

Transfer Learning for the Behavior Prediction of Microwave Structures

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

25-10-2022 / 02-11-2022

CITATION

Ma, Jiteng; Dang, Shuping; Li, Peizheng; Watkins, Gavin; Morris, Kevin; Beach, Mark (2022): Transfer Learning for the Behavior Prediction of Microwave Structures. TechRxiv. Preprint.
<https://doi.org/10.36227/techrxiv.21397347.v1>

DOI

[10.36227/techrxiv.21397347.v1](https://doi.org/10.36227/techrxiv.21397347.v1)

Transfer Learning for the Behavior Prediction of Microwave Structures

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

Abstract—Microwave structures behavior prediction is an important research topic in radio frequency (RF) design. In recent years, deep-learning-based techniques have been widely implemented to study microwaves, and they are envisaged to revolutionize this arduous and time-consuming work. However, empirical data collection and neural network training are two significant challenges of applying deep learning techniques to practical RF modeling and design problems. To this end, this letter investigates a transfer-learning-based approach to improve the accuracy and efficiency of predicting microwave structure behaviors. Through experimental comparisons, we validate that the proposed approach can reduce the amount of data required for training while shortening the neural network training time for the behavior prediction of microwave structures.

Index Terms—Transfer learning, deep neural network, microwave structure, behavior prediction.

I. INTRODUCTION

DEEP learning has been widely acknowledged as an effective paradigm for the behavior prediction and design of microwave structures due to the ability to discover complicated and non-linear relationships [1], [2]. To model electromagnetic (EM) behaviors of microwave structures, radio frequency (RF) engineers often analyze the wave propagation characteristics based on Maxwell’s equations through the electronic design automation (EDA) software. Deep learning has been well explored in the microwave structure modeling to get the RF designers away from complicated and tedious simulation and optimization process [3], [4].

However, training a qualified deep learning model requires a massive amount of empirical data. In addition, both modeling structures through the EDA software and conducting EM simulation are time-consuming. Also, the behavior modeling of microwave structures often depends on massive variables, including the operation frequency, material, and components geometries. A single neural network can hardly cover all these variables. As a result, the training data for the behavior modeling of microwave structures require to be recollected, and the model needs to be completely retrained from scratch once any of these parameters change [5]. Space mapping is a well-known technique to deal with these variations

This work was supported by the SWAN Prosperity Partnership (UKRI/EPSC) under Grant EP/T005572/1.

Jiteng Ma, Shuping Dang, Peizheng Li, Kevin Morris, and Mark Beach are with the Department of Electrical and Electronic Engineering, University of Bristol, Bristol BS8 1UB, U.K. (e-mail: {jiteng.ma, shuping.dang, peizheng.li, kevin.morris, 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).

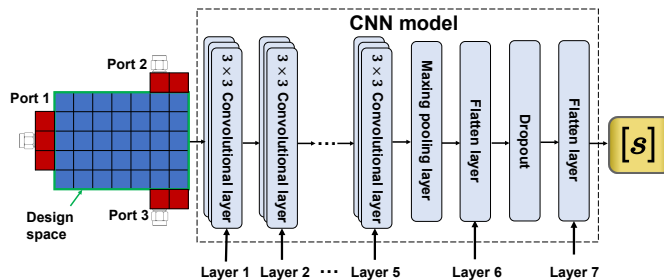


Fig. 1. Implementation block diagram of the proposed neural network model.

by exploiting the knowledge of the coarse model [6], [7]. However, the performance of space mapping depends on the empirical knowledge of the course model, which increases the complexity of the microwave behavior prediction model.

To address these challenges, we introduce transfer learning to improve the efficiency of microwave structures behavior prediction. Motivated by insufficient training data, transfer learning has been applied to applications of bioinformatics, robotics, and communications [8], [9]. Recently, transfer learning has also been successfully used to model the non-linear features of power amplifiers [10], [11]. We carry out extensive experiments to make comprehensive comparisons in operation frequency and structure size and examine its scalability. Systematic simulation results presented in this letter show that transfer learning is able to significantly improve the neural network accuracy in predicting microwave structures’ behaviors. It is also revealed that the source task is beneficial for a target task since fewer data and shorter time are required for training. We subsequently validate that the performance of transfer learning depends on the relevancy of the source and target tasks by comparing two different transfer learning tasks.

II. DEEP-LEARNING-BASED ON TRANSFER LEARNING

A. Deep Learning Model and Data Collection

Convolutional neural network (CNN) models have been validated to be influential in the field of image processing since image pixels are highly correlated to the neighboring [12], [13], [14]. Sharing a similar mathematical nature to image processing, CNN models can be used for precisely predicting the response for microwave structure geometries [15].

The diagram of the CNN model in this letter is illustrated in Fig. 1. The first convolutional block contains five convolutional layers to capture the spacial information of input structures. Each layer uses a 3×3 kernel and applies the ReLU activation

function. The dense block with two fully connected layers rearranges the output from the convolutional block. A dropout layer with ratio of 20% is introduced to the proposed model to prevent the model from overfitting [14]. The aforementioned neural network topology and its hyperparameters are all fine-tuned results using domain knowledge. By the CNN model, empirically, any number of convolutional layers and fully connected layers should work. The deeper network in generally performs better [16]. In this paper, a fine-tuned five convolutional layers CNN is chosen for having a decent microwave structure behavior prediction, while the network is not too large to run out of the GPU memory.

Inspired by the universal mesh solution suited for any given structure in EDA software, we characterize the microwave structures by dividing them into multiple square cells and describe them with binary matrices. Consequently, the topological resolution of input depends on the number of cells covering the structure area. As shown in Fig. 1, the design space is split into 35 squaring cells, which can be quantified by 7×5 binary matrices. The microwave structures are randomly generated within the design space. Totally, there are $2^{35} = 3.4 \times 10^{10}$ possible structures. The outputs are the estimated complex scattering parameters (S-parameters) vectors. A CNN model is developed to predict any structure's S-parameter behavior quickly and precisely. The structure is connected with three fixed 50Ω connectors and labeled by the Keysight ADS software. This characterizing method can be easily tailored and extended to application scenarios requiring higher design resolutions and different number of ports.

B. Transfer Learning for Microwave Behavior Prediction

Transfer learning can initiate the learning process for a new target task by transferring knowledge from a previously well-learned source task. The more related the source task is to the target task, the better performance the transferred model can produce [17]. For most microwave behavior modeling tasks, existing learning-based approaches generally follow specific templates corresponding to different frequencies or sizes [4], [18]. These parameter-varying tasks share the identical nature and similar input-output patterns. This sets the foundation of applying transfer learning for microwave behavior prediction.

The dataset with U groups of the input features and output labels are denoted as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_U]$ and $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_U]$, respectively. The training process is defined to solve the following optimization problem:

$$\min_{\omega, z} \sum_{i=1}^U \mathcal{L}(f_{\omega, z}(\mathbf{X}_i), \mathbf{Y}_i), \quad (1)$$

where $\mathcal{L}(\cdot)$ is the loss function, and $f_{\omega, z}(\cdot)$ is the predicting model. The structures behavior prediction can be treated as a regression problem, where we choose the mean absolute error (MAE) as the loss function, which is explicitly given by

$$\mathcal{L}(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^N |x_i - y_i|}{N}. \quad (2)$$

The output \mathbf{a}_{i+1} from fully connected layer $i + 1$ is equal to

$$\mathbf{a}_{i+1} = \mathcal{A}(\omega_i^T \cdot \mathbf{a}_i + z_i), \quad i = 1, 2, \dots, m, m + 1, \dots, n \quad (3)$$

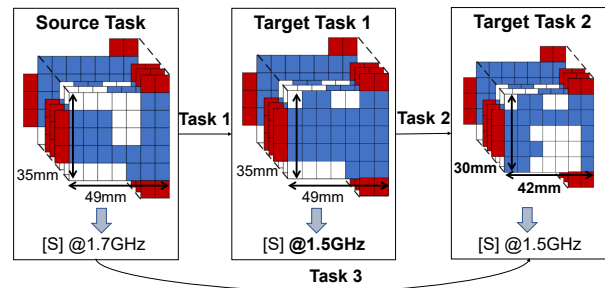


Fig. 2. Transfer learning process across different frequencies.

where $\mathcal{A}(\cdot)$ is the activation function; ω_i and z_i are the weight and bias vectors in layer i . In the convolutional layers, ω represents the kernel, and the dot product operation in (3) is replaced by the cross-correlation operation. To apply transfer learning, the first step is to develop a well-trained model for the source task, by which we can fix the parameters $[\omega_i, z_i], 1 \leq i \leq m$, are fixed. This model is then retrained with the target task dataset. The first m layers are called fixed layers, while the rest layers are called adaptation layers. In this way, the knowledge extracted from the source task is transferred to the model designed for the target task through the fixed layers and the initial network parameters $[\omega_i, z_i], m < i \leq n$, in the adaptation layers.

III. PERFORMANCE EVALUATION AND DISCUSSION

To comprehensively evaluate the performance of transfer learning for the behavior prediction of microwave structures, we designed and carried out three transfer tasks. As shown in Fig. 2, task 1 is designed to transfer knowledge between two adjacent frequencies; task 2 leverages the knowledge to train a microwave simulation model with a different structure size; in task 3, both frequencies and sizes are set to be different from the source task. To label these randomly generated structures, we simulate the S-parameter response of the 40,000 structures using EM simulation tools with the size of $49 \text{ mm} \times 35 \text{ mm}$, and 14,000 structures of $42 \text{ mm} \times 30 \text{ mm}$. Among them, 70% of the data was used for training, while the rest for testing. The computing platform comprises an Intel i9-9900X CPU @ 3.50 GHz and an Nvidia RTX2080 GPU with 12 GB memory.

A. Transfer Knowledge to Different Frequencies

As shown in Fig. 2, the source task refers to the prediction of the microwave structures' S-parameter response at the operation frequency of 1.7 GHz. The target task is to predict the response at the operation frequency of 1.5 GHz.

Fig. 3 shows the experimental results corresponding to different numbers of fixed layers. Transfer learning performs accurately and efficiently when the first two convolutional layers are fixed for the target training process. Thus, we choose to fix the first two layers in the following experiments. Fig. 4 compares the MAE rates and training time for cases with different amounts of training data. The results signify that the model entailing transferred knowledge is less impacted when the number of datasets is reduced. In contrast, the network

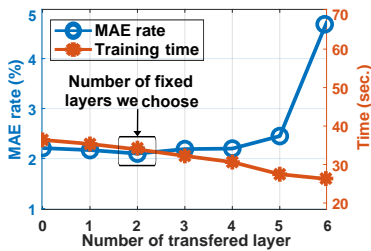


Fig. 3. MAE and training time with different numbers of fixed layers.

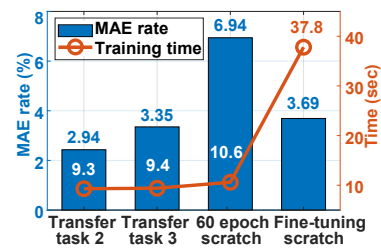


Fig. 5. MAE and training time with different template sizes.

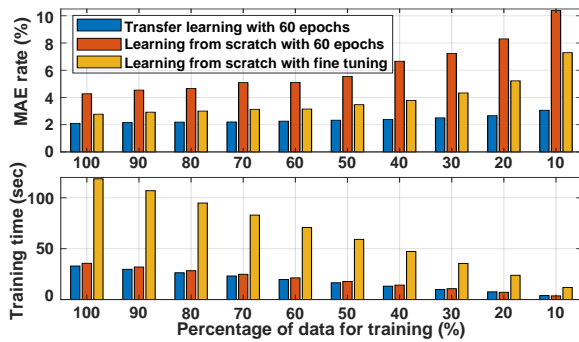


Fig. 4. MAE and training time with different amounts of training data.

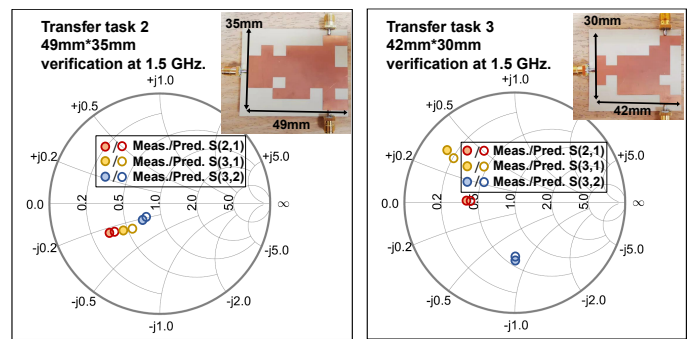


Fig. 6. Microwave structure prototype and measurement comparison.

without prior knowledge (learning from scratch) significantly suffers from data reduction. Specifically, using only 30% of the training data to train the model, the MAE testing result decreases by 20%, which is still better than the model trained by full training data. In comparison, the model without source knowledge drops by 63%. Even performing fine tuning for the learning from scratch model with a triple of training time, the MAE rate is still higher than the transfer learning model.

B. Transfer Knowledge to Different Template Sizes

The source knowledge extracted from one model with a specific size is also supposed to be constructive to train other models of different sizes. Following this rationale, the knowledge from task 1 can be leveraged for training networks and simulating microwave structures with different sizes in task 2. As shown in Fig. 2, we transfer the template from 49 mm × 35 mm to 42 mm × 30 mm. Task 3 is implemented to make comparisons between two different transfer tasks. Based on the results from task 1, 14,000 (30%) dataset are labeled by the S-parameter response of the 42 mm × 30 mm structures.

The results for the 42 mm × 30 mm structure are compared in Fig. 5. Tasks 2 and 3 achieve lower MAE rates than the network trained from scratch. Similar to the results in task 1, training a model with source knowledge significantly improves the performance of the target task. Compared with a fine-tuning model with much longer training time, the MAE rate is still 50% poorer than the model derived from task 2.

The comparison between the two transfer tasks shows that task 2 achieves an 18% lower MAE rate than task 3 within the same training time. This comparison substantiates that in the context of the behavior prediction of microwave structures, the more related the target task is to the source task, the more

advantageous performance will be yielded by the proposed transfer-learning-based approach.

In order to examine the feasibility of transfer learning in practice, two microstrip line structure prototypes were fabricated on an ISOLA substrate of 0.762 mm thickness with a dielectric constant of 2.8 to evaluate the prediction results from the transferring models. Fig. 6 shows the fabricated prototypes and the measurement results from the vector network analyzer. Our model achieves a relatively small MAE of 3.7% between the prediction yielded by the transfer learning model and fabrication validation, which proves the effectiveness and accuracy of our proposed transfer-learning-based approach. From experimental results, it is clear that source knowledge can be leveraged for facilitating model training for the target task, as the source task and target task share similarities. Consequently, only part of the neural network needs to be trained to develop high-performance models for similar tasks of microwave structure behavior prediction.

IV. CONCLUSION

In summary, we proposed a transfer-learning-based CNN model to predict microwave structures' behaviors. This model is generic and thus can be tailored or extended to multiple application scenarios. The proposed model significantly reduces the training time and the amount of training data compared to a learning from scratch model. Through extensive experimental results, we also validated that the performance yielded by transfer learning depends on the similarity between the source task and the target task. The research outcomes in this letter well answered several key questions pertaining to automatic RF behavior prediction and opened up a new possibility for reaching a compromise between accuracy and efficiency.

REFERENCES

- [1] Q.-J. Zhang, K. Gupta, and V. Devabhaktuni, "Artificial neural networks for RF and microwave design - from theory to practice," *IEEE Trans. Microw. Theory Techn.*, vol. 51, no. 4, pp. 1339–1350, 2003.
- [2] Z. Zhang, Q. S. Cheng, H. Chen, and F. Jiang, "An efficient hybrid sampling method for neural network-based microwave component modeling and optimization," *IEEE Microw. Wirel. Compon. Lett.*, vol. 30, no. 7, pp. 625–628, 2020.
- [3] J. Jin, C. Zhang, F. Feng, W. Na, J. Ma, and Q.-J. Zhang, "Deep neural network technique for high-dimensional microwave modeling and applications to parameter extraction of microwave filters," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 10, pp. 4140–4155, 2019.
- [4] 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.
- [5] 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 Microw. Mag.*, vol. 1, no. 1, pp. 494–507, 2021.
- [6] S. Koziel and J. W. Bandler, "Space mapping with multiple coarse models for optimization of microwave components," *IEEE Microw. Wirel. Compon. Lett.*, vol. 18, no. 1, pp. 1–3, 2008.
- [7] F. Feng, J. Zhang, W. Zhang, Z. Zhao, J. Jin, and Q.-J. Zhang, "Coarse-and fine-mesh space mapping for em optimization incorporating mesh deformation," *IEEE Microw. Wirel. Compon. Lett.*, vol. 29, no. 8, pp. 510–512, 2019.
- [8] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *Proc. Int. Jt. Conf. Neural Netw.* Springer, 2018, pp. 270–279.
- [9] P. Li, H. Cui, A. Khan, U. Raza, R. Piechocki, A. Doufexi, and T. Farnham, "Deep transfer learning for WiFi localization," in *Proc. IEEE Radar Conf.*, 2021, pp. 1–5.
- [10] 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.
- [11] S. Zhang, X. Hu, Z. Liu, L. Sun, K. Han, W. Wang, and F. M. Ghannouchi, "Deep neural network behavioral modeling based on transfer learning for broadband wireless power amplifier," *IEEE Microw. Wirel. Compon. Lett.*, 2021.
- [12] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *Proc. Int. Conf. Eng. Technol.*, 2017, pp. 1–6.
- [13] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [14] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, "Dive into deep learning," *arXiv preprint arXiv:2106.11342*, 2021.
- [15] H. M. Yao, M. Li, and L. Jiang, "Applying deep learning approach to the far-field subwavelength imaging based on near-field resonant metalens at microwave frequencies," *IEEE Access*, vol. 7, pp. 63 801–63 808, 2019.
- [16] G. Urban, K. J. Geras, S. E. Kahou, O. Aslan, S. Wang, R. Caruana, A. Mohamed, M. Philipose, and M. Richardson, "Do deep convolutional nets really need to be deep and convolutional?" *arXiv preprint arXiv:1603.05691*, 2016.
- [17] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Noguees, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [18] G. Zhang, H. He, and D. Katabi, "Circuit-GNN: Graph neural networks for distributed circuit design," in *Proc. Int. Conf. Mach. Learn.* PMLR, 2019, pp. 7364–7373.