Article:

Ma, J. orcid.org/0000-0003-0583-4349, Dang, S. orcid.org/0000-0002-0018-815X, Watkins, G. orcid.org/0000-0001-9358-5041 et al. (2 more authors) (2023) A High-Performance Transfer Learning-Based Model for Microwave Structure Behavior Prediction. IEEE Transactions on Circuits and Systems II: Express Briefs. ISSN 1549-7747

https://doi.org/10.1109/tcsii.2023.3296454

This article has been accepted for publication in IEEE Transactions on Circuits and Systems--II: Express Briefs. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCSII.2023.3296454

A High-Performance Transfer Learning-Based Model for Microwave Structure Behavior Prediction

Jiteng Ma, *Graduate Student Member, IEEE*, Shuping Dang, *Member, IEEE*, Gavin Watkins, *Member, IEEE*, Kevin Morris, *Member, IEEE*, and Mark Beach, *Senior Member, IEEE*

***Abstract*—Microwave structure behavior prediction enables the estimation of circuit response over a frequency range, playing a crucial role in the design of radio frequency (RF) structures. Deep neural network (DNN) approaches have demonstrated their capability to simulate microwave structure behaviors. Nonetheless, the quality and utility of the model are constrained by the availability of data and computational capabilities. These inherent disadvantages hinder the extensive application of DNN in microwave structure behavior prediction. Transfer learning has recently been produced as a method offering improved accuracy and speed for predicting microwave circuit behavior. This paper proposes a novel transfer learning-based model to expedite the prediction process for a sequence of frequency samples. Through experimental validation, it is illustrated that the proposed methodology outperforms the conventional DNN techniques for microwave structure behavior prediction by effectively reducing the required data and shortening the training time. The proposed model also facilitates the fine-tuning of hyperparameters and reduces the simulator computing load.**

***Index Terms*—Transfer learning, deep neural network, microwave behavior prediction, frequency response.**

I. Introduction

MICROWAVE structure simulation plays an essential role in the current radio frequency (RF) design process. As the structure becomes highly integrated while operating at higher frequencies, microwave structure verification and optimization rely on fast and accurate simulation to estimate behaviors before fabrication. Microwave structures were initially analyzed using electromagnetic or microwave theories. However, these approaches become impractical when dealing with complex structures in practice [1], [2]. Computer-based circuit simulation and electromagnetic (EM) simulation software were developed to address this challenge. However, circuit simulations often lack the desired level of accuracy, while EM simulations can be excessively time-consuming, hindering design and manufacturing processes [3]. DNN has been successfully implemented in various microwave applications owing to its exceptional ability to solve multi-dimensional and non-linear problems. Research has been widely conducted to explore microwave structure behavior prediction using DNN [4], [5] Preliminary results have demonstrated that well-trained deep neural networks (DNN) can promptly and accurately predict the behaviors of microwave structures sharing similar characteristics [6]–[10]. RF engineers can thus save much time by doing electromagnetic (EM) simulations by reusing well-trained models. Additional research substantiates the capability of machine learning technologies in various applications involving circuit components [11], [12].

Additionally, the neural network simulator can assist in training neural network models for microwave design tasks [13]–[15]. This could help with mitigating the non-uniqueness design problem in microwave design [16]. However, one significant challenge of using DNN-based methods is the limited scalability. When faced with slightly modified tasks, the entire process of updating data and training models may need to start from scratch. The absence of scalability in the DNN-based approach could introduce rigidity and inefficiencies.

Some methods have been proposed to accelerate the behavior modeling of microwave components. Knowledge-based neural network (KBNN) was proposed, where the neural networks are offered additional information that may not be sufficiently captured in a limited training dataset [17]. KBNN greatly enhances the learning and generalization capabilities of the neural network, particularly when the provided knowledge is beyond the boundaries of the training data. However, prior knowledge is often derived from extrapolation and requires a specific format to be leveraged by the neural network, which generally requires specific domain knowledge and microwave circuit design experience. Another well-known method to enhance EM modeling is space mapping, combining coarse and fine models [18]. Coarse models are typically generated using neural network models, and this approach helps reduce the required computational resources. However, space mapping still relies on empirical knowledge, and introducing the coarse-fine model structures increases the complexity when fine-tuning the models.

In light of the prior work, we propose a novel DNN model relying on transfer learning in this paper. The basic idea is to split a complex simulation task into multiple simple tasks in different frequencies. We start by training a model for a simple task. Similar tasks can then leverage the knowledge from the well-trained source model. The simulation results show that fewer computation resources are needed for training models than conventional approaches. The measurement results verify that the proposed model dramatically improves training efficiency and accuracy.

This work was supported by Toshiba's Bristol Research and Innovation Laboratory Contribution to the United Kingdom Research and Innovation (UKRI)/ Engineering and Physical Sciences Research Council (EPSRC) Prosperity Partnership in Secure Wireless Agile Networks (SWAN) under Grant EP/T005572/1.

Jiteng Ma, Shuping Dang, and Mark Beach are with the Department of Electrical and Electronic Engineering, University of Bristol, Bristol BS8 1UB, U.K., (e-mail: jiteng.ma@bristol.ac.uk; shuping.dang@bristol.ac.uk; m.a.beach@bristol.ac.uk).

Gavin Watkins is with Toshiba Europe Ltd., Bristol BS1 4ND, U.K., (e-mail: gavin.watkins@toshiba-bril.com).

Kevin Morris is with the School of Electronic and Electrical Engineering, University of Leeds, Leeds, LS2 9JT U.K., (e-mail: k.morris1@leeds.ac.uk).

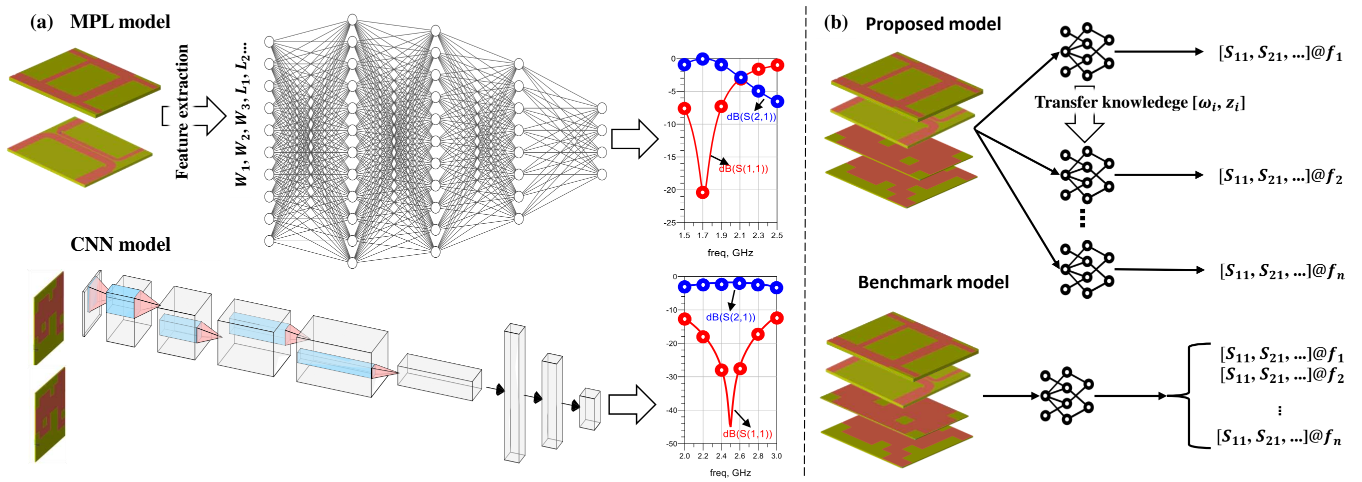

Fig. 1. (a) Comparison of the classical DNN architectures for microwave structure behavior prediction. Two DNN architectures were generated by [19]. (b) Implementation of the proposed transfer learning-based model and the benchmark (conventional) model.

The rest of this paper is organized as follows. Section II illustrates the fundamental DNN architecture for microwave component simulation and our proposed transfer learning-based model. In Section III, we compare our proposed model with the benchmark model in terms of the amount of training data, training time, and neural network size. We also fabricate circuits to validate the performance advantage brought by our model. This work is finally concluded in Section IV.

II. Transfer Learning-Based Model for Microwave Structure Behavior Prediction

A. DNN Architectures

Multi-layer perceptron (MLP) is the simplest DNN with multiple fully connected layers. Each layer is connected with an activation function for non-linear affine transformation. MLP is widely used in different microwave structure simulations, especially when the structure parameters are heterogeneous. However, the mutual position information in the structure's geometries is lost since all the features are extracted in an array as the input of the MLP network. An example of simulating via MLP is illustrated at the top in Fig. 1(a).

In recent years, convolutional neural network (CNN) has become a well-known architecture due to its ability to extract mutual position information. The convolutional operations with kernels provide translation-equivariant responses known as feature maps. CNN is apt for microwave structure simulation since the position features are vital to predict behavior [15]. However, the structure parameters must be homogeneous when applying CNN. An example of simulating a grid-like structure using CNN is shown at the bottom in Fig. 1(a).

B. Proposed Transfer Learning-Based Model

Transfer learning is a machine learning technique where a model trained on one task is re-purposed on a second related task. Instead of restarting the learning process from scratch, the model leverages knowledge gained from the first task to inform its decisions on the second task. In this way, transfer learning can save time and computational resources and improve the model's performance on the second task [20].

In the microwave structure simulation task, the S-parameter is always a continuous argument that varies over the operating frequency. Because of the continuity, transfer learning can leverage the source knowledge to predict adjacent frequency responses within limited training epochs. In addition, the source knowledge enables the model to be generalized for being adaptive depending on new tasks.

The transferred knowledge in this work corresponds to the weight and bias vectors obtained from the well-trained source models for the first frequency point. The output, denoted as $\mathbf{a}_{i+1}$, from fully connected layer $i+1$ can be defined as

$$\mathbf{a}_{i+1} = \mathcal{A}(\omega_i^T \cdot \mathbf{a}_i + z_i), \quad i = 1, 2, \ldots, m, m+1, \ldots, n \quad (1)$$

where $\mathcal{A}(\cdot)$ is the activation function; ω_i and z_i are the weight and bias vectors of layer i. For convolutional layers, ω_i represents the convolutional kernel, and z_i is the scalar bias; the dot product operation in (1) is replaced by the cross-correlation operation. In this case, $[\omega_i, z_i]$ is the source knowledge, which is transferred to the models trained for target tasks as initial parameters.

The processes of the proposed method and the benchmark (conventional) method are pictorially illustrated in Fig. 1(b) for comparison purposes. The benchmark model develops a single DNN to predict multiple outputs corresponding to multiple microwave structures' frequency response samples [14], [18]. In contrast, the proposed model divides microwave simulation tasks into multiple sub-tasks with a similar nature. Then, a DNN model is trained to predict a single frequency sampling point. Consequently, given a well-trained DNN model pertaining to a single sampling point, transfer learning can be implemented to extract the knowledge from this well-trained model and facilitate the model training process for the rest of the sampling points. Accordingly, the proposed model training process is formulated and explained in Algorithm 1.

Algorithm 1 Proposed transfer learning method.

Require: The labeled training and testing data for different frequencies and a suitable DNN.
Ensure: Test loss is below a preset threshold.

Split the whole task training data into multiple sub-tasks
while the whole task test loss is above the threshold **do**
    while the first sub-task loss is above the threshold **do**
        Import the data and fine-tune the hyperparameters
        Train DNN for the first sub-task and validate loss
    end while
    while untrained sub-tasks exist **do**
        Import the data and reuse the fine-tuned hyperparameters with fewer training epochs
        Train DNN for the sub-task and validate test loss
        if the sub-task test loss is above the threshold **then**
            Add more epochs and retrain the network
        end if
        Save the hyperparameters of the sub-task
    end while
    Reduce the sub-task threshold and train the DNN
end while

III. Performance Evaluation and Discussion

In this section, we compare the performance of the proposed transfer learning-based method with other classical methods using both MLP and CNN architectures. We adopt ReLU as the activation function and connect a dropout layer after the activation function to prevent the model from overfitting [21]. The implemented CNN is composed of multiple convolutional blocks and fully connected layers [22]. As microwave structure behavior prediction is a regression problem, the mean squared error (MSE) is selected as the loss function, and the coefficient of determination (commonly known as R^2) is employed as a quantitatively evaluate the fitting precision between the predicted vector $\hat{\mathbf{y}} \in \{\hat{y}_1, \cdots, \hat{y}_n\}$ and the observed vector $\mathbf{y} \in \{y_1, \cdots, y_n\}$. R^2 is a statistical measure for regression models that determines the proportion of variance defined by

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (y_i - \overline{y})}, \tag{2}$$

$\overline{y}$ is the mean of the observed data, i.e.,

$$\overline{y} = \frac{1}{n} \sum_{i-1}^{n} y_i. \tag{3}$$

The range of R^2 is normalized between 0 and 1. The higher the value, the better the linear regression fits the data.

Unlike the conventional DNN network architecture, the parameters of the saved sub-task network must be loaded to each sub-task to implement this model. The whole task output is derived from combining the outputs of all the sub-task networks. The implementation of the model may lead to increased latency. However, in the context of microwave structure behavior prediction and its associated designs, the latency is tolerable as long as it is below a certain threshold.

To thoroughly evaluate the performance of the proposed method, we design and conduct comparisons in three dimensions: the amount of training data, training time, and neural network size. The computing platform for the following experiments is an Intel i9-9900X CPU @3.50 GHz, and an Nvidia RTX2080 GPU with 12 GB memory. The DNN models are implemented on the PyTorch framework. The Keysight ADS EM simulation labels the dataset, and we use an 80/20 training/test split on the dataset.

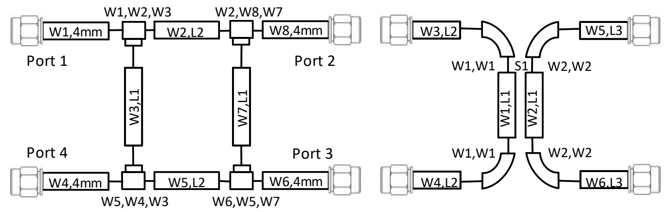

Fig. 2. Templates of the branchline coupler and coupled-line coupler.

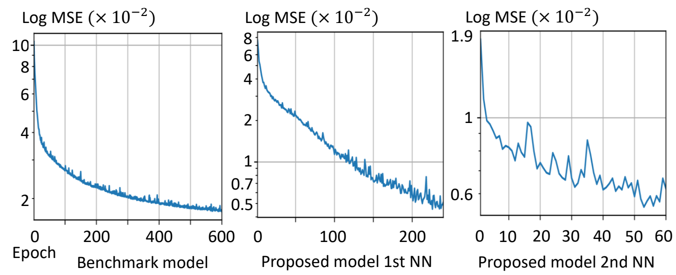

Fig. 3. Training loss comparison between the benchmark and proposed model.

A. Transmission Line Structure Simulation Using MLP

1) Branchline Coupler: The first case is to predict the branchline couplers' behavior. A typical branchline coupler is the quadrature coupler that provides an equal splitting ratio of 3 dB at the center frequency. Without loss of generality, the unequal ratio can be developed by varying the impedance of the arms of a branchline.

The template of the branchline coupler is shown in Fig. 2, where the width W_i and length L_j of each variable component are annotated. The structure is connected with fixed $50\,\Omega$ connectors for each port. The widths and lengths of transmission lines signify the coupler features. 30,000 randomly generated structures are labeled. As shown in Fig 1, we develop a MLP model with five fully connected layers in this case.

The advantage of the proposed model can be explicitly demonstrated in Fig. 3, which compares the training loss of the benchmark model and the proposed model. The benchmark model converges after 600 training epochs with a training MSE of 1.01% for six different frequency points. In the proposed model, the first DNN converges after only 240 epochs of training with an even lower MSE of 0.63% for the first frequency point. The second and the rest models leverage the knowledge of the adjacent DNN and coverage within much fewer training epochs of around 60. As a result, the average training MSE for the benchmark model is 1.01%, while the proposed model reaches a much superior result of 0.77%.

The validation results of R^2 using the testing data are presented and compared in Fig. 4. Firstly, the amounts of training data are compared on the left bar chart, while the training time and the neural network size are identical. The

This article has been accepted for publication in IEEE Transactions on Circuits and Systems--II: Express Briefs. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCSII.2023.3296454

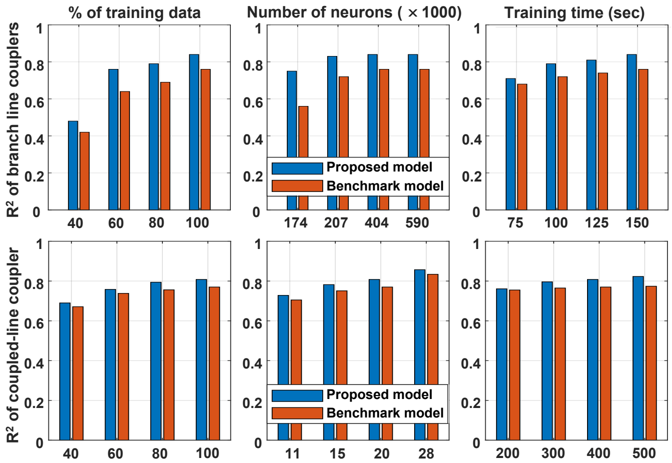

Fig. 4. Performance comparison of the branchline and coupled-line couplers.

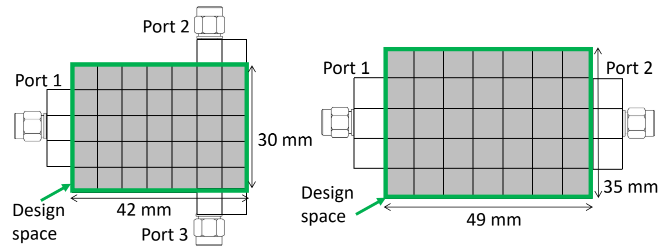

Fig. 5. The design template of the grid-like structures includes the three-port combiner (left) and two-port matching network (right).

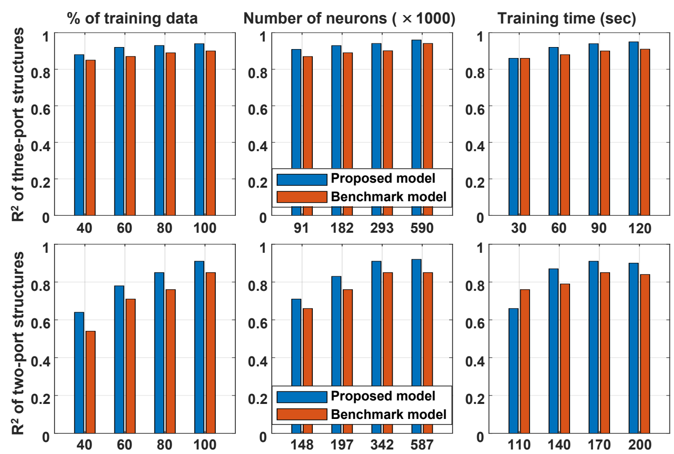

Fig. 6. Performance comparison for three-port and two-port structures.

results show that the R^2 of the proposed model is significantly lower than that of the benchmark when the training data amount is larger than 60% (14,400). However, if the training data is less than 40% (9,600), the performance of two models gets close. The main reason is that the transferred knowledge from insufficient training data cannot benefit the target tasks.

We then compare the R^2 rate by sweeping the total training time with a neural network consisting of 4.04×10^5 neurons. As shown in the middle of Fig. 4, the R^2 of the proposed model is 14% higher than that of the benchmark model in terms of the rate when the training time is 150 s. However, the performance of the two models gets close when the training time reduces to 75 s. These results signify that sufficient training time is required to exploit the proposed model.

The performance of different network sizes are also compared. These neural networks are trained for 150 s and share the same MLP structure as shown in Fig. 1, but each layer is deployed with different numbers of neurons. As illustrated in the bar chart on the right side of Fig. 4, the proposed model predicts more accurately than the benchmark model, regardless of the number of neurons deployed in the network.

2) Coupled-Line Coupler: In the case of coupled-line couplers, we gathered a dataset comprising 20,000 structures. The design template is depicted on the right side of Fig. 2. The results in Fig. 4 demonstrate the exceptional performance of the proposed model in various aspects. Compared to the benchmark model, the proposed model consistently exhibits a lower R^2 value, regardless of the amount of training data and the number of neurons in MLP. However, the performance of the two models gets close when the training time is 200 s, which indicates that the proposed model requires a minimum training time to surpass the benchmark model.

B. Grid-Like Structure Simulation Using CNN

Instead of considering the microwave structure as the width and length of several predefined elements, an all-inclusive way of modeling a geometric structure is to use a finite number of grid-like patterns to tessellate the entire design area. This finite element method is widely implemented by the numerical EM structure simulation software. The higher the number we choose, the more preciously the model can represent an actual structure. Accordingly, more labeled data is required to train the neural network. As shown in fig. 1, the design space is partitioned into identical squares, through which we develop a CNN network including six convolutional layers and three fully connected layers to grasp the electromagnetic propagation properties.

1) Three-Port Structures: In this case, the model predicts the performance of three-port structures, whose design template is illustrated on the left side of Fig. 5. The design space is comprised of 35 identical 6 mm×6 mm cells, which can be described by 5×7 binary matrices. In total, 40,000 randomly generated structures are labeled.

The performance is compared on three bar charts at the top of Fig. 6. By sweeping the proportion of training data, it shows that the R^2 of the proposed model is significantly lower than that of the benchmark model. The different neural network sizes are also compared within 90 s of training time. The results show that the proposed model predicts more accurately than the benchmark model, regardless of the number of neurons deployed in the network. By sweeping the total training time with a neural network of 3.42×10^5 neurons, the R^2 of the proposed model is higher than that of the benchmark model when the training time is longer than 60 s. However, their performance gets closer when the training time reduces to 30 s. The results signify that sufficient training data and time are both required to exploit the proposed model.

2) Two-Port Structures: Similar comparisons are carried out for two-port structures, where we collect 20,000 structures. The design template is shown on the right side of Fig. 5, where

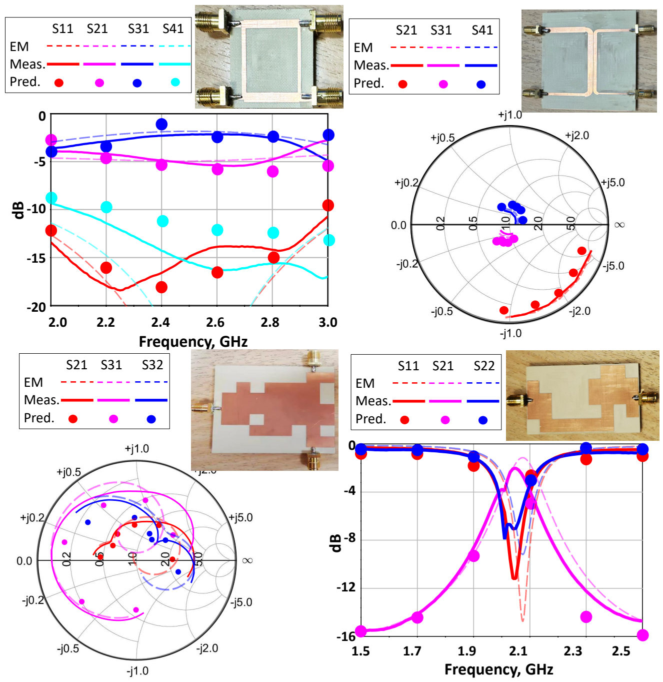

Fig. 7. Microwave structure prototypes and comparison among EM simulation, measurement, and DNN prediction.

each squaring cell is 7 mm×7 mm. The result in Fig. 6 clearly shows the superiority of the proposed model. The R^2 of the proposed model is lower than that of the benchmark model even if 40% of training data or time is removed. The proposed model performs better than the benchmark using networks with different numbers of neurons. However, if the training time is dropped to 110 s, the benchmark outperforms.

C. Fabrication Validation

To evaluate the prediction results, four microwave structure prototypes were fabricated on an Isola substrate of 0.762 mm thickness with a dielectric constant of 2.8. Fig. 7 visualizes the prototypes and presents the EM and measurement results from a vector network analyzer. All structures achieve a relatively small difference between the DNN prediction and EM simulation. The differences between DNN prediction and measurement results are slightly higher due to the variation between the EM and measurement results.

IV. Conclusion

In summary, this paper proposed a transfer-learning-based method to enhance the microwave structure prediction ability by splitting a complex simulation task into multiple simple sub-tasks and solving them by leveraging transfer learning. The experimental results showed that the proposed method could help with the reduction in the required time and computing resources for training neural networks. The experimental results corroborated that there exist a minimum amount of training data and required training time for fully exploiting the proposed method based on transfer learning. The proposed method has the potential to be used in other circuits and systems-related research topics, especially cases that require predicting analog and continuous waveforms.

References

[1] D. M. Pozar, *Microwave engineering*. John wiley & sons, 2011.

[2] G. Huang, J. Hu, Y. He, J. Liu, M. Ma, Z. Shen, J. Wu, Y. Xu, H. Zhang, K. Zhong *et al.*, “Machine learning for electronic design automation: A survey,” *ACM Trans. Des. Autom. Electron. Syst.*, vol. 26, no. 5, pp. 1–46, 2021.

[3] M. B. Yelten, T. Zhu, S. Koziel, P. D. Franzon, and M. B. Steer, “Demystifying surrogate modeling for circuits and systems,” *IEEE Circuits and Systems Magazine*, vol. 12, no. 1, pp. 45–63, 2012.

[4] Q.-J. Zhang, E. Gad, B. Nouri, W. Na, and M. Nakhla, “Simulation and automated modeling of microwave circuits: State-of-the-art and emerging trends,” *IEEE J. Microwaves*, vol. 1, no. 1, pp. 494–507, 2021.

[5] Z. Liu, X. Hu, T. Liu, X. Li, W. Wang, and F. M. Ghannouchi, “Attention-based deep neural network behavioral model for wideband wireless power amplifiers,” *IEEE Microw. Wirel. Compon. Lett.*, vol. 30, no. 1, pp. 82–85, 2020.

[6] H. Kabir, Y. Wang, M. Yu, and Q.-J. Zhang, “Neural network inverse modeling and applications to microwave filter design,” *IEEE Trans. Microw. Theory Tech.*, vol. 56, no. 4, pp. 867–879, 2008.

[7] H. T. Nguyen and A. F. Peterson, “Machine learning for automating millimeter-wave directional coupler designs,” *IEEE Trans. Circuits Syst. II: Express Br.*, vol. 69, no. 2, pp. 549–553, 2021.

[8] H. Cai, Y. Shen, and S. Hu, “Cascaded deep Q-network for optimization of substrate integrated waveguide couplers,” *IEEE Trans. Circuits Syst. II: Express Br.*, 2023.

[9] J. E. Rayas-Sánchez, S. Koziel, and J. W. Bandler, “Advanced RF and microwave design optimization: A journey and a vision of future trends,” *IEEE J. Microwaves*, vol. 1, no. 1, pp. 481–493, 2021.

[10] Y. Shen, J. Xu, J. Yi, E. Chen, and V. Chen, “Class-E power amplifiers incorporating fingerprint augmentation with combinatorial security primitives for machine-learning-based authentication in 65 nm CMOS,” *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 69, no. 5, pp. 1896–1909, 2022.

[11] Y. Zhao and Y. Lian, “Event-driven circuits and systems: A promising low power technique for intelligent sensors in AIoT era,” *IEEE Trans. Circuits Syst. II: Express Br.*, vol. 69, no. 7, pp. 3122–3128, 2022.

[12] A. Fathollahi, M. Gheisarnejad, B. Andresen, H. Farsizadeh, and M.-H. Khooban, “Robust artificial intelligence controller for stabilization of full-bridge converters feeding constant power loads,” *IEEE Trans. Circuits Syst. II: Express Br.*, 2023.

[13] H. T. Nguyen and A. F. Peterson, “Machine learning for automating the design of millimeter-wave baluns,” *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 68, no. 6, pp. 2329–2340, 2021.

[14] G. Zhang, H. He, and D. Katabi, “Circuit-GNN: Graph neural networks for distributed circuit design,” in *Proc. IEEE International Conference on Machine Learning*. PMLR, 2019, pp. 7364–7373.

[15] J. Ma, S. Dang, P. Li, G. T. Watkins, K. Morris, and M. Beach, “A learning-based methodology for microwave passive component design,” *IEEE Trans. Microw. Theory Tech.*, 2023.

[16] H. Kabir, Y. Wang, M. Yu, and Q.-J. Zhang, “Neural network inverse modeling and applications to microwave filter design,” *IEEE Trans. Microw. Theory Tech.*, vol. 56, no. 4, pp. 867–879, 2008.

[17] F. Wang and Q.-J. Zhang, “Knowledge-based neural models for microwave design,” *IEEE Trans. Microw. Theory Tech.*, vol. 45, no. 12, pp. 2333–2343, 1997.

[18] F. Feng, W. Na, J. Jin, W. Zhang, and Q.-J. Zhang, “ANNs for fast parameterized EM modeling: The state of the art in machine learning for design automation of passive microwave structures,” *IEEE Microw. Mag.*, vol. 22, no. 10, pp. 37–50, 2021.

[19] A. LeNail, “NN-SVG: Publication-ready neural network architecture schematics.” *J. Open Source Softw.*, vol. 4, no. 33, p. 747, 2019.

[20] J. Ma, S. Dang, P. Li, G. Watkins, K. Morris, and M. Beach, “Transfer learning for the behavior prediction of microwave structures,” *IEEE Microw. Wirel. Compon. Lett.*, pp. 1–4, 2022.

[21] L. Wan, M. Zeiler, S. Zhang, Y. Le Cun, and R. Fergus, “Regularization of neural networks using dropconnect,” in *Proc. IEEE International Conference on Machine Learning*. PMLR, 2013, pp. 1058–1066.

[22] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, “Dive into deep learning,” *arXiv preprint arXiv:2106.11342*, 2021.