Dataset Shrinking for Accelerated Deep Learning-Based Metamaterial Absorber Design

Qimin Ding, Guobin Wan, Nan Wang, and Xin Ma

***Abstract*— The design of metamaterial absorbers (MMAs) using deep learning (DL) models typically involves the construction of a large dataset through time-consuming full-wave simulations (FSs). In this letter, an efficient dataset shrinking method is proposed that employs both electromagnetic (EM) theory and DL theory to enhance design efficiency. The impact of each MMA on training the DL model is estimated through the equivalent circuit model (ECM) of the MMA by analyzing the gradients of the MMA parameters with respect to DL model training. Impact sampling is then used to construct a smaller dataset while still maintaining DL model performance and reducing time on FS. The proposed approach is validated through the design of single- and multilayer absorbers using the inverse-DL algorithm and DL integrated into the genetic algorithm (GA). Numeric results demonstrate that the proposed method achieves a much faster design than those without optimized dataset construction processes.**

***Index Terms*— Dataset shrinking, deep learning (DL), impact sampling, inverse design, metamaterial absorber (MMA).**

I. Introduction

METAMATERIAL absorbers (MMAs) are periodic structures engineered to efficiently absorb electromagnetic (EM) radiation within specific frequency ranges [1], [2], [3]. Deep learning (DL) technology has been recently used to expedite the MMA design process [4], [5], [6], [7].

There are primarily two types of DL models for designing MMAs with vectorized geometric parameters. The first type of DL model takes the design requirements as input and directly outputs the geometric parameters of the MMA structure [6], [8]. The second type of DL model takes the geometric parameters of the MMA as input and outputs its EM responses, such as the absorptivity spectrum. These DL models can be integrated into other optimization algorithms [e.g., the genetic algorithm (GA)] to operate as a faster estimator of the fitness function [9], [10], [11]. Regardless of the approach used, the design performance heavily depends on the size of the dataset used to train the DL models, and the decreased design efficiency caused by the time-consuming dataset construction process is usually ignored in previous works.

Manuscript received 2 May 2023; accepted 11 May 2023. Date of publication 25 May 2023; date of current version 9 August 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62001389. *(Corresponding author: Guobin Wan.)*

The authors are with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: dingqimin@mail.nwpu.edu.cn; gbwan@nwpu.edu.cn; wangnan_nwpu@mail.nwpu.edu.cn; maxin1105@nwpu.edu.cn).

Color versions of one or more figures in this letter are available at https://doi.org/10.1109/LMWT.2023.3276892.

Digital Object Identifier 10.1109/LMWT.2023.3276892

To enhance dataset construction efficiency, it is crucial to identify valuable MMA samples that are more useful for training DL models. The performance-driven modeling of high-frequency structures in confined domains is considered in [12] and [13] to enhance the design efficiency. An equivalent circuit theory-assisted method is proposed in [10], where the DL model is trained using more valuable MMA samples from specific regions of the global parameter space. According to our understanding, the combination of sensitivity analysis (SA) technology [14], [15], [16], [17] and the EM theory has never been used to shrink the dataset used in MMA design problems.

In this letter, an adjoint dataset is constructed with the assistance of equivalent circuit theory before the real dataset construction by full-wave simulation (FS), and the impact indexes of MMA samples with respect to the training process are analyzed by SA theory. A smaller dataset is then constructed by impact sampling, resulting in an enhanced design efficiency.

II. Efficient Design Process

A. Problem Formulation

Let $\mathcal{F}_{\mathrm{NN}}(\boldsymbol{y};\boldsymbol{w})$ denote the first type of DL model to design the MMA according to the design requirements $\boldsymbol{y}$, where $\boldsymbol{w}$ is the weight of the DL model. The output of $\mathcal{F}_{\mathrm{NN}}$ is a vector $\boldsymbol{x}$ represents the geometric parameters of the MMA structure. Consider a general dataset $\mathcal{Z} = \{(\mathcal{X}, \mathcal{Y})\}$, where $\mathcal{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$, $\mathcal{Y} = \{\mathcal{F}_{\mathrm{FS}}(\boldsymbol{x}_1), \mathcal{F}_{\mathrm{FS}}(\boldsymbol{x}_2), \ldots, \mathcal{F}_{\mathrm{FS}}(\boldsymbol{x}_n)\}$, n is the size of this full-size dataset, and $\mathcal{F}_{\mathrm{FS}}(\boldsymbol{x})$ gets the EM responses of the MMA by the FS. To train $\mathcal{F}_{\mathrm{NN}}(\boldsymbol{y};\boldsymbol{w})$, $\mathcal{Y}$ serves as the inputs and $\mathcal{X}$ serves as the outputs of $\mathcal{F}_{\mathrm{NN}}$. It is expected to find a smaller $\mathcal{X}^*$ to construct a dataset $\mathcal{Z}^* = \{(\mathcal{X}^*, \mathcal{Y}^*)\}$ with the size of $m (m < n)$, while the performances of the DL models are similar.

The performance of the DL model is only related to its weight $\boldsymbol{w}$, the similarity of the performance can be regarded as the similarity of the weights trained by different datasets

$$\mathcal{X}^* = \underset{|\mathcal{X}^*|=m,\, \mathcal{Z}^*=\{(\mathcal{X}^*,\mathcal{Y}^*)\}}{\arg\min} \|\boldsymbol{w}^*_{\mathcal{Z}} - \boldsymbol{w}^*_{\mathcal{Z}^*}\|_2 \tag{1}$$

where $|\cdot|$ measures the size of a set, $\boldsymbol{w}^*$ is the optimized weight of the DL model, and $\|\cdot\|_2$ is the two-norm of a vector.

$\mathcal{X}^*$ can be obtained by exhausting C_n^m different subsets when m and $\mathcal{Z}$ are available. However: 1) the FS is the preprocess to obtain $\mathcal{Z}$, which means the time for FS cannot be saved anymore by shrinking the dataset and 2) the cost caused by training the DL model for C_n^m times is unaffordable.

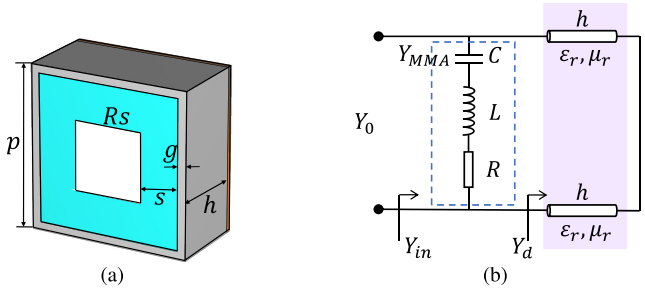

Fig. 1. (a) Geometric structure of the MMA. (b) ECM of the MMA.

B. ECM of the MMA

The equivalent circuit model (ECM) of the MMA is a reasonable surrogate model to construct an adjoint dataset $\mathcal{Z}'$ (though error exists compared with FS) for MMA sampling before generating $\mathcal{Y}$ by FS

$$\mathcal{Z}' = \{z_i'\} = \{(\boldsymbol{x}_i, \boldsymbol{y}_i')\}, \quad \boldsymbol{y}_i' = \mathcal{F}_{\text{ECM}}(\boldsymbol{x}_i) \tag{2}$$

where $\mathcal{F}_{\text{ECM}}$ gets the EM responses of the MMA by the ECM. $\mathcal{F}_{\text{ECM}}$ is usually a simple closed-form expression, so, the cost to construct $\mathcal{Z}'$ is negligible when compared with FS.

For example, the MMA in Fig. 1 has a square-loop-shaped resistance layer on a polymethacrylimide dielectric ($\varepsilon_r = 1.1$, $\mu_r = 1$) with a metal plate at the bottom. With its period p fixed as 10 mm, this type of MMAs can be represented by four parameters, $\boldsymbol{x} = (Rs, g, s, h)$. The absorptivity spectrum of this MMA at discrete frequency points can be acquired by the ECM as

$$\boldsymbol{A}|_\omega = 1 - |\Gamma|^2 = 1 - \left| \frac{Y_0 - (Y_d + Y_{\text{MMA}})}{Y_0 + (Y_d + Y_{\text{MMA}})} \right|^2 \tag{3}$$

where $Y_d = -jY_0(\varepsilon_r/\mu_r)^{1/2} \cot 2\pi(\varepsilon_r\mu_r)^{1/2}(h/\lambda)$, $Y_{\text{MMA}} = 1/(R + j\omega L + 1/(j\omega C))$, and ω is the angular frequency. The RLC values with respect to Rs, g, and s are calculated by equations in [18] and [19]. Thus, the adjoint dataset $\mathcal{Z}'$ can be constructed by simply iterating (3) for n times.

C. Estimate the Impact Index of Each MMA

An alternative solution for 2) raised in Section II-A is to find MMAs that have higher impacts on $\boldsymbol{w}_{\mathcal{Z}}^*$, then $\mathcal{X}^*$ can be generated by sampling these MMAs from $\mathcal{X}$.

The universal formula for optimizing $\boldsymbol{w}_{\mathcal{Z}}^*$ is the following equation, where L adopts the mean square error (mse) loss function expressed as $L(z; \boldsymbol{w}) = \|\boldsymbol{x} - \mathcal{F}_{\text{NN}}(\boldsymbol{y}; \boldsymbol{w})\|_2^2/\dim(\boldsymbol{x})$ in this letter:

$$\boldsymbol{w}_{\mathcal{Z}}^* = \arg\min_{\boldsymbol{w}} \frac{1}{n} \sum_{z_i \in \mathcal{Z}} L(z_i; \boldsymbol{w}). \tag{4}$$

The weight for each sample to optimize $\boldsymbol{w}$ in (4) is $(1/n)$. If the weight of a specific MMA structure $\hat{\boldsymbol{x}}$ is 0, the corresponding sample $\hat{z}$ has no impact on $\boldsymbol{w}_{\mathcal{Z}}^*$. Specifically, when the weight of $\hat{\boldsymbol{x}}$ is $(1/n) + \xi$, (4) is rephrased as

$$\boldsymbol{w}_{\mathcal{Z}-\hat{z};\xi}^* = \arg\min_{\boldsymbol{w}} \frac{1}{n} \sum_{z_i \in \mathcal{Z}} L(z_i; \boldsymbol{w}) + \xi L(\hat{z}; \boldsymbol{w}). \tag{5}$$

When $\xi = -(1/n)$, $\hat{z}$ has no impact on $\boldsymbol{w}_{\mathcal{Z}}^*$. Luckily, the differentiation of (5) with respect to ξ has a closed-form solution when $\xi \to 0$ [20]

$$\Delta \triangleq \left. \frac{d\boldsymbol{w}_{\mathcal{Z}-\hat{z};\xi}^*}{d\xi} \right|_{\xi \to 0} = -H_{\boldsymbol{w}_{\mathcal{Z}}^*}^{-1} \nabla_{\boldsymbol{w}} L(\hat{z}, \boldsymbol{w}_{\mathcal{Z}}^*) \tag{6}$$

where $H_{\boldsymbol{w}_{\mathcal{Z}}^*} = (1/n) \sum_{z_i \in \mathcal{Z}} \nabla_{\boldsymbol{w}}^2 L(z_i; \boldsymbol{w}_{\mathcal{Z}}^*)$, $\nabla_{\boldsymbol{w}}$ and $\nabla_{\boldsymbol{w}}^2$ are the first- and second-gradient operator. Usually, $(1/n)$ is small enough so that the impact index of $\hat{z}$ is defined as the perturbation of $\boldsymbol{w}_{\mathcal{Z}}^*$ caused by removing $\hat{z}$, and is expressed as

$$\left\| \boldsymbol{w}_{\mathcal{Z}-\hat{z};\xi}^* - \boldsymbol{w}_{\mathcal{Z}}^* \right\|_2 \approx \left\| -\frac{1}{n}\Delta \right\|_2 = \left\| \frac{1}{n} H_{\boldsymbol{w}_{\mathcal{Z}}^*}^{-1} \nabla_{\boldsymbol{w}} L(\hat{z}, \boldsymbol{w}_{\mathcal{Z}}^*) \right\|_2. \tag{7}$$

The higher impact index means the MMA sample has a higher probability to be sampled from $\mathcal{X}$. $H_{\boldsymbol{w}_{\mathcal{Z}}^*}$ and $\boldsymbol{w}_{\mathcal{Z}}^*$ in (7) are generally applicable in the dataset, so the impact indexes of all samples can be calculated by n times of gradient operation after training the DL model for only once. This impact sampling process is much faster than C_n^m times training of the DL model mentioned in Section II-A. As a common issue, however, the performance of the DL model is unavailable before training, the final size of $\mathcal{X}^*$ could only be determined by adding more data to $\mathcal{X}^*$ in a progressive manner [21].

Thus, the proposed design process can be summarized as follows.

1) Construct the adjoint dataset $\mathcal{Z}'$ by (2) with ECM.
2) Train $\mathcal{F}_{\text{NN}}$ for the first time with $\mathcal{Z}'$ to get $\boldsymbol{w}_{\mathcal{Z}'}^*$ by (4).
3) Estimate the impact of each MMA in $\mathcal{X}$ with the trained $\mathcal{F}_{\text{NN}}$ by (7).
4) Sample $\mathcal{X}^*$ according to the impact indexes from $\mathcal{X}$, and generate $\mathcal{Y}^*$ by conducting FS on $\mathcal{X}^*$.
5) Train $\mathcal{F}_{\text{NN}}$ from scratch again with $\mathcal{Z}^*$ to get $\boldsymbol{w}_{\mathcal{Z}^*}^*$ by (4), and use $\mathcal{F}_{\text{NN}}$ to design MMAs.

III. Numerical Results

A. Design the Single-Layer Absorber With Inverse-DL

Consider designing the absorber shown in Fig. 1(a) using $\mathcal{F}_{\text{NN}}$ with eight hidden layers by $\boldsymbol{x}^{\text{opt}} = \mathcal{F}_{\text{NN}}(\boldsymbol{y}; \boldsymbol{w})$, where $\boldsymbol{y}$ is an absorptivity spectrum represents specific boundaries of the 90% absorption band.

After eliminating outliers, the full dataset of 2476 samples is constructed through FS by random sampling in the parameter space. The optimizer of the DL model is Rmsprop, with an initial learning rate of 10^{-3} [22]. It is guaranteed that the selection of the structure of DL models can converge at least on the full-size dataset, and the training strategy for all the models is the same so that the experiments are reliable. For convenience, the models trained by the dataset constructed by sampling randomly, sampling by impact indexes calculated by ECM and FS are represented as DL-random, DL-ECM, and DL-FS, respectively.

Fig. 2(a) shows the histogram of the impact indexes of MMAs in the full-size dataset, which is significantly different from the conventional case where each MMA structure has the same impact to $\boldsymbol{w}_{\mathcal{Z}}^*$. Fig. 2(b) shows an example of the selection of MMA samples based on their impact indexes.

The performance of the DL model to design the MMA is shown in Fig. 2(c) when a specific percentage of MMAs are selected to train the model. The performance of DL-FS is the upper bound of the proposed method since the impact

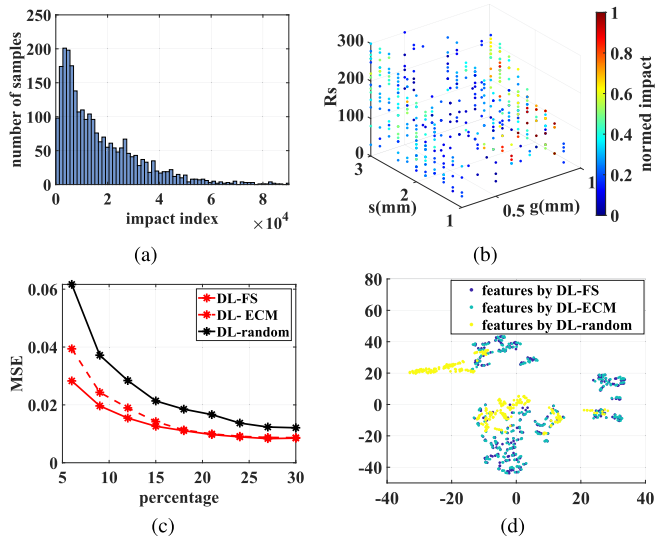

Fig. 2. (a) Histogram of the impacts of MMAs in the dataset. (b) Selected samples by the proposed method. (c) Performance of designing the MMA. (d) Latent features of MMA samples visualized by t-SNE.

indexes are calculated by the accurate dataset constructed by the FS. In practice, however, only the performance of DL-ECM is available, and the performance gap in Fig. 2(c) can be explained by visualizing the distribution of latent features with the help of t-distributed stochastic neighbor embedding (t-SNE) [23], [24]. After the DL model is trained, another 250 samples are used to visualize the outputs of the last hidden layer, which is shown in Fig. 2(d). The error of ECM causes feature perturbation between DL-ECM and DL-FS, resulting in the need for more data for DL-ECM to match DL-FS's performance. The significant deviation of the features generated by DL-random results in its severe performance deterioration, and the data needed for DL-random is much more than that for DL-ECM and DL-FS.

B. Design the Multilayer Absorber With GA-DL

The impact sampling process is also applicable to the second type of DL model denoted by $\mathcal{F}_{\mathrm{NN}}^{-1}(\boldsymbol{x};\boldsymbol{w})$, where -1 indicates the symmetry of inputs and outputs of $\mathcal{F}_{\mathrm{NN}}^{-1}$ and $\mathcal{F}_{\mathrm{NN}}$. Therefore, $\mathcal{Z}$ can still be used for $\mathcal{F}_{\mathrm{NN}}^{-1}$ during training process (4) and sampling process (7) by exchanging $\mathcal{X}$ and $\mathcal{Y}$ in $\mathcal{Z}$.

The GA integrated with DL (GA-DL) is then formulated as $\boldsymbol{x}^{\mathrm{opt}} = \min_{\boldsymbol{x}} f(\mathcal{F}_{\mathrm{NN}}^{-1}(\boldsymbol{x};\boldsymbol{w}),\boldsymbol{y})$, where $f(\cdot)$ is the cost function to be minimized by GA

$$f(\boldsymbol{A},\boldsymbol{y}) = \sum_i \max\left\{(\boldsymbol{A}_i - 0.9)\times\left(2\times\mathbb{1}_{f_i\notin[\boldsymbol{y}_l,\boldsymbol{y}_h]} - 1\right), 0\right\}. \quad (8)$$

In (8), $\boldsymbol{A}$ is the output of $\mathcal{F}_{\mathrm{NN}}^{-1}$, i is the index of discrete frequency points, $\boldsymbol{y}_l$ and $\boldsymbol{y}_h$ are bounds of the absorption band in $\boldsymbol{y}$ (which represents the design requirements).

Consider designing the MMA structure shown in Fig. 3(a) [the flowchart shown in Fig. 3(b)]. The geometric parameters of the MMA is $\boldsymbol{x} = (Rs_1, g_1, h_1, Rs_2, g_2, s_2, h_2, Rs_3, g_3, h_3)$ (unit for Rs: $\Omega/\square$, unit for g, s, h: mm).

4998 MMA samples are sampled randomly from the parameter space after eliminating outliers to form the full-size dataset. Fig. 4(a) shows the minimal size of the dataset to

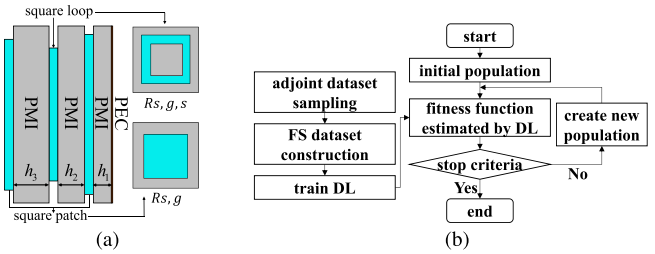

Fig. 3. (a) Structure of the MMA with 10 degrees of freedom. (b) Flowchart of the GA-DL design method.

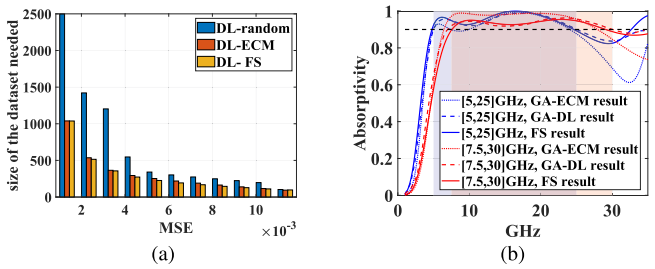

Fig. 4. (a) Minimal size of the dataset versus the performance of the DL model. (b) Design examples with different absorption bands. Geometric parameters are $\boldsymbol{x} = (173, 0.68, 4.30, 133, 0.24, 1.01, 2.77, 205, 1.39, 2.95)$ and $\boldsymbol{x} = (295, 1.17, 2.56, 135, 0.59, 1.13, 2.66, 400, 1.85, 3.17)$.

TABLE I
RUNNING TIME OF THE DESIGN PROCESS

stage	conventional*	proposed
preprocessing	–	$1+5+3=9$min
dataset construction	55hour	11hour
training and optimization	$5+2=7$min	$1+2=3$min
total	$\approx$ 55hour	$\approx$ 11hour

* The method without dataset shrinking refers to [5]–[7], [11].

achieve the specific performance of the DL model by different dataset construction methods, and the proposed method saves lots of time for FS but still achieves the same performance. As a comparison, the mse error between the ECM and the FS is 0.0661, which is much worse than the performance of the DL model (an mse error of 0.0114 can be achieved with only 120 training samples). Fig. 4(b) shows two design examples of the GA-DL algorithm, where GA-ECM fails to satisfy the requirements due to the error between ECM and the FS but GA-DL meets the design requirements effectively.

The running time of the design process with and without dataset shrinking are compared in Table I. The preprocessing stage contains the construction of $\mathcal{Z}'$, training the DL model on $\mathcal{Z}'$, and the calculation of impact indexes. It can be observed in Table I that since the time of FS is much saved, the design efficiency is significantly enhanced.

IV. CONCLUSION

A novel dataset-shrinking method is proposed to enhance the efficiency of DL-based MMA design algorithms. The study provides a detailed explanation of the necessity of using ECM and SA theory in the dataset-shrinking process and discusses the affordable side effect introduced by ECM. The proposed method is applicable to both inverse-DL and DL-GA design algorithms. The accuracy and efficiency of the proposed method are demonstrated through numerical validations on both single-layer and multilayer MMA absorbers.

REFERENCES

[1] G. Deng, Z. Yu, J. Yang, Z. Yin, Y. Li, and B. Chi, "A miniaturized 3-D metamaterial absorber with wide angle stability," *IEEE Microw. Wireless Compon. Lett.*, vol. 32, no. 9, pp. 1111–1114, Sep. 2022.

[2] N. I. Landy, S. Sajuyigbe, J. J. Mock, D. R. Smith, and W. J. Padilla, "Perfect metamaterial absorber," *Phys. Rev. Lett.*, vol. 100, no. 20, p. 207402, May 2008.

[3] P. Zhou et al., "A stretchable metamaterial absorber with deformation compensation design at microwave frequencies," *IEEE Trans. Antennas Propag.*, vol. 67, no. 1, pp. 291–297, Jan. 2019.

[4] 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.

[5] J. Chen et al., "Absorption and diffusion enabled ultrathin broadband metamaterial absorber designed by deep neural network and PSO," *IEEE Antennas Wireless Propag. Lett.*, vol. 20, no. 10, pp. 1993–1997, Oct. 2021.

[6] J. Hou et al., "Customized inverse design of metamaterial absorber based on target-driven deep learning method," *IEEE Access*, vol. 8, pp. 211849–211859, 2020.

[7] L. Yuan, L. Wang, X. Yang, H. Huang, and B. Wang, "An efficient artificial neural network model for inverse design of metasurfaces," *IEEE Antennas Wireless Propag. Lett.*, vol. 20, no. 6, pp. 1013–1017, Jun. 2021.

[8] T. Qiu et al., "Deep learning: A rapid and efficient route to automatic metasurface design," *Adv. Sci.*, vol. 6, no. 12, Jun. 2019, Art. no. 1900128.

[9] M. Ohira, H. Deguchi, M. Tsuji, and H. Shigesawa, "Multiband single-layer frequency selective surface designed by combination of genetic algorithm and geometry-refinement technique," *IEEE Trans. Antennas Propag.*, vol. 52, no. 11, pp. 2925–2931, Nov. 2004.

[10] Z. Wei et al., "Equivalent circuit theory-assisted deep learning for accelerated generative design of metasurfaces," *IEEE Trans. Antennas Propag.*, vol. 70, no. 7, pp. 5120–5129, Jul. 2022.

[11] R. Zhu et al., "Multiplexing the aperture of a metasurface: Inverse design via deep-learning-forward genetic algorithm," *J. Phys. D: Appl. Phys.*, vol. 53, no. 45, Nov. 2020, Art. no. 455002.

[12] S. Koziel, A. Pietrenko-Dabrowska, and U. Ullah, "Low-cost modeling of microwave components by means of two-stage inverse/forward surrogates and domain confinement," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 12, pp. 5189–5202, Dec. 2021.

[13] A. Pietrenko-Dabrowska and S. Koziel, "Surrogate modeling of impedance matching transformers by means of variable-fidelity electromagnetic simulations and nested cokriging," *Int. J. RF Microw. Comput.-Aided Eng.*, vol. 30, no. 8, p. e22268, Aug. 2020.

[14] P. Zhang, "A novel feature selection method based on global sensitivity analysis with application in machine learning-based prediction model," *Appl. Soft Comput.*, vol. 85, Dec. 2019, Art. no. 105859.

[15] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, Dec. 2017, pp. 4768–4777.

[16] B. Mirzasoleiman, J. Bilmes, and J. Leskovec, "Coresets for data-efficient training of machine learning models," in *Proc. 37th Int. Conf. Mach. Learn.*, 2020, pp. 6950–6960.

[17] M. Wojnowicz et al., "'Influence sketching': Finding influential samples in large-scale regressions," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, 2016, pp. 3601–3612.

[18] M. I. Hossain, N. Nguyen-Trong, and A. M. Abbosh, "Broadband magnetic absorber based on double-layer frequency-selective surface," *IEEE Trans. Antennas Propag.*, vol. 70, no. 1, pp. 410–419, Jan. 2022.

[19] R. J. Langley and E. A. Parker, "Equivalent circuit model for arrays of square loops," *Electron. Lett.*, vol. 18, no. 7, pp. 294–296, Apr. 1982.

[20] R. D. Cook and S. Weisberg, *Residuals and Influence in Regression*. London, U.K.: Chapman & Hall, 1982.

[21] F. Wen, J. Jiang, and J. A. Fan, "Robust freeform metasurface design based on progressively growing generative networks," *ACS Photon.*, vol. 7, no. 8, pp. 2098–2104, Aug. 2020.

[22] M. C. Mukkamala and M. Hein, "Variants of RMSProp and Adagrad with logarithmic regret bounds," in *Proc. 34th Int. Conf. Mach. Learn.*, 2017, pp. 2545–2553.

[23] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[24] L. van der Maaten, "Accelerating t-SNE using tree-based algorithms," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3221–3245, 2014.