved BPSO and hybrid BPSO are estimated using the mean iterations given in [15] and the mean full-wave simulation time used in MLAO-AGD.

The performance of the proposed MLAO-AGD is compared with that of the conventional BEAs. The MLAO-AGD algorithm is repeated five times, and the convergence rates and mean convergence rates are shown in Fig. 6. The performance of the improved BPSO and hybrid BPSO is given in [15], with their computational time calculated using the given mean iteration number and the mean full-wave simulation time, the same as that in MLAO-AGD. The BPSO with the improved TF given in Section III is implemented and repeated five times to make a comparison, with a maximum iteration number of 1000. MLAO-AGD can find the optimal antenna geometry with a performance exceeding -10 dB-$|S_{11}|$ on average at iteration number 187, signifying that only 237 full-wave simulations are needed. Moreover, the proposed algorithm reaches a mean fitness value of -13.09 within 500 iterations, which represents the worst $|S_{11}|$-in-dB value within the predefined bands. The mean number of iterations and time consumption are presented in Table I. In this case, the mean computational time costs for full-wave simulation, optimization, and training in one iteration are 5.28, 0.61, and 1.23 min, respectively. MLAO-AGD can find designs with fitness values lower than -10 within a mean time of 26.60 h.

B. Wideband Coaxial-Fed Microstrip Antenna Design

With the rapid development of modern communications, the growing demand for wideband antennas has become a top requirement in the communication system design process. The conventional wideband antenna design methodology searches for methods to modify the classic antenna geometry to create and combine multiple resonant frequency points. However, this approach is prior-knowledge-based and cannot ensure the performance of the selected designed geometry, especially when facing design limitations and constraints in practical cases.

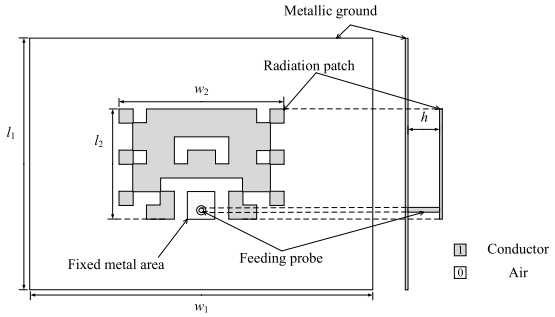

Fig. 7. Optimized configuration of the wideband coaxial-fed microstrip antenna. $w_1 = 150, l_1 = 110, w_2 = 96, l_2 = 72,$ and $h = 15$ (units: mm).

A typical case is the wideband coaxial-fed microstrip antenna design illustrated in Fig. 7, which is investigated in [35] and [58] using parallel EM GA optimization (EGO) and BBSO, respectively. While the fitness functions used in both the studies are defined using the reflection coefficient performance at 1.9 and 2.4 GHz, the two algorithms exhibit different tendencies. In [58], the final antenna geometry is a dual-band design, while in [35], the optimized geometry exhibits wideband performance from 1.9 to 2.4 GHz. To verify the effectiveness of the proposed MLAO-AGD, the objective function used here is the same as (24), in which the frequency points of concern cover the entire frequency band from 1.9 to 2.4 GHz, a design goal that is certainly harder to achieve than the conventional dual-band tasks. To ensure broadband performance, the frequency points of concern are sampled every 0.05 GHz within the entire operating bandwidth.

In [58], the EGO needs nearly 50 000 evaluations to accomplish the design goals. The BBSO effectively reduces the evaluation number to approximately 2500, which is only 5% of the number used in [58]. As shown in Fig. 8 and Fig. 9, BBSO can achieve a satisfactory design within 41 iterations, which takes approximately 1000 full-wave simulations. The BPSO with an improved TF is also implemented, with a maximum simulation number of 1000, which exhibits comparable performance to BBSO. Compared with the above-mentioned two algorithms, MLAO-AGD can find satisfactory results within an average iteration number of 204. Combined with the 50 simulations to build the initial dataset, the proposed algorithm can achieve the predefined goal using only 254 simulations. Moreover, the fitness function continues to improve very quickly after achieving the design goal, reaching approximately -11.63 dB $|S_{11}|$ within the design band.

The reflection coefficients of the initial design and the optimized designs in five optimization cases are shown in Fig. 8. In the initial design with the full-metal geometry, the antenna is resonant at 2.32 GHz with the best $|S_{11}|$ of -4.26 dB. After optimization, the antennas are resonant at approximately 2.0 and 2.40 GHz and achieve a relative bandwidth varying from 26.2% to 23.0%. The optimized results could exhibit slight perturbations under different simulation settings. Radiation nulls could exist in the broadside direction at some operating points due to high-order modes.

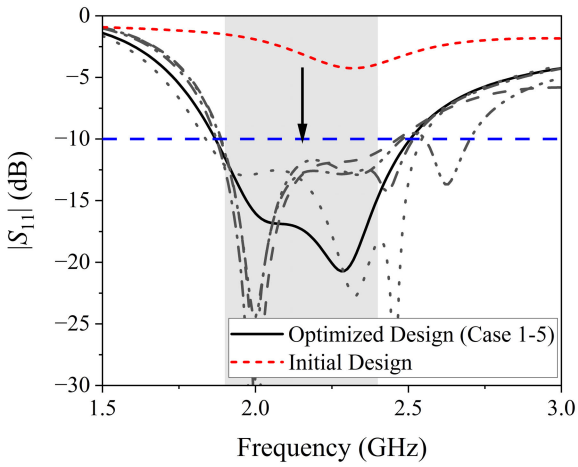

Fig. 8. Performance of the broadband microstrip antenna: initial design and optimized design with predicted and verified results.

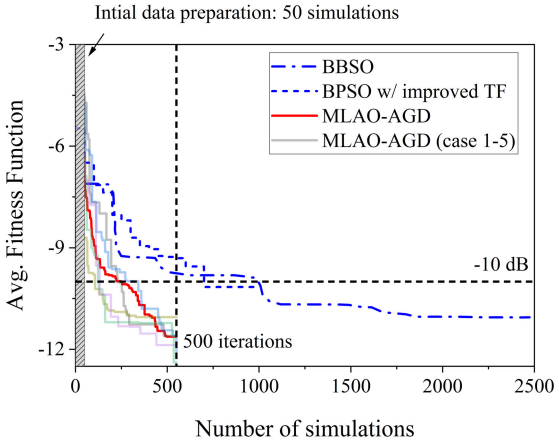

Fig. 9. Convergence rates for the broadband microstrip AGD using different methods and different fitness functions.

TABLE II

STATISTICS OF DIFFERENT ALGORITHM RESULTS FOR THE WIDEBAND MICROSTRIP ANTENNA DESIGN

Algorithm	Worst iter.	Best iter.	Mean iter.	Time*
BPSO w/ imp. TF	>1000	250	> 680	> 17.94 h
MLAO-AGD	342	107	254	9.62 h

* The time for the BPSO with an improved TF is estimated using the mean full-wave simulation time used in MLAO-AGD.

Further investigations could be implemented by considering directivity in desired radiation directions as an optimization goal.

Crucially, while the prediction ability of the CNN model improves with the number of iterations, in most cases, the algorithm is implemented under cases in which the CNN does not contain enough data to make accurate predictions. Normally, tens of thousands or even hundreds of thousands of simulations and deep network architectures are necessary to establish surrogate models with high prediction accuracy within the entire parameter domain. However, the CNN models established in MLAO-AGD are based on a small dataset and are defined with a shallow net architecture, which contributes to directing the optimization toward the predefined design goals. The computational cost of the proposed MLAO-AGD compared with that of the BPSO with an improved TF is shown in Table II. In this case, the mean computational time costs for full-wave simulation, optimization, and training in one iteration are 1.58, 0.43, and 0.43 min, respectively. The time for the BPSO with an improved TF is estimated using the mean full-wave simulation time. MLAO-AGD can find optimal designs with a mean of 254 iterations, which takes approximately 9.62 h.

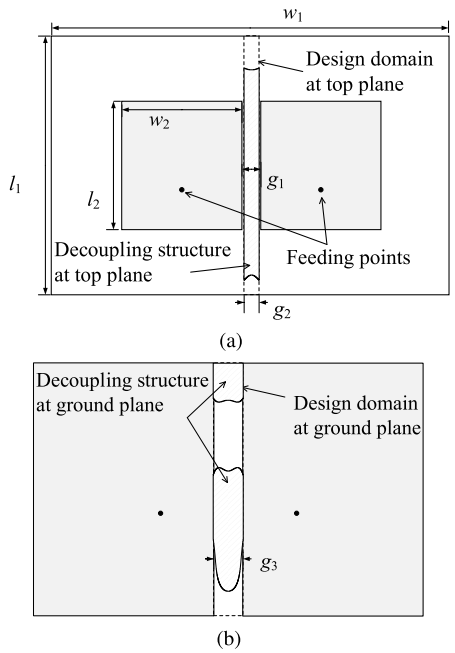

Fig. 10. Optimized DS of the MIMO antenna: (a) top plane and (b) ground plane. $w_1 = 36.9, l_1 = 24, w_2 = 11.15, g_1 = 1.75, g_2 = 1.4$, and $g_3 = 2.7$ (units: mm).

C. Mutual Coupling Reduction Design

The proposed algorithm is then implemented to design a DS for an MIMO system. Recently, many techniques, including the use of a decoupling network [59], a shielding wall [60], a defective ground structure (DGS) [61], an EM bandgap (EBG) structure [62], and a metasurface-based structure [63], have been introduced to reduce the mutual coupling between different antenna elements and achieve greater channel capacity. Many of the above-mentioned techniques can be viewed as types of knowledge-guided geometry design and can be easily integrated with the proposed MLAO-AGD to achieve great decoupling performance using very fast optimization. While some excellent geometry optimization methods have been introduced to design DSs [2], [16], only a few of them can simultaneously optimize the geometry and design parameters [16]. By introducing NGnet into the MLAO-AGD scheme, the algorithm can simultaneously find the optimal DS with smooth boundaries and optimal design parameters.

A typical example is to use the proposed MLAO-AGD to design DSs on both the ground plane and top plane of a two-element MIMO system, as shown in Fig. 10. The antennas are designed to operate within the frequency band of 5.725–5.825 GHz. The substrate has $\varepsilon_r = 4.6$ and tan $\delta = 0.001$. The thickness of the metallic layers is set as 0 to accelerate the simulation process. The center-to-center distance of the two closely placed antennas is 12.9 mm, with

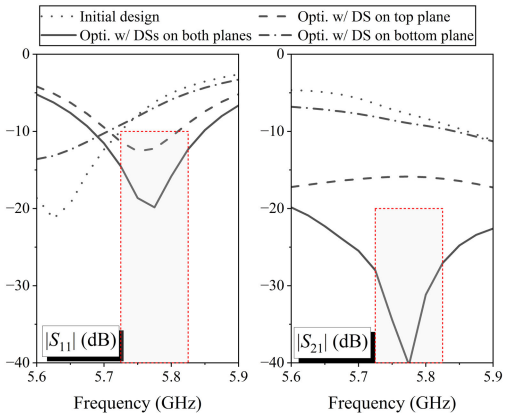

Fig. 11. Performance of different MIMO antennas with and without DSs: $|S_{11}|$ and $|S_{21}|$.

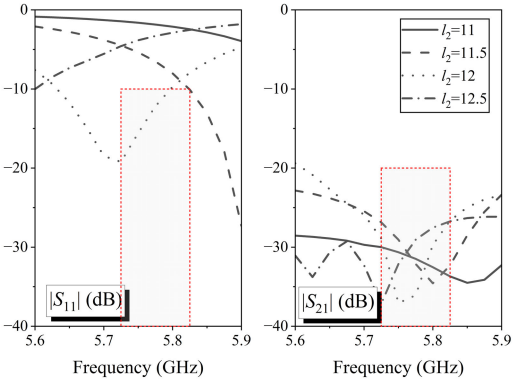

Fig. 12. Performance of MIMO antennas with DSs: S-parameter performance variation with l_2.

a gap of 1.75 mm between them. The simulated S parameters of the four antennas are shown in Fig. 11. The impedance matching of the initial design (without the DSs) is acceptable, with all $|S_{11}|$ parameter values below -9.91 dB. However, regarding mutual coupling, the initial design suffers from the $|S_{21}|$ parameter reaching -5.26 dB. The performance of MIMO antennas also varies with l_2, as shown in Fig. 12.

To improve the isolation performance of the MIMO antenna array, DSs are generated within the two design domains on the ground plane and top plane. The design domains on both the planes are illustrated by dashed lines. On the top plane, the design domain to be optimized is set as a rectangular region with a size of 24×1.4 mm^2. On the ground plane, the design domain is set as a rectangular region with a size of 24×2.7 mm^2. The optimization regions are divided into two symmetric parts, i.e., the L-part and the R-part. The DSs are generated by NGnet using two real-value matrices with sizes of 2×5 and 4×5, which represent the DSs for the top plane and ground plane, respectively. After mirror flip, the DSs on the two planes are generated using 20 and 40 Gaussian functions. The Gaussian functions are made anisotropic to tune the affection of $\boldsymbol{w}$ on the geometries, which can also be modified in future investigations. The length of the patch antenna l_2 is introduced into the modeling and optimization processes due to its importance for the reflection coefficient behavior of the two antennas. The variable range for the geometric hyperparameters $\boldsymbol{w}$ is $[-1, 1]^{30}$, and the range for the structural parameter l_2 is [11, 12.5].

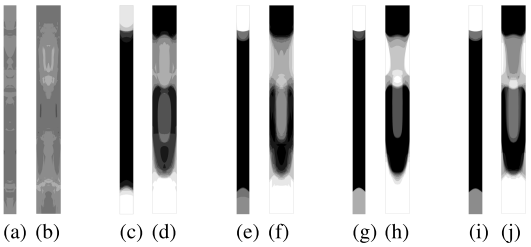

Fig. 13. Evolution of the antenna geometry in a typical MLAO-AGD process. Mean topologies within (a) and (b) initial 50 simulations, (c) and (d) iterations 1–50, (e) and (f) iterations 51–100, (g) and (h) iterations 101–150, and (i) and (j) iterations 151–200.

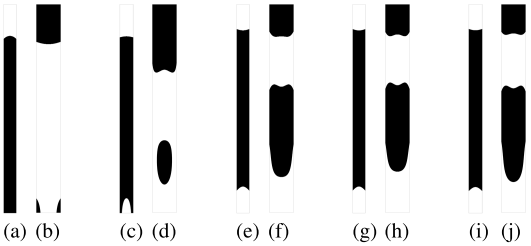

Fig. 14. Evolution of the antenna geometry in a typical MLAO-AGD process. Best topologies within (a) and (b) initial 50 simulations: simulation number of 15 with $l_2 = 11.88$ mm, (c) and (d) iterations 1–50: iteration 4 with $l_2 = 11.88$ mm, (e) and (f) iterations 51–100: iteration 99 with $l_2 = 11.86$ mm, (g) and (h) iterations 101–150: iteration 115 with $l_2 = 11.88$ mm, and (i) and (j) iterations 151–200: iteration 199 with $l_2 = 11.87$ mm.

Several different stages in the optimization process of both the DS geometry and structure parameters are shown in Figs. 13 and 14. Fig. 13(a) shows the best design in the initial dataset. The initial dataset is generated using 46 randomly sampled topologies and four special topologies with or without a full-metal strip on each plane. While the MIMO system without DSs can achieve acceptable reflection coefficient performance within the operating band, the $|S_{11}|$ performance can be heavily affected by DSs during the optimization process. Therefore, the optimization algorithm is introduced to minimize the following fitness function:

$$\text{fitness} = -\left(\beta_1 \max\left\{\left(f_1(\boldsymbol{w}, \boldsymbol{l}), f_1^{\text{ref}}\right)\right\} + \beta_2 \max\left\{\left(f_2(\boldsymbol{w}, \boldsymbol{l}), f_2^{\text{ref}}\right)\right\}\right) \tag{26}$$

in which f_1 and f_2 represent the worst $|S_{11}|$ and $|S_{21}|$ performance within the operating bandwidth, respectively, with their reference points f_1^{ref} and f_2^{ref} set as -12 and -30 dB, with $\beta_1 = 5$ and $\beta_2 = 1$ set to emphasize $|S_{11}|$ performance to avoid deterioration of the reflection coefficient performance when adding the DSs. The optimization process is illustrated in Fig. 15, in which the fitness function value of the proposed MLAO-AGD is compared with that of the classic GA under the same initial dataset. MLAO-AGD provides a fast convergence speed once the initial data collection process is finished. In a typical case, after only 151 full-wave simulations, the algorithm achieves a fitness function value of -84.5, with $|S_{11}|$ and $|S_{21}|$ performing better than -12.27 and -24.50 dB within the entire operating bandwidth.

Fig. 14 illustrates five stages in the optimization process that achieve the best performance considering the fitness function value after every 50 simulations. The optimal topologies

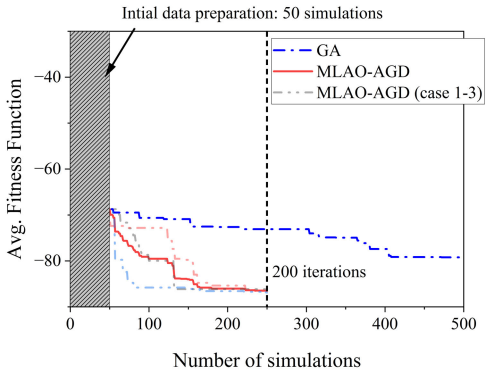

Fig. 15. Convergence rates for the decoupling geometry design using the proposed MLAO-AGD and GA.

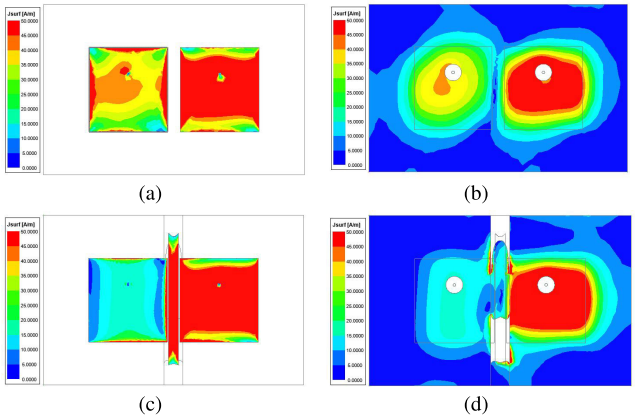

Fig. 16. Current distributions of the MIMO antenna. (a) W/o the DS on the top plane and (b) ground plane. (c) With the DS on the top plane and (d) ground plane.

intensively vary in the early iterations, changing from a two-sided ground DS geometry to a single-sided ground DS geometry in only four iterations. In the following iterations, the algorithm searches for the proper size and shape of the two-element geometry in the ground plane and the length and shape of the DS geometry in the top plane. At the same time, the length of the antenna l_2 converges to 11.87 mm. The $S-$parameter performance of the optimized MIMO antenna array is shown in Fig. 11. The $|S_{11}|$ performance within the operating bandwidth is better than -12.27 dB, which is an increase of 2.36 dB compared with the initial design without DSs. Moreover, the $|S_{21}|$ performance within the operating bandwidth is better than -24.50 dB, which is an increase of 19.24 dB compared with the initial design without DSs. Another investigation shows that the performance of the final design is achieved by implementing DSs on both sides of the substrate, as shown in Fig. 11, and implementation of only one DS on either plane cannot achieve the final performance.

Fig. 12 shows that the reflection coefficient performance of the antennas is very sensitive to the variation of l_2. In addition, as shown in Fig. 11, both S parameters are very sensitive to the DSs with fixed l_2. The optimization of the geometries of the DSs will inevitably deteriorate the reflection coefficient of the two antennas when fixing the antenna dimensions, which means that the dimensions of the antenna elements cannot be selected before DS geometry optimization. Therefore, the predefined requirements cannot be achieved if the geometry or the dimensions are individually optimized. MLAO-AGD can achieve hybrid modeling and optimization considering both the antenna geometry and design parameters, therefore greatly extending the boundaries of the ability of the conventional MLAO methods. To gain insight into the working mechanism of the optimized DSs, the current distributions with and without the DSs are simulated and shown in Fig. 16. Clearly, after the addition of the DSs, when the antenna on the right side is excited and another antenna is matched, the current distributions on the inactivated antenna and on the ground under the inactivated antenna are greatly reduced.

VI. Conclusion

An MLAO-AGD method was introduced and demonstrated to achieve fast and stable convergence for practical antenna design cases. ML methods, including a CNN and GPR, were introduced into the MLAO scheme to build surrogate models between the antenna geometry and performance. The ML-based models were then integrated with optimization algorithms to predict potential antenna topologies. Several antenna design examples, including multiband and broadband antenna design and mutual coupling reduction design, were executed to verify the proposed MLAO-AGD, which achieved great performance considering both the convergence speed and antenna performance.

References

[1] J. Wang, X.-S. Yang, X. Ding, and B.-Z. Wang, "Antenna radiation characteristics optimization by a hybrid topological method," *IEEE Trans. Antennas Propag.*, vol. 65, no. 6, pp. 2843–2854, Jun. 2017.

[2] S.-H. Zhu, X.-S. Yang, J. Wang, and B.-Z. Wang, "Design of MIMO antenna isolation structure based on a hybrid topology optimization method," *IEEE Trans. Antennas Propag.*, vol. 67, no. 10, pp. 6298–6307, Oct. 2019.

[3] A. R. Diaz and O. Sigmund, "A topology optimization method for design of negative permeability metamaterials," *Structural Multidisciplinary Optim.*, vol. 41, no. 2, pp. 163–177, Mar. 2010.

[4] S. Liu, Q. Wang, and R. Gao, "MoM-based topology optimization method for planar metallic antenna design," *Acta Mechanica Sinica*, vol. 32, no. 6, pp. 1058–1064, Dec. 2016.

[5] Q. Wang, R. Gao, and S. Liu, "A novel parameterization method for the topology optimization of metallic antenna design," *Acta Mechanica Sinica*, vol. 33, no. 6, pp. 1040–1050, Dec. 2017.

[6] S. Zhou, W. Li, and Q. Li, "Level-set based topology optimization for electromagnetic dipole antenna design," *J. Comput. Phys.*, vol. 229, no. 19, pp. 6915–6930, Sep. 2010.

[7] S. Liu, Q. Wang, and R. Gao, "A topology optimization method for design of small GPR antennas," *Structural Multidisciplinary Optim.*, vol. 50, no. 6, pp. 1165–1174, Dec. 2014.

[8] E. Hassan et al., "Multilayer topology optimization of wideband SIW-to-waveguide transitions," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 4, pp. 1326–1339, Apr. 2020.

[9] T. Mori, R. Murakami, Y. Sato, F. Campelo, and H. Igarashi, "Shape optimization of wideband antennas for microwave energy harvesters using FDTD," *IEEE Trans. Magn.*, vol. 51, no. 3, pp. 1–4, Mar. 2015.

[10] G. Kiziltas, D. Psychoudakis, J. L. Volakis, and N. Kikuchi, "Topology design optimization of dielectric substrates for bandwidth improvement of a patch antenna," *IEEE Trans. Antennas Propag.*, vol. 51, no. 10, pp. 2732–2743, Oct. 2003.

[11] E. Hassan, E. Wadbro, and M. Berggren, "Topology optimization of metallic antennas," *IEEE Trans. Antennas Propag.*, vol. 62, no. 5, pp. 2488–2500, May 2014.

[12] E. Hassan, D. Noreland, R. Augustine, E. Wadbro, and M. Berggren, "Topology optimization of planar antennas for wideband near-field coupling," *IEEE Trans. Antennas Propag.*, vol. 63, no. 9, pp. 4208–4213, Sep. 2015.

[13] A. Erentok and O. Sigmund, "Topology optimization of sub-wavelength antennas," *IEEE Trans. Antennas Propag.*, vol. 59, no. 1, pp. 58–69, Jan. 2011.

[14] J. Wang, X.-S. Yang, X. Ding, and B.-Z. Wang, "Topology optimization of conical-beam antennas exploiting rotational symmetry," *IEEE Trans. Antennas Propag.*, vol. 66, no. 5, pp. 2254–2261, May 2018.

[15] X. Jia and G. Lu, "A hybrid Taguchi binary particle swarm optimization for antenna designs," *IEEE Antennas Wireless Propag. Lett.*, vol. 18, no. 8, pp. 1581–1585, Aug. 2019.

[16] A. Ghadimi, V. Nayyeri, M. Khanjarian, M. Soleimani, and O. M. Ramahi, "A systematic approach for mutual coupling reduction between microstrip antennas using pixelization and binary optimization," *IEEE Antennas Wireless Propag. Lett.*, vol. 19, no. 12, pp. 2048–2052, Dec. 2020.

[17] S. Goudos, "Antenna design using binary differential evolution: Application to discrete-valued design problems," *IEEE Antennas Propag. Mag.*, vol. 59, no. 1, pp. 74–93, Feb. 2017.

[18] Y.-S. Chen and Y.-H. Chiu, "Application of multiobjective topology optimization to miniature ultrawideband antennas with enhanced pulse preservation," *IEEE Antennas Wireless Propag. Lett.*, vol. 15, pp. 842–845, 2016.

[19] K. Itoh, H. Nakajima, H. Matsuda, M. Tanaka, and H. Igarashi, "Development of small dielectric lens for slot antenna using topology optimization with normalized Gaussian network," *IEICE Trans. Electron.*, vol. E101.C, no. 10, pp. 784–790, 2018.

[20] N. Jin and Y. Rahmat-Samii, "Hybrid real-binary particle swarm optimization (HPSO) in engineering electromagnetics," *IEEE Trans. Antennas Propag.*, vol. 58, no. 12, pp. 3786–3794, Dec. 2010.

[21] H. Choo and H. Ling, "Design of multiband microstrip antennas using a genetic algorithm," *IEEE Microw. Wireless Compon. Lett.*, vol. 12, no. 9, pp. 345–347, Sep. 2002.

[22] N. Herscovici, M. F. Osorio, and C. Peixeiro, "Miniaturization of rectangular microstrip patches using genetic algorithms," *IEEE Antennas Wireless Propag. Lett.*, vol. 1, pp. 94–97, 2002.

[23] P. Soontornpipit, C. M. Furse, and Y. C. Chung, "Miniaturized biocompatible microstrip antenna using genetic algorithm," *IEEE Trans. Antennas Propag.*, vol. 53, no. 6, pp. 1939–1945, Jun. 2005.

[24] R. O. Ouedraogo, E. J. Rothwell, A. Diaz, S.-Y. Chen, A. Temme, and K. Fuchi, "In situ optimization of metamaterial-inspired loop antennas," *IEEE Antennas Wireless Propag. Lett.*, vol. 9, pp. 75–78, 2010.

[25] A. A. Minasian and T. S. Bird, "Particle swarm optimization of microstrip antennas for wireless communication systems," *IEEE Trans. Antennas Propag.*, vol. 61, no. 12, pp. 6214–6217, Dec. 2013.

[26] R. Li, D. McNamara, G. Wei, and J. Li, "Increasing radiation efficiency using antenna shape optimization approach," *IEEE Antennas Wireless Propag. Lett.*, vol. 17, no. 3, pp. 393–396, Mar. 2018.

[27] L. Wang, G. Wang, and J. Sidén, "Design of high-directivity wideband microstrip directional coupler with fragment-type structure," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 12, pp. 3962–3970, Dec. 2015.

[28] D. Lu, L. Wang, E. Yang, and G. Wang, "Design of high-isolation wideband dual-polarized compact MIMO antennas with multiobjective optimization," *IEEE Trans. Antennas Propag.*, vol. 66, no. 3, pp. 1522–1527, Mar. 2018.

[29] J. Dong, Q. Li, and L. Deng, "Design of fragment-type antenna structure using an improved BPSO," *IEEE Trans. Antennas Propag.*, vol. 66, no. 2, pp. 564–571, Feb. 2018.

[30] E. B. Whiting et al., "Adjoint sensitivity optimization of three-dimensional directivity-enhancing, size-reducing GRIN lenses," *IEEE Antennas Wireless Propag. Lett.*, vol. 21, no. 11, pp. 2166–2170, Nov. 2022.

[31] C. M. Lalau-Keraly, S. Bhargava, O. D. Miller, and E. Yablonovitch, "Adjoint shape optimization applied to electromagnetic design," *Opt. Exp.*, vol. 21, no. 18, pp. 21693–21701, 2013.

[32] D. Sell, J. Yang, S. Doshay, R. Yang, and J. A. Fan, "Large-angle, multifunctional metagratings based on freeform multimode geometries," *Nano Lett.*, vol. 17, no. 6, pp. 3752–3757, Jun. 2017.

[33] M. P. Bendsoe and O. Sigmund, *Topology Optimization: Theory, Methods, and Applications*. Berlin, Germany: Springer, 2003.

[34] S. Osher, R. Fedkiw, and K. Piechor, "Level set methods and dynamic implicit surfaces," *Appl. Mech. Rev.*, vol. 57, no. 3, p. B15, 2004.

[35] A. Aldhafeeri and Y. Rahmat-Samii, "Brain storm optimization for electromagnetic applications: Continuous and discrete," *IEEE Trans. Antennas Propag.*, vol. 67, no. 4, pp. 2710–2722, Apr. 2019.

[36] B. Liu, H. Aliakbarian, Z. Ma, G. A. E. Vandenbosch, G. Gielen, and P. Excell, "An efficient method for antenna design optimization based on evolutionary computation and machine learning techniques," *IEEE Trans. Antennas Propag.*, vol. 62, no. 1, pp. 7–18, Jan. 2014.

[37] Q. Wu, W. Chen, H. Wang, and W. Hong, "Machine learning-assisted tolerance analysis and its application to antennas," in *Proc. IEEE Int. Symp. Antennas Propag. North Amer. Radio Sci. Meeting*, Jul. 2020, pp. 1853–1854.

[38] Q. Wu, H. Wang, and W. Hong, "Multistage collaborative machine learning and its application to antenna modeling and optimization," *IEEE Trans. Antennas Propag.*, vol. 68, no. 5, pp. 3397–3409, May 2020.

[39] Q. Wu, W. Chen, C. Yu, H. Wang, and W. Hong, "Multilayer machine learning-assisted optimization-based robust design and its applications to antennas and array," *IEEE Trans. Antennas Propag.*, vol. 69, no. 9, pp. 6052–6057, Sep. 2021.

[40] W. Chen, Q. Wu, C. Yu, H. Wang, and W. Hong, "Multibranch machine learning-assisted optimization and its application to antenna design," *IEEE Trans. Antennas Propag.*, vol. 70, no. 7, pp. 4985–4996, Jul. 2022.

[41] Q. Wu, W. Chen, C. Yu, H. Wang, and W. Hong, "Machine learning-assisted array synthesis using active base element modeling," *IEEE Trans. Antennas Propag.*, vol. 70, no. 7, pp. 5054–5065, Jul. 2022.

[42] Q. Wu, W. Chen, C. Yu, H. Wang, and W. Hong, "Knowledge-guided active base element modeling in machine learning-assisted antenna array design," *IEEE Trans. Antennas Propag.*, vol. 71, no. 2, pp. 1578–1589, Dec. 2022.

[43] Y. Sharma, H. H. Zhang, and H. Xin, "Machine learning techniques for optimizing design of double T-shaped monopole antenna," *IEEE Trans. Antennas Propag.*, vol. 68, no. 7, pp. 5658–5663, Jul. 2020.

[44] J. Tak, A. Kantemur, Y. Sharma, and H. Xin, "A 3-D-printed W-band slotted waveguide array antenna optimized using machine learning," *IEEE Antennas Wireless Propag. Lett.*, vol. 17, no. 11, pp. 2008–2012, Nov. 2018.

[45] L. Cui, Y. Zhang, R. Zhang, and Q. H. Liu, "A modified efficient KNN method for antenna optimization and design," *IEEE Trans. Antennas Propag.*, vol. 68, no. 10, pp. 6858–6866, Oct. 2020.

[46] S. Koziel and A. Pietrenko-Dabrowska, "Performance-based nested surrogate modeling of antenna input characteristics," *IEEE Trans. Antennas Propag.*, vol. 67, no. 5, pp. 2904–2912, May 2019.

[47] Y. Kim, S. Keely, J. Ghosh, and H. Ling, "Application of artificial neural networks to broadband antenna design based on a parametric frequency model," *IEEE Trans. Antennas Propag.*, vol. 55, no. 3, pp. 669–674, Mar. 2007.

[48] S. Koziel and S. Ogurtsov, "Multi-objective design of antennas using variable-fidelity simulations and surrogate models," *IEEE Trans. Antennas Propag.*, vol. 61, no. 12, pp. 5931–5939, Dec. 2013.

[49] D. R. Prado, J. A. López-Fernández, M. Arrebola, and G. Goussetis, "Support vector regression to accelerate design and crosspolar optimization of shaped-beam reflectarray antennas for space applications," *IEEE Trans. Antennas Propag.*, vol. 67, no. 3, pp. 1659–1668, Mar. 2019.

[50] J. P. Jacobs, "Accurate modeling by convolutional neural-network regression of resonant frequencies of dual-band pixelated microstrip antenna," *IEEE Antennas Wireless Propag. Lett.*, vol. 20, no. 12, pp. 2417–2421, Dec. 2021.

[51] M. R. Khan, C. L. Zekios, S. Bhardwaj, and S. V. Georgakopoulos, "Data-driven electromagnetic scalar field estimation of a patch antenna using deep convolutional neural network," in *Proc. IEEE Int. Symp. Antennas Propag. USNC-URSI Radio Sci. Meeting (APS/URSI)*, Dec. 2021, pp. 1497–1498.

[52] H.-Y. Luo, Y. Hong, Y.-H. Lv, and W. Shao, "Parametric modeling of UWB antennas using convolutional neural networks," in *Proc. IEEE Int. Symp. Antennas Propag. North Amer. Radio Sci. Meeting*, Jul. 2020, pp. 2055–2056.

[53] X. Zhang, Y. Tian, and X. Zheng, "Optimal design of fragment-type antenna structure based on PSO-CNN," in *Proc. Int. Appl. Comput. Electromagn. Soc. Symp. China (ACES)*, vol. 1, Aug. 2019, pp. 1–2.

[54] K. Watanabe, F. Campelo, and H. Igarashi, "Topology optimization based on immune algorithm and multi-grid method," in *Proc. 12th Biennial IEEE Conf. Electromagn. Field Comput.*, Apr./May 2006, p. 480.

[55] C. E. Rasmussen, "Gaussian processes in machine learning," in *Proc. Summer School Mach. Learn.* Berlin, Germany: Springer, 2003, pp. 63–71.

[56] J. Branke and C. Schmidt, "Faster convergence by means of fitness estimation," *Soft Comput.*, vol. 9, no. 1, pp. 13–20, Jan. 2005.

[57] M. T. M. Emmerich, K. C. Giannakoglou, and B. Naujoks, "Single- and multiobjective evolutionary optimization assisted by Gaussian random field metamodels," *IEEE Trans. Evol. Comput.*, vol. 10, no. 4, pp. 421–439, Aug. 2006.
[58] F. J. Villegas, T. Cwik, Y. Rahmat-Samii, and M. Manteghi, "A parallel electromagnetic genetic-algorithm optimization (EGO) application for patch antenna design," *IEEE Trans. Antennas Propag.*, vol. 52, no. 9, pp. 2424–2435, Sep. 2004.
[59] S.-C. Chen, Y.-S. Wang, and S.-J. Chung, "A decoupling technique for increasing the port isolation between two strongly coupled antennas," *IEEE Trans. Antennas Propag.*, vol. 56, no. 12, pp. 3650–3658, Dec. 2008.
[60] H. Qi, L. Liu, X. Yin, H. Zhao, and W. J. Kulesza, "Mutual coupling suppression between two closely spaced microstrip antennas with an asymmetrical coplanar strip wall," *IEEE Antennas Wireless Propag. Lett.*, vol. 15, pp. 191–194, 2016.
[61] J. OuYang, F. Yang, and Z. M. Wang, "Reducing mutual coupling of closely spaced microstrip MIMO antennas for WLAN application," *IEEE Antennas Wireless Propag. Lett.*, vol. 10, pp. 310–313, 2011.
[62] X. Yang, Y. Liu, Y.-X. Xu, and S.-X. Gong, "Isolation enhancement in patch antenna array with fractal UC-EBG structure and cross slot," *IEEE Antennas Wireless Propag. Lett.*, vol. 16, pp. 2175–2178, 2017.
[63] X. M. Yang, X. G. Liu, X. Y. Zhou, and T. J. Cui, "Reduction of mutual coupling between closely packed patch antennas using waveguided metamaterials," *IEEE Antennas Wireless Propag. Lett.*, vol. 11, pp. 389–391, 2012.

Qi Wu (Member, IEEE) was born in 1993. He received the B.S. degree in communication engineering from the Communication University of China, Beijing, China, in 2012, and the M.S. and Ph.D. degrees in electrical engineering from Southeast University, Nanjing, China, in 2015 and 2019, respectively.

He joined the State Key Laboratory of Millimeter Waves, Southeast University, in March 2019, where he is currently an Associate Professor. From October 2016 to September 2017, he was a Junior Visiting Fellow at Ando and Hirokawa Laboratory, Tokyo Institute of Technology, Tokyo, Japan. He has authored and coauthored more than 50 technical publications in peer-reviewed international journals and conference proceedings, and more than 20 patents. His current research interest is AI-powered antenna and radio frequency technologies (iART).

Dr. Wu was a recipient of the Honorable Mention at the Student Contest of IEEE APS-URSI Conference in 2018, selected as the Finalist in the IEEE EuCAP in 2016, the student paper contest of iWAT in 2018, and the best poster contest of iWAT in 2019. He has served as a Reviewer for IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS, and IEEE MICROWAVE AND WIRELESS COMPONENTS LETTERS.

Weiqi Chen (Graduate Student Member, IEEE) was born in 1995. She received the B.S. degree in electronic information science and technology from Nankai University, Tianjin, China, in 2018. She is currently pursuing the Ph.D. degree in electrical engineering with Southeast University, Nanjing, China.

Her current research interests include artificial intelligence (AI)-powered antenna and radio frequency technologies (iART).

Chen Yu (Member, IEEE) was born in 1979. She received the M.S. and Ph.D. degrees in electrical engineering from Southeast University, Nanjing, China, in 2004 and 2011, respectively.

She joined the State Key Laboratory of Millimeter Waves, Southeast University, in April 2004, where she is currently an Associate Professor. In 2007, she was a short-term Visiting Scholar at the Kalmar Institute of Technology, Kalmar, Sweden. Her current research interests include antennas and antenna arrays for wireless mobile communications and radars.

Dr. Yu received the first-class Science and Technology Progress Award of Jiangsu Province of China in 2009.

Haiming Wang (Member, IEEE) was born in 1975. He received the B.S., M.S., and Ph.D. degrees in electrical engineering from Southeast University, Nanjing, China, in 1999, 2002, and 2009, respectively.

He joined the School of Information Science and Engineering, Southeast University, in 2002, and is currently a Distinguished Professor. He is also a part-time Professor with the Purple Mountain Laboratories, Nanjing. He has authored and coauthored more than 50 technical publications in IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION and other peer-reviewed academic journals and more than 70 and 52 patents have been granted. His current research interests include AI-powered antenna and radio frequency technologies (iART), AI-powered channel measurement and modeling technologies (iCHAMM), and integrated communications, sensing, and positioning (iCSAP).

Prof. Wang was awarded twice for contributing to the development of IEEE 802.11aj by the IEEE Standards Association in 2018 and 2020.

Wei Hong (Fellow, IEEE) received the B.S. degree from the University of Information Engineering, Zhengzhou, China, in 1982, and the M.S. and Ph.D. degrees from Southeast University, Nanjing, China, in 1985 and 1988, respectively, all in radio engineering.

He is currently a Professor of the School of Information Science and Engineering, Southeast University. In 1993, 1995, 1996, 1997, and 1998, he was a short-term Visiting Scholar with the University of California at Berkeley, Berkeley, CA, USA, and University of California at Santa Cruz, Santa Cruz, CA, USA, respectively. He has been engaged in numerical methods for electromagnetic problems, millimeter-wave theory and technology, antennas, and RF technology for wireless communications. He has authored and coauthored more than 400 technical publications and two books.

Dr. Hong is a fellow of CIE. He was the Vice Presidents of the CIE Microwave Society and Antenna Society and was an elected IEEE MTT-S AdCom Member from 2014 to 2016. He was twice awarded the National Natural Prizes and four times awarded the First-Class Science and Technology Progress Prizes issued by the Ministry of Education of China and Jiangsu Province Government. Besides, he also received the Foundations for China Distinguished Young Investigators and for "Innovation Group" issued by the National Natural Science Foundation of China. He served as the Associate Editor of IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES (MTT) from 2007 to 2010.