

RESEARCH ARTICLE

Attention-Based Deep Neural Network for High-Dimensional Microwave Modeling of Non-Reciprocal Bandpass Filters

GIRDHARI CHAUDHARY¹, (Member, IEEE), YAO MENG²,
DONG SUN PARK², AND YONGCHAE JEONG^{1,3}, (Senior Member, IEEE)

¹Jeonbuk RICE Intelligence Innovation Research Center, Jeonbuk National University, Jeonju-si, Jeollabuk-do 54896, Republic of Korea

²Division of Electronic Engineering, Core Research Institute of Intelligent Robots, Jeonbuk National University, Jeonju-si, Jeollabuk-do 54896, Republic of Korea

³Division of Electronic Engineering, IT Convergence Research Center, Jeonbuk National University, Jeonju-si, Jeollabuk-do 54896, Republic of Korea

Corresponding author: Yongchae Jeong (ycjeong@jbnu.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by Korean Government [Ministry of Science and ICT (MSIT)] under Grant (RS-2023-00209081, 40%); in part by the Basic Science Research Program through NRF of Korea funded by the Ministry of Education under Grant (RS-2019-NR040079, 40%); in part by the Institute of Information & Communications Technology Planning & Evaluation (IITP)-Innovative Human Resource Development for Local Intellectualization Program grant funded by Korean Government (MSIT) under Grant (IITP-2024-RS-2024-00439292, 10%); and in part by the Research Base Construction Fund Support Program funded by Jeonbuk National University (10%), in 2024.

ABSTRACT This paper presents attention-based deep neural networks for high-dimensional microwave modeling to predict behavior of spatio-temporal modulated (STM) non-reciprocal bandpass filters (NR-BPFs). The proposed method integrates convolutional neural networks (CNN), joint channel-spatial attention, and residual connections to enhance prediction accuracy and model generalization. The combined channel-spatial attention approach allows the model to concentrate on essential features, while residual connections improve learning by reducing overfitting and preventing vanishing gradients, ensuring efficient gradient flow. For experimental validation of the proposed model, high-dimensional microwave modeling of four types of NR-BPF (example 1: third-order single-band NR-BPF, example 2: fourth-order single-band NR-BPF, example 3: third-order dual-band NR-BPF, example 4: microstrip line third-order single-band NR-BPF) are performed. Experimental results show that the proposed model exhibits superior performances for four types of NR-BPFs, achieving lower root mean square error (RMSE) and higher R-squared as compared to conventional deep multilayer perceptron (MLP) and standard CNN models. The results indicate that the proposed approach provides a robust solution for accurately predicting complex input-output relationships in STM-based NR-BPFs, outperforming conventional deep MLP and CNN models in both prediction accuracy and model robustness.

INDEX TERMS Attention-based deep learning, harmonic balance, high-dimensional, microwave modeling, non-reciprocal filters.

I. INTRODUCTION

Non-reciprocal circuits are one of the crucial components of RF wireless communication systems and radars [1], [2]. Non-reciprocal circuits break the principle of reciprocity and allow signal transmission in forward direction and block in opposite direction. To break reciprocity theory

in the microwave/millimeter domain, one of the common methods is to use ferrite-based magnetic material [3], [4]. However, ferrite-based non-reciprocal circuits are bulky, expensive, and not compatible with integrated circuit (IC) design. To alleviate the use of ferrite-based resonators, spatio-temporal modulation (STM) is an alternative efficient method to implement microwave/millimeter wave non-reciprocal circuits [5], [6]. In the STM approach, non-reciprocal circuits such as non-reciprocal bandpass filters (NR-BPFs) and

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

circulators are designed by modulating capacitor values of non-linear resonators with progressive phase shift modulation signal [7], [8], [9], [10], [11], [12], [13].

Predicting the performances of STM-based non-reciprocal circuits and capturing the process behind non-reciprocity are vital because they provide insight into ways to minimize forward insertion loss and enhance reverse isolation. However, the performance of STM-based non-reciprocal circuits cannot be captured by traditional scattering parameters (*S*-parameters) analysis. A classical approach for the predicting behavior of STM-based non-reciprocal circuits is possible by performing a time-domain transient analysis followed by a Fourier analysis. However, this process requires a long simulation time with a very short time step. Another classical approach for predicting the behavior of STM-based non-reciprocal circuits is the harmonic balance (HB) method [14], where the circuit is split into a linear subcircuit and a nonlinear subcircuit. In the HB method, the elements in the linear section are expressed with their frequency domain admittance or impedance, and the analysis is done in the Fourier domain [15]. The nonlinear section is solved in the time domain by algebraic methods or by solving time-domain differential equations [15]. Therefore, the HB method may suffer from several challenges such as convergence issues and computational complexity. In addition, the HB method uses an iterative process to solve non-linear equations in Fourier and time-domain analysis, therefore, the solution of nonlinear equations may not always converge, and as a result, the performance of non-reciprocal circuits cannot be accurately predicted.

In recent years, artificial neural networks (ANN) are emerged as powerful tools to solve complex problems and also used in forward/inverse modeling of microwave circuits [16], [17]. Since there is no analytical formula or function to solve forward/inverse modeling problems of microwave circuits, ANN has been explored in learning the input and output relationship of microwave circuits such as predicting the frequency response of reciprocal microwave circuits [18]. For predicting the performance of STM-based NR-BPF, typically ANN models require circuit parameters of non-reciprocal circuits as inputs and scattering parameters (*S*-parameters) at different frequency samples as output, resulting in model input and outputs as high-dimensional nonlinear relationships. Therefore, it is necessary to develop efficient high-dimensional microwave modeling to satisfy the requirements of STM-based microwave NR-BPF. As a result, the predicting performance of the NR-BPF problem becomes more challenging and it may be beyond the capability of conventional shallow ANN (such as networks containing one hidden layer) [19].

To solve high-dimensional microwave modeling problems of microwave circuits, deep neural networks such as feedforward deep neural networks [20], [21], regularized deep belief networks (R-DBN) [22], and convolutional neural network (CNN) [23] have been presented for reciprocal microwave

filters. One of the widely used deep neural networks for microwave modeling problems is the deep multilayer perceptron (MLP) [20], which is a fully connected neural network consisting of an input layer, several hidden layers, and an output layer. Selecting an activation function in deep MLP can influence the training error and can be easily overfitted. In addition, the activation functions such as sigmoid and tanh functions pose challenges in training deep MLP due to the vanishing gradient problem. When neural networks become deeper and deeper (i.e. large number of hidden layers), the gradients become increasingly small as they propagate back through many layers, primarily due to repeated multiplication by small values from weights and activation functions derivatives. This results in insufficient updates to the weights of the early layers and causes slow learning or even stalled learning in deep neural networks. In multilayer feedforward networks [21], the inputs to each layer are influenced by the outputs of the preceding layers, therefore, the training of these neural networks is complex and time-consuming. Due to the dependency of input for one layer to preceding layers, even small changes in the lower layers can make training deeper networks more difficult.

CNN is one of the other types of deep neural networks that have been used for learning high-dimensional microwave modeling problems, in which inputs and/or outputs can be processed in an image-like representation [22], [23], [24]. CNNs are effective for handling high-dimensional modeling problems of microwave circuits as CNN exploits local connections and shared weights, which significantly results in reducing the number of parameters. However, conventional CNN is less robust and sensitive to changes in architecture and weight initialization. Therefore, it is harder to consistently train models and requires significant hyperparameter tuning for optimal performance. In addition, the numerous training data intensifies this problem. There are other types of deep neural networks such as deep recurrent neural networks (RNNs) [25], [26], [27], [28], and long short-term memory (LSTM) [29] which have been applied for modeling nonlinear electronic circuits. In [30], researchers applied LSTM based on recurrent neural network (RNN) to predict short-term attenuation in microwave link. LSTM networks are particularly valued for their ability to capture temporal dependencies in sequential data through gated mechanisms that regulate information flow. However, implementation of LSTM networks faces significant challenges arises from multi-high-order matrix vector multiplications [31], [32]. It should be noted that excessive retraining or fine-tuning of deep neural networks increase computational costs and often fails to maintain stable performance across various conditions. Therefore, developing a robust ANN model that can accurately address complex, high-dimensional input-output relationships in microwave circuits, while enhancing model generalization, is highly beneficial for advancing the efficiency and precision of circuit design and analysis.

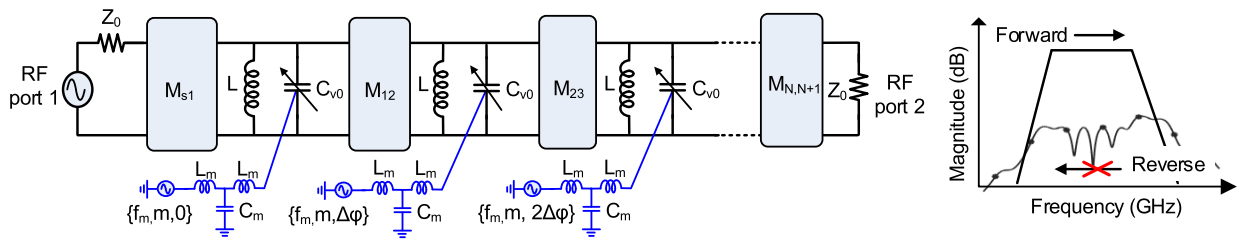


FIGURE 1. Circuit diagram of spatio-temporal modulated (STM) based non-reciprocal bandpass filter (NR-BPF) and conceptual frequency response.

In STM-based microwave non-reciprocal circuits, specific regions of frequency response are crucial as they contain more critical features needed to accurately predict circuit behavior. In addition, these regions can significantly influence non-reciprocity behavior of the circuit that can become challenge for deep learning microwave modeling. To address complexities for precise and robust deep learning modeling, accurately capturing these features are essential.

To overcome this challenge and the shortcomings of previously presented deep neural networks, this paper proposed an attention-based deep neural network by combining CNN, joint channel-spatial attention, and residual connection. The proposed attention-based deep neural network is capable of handling complex input-output relationships of high-dimensional microwave modeling problems of STM-based non-reciprocal microwave circuits with enhanced prediction accuracy and better generalization across different non-reciprocal microwave circuits. Attention-based mechanisms can dynamically adjust the focus to extract the most important features in different datasets. To improve the robustness of the proposed model, we used residual connection which allows information to bypass poorly optimized layers and enhance the model's ability to learn consistently and efficiently.

II. PROPOSED MICROWAVE MODELING OF NR-BPF

A. PRINCIPLE AND CHALLENGE

Fig. 1 shows the circuit diagram of STM-based NR-BPF, where time-varying LC resonators with self-resonant coupling (M_{ii}) are coupled through coupling element values (M_{ij}). NR-BPF integrates bandpass filter characteristics and isolator functionality in one circuit, as a result, it provides the signal transmission in forward direction while blocking in reverse direction. To achieve a non-reciprocal filter response, constitute capacitors of resonators are modulated with progressive phase shift sinusoidal modulation signal as shown in (1).

$$C_{v,i}(t) = C_{v0} [1 + m \cos \{2\pi f_m t (i - 1) \Delta\phi\}] \quad (1)$$

where f_m is modulation frequency, m is modulation index and $\Delta\phi$ is progressive phase shift [7]. Similarly, C_{v0} represents nominal capacitance when a modulation signal is not applied. For achieving bandpass filtering response,

the time-varying resonators are coupled through coupling element values (such as M_{s1} , M_{12} , ...) as depicted in Fig. 1.

Accurately predicting the frequency response of NR-BPFs is crucial for understanding non-reciprocity behavior, as well as for rapid simulation and fabrication. Traditionally, the frequency response of NR-BPF is simulated using the HB method, where the circuit is split into a linear subcircuit and a nonlinear subcircuit, and circuit behavior is extracted by using time and Fourier domain analysis. However, the solution to the nonlinear equations in the HB method may not always converge, leading to inaccurate prediction of NR-BPF behavior.

In recent years, ANN has emerged as one of the power tools to solve complex input and output relationships of microwave circuits, making them suitable for predicting NR-BPF behavior. In high-dimensional microwave modeling, circuit parameters of the NR-BPFs are used as input, while scattering parameters (S -parameters) at various frequency points can be modeled as output. Fig. 2 depicts the real and imaginary parts of S -parameters (input reflection coefficient: S_{11} , forward transmission coefficient: S_{21} , reverse transmission coefficient: S_{12} , output reflection coefficient: S_{22}) across 350 frequency points ranging from 1.50 GHz to 2.10 GHz. These real and imaginary S -parameter values can be categorized into three frequency regions such as Region 1, Region 2, and Region 3. As shown in the figure, Region 2 contains the most critical features for accurately predicting the behavior of STM-based NR-BPFs. Specifically, the real and imaginary values of reverse transmission coefficient (*i.e.* $\text{re}(S_{12})$ and $\text{im}(S_{12})$) are very small and change abruptly within Region 2 when small changed in the input circuit parameters occur. Even slight variation in the value of $\text{re}(S_{12})$ and $\text{im}(S_{12})$ in region 2 significantly impacts on non-reciprocity performance of NR-BPFs. Therefore, it remains a challenge for high-dimensional microwave modeling of NR-BPFs to accurately predict non-reciprocity behavior. To overcome this challenge, we propose an attention-based deep neural network for high-dimensional microwave modeling of NR-BPFs. This approach can enable the proposed model to focus on the most critical features such as abrupt changes in S -parameters and provide a robust solution for accurate prediction of non-reciprocity behavior of NR-BPFs.

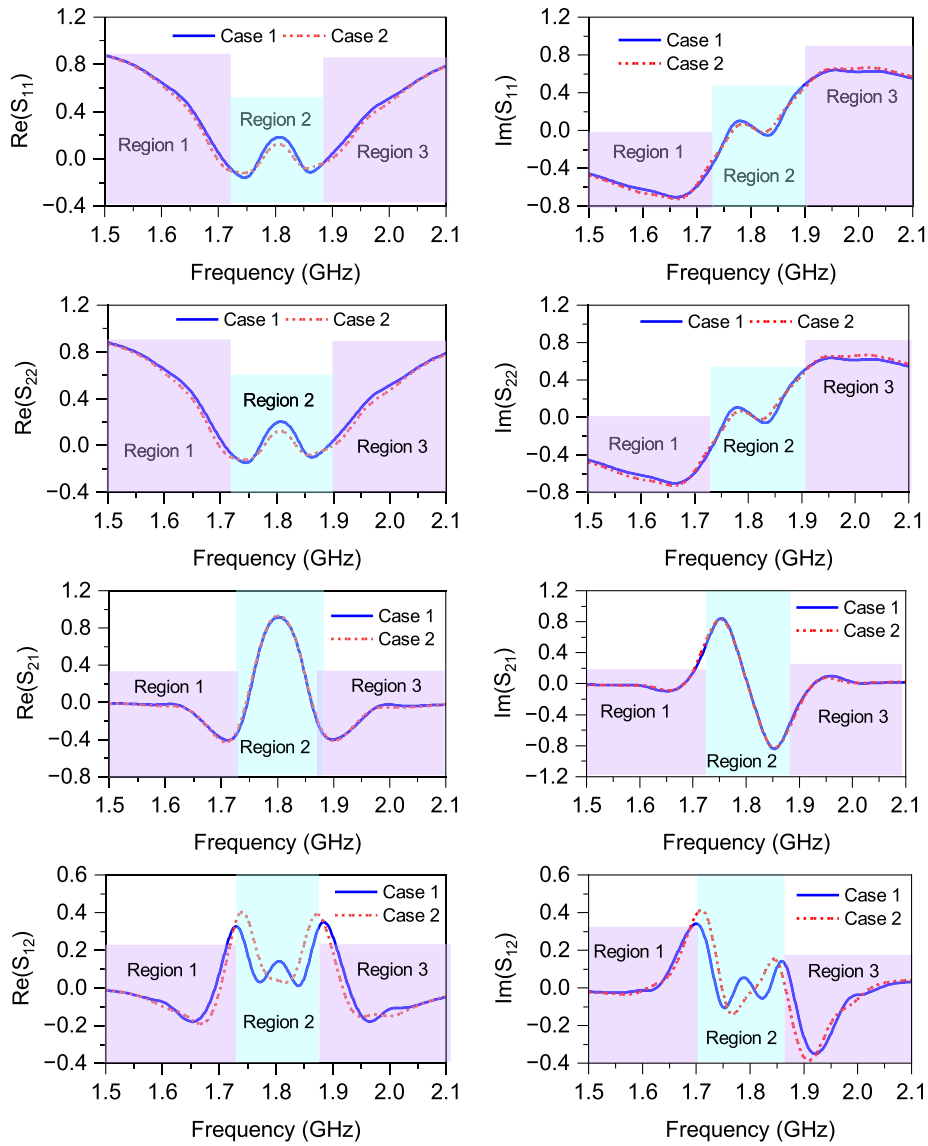


FIGURE 2. Real and imaginary parts of scattering parameters (S -parameters) of STM-based third-order single-band NR-BPF with different circuit parameter values such as Case 1: $M_{51} = 1.2214$, $M_{12} = 1.2045$, $M_{23} = 1.2197$, $M_{34} = 1.2128$, $f_m = 125$ MHz, $m = 0.1184$, $\Delta\varphi = 60^\circ$, $BW = 100$ MHz and Case 2: $M_{51} = 1.2022$, $M_{12} = 1.1822$, $M_{23} = 1.1945$, $M_{34} = 1.2128$, $f_m = 110$ MHz, $m = 0.1090$, $\Delta\varphi = 60^\circ$, $BW = 100$ MHz.

B. PROPOSED ATTENTION-BASED DEEP LEARNING MODEL

Fig. 3 shows the proposed attention-based deep learning model for high-dimensional microwave modeling of STM-based NR-BPFs. In the proposed model, circuit parameters serve as input, while real and imaginary parts of S -parameters across various frequency points are outputs. The proposed method combines CNN, the residual network [33], and the channel-spatial attention model [34], as illustrated in Fig. 3. The input signals pass through three convolutional blocks, each comprising of 1D batch normalization, 1D convolutional, rectified linear unit (ReLU) activation, and 1D max pooling.

In three convolutional blocks, the first layer utilizes a 7×7 filter to effectively capture longer dependencies and complex patterns, while a 3×3 filter is applied in the last two convolutional layers. The extracted features then flow sequentially through the channel-spatial attention block and residual blocks. Finally, the features are processed by a fully connected layer to generate the real and imaginary parts of S -parameters.

i) Joint channel-spatial attention and 1D residual model

Although CNNs are primarily used for image recognition and computer vision tasks, their capability to automatically learn complex patterns without the need for explicit feature engineering also makes them well-suited for processing

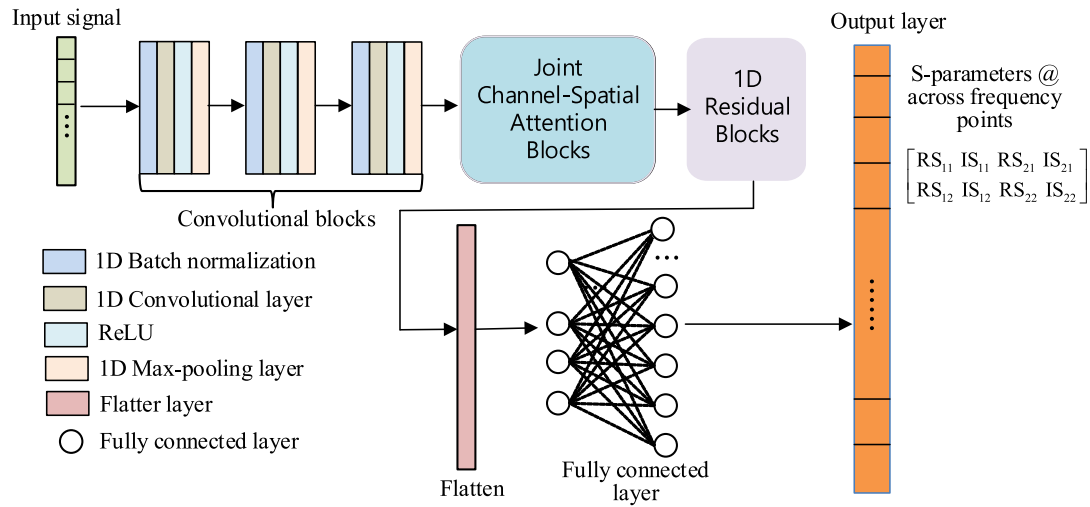


FIGURE 3. Architecture of proposed attention-based deep neural network for high-dimensional microwave modeling of NR-BPFs.

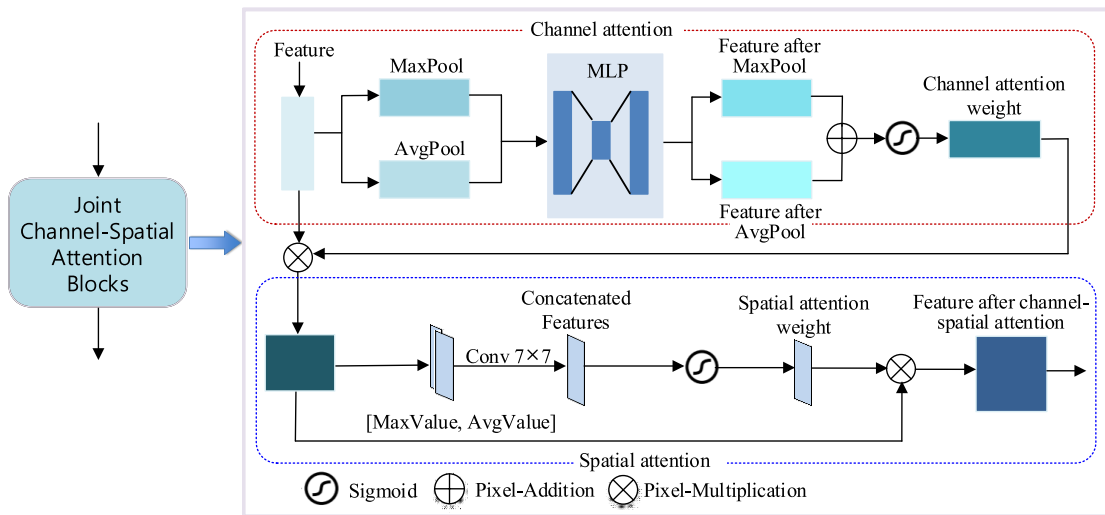


FIGURE 4. Detailed architecture of joint channel-spatial attention model to capture significant features in specific frequency region with substantial variations.

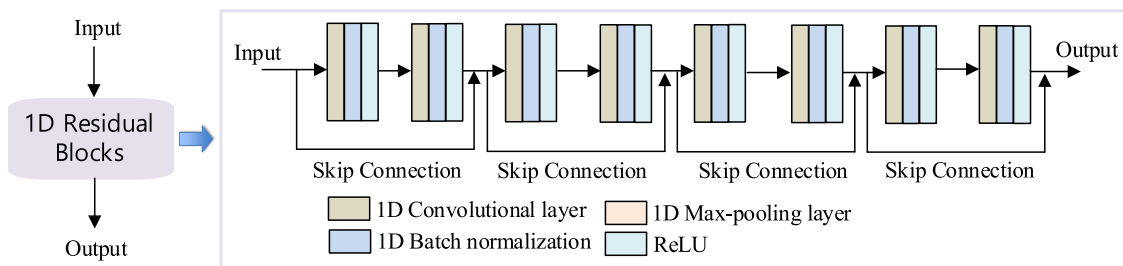


FIGURE 5. Detailed architecture of 1D residual block with skip connections to ensure effective gradient information flow.

one-dimensional (1D) sequence data. However, CNNs alone are not enough to handle the high-dimensional modeling challenge of NR-BPFs, therefore, we integrate the channel-spatial attention mechanism along with CNN to

capture the most relevant information at a specific frequency region with substantial variations. Specifically, channel attention helps the model in determining *what* information to focus on, while spatial attention guides the model in learning *where*

in the position that information is most relevant. In CNN, different channels extract various features from the input signal, where some of these features may be more important for specific tasks. Channel attention mechanisms dynamically assign weights to channel that provide valuable information for a task while suppressing less relevant channels. This selective focus improves the feature representation, enabling the model to adaptively highlight the most crucial aspects of the input signal. To implement channel attention, we utilize a 1D adaptive average pooling and 1D adaptive max pooling operations to aggregate spatial information of a feature map, as shown in Fig. 4. Then, the average pool feature and max-pooled features are fed into a shared MLP network to generate a final channel attention weight. This MLP network consists of two 1D convolutional layers: the first layer with 16 filters of size 1×1 and ReLU activation function, followed by the second layer with 256 filters of size 1×1 . After MLP, the two types of features are combined pixel-wise and passed through the sigmoid activation function.

Specifically, given a feature $F \in \mathbb{R}^{L \times C}$, we first apply 1D adaptive average pooling and a 1D adaptive max pooling along the signal length dimension to reduce each channel's time series data to a single descriptor and results are given as (2).

$$f_{avg}^c = \frac{1}{L} \sum_{l=1}^L F(l, c) \quad (2a)$$

$$f_{max}^c = \max_{1 \leq l \leq L} F(l, c) \quad (2b)$$

where $f_{avg}^c \in \mathbb{R}^C$ and $f_{max}^c \in \mathbb{R}^C$ are the average and maximum pooled descriptors for each channel across the entire signal. The pooled descriptors are then passed through a shared MLP consisting of two fully connected layers to generate the channel attention. The output of the channel attention map $M_{channel} \in \mathbb{R}^C$ with an attention weight for each channel can be given as (3).

$$M_{channel}(F) = \sigma \left[W_2 \{ \text{ReLU}(W_1 f_{avg}) \} + W_2 \{ \text{ReLU}(W_1 f_{max}) \} \right] \quad (3)$$

where $W_1 \in \mathbb{R}^{C \times C/r}$ and $W_2 \in \mathbb{R}^{C/r \times C}$ are weight metrics of the MLP with a reduction ratio r . Similarly, σ is the sigmoid function to constrain the values between 0 and 1. Then the channel attention map is applied to the input signal by element-wise multiplication, broadcasting the attention weights across the temporal dimension as given in (4).

$$F_{channel_att} = F \odot M_{channel} \quad (4)$$

where $M_{channel} \in \mathbb{R}^{1 \times C}$ is extending along the length of the signal L . Spatial attention emphasizes most relevant features in the interesting frequency region to the prediction task, enabling the model to focus on key spatial areas. This helps the model identify *where* important information is located. As illustrated in Fig. 4, average-pooling and max-pooling are first applied along the channel axis, and their outputs are

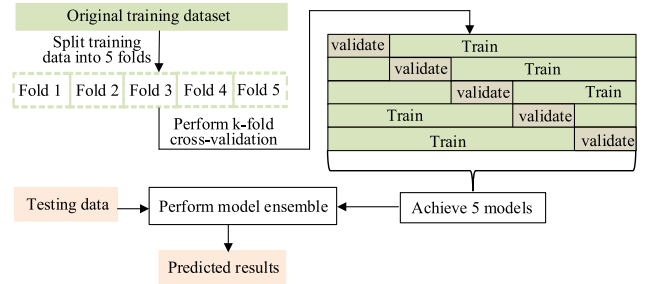


FIGURE 6. Cross-validation procedure for generalization of model.

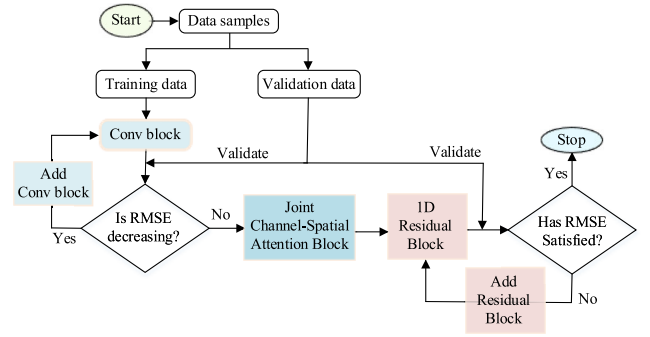


FIGURE 7. Flowchart of training and validation procedures of proposed attention-based deep neural network.

concatenated to create a more informative feature descriptor. A larger 7×7 filter is then used to capture broader patterns in the input, allowing the model to learn more generalized features, which enhances generalization performance and reduces sensitivity to isolated noise. Similar to channel attention, a Sigmoid activation function is utilized to produce spatial attention weight. The spatial attention is computed by applying both average pooling and max pooling across the channel dimension C , creating two temporal descriptors as given by (5).

$$F_{avg}^S = \frac{1}{C} \sum_{c=1}^C F_{channel_att}(l, c) \quad (5a)$$

$$F_{max}^S = \max_{1 \leq c \leq C} F_{channel_att}(l, c), \quad (5b)$$

where $F_{avg}^S \in \mathbb{R}^L$ and $F_{max}^S \in \mathbb{R}^L$ represent the average and maximum values across all channels for each signal. Next, the average and maximum values are concatenated and passed through a 1D convolution, and the resulting feature can be written as (6).

$$F_{concat}^S = \left[F_{avg}^S; F_{max}^S \right] \in \mathbb{R}^{L \times 2} \quad (6)$$

This concatenated feature is passed through a 1D convolution with a kernel size of 7 to generate the spatial attention map as follows:

$$M_{spatial}(F) = \sigma \left\{ \text{Conv}_{1D} \left(F_{concat}^S \right) \right\} \quad (7)$$

The resulting spatial attention map provides an attention weight for each element in the signal. Furthermore, the temporal attention map is applied to the channel-attended feature map by element-wise multiplication, broadcasting the attention weights across the channel dimension:

$$F_{spatial_att} = F_{channel_att} \odot M_{spatial} \quad (8)$$

where $M_{spatial} \in \mathbb{R}^{L \times 1}$ is extending along the channel dimension.

Studies have shown that deep layers can learn more relation information between input and output [33], [34]. To deepen the model and improve the performance, we employ residual blocks in this work as shown in Fig. 5. As shown in Fig. 5, the four 1D residual blocks are utilized where each block contains two convolutional blocks. Each sub-block consists of a convolution layer, batch normalization, and a ReLU activation function. The key point in residual blocks is using skip connections (shortcuts) to ensure that gradient information flows more effectively through the network during training, leading to better updates for earlier layers. Furthermore, skip connections provide flexibility to the network for choosing whether to apply transformations or pass the input directly to the next layer. This means that if certain transformations are unnecessary, the network can essentially skip them, making it more efficient. These characters are especially important in deep architecture.

ii) Cross-validation and model ensemble

To improve the generalization and prediction accuracy of the model, we incorporate the k -fold cross-validation technique with the proposed attention-based deep neural network as shown in Fig. 6. Cross-validation is a fundamental technique for assessing how a model will generalize to an independent dataset. It helps in validating the performance of model and selecting the best model. Cross-validation is also crucial for preventing overfitting and ensuring that the model performs well on unseen data. In this work, the k -fold cross-validation is introduced, which is a simple, but powerful technique. The idea is to divide the training dataset into k -equal folds. The model would be trained k times, each time using $k-1$ folds for training and the remaining fold for validation. By rotating through each fold as the validation set, every data point would get a chance to be in the validation set once, ensuring a fair and comprehensive evaluation. After training, k models will be achieved. In the testing process, model ensemble technique is employed to predict the high dimensional S -parameters. The model ensemble combines multiple models to improve overall performance. It aims to reduce variance, and bias, and improve prediction accuracy by leveraging the strengths of each model. Specifically, the predicted results from each model are averaged to provide a final overall predicted result.

iii) Evaluation matrices

In this work, two types of evaluation matrices are utilized to assess the effectiveness of the proposed model as described follows:

1) ROOT MEAN SQUARED ERROR (RMSE)

The RMSE is a common metric used to evaluate the accuracy of a regression model. It measures the average magnitude of the errors between predicted and actual values, providing a sense of how well the model predictions agree with the true values. The expression for RMSE is given as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}, \quad (9)$$

where n is the number of observations in the dataset, y_i is the ideal value for the i^{th} observation and $\hat{y}_i$ is the predicted value. It should be noted that lower RMSE values indicate that the predictions of model are closer to the actual values, suggesting better accuracy.

2) R-SQUARED (R^2)

R-squared is a statistical metric that measures the proportion of the variance in the dependent variable (target) that is explained by the independent variables (features) in a regression model. It provides insights into how well the model captures the variability of the target variable. The expression for R-squared is given by (10).

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\bar{y} - y_i)^2}, \quad (10)$$

where $\bar{y}$ is the mean of the observed values. R^2 ranges from 0 to 1. The higher value R^2 such as 1 represents that the model perfectly predicts the dependent variable based on the independent variables.

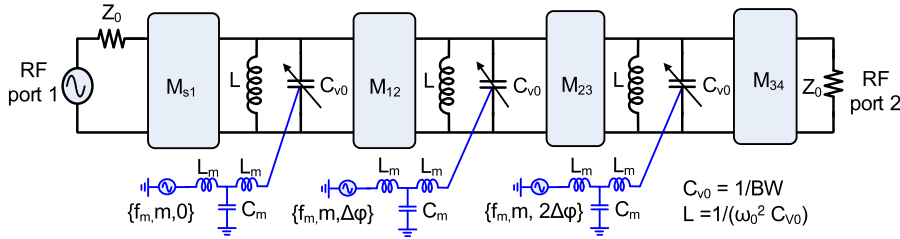
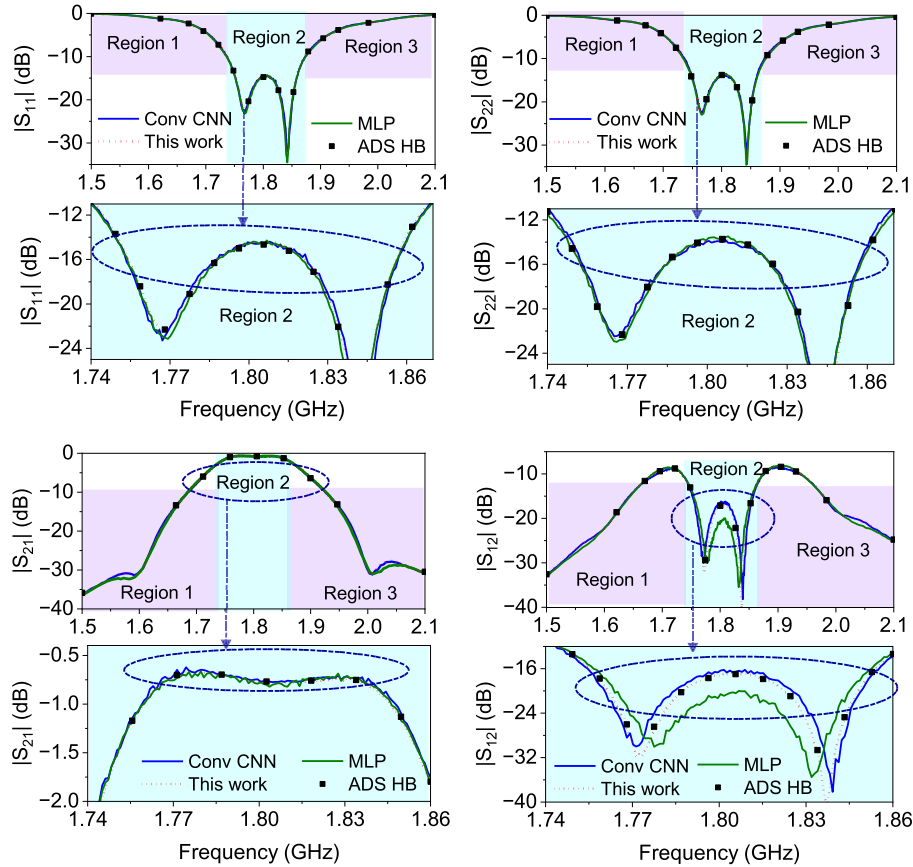
III. RESULTS AND DISCUSSION

For experimental validation, the proposed attention-based deep neural network is applied to the high-dimensional microwave modeling of four types of NR-BPF (*i.e.* Example 1: third-order single-band NR-BPF, Example 2: fourth-order single-band NR-BPF, Example 3: third-order dual-band NR-BPF, Example 4: microstrip line third-order single band NR-BPF). The goal of microwave modeling is to predict S -parameters of NR-BPFs. In this work, we utilize the Keysight Advanced Design System (ADS) simulator to generate training data sets. These data sets are created by using 32,500 uniformly distributed random samples, spanning a frequency range of 1.5 GHz to 2.1 GHz with 350 points. The total number of outputs is 2,800.

Because of the significant disparity in the input data, it is essential to ensure effective learning and convergence during the training of the deep learning model. Therefore, it is crucial to appropriately normalize or scale the data. In this work, a standardization normalization technique is employed to ensure uniform scaling for both the input and output data. To enhance the effectiveness of learning and convergence during the training of the deep learning model, we normalize input circuit parameters and corresponding output S -parameters using the standardization normalization technique, often referred to as z -score normalization where

TABLE 1. Comparison of conventional deep learning models and proposed attention-based deep learning model for the third-order single-band NR-BPF as example 1. (Δ : skip values compared to conventional deep learning models).

Modeling Example	Training data	Modeling methods	RMSE @ different frequency region separately			Overall frequency region	
			Region 1	Region 2	Region 3	RMSE	R ²
Modeling example 1: Third-order single-band NR-BPF	32,000	Deep MLP	0.42 %	1.18 %	0.63 %	8.87 %	99.91 %
		Conventional CNN	0.47 %	0.70 %	0.49 %	4.29 %	99.97 %
		This work	0.24 %	0.41 %	0.28 %	2.38 %	99.99 %
		Δ	-0.18 %	-0.77 %	-0.35 %	-6.49 %	+0.08 %
			-0.23 %	-0.29 %	-0.21 %	-1.91 %	+0.02 %

**FIGURE 8.** Circuit diagram of third-order single-band NR-BPF for high-dimensional microwave modeling as example 1.**FIGURE 9.** Comparison of results from various deep learning models and Keysight advanced system design harmonic balance (ADS HB) method for third-order single-band NR-BPF as example 1.

the standardized feature X is given as (11).

$$Z = \frac{X - \mu}{\delta}, \quad (11)$$

where Z is the standardized value of the feature X , μ is the mean of the feature, and δ is the standard deviation of the feature. This normalization technique ensures that features of

TABLE 2. Comparison of conventional deep learning models and proposed attention-based deep learning model for the fourth-order single-band NR-BPF as example 2. (Δ : skip values compared to conventional deep learning models).

Modeling Example	Training data	Modeling Methods	RMSE @ different frequency region separately			@ Overall frequency region	
			Region 1	Region 2	Region 3	RMSE	R ²
Example 2: Fourth-order single-band NR-BPF	32,000	Deep MLP	0.27 %	1.41 %	0.37 %	13.89 %	99.88 %
		Conventional CNN	0.40 %	1.19 %	0.38 %	8.06 %	99.92 %
		This work	0.08 %	0.26 %	0.11 %	2.07 %	99.99%
		Δ	-0.19 % -0.32 %	-1.15 % -0.93 %	-0.26 % -0.27 %	-11.82 % -5.99 %	+0.11 % +0.07 %

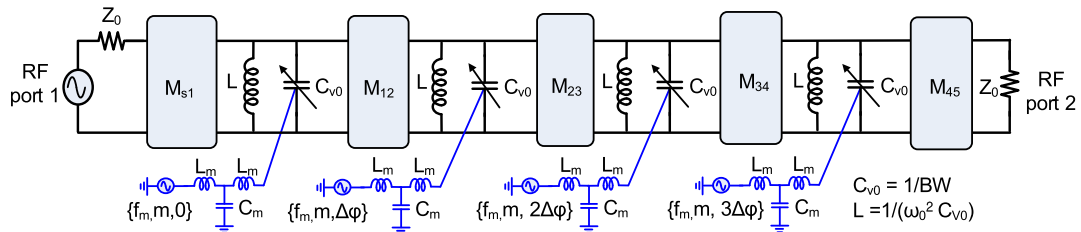
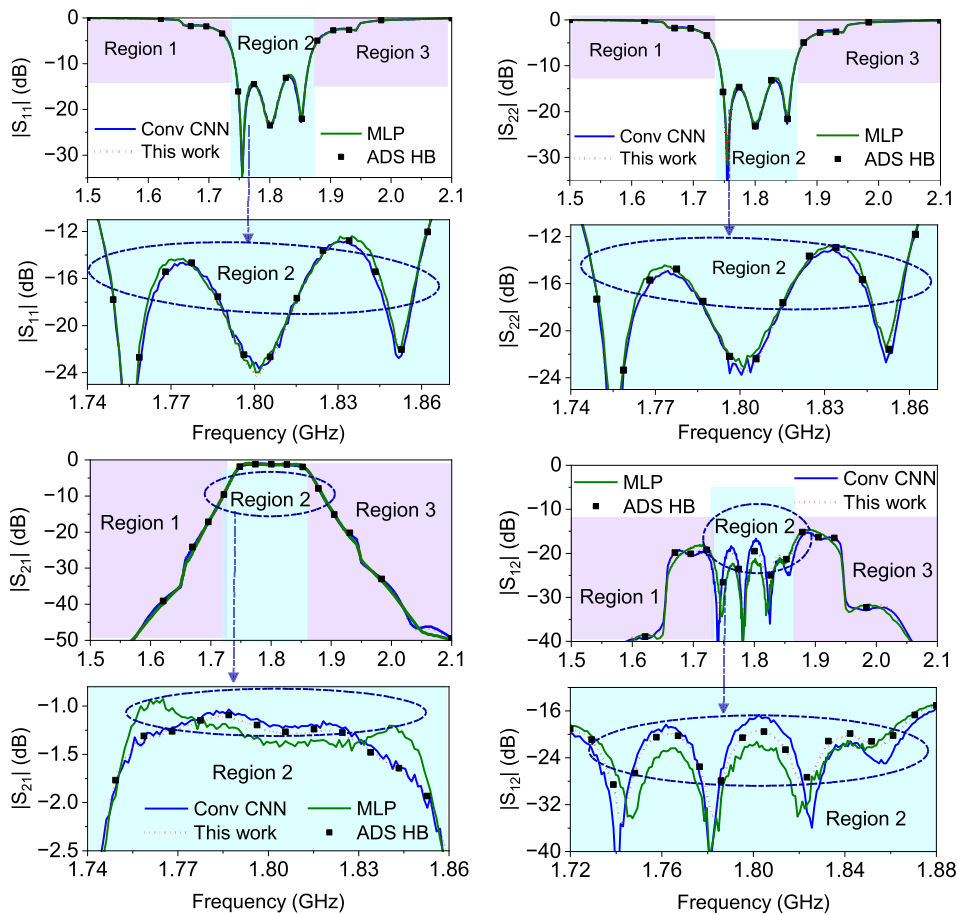
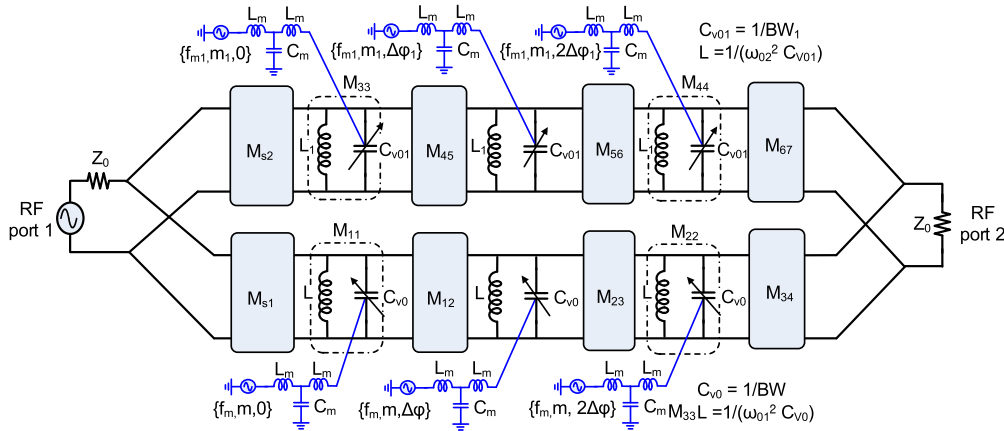
**FIGURE 10.** Circuit diagram of fourth-order single-band NR-BPF for high-dimensional microwave modeling as example 2.**FIGURE 11.** Comparison of results from various deep learning models and Keysight advanced system design harmonic balance (ADS HB) method for fourth-order single-band NR-BPF as example 2.

TABLE 3. Comparison of conventional deep learning models and proposed attention-based deep learning model for third-order dual-band NR-BPF as example 3. (Δ : Skip values compared to conventional deep learning models.)

Modeling Example	Training data	Modeling methods	RMSE @ different frequency regions separately					@ Overall frequency region	
			Region 1	Region 2	Region 3	Region 4	Region 5	RMSE	R ²
Example 3: Third-order dual-band NR-BPF	32,000	Deep MLP	0.51 %	1.50 %	0.98 %	1.70 %	0.46 %	12.59 %	0.9987
		Conventional CNN	0.40 %	0.75 %	0.63 %	0.48 %	0.34 %	6.24 %	0.9997
		This work	0.34 %	0.26 %	0.22 %	0.32 %	0.33 %	3.41 %	0.9999
		Δ	-0.17 %	-1.24 %	-0.76 %	-1.38 %	-0.13 %	-9.18 %	+0.12 %
			-0.06 %	-0.49 %	-0.41 %	-0.16 %	-0.01 %	-2.83 %	+0.02 %

**FIGURE 12.** Circuit diagram of third-order dual-band NR-BPF for high-dimensional microwave modeling as example 3.

the datasets have a mean of zero and a standard deviation of one, ensuring consistent scaling for both training and validation data sets. This preserves the original distribution's shape and ensures that the data retains its statistical properties. This technique enhances model performance by preventing features with larger scales from dominating the learning process.

Fig. 7 shows the flow chart of training and validation procedures of the proposed model. The adaptive moment estimation (ADAM) optimization algorithm is employed to update the network weights. During training of network, the mean squared error (MSE) is utilized as the loss function. To make the proposed model adaptable to various types of NR-BPF structures, we used the same hyperparameters for all four types of NR-BPF structures (Examples: 1, 2, 3, and 4). Since learning rate hyperparameter is crucial, therefore, adaptive optimizers such as ADAM is employed in this work so that learning rate can adjust dynamically. If the loss oscillates or diverges, the learning rate should be decreased. Conversely, if loss is decreasing very slowly, then learning rate should be increased.

The initial learning rate is set to 0.001. During the training process, if there is no decrease in the validation loss value after 5 epochs, the learning rate should be reduced by a factor of 0.1. The batch size is set to 2000, and the number of workers is set to 8 with one RTX 3090 GPU (24 GB memory). After 100 epochs, the proposed attention-based deep learning model demonstrated effective convergence, indicating the

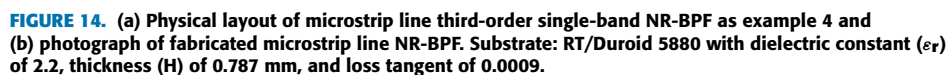
proposed model successfully learns from training data. Once the model has been successfully trained, it can be used to predict S -parameters of NR-BPFs.

A. EXAMPLE 1: MODELING OF THIRD-ORDER SINGLE-BAND NR-BPF

In this first example, the proposed attention-based deep neural network is applied to the high-dimensional microwave modeling of a third-order NR-BPF as shown in Fig. 8. The inputs of model are the circuit parameter of NR-BPF, and outputs of model are real and imaginary parts of S -parameters at various frequency points as shown below.

- Input : $[M_{s1} M_{12} M_{23} M_{34} f_m m \Delta \varphi BW]^T$
- Output $\begin{bmatrix} RS_{11} \cdots IS_{11} \cdots RS_{21} \cdots IS_{21} \cdots \\ RS_{12} \cdots IS_{12} \cdots RS_{22} \cdots IS_{22} \cdots \end{bmatrix}^T$ @ various frequency 350 points between 1.5 GHz and 2.1 GHz.

Once the model is successfully trained, we test the model providing model input parameters and S -parameters are extracted as model output. Table 1 demonstrates prediction accuracy of the deep learning models for the third-order single-band NR-BPF. For comparison purposes, we also implemented deep MLP and conventional CNN using the same datasets. The table clearly shows that the proposed methodology significantly outperforms as compared to other deep neural network methods such as MLP and conventional CNN. Notably, the proposed method exhibits superior perfor-



proposed model achieves accurate prediction results, particularly within the specific frequency range of 1.74 to 1.86 GHz (Region 2), where the values change significantly. This indicates that the attention mechanism effectively learns the most useful relational information between input signals and output parameters. The significance of the improvements can be

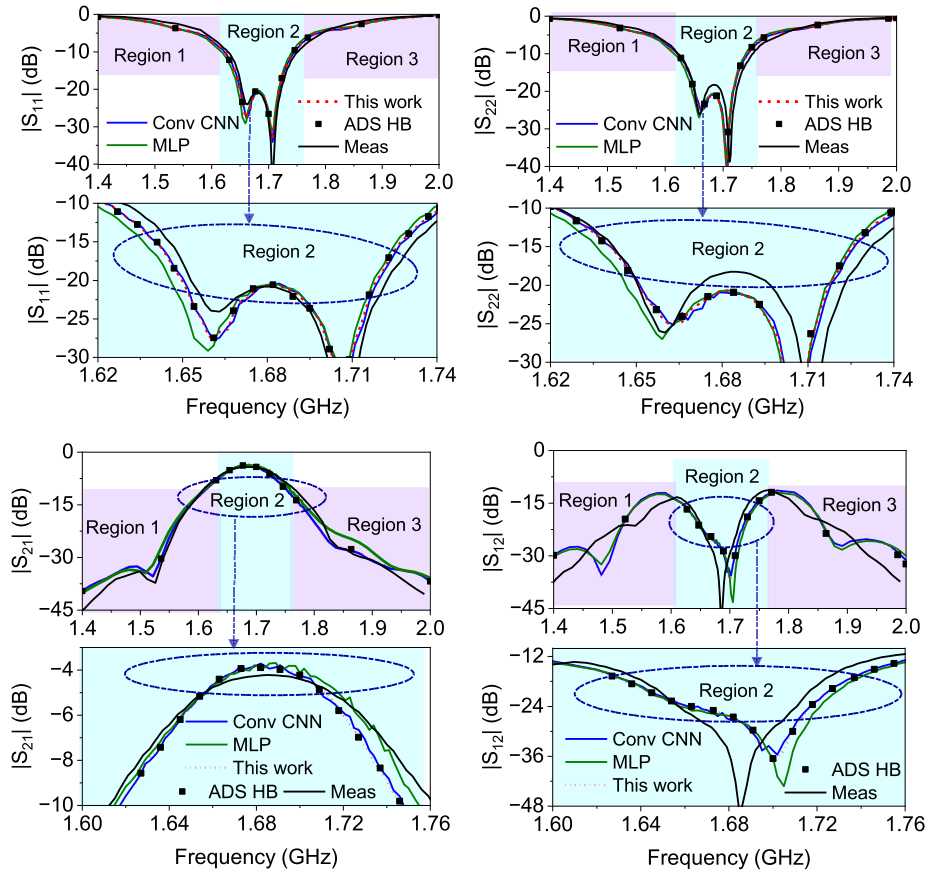


FIGURE 15. Comparison of results from various deep learning models with Keysight advanced system design harmonic balance (ADS HB) method and measurement results for microstrip line NR-BPF as example 4.

TABLE 4. comparison of conventional deep learning models and proposed attention-based deep learning model for microstrip line third-order NR-BPF as example 4. (Δ : Skip values compared to conventional deep learning models.)

Modeling Example	Training data	Modeling Methods	RMSE @ different frequency region separately			@ Overall frequency region	
			Region 1	Region 2	Region 3	RMSE	R ²
Example 4: Microstrip line third-order single- band NR-BPF	12,000	Deep MLP	36.38 %	20.56 %	36.38 %	23.45 %	99.37 %
		Conventional CNN	7.08 %	8.59 %	7.08 %	8.77 %	99.90 %
		This work	6.17 %	4.12 %	6.17 %	4.71 %	99.98 %
		Δ	-30.21 %	-16.44 %	-13.97 %	-18.74 %	+0.61 %
			-0.91 %	-4.47 %	-5.34 %	-4.06 %	+0.08 %

observed in S_{21} and S_{12} graphs, where the predicted results from the proposed model closely match the true desired values obtained from the HB method.

B. MODELING OF FOURTH-ORDER SINGLE-BAND NR-BPF

In this second example, the proposed attention-based deep neural network is applied to predict the behavior of fourth-order single-band NR-BPF as shown in Fig. 10. As like the first example, the model's inputs are the circuit parameter of NR-BPF, and outputs are real and imaginary parts of S -parameters at various frequency points as shown below. As compared to the first example, the number of input parameters has increased from 8 to 9.

- Input : $[M_{51} \ M_{12} \ M_{23} \ M_{34} \ M_{45} \ f_m \ m \ \Delta\varphi \ BW]^T$
- Output: $\begin{bmatrix} RS_{11} \cdots IS_{11} \cdots RS_{21} \cdots IS_{21} \cdots \\ RS_{12} \cdots IS_{12} \cdots RS_{22} \cdots IS_{22} \cdots \end{bmatrix}^T$ @ various frequency 350 points between 1.5 GHz and 2.1 GHz.

Table 2 demonstrates RMSE and R² of the generated by model of the fourth-order single-band NR-BPF. As seen from this table, the proposed model predicts S -parameters with better accuracy as compared to MLP and standard CNN. As observed from Table, the proposed method achieved significant improvements in model performance, particularly in frequency Region 2. Fig. 11 shows the comparison between the S -parameters computed from different deep neural network models and ADS HB method. As seen from the figure,

TABLE 5. Ablation study for impact of joint channel-spatial attention block and 1D residual block.

Modeling Example	Baseline	Attention block	Residual block	RMSE	R ²
Example 1: Third-order single-band NR-BPF	✓	×	×	4.29 %	99.97%
	✓	✓	×	2.55 %	99.98%
	✓	×	✓	2.62 %	99.98%
	✓	✓	✓	2.38 %	99.99%
Example 2: Fourth-order single-band NR-BPF	✓	×	×	8.06 %	99.92%
	✓	✓	×	3.35 %	99.98%
	✓	×	✓	3.45 %	99.98%
	✓	✓	✓	2.07 %	99.99%
Example 3: Third-order dual-band NR-BPF	✓	×	×	6.24 %	99.97%
	✓	✓	×	4.23 %	99.97%
	✓	×	✓	4.45 %	99.97%
	✓	✓	✓	3.41 %	99.99%

TABLE 6. Comparison of different deep neural network methods of high-dimensional NR-BPF modeling with distinct training data samples and different testing cases.

Testing Cases	Modeling methods	Single-band NR-BPF				Dual-band NR-BPF	
		Third-order		Fourth-order		Third-order	
		RMSE	R ²	RMSE	R ²	RMSE	R ²
Case 1	Deep MLP	11.17 %	99.81%	17.73 %	99.38 %	15.96 %	99.83 %
	Conventional CNN	4.37 %	99.94%	6.89 %	99.91 %	6.85 %	99.91 %
	Proposed attention-based DNN	2.78 %	99.98 %	3.87 %	99.98 %	3.75 %	99.98 %
	RMSE reduced as compared to conventional deep learning models	-8.39 % -1.59 %		-13.86 % -3.02 %		-12.21 % -3.10 %	
Case 2	Deep MLP	9.66 %	99.84 %	17.81 %	99.33 %	14.69 %	99.83 %
	Conventional CNN	4.58 %	99.92 %	6.59 %	99.91 %	5.74 %	99.95 %
	Proposed attention-based DNN	3.84 %	99.96 %	3.88 %	99.98 %	4.16 %	99.98 %
	RMSE reduced as compared to conventional deep learning models	-5.82 % -0.74 %		-13.93 % -2.71 %		-10.53 % -1.58 %	
Case 3	Deep MLP	9.50 %	99.85 %	12.86 %	99.68 %	10.23 %	99.92 %
	Conventional CNN	4.38 %	99.93 %	5.82 %	99.92 %	5.83 %	99.95 %
	Proposed attention-based DNN	2.28 %	99.99 %	4.77 %	99.93 %	3.97 %	99.98 %
	RMSE reduced as compared to conventional deep learning models	-7.22 % -2.10 %		-8.09 % -1.05 %		-6.26 % -1.86 %	
Case 4	Deep MLP	9.87 %	99.82 %	11.27 %	99.74 %	18.65 %	99.71 %
	Conventional CNN	4.18 %	99.93 %	4.56 %	99.94 %	6.22 %	99.93 %
	Proposed attention-based DNN	2.88 %	99.98 %	3.31 %	99.98 %	4.32 %	99.97 %
	RMSE reduced as compared to conventional deep learning models	-6.99 % -1.3 %		-7.96 % -1.25 %		-14.33 % -1.90 %	
Case 5	Deep MLP	9.49 %	99.86 %	10.16 %	99.82 %	17.86 %	99.72 %
	Conventional CNN	4.11 %	99.93 %	5.65 %	99.90 %	8.22 %	99.88 %
	Proposed attention-based DNN	3.29 %	99.97 %	3.79 %	99.97 %	4.03 %	99.98 %
	RMSE reduced as compared to conventional deep learning models	-6.20 % -0.82 %		-6.37 % -1.86 %		-13.83 % -4.19 %	

DNN: Deep neural network, MLP: Multilayer perceptron, CNN: Convolutional neural network

Testing cases refer to different combination of input parameters of model.

the S -parameters predicted using the proposed model are well-matched with the S -parameters obtained from the ADS HB method. Although the results of $|S_{11}|$ and $|S_{22}|$ are well matched for the three models, the significance of the improvements can be observed in $|S_{21}|$ and $|S_{12}|$ in the case of the proposed attention-based deep neural network model.

C. MODELING OF THIRD-ORDER DUAL-BAND NR-BPF

In this third example, the proposed attention-based deep neural network was employed to predict behavior of the third-order dual-band NR-BPF as shown in Fig. 12. As compared

to the examples 1 and 2, the input and output relationship of predicting the behavior of third-order dual-band NR-BPF is complex. In addition, the number of inputs of third-order dual-band NR-BPF has increased to 20. The input and output of the model are summarized as follows:

- Input: $\begin{bmatrix} M_{s1} & M_{12} & M_{23} & M_{34} & M_{s2} & M_{45} & M_{56} & M_{67} & M_{11} & M_{22} \\ M_{33} & M_{44} & f_m & m\Delta\varphi & BWf_{m1} & m_1\Delta\varphi_1 & BW_1 \end{bmatrix}^T$
- Output: $\begin{bmatrix} RS_{11} \cdots IS_{11} \cdots RS_{21} \cdots IS_{21} \cdots \\ RS_{12} \cdots IS_{12} \cdots RS_{22} \cdots IS_{22} \cdots \end{bmatrix}^T @ \text{various frequency 350 points between 0.8 GHz and 2.2 GHz.}$

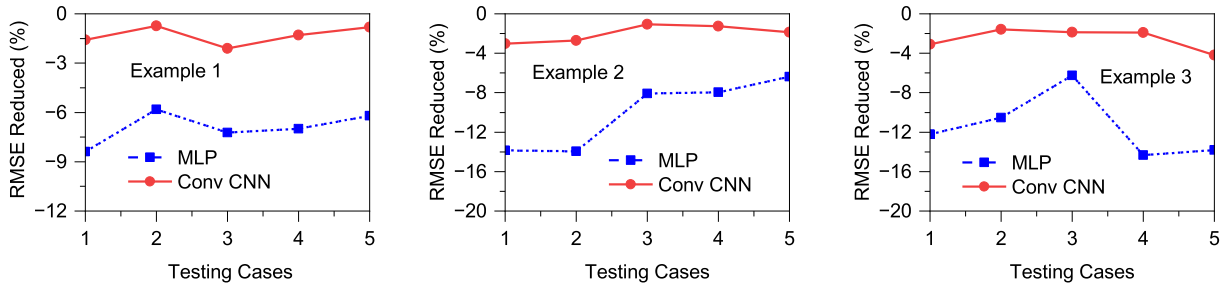


FIGURE 16. Illustration of reduction in RMSE of proposed attention-based deep neural network as compared to MLP and standard CNN with distinct training data samples and different testing cases. Example 1: third-order single-band NR-BPF, Example 2: fourth-order single-band NR-BPF, Example 3: third-order dual-band NR-BPF.

Table 3 shows the results of the prediction accuracy of third-order dual-band NR-BPF that are obtained from different deep neural network models. As demonstrated in the table, the proposed attention-based deep neural network achieved lowest RMSE and highest R^2 as compared to MLP and the conventional CNN model. Notably, the proposed attention-based method achieved a 9.18% reduction in RMSE and a 0.12% improvement in R^2 compared to the deep MLP method, a 2.83% reduction in RMSE and a 0.02% improvement in R^2 compared to the conventional CNN model. Fig. 13 shows comparison of S -parameters results which are computed from the proposed attention-based deep neural network, MLP, conventional CNN, and ADS HB method. As seen from the figure, the S -parameters predicted from the proposed attention-based deep neural network agreed well with desired S -parameters and better than MLP and conventional CNN method as demonstrated in $|S_{21}|$ and $|S_{12}|$ graphs.

D. MODELING OF MICROSTRIP LINE THIRD-ORDER SINGLE-BAND NR-BPF

To further evaluate effectiveness of the high-dimensional microwave modeling of STM-based NR-BPFs, the proposed attention-based deep learning model is applied to develop parametric ANN model of microstrip line third-order single-band NR-BPF as shown in Fig. 14 as example 4. The model inputs are $x = [W_s, L_s, W_{12}, L_{12}, W_{R1}, L_{R1}, W_{R2}, L_{R2}, V_{dc}, f_m, V_m, \Delta\phi]^T$, which are widths and lengths of microstrip line, dc-bias voltage of varactor, and modulation parameters (*i.e.* f_m , V_m , and $\Delta\phi$) [13]. It should be noted that the widths and lengths are defined in millimeter (mm), whereas V_{dc} and V_m are in volts, f_m is in MHz and $\Delta\phi$ is in degree. The model outputs are real and imaginary parts of S -parameters at various frequency points as described as follows:

- Output: $y = \begin{bmatrix} RS_{11} \cdots IS_{11} \cdots RS_{21} \cdots IS_{21} \cdots \\ RS_{12} \cdots IS_{12} \cdots RS_{22} \cdots IS_{22} \cdots \end{bmatrix}^T @$ various frequency 101 points between 1.4 GHz and 2.0 GHz.

The training and testing data are generated by using 3D-electromagnetic (EM) simulation tool (*i.e.* ANSYS HFSS) along with HB simulator (*i.e.* Keysight ADS) where input and output relationships of microstrip line NR-BPF are

captured. Once deep learning model with an attention mechanism is trained with given data and tested with the input *i.e.* $x = [6.585, 21.4871, 21.4523, 3.3456, 1.8343, 30.4366, 0.3245, 30.6507, 3.5352, 112.0744, 0.9674, 53.2242]$, and predicted S -parameters are extracted. For comparison purposes, the performances of the proposed attention-based deep learning model are compared with MLP and conventional CNN model and results are summarized in Table 4. As depicted in table, the proposed attention-based deep learning model achieved the lowest RMSE and highest R^2 as compared to MLP and conventional CNN model, demonstrating superior predictive accuracy.

Fig. 15 shows the S -parameters results which are computed from different deep neural network models and ADS HB method. In addition, the proposed model results are also compared with measured S -parameters. The measured S -parameter results are extracted using Keysight N5224A vector network analyzer. As seen from the figure, the S -parameters predicted using the proposed model are well-matched with the S -parameters obtained from the ADS HB method as well as measured S -parameters.

E. ABLATION STUDY

To evaluate the impact of the joint channel-spatial attention block and the 1D residual block in the proposed model, its performance is compared against the baseline method and results are shown in Table 5. As depicted in Table 5, both the joint channel-spatial attention block and the 1D residual block significantly reduce the RMSE value, highlighting their individual contributions to enhancing performance. Notably, when these components are used in combination, the model consistently achieves the best performance across all examples, indicating that their synergistic integration enhances the model's ability to capture specific features and improve predictive accuracy.

F. GENERALIZATION OF MODEL

To evaluate the generalization capability of the proposed method, we maintained the same model structure while generating distinct training data samples to train the model across three different microwave circuits (*i.e.*,

TABLE 7. Trade-off analysis between accuracy and computational efficiency.

Deep Learning models	Computational complexity			
	NOP (MB)	FLOPS (MB)	Training time (Minute)	FPS
Deep MLP	0.754	0.751	9.73	2459.4
Conventional CNN	3.519	3.665	10.32	823.8
Proposed Attention-based deep learning model	5.107	6.835	10.73	274.91

NOP: Number of parameters of model, FLOPS: Floating point operations, FPS: frame per second in testing of model, MB: Megabytes

Example 1: third-order single-band NR-BPF, Example 2: fourth-order single-band NR-BPF, and Example 3: third-order dual-band NR-BPF). Additionally, we prepared more testing cases with different combination of circuit parameters to further assess the performance of model and results are shown in Table 6. As can be seen from the table, the attention-based deep neural network achieved lowest RMSE and highest R^2 for all testing cases. Furthermore, the proposed method demonstrated stable performance across all three types of NR-BPFs, even with varying training data. Specifically, the RMSE and R^2 values remained consistent with those obtained from previous datasets. As indicated in the table, the proposed model shows no significant variation in RMSE values across different cases for all three NR-BPFs. In contrast, the MLP and CNN methods exhibited RMSE large variations across different testing cases. The proposed attention-based deep neural network model significantly outperformed the other approaches, particularly in comparison to the MLP method, showing substantial improvements. Fig. 16 shows the reduction of RMSE as compared to previous methods. As seen from the figures, it is clear that the proposed model achieved a reduction in RMSE for various testing cases as compared to previous methods. This improvement is observed due to the better generalization of the proposed attention-based method which leads to predicting the S -parameters with smaller error.

To investigate the trade-off between computational efficiency for different deep learning models, we computed computational complexity by considering number of parameters (NOP), floating-point operations (FLOPS), training time, and frame per second (FPS) in testing model and results are summarized in Table 7. Although the proposed attention-based model has more parameter as compared to MLP and conventional CNN, the training time of the proposed model does not increase significantly as well as it predicts outputs in real time with high accuracy. Therefore, a trade-off between computational complexity and prediction efficiency is essential.

IV. CONCLUSION

This paper demonstrated a new attention-based deep neural network technique for high-dimensional microwave modeling of spatio-temporal modulated non-reciprocal bandpass filters. The proposed method uses joint channel-spatial attention and residual connection for improving prediction accuracy and generalization of high-dimensional modeling

of non-reciprocal bandpass filters. The prediction accuracy is enhanced by employing the attention mechanism, which emphasizes the most important features, and by the residual connection, which helps to prevent overfitting. The effectiveness and generalization of the proposed method are validated through four examples of high-dimensional S -parameters extraction modeling for non-reciprocal bandpass filters, utilizing various training data and testing scenarios. The experimental results indicate that the proposed method achieved higher prediction accuracy with lower root mean square error (RMSE) as compared to conventional deep multilayer perceptron and convolutional neural network models. Improvement in prediction accuracy is due to the impact of an attention-based mechanism that effectively learns the most useful relational information between input signals and output parameters. In addition, the proposed work provides novel insights into circuit design and analysis of non-reciprocal microwave circuits.

ACKNOWLEDGMENT

(Girdhari Chaudhary and Yao Meng contributed equally to this work.)

REFERENCES

- [1] J. Zhou, N. Reiskarimian, J. Diakonikolas, T. Dinc, T. Chen, G. Zussman, and H. Krishnaswamy, "Integrated full duplex radios," *IEEE Commun. Mag.*, vol. 55, no. 4, pp. 142–151, Apr. 2017.
- [2] N. Reiskarimian, A. Nagulu, T. Dinc, and H. Krishnaswamy, "Nonreciprocal electronic devices: A hypothesis turned into reality," *IEEE Microw. Mag.*, vol. 20, no. 4, pp. 94–111, Apr. 2019.
- [3] C. E. Fay and R. L. Comstock, "Operation of the ferrite junction circulator," *IEEE Trans. Microw. Theory Techn.*, vol. MTT-13, no. 1, pp. 15–27, Jan. 1965.
- [4] C. K. Seewald and J. R. Bray, "Ferrite-filled antisymmetrically biased rectangular waveguide isolator using magnetostatic surface wave modes," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 6, pp. 1493–1501, Jun. 2010.
- [5] N. A. Estep, D. L. Sounas, and A. Alù, "Magnetless microwave circulators based on spatiotemporally modulated rings of coupled resonators," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 2, pp. 502–518, Feb. 2016.
- [6] A. Kord, D. L. Sounas, and A. Alù, "Pseudo-linear time-invariant magnetless circulators based on differential spatiotemporal modulation of resonant junctions," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 6, pp. 2731–2745, Jun. 2018.
- [7] X. Wu, X. Liu, M. D. Hickie, D. Peroulis, J. S. Gómez-Díaz, and A. Á. Melcón, "Isolating bandpass filters using time-modulated resonators," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 6, pp. 2331–2345, Jun. 2019.
- [8] A. Alvarez-Melcon, X. Wu, J. Zang, X. Liu, and J. S. Gómez-Díaz, "Coupling matrix representation of nonreciprocal filters based on time-modulated resonators," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 12, pp. 4751–4763, Dec. 2019.

- [9] G. Chaudhary and Y. Jeong, "Frequency tunable impedance matching non-reciprocal bandpass filter using time-modulated quarter-wave resonators," *IEEE Trans. Ind. Electron.*, vol. 69, no. 8, pp. 8356–8365, Aug. 2022.
- [10] G. Chaudhary and Y. Jeong, "Nonreciprocal bandpass filter using mixed static and time-modulated resonators," *IEEE Microw. Wireless Compon. Lett.*, vol. 32, no. 4, pp. 297–300, Apr. 2022.
- [11] G. Chaudhary and Y. Jeong, "A magnetless filtering circulator with enhanced isolation bandwidth using mixed static and time-modulated resonators," *AEU-Int. J. Electron. Commun.*, vol. 177, Apr. 2024, Art. no. 155198.
- [12] G. Chaudhary and Y. Jeong, "Non-magnetic non-reciprocal bandpass filter with quasi-elliptic response and tunable center frequency," *AEU - Int. J. Electron. Commun.*, vol. 175, Feb. 2024, Art. no. 155111.
- [13] G. Chaudhary and Y. Jeong, "Rigorous design and experimental validation of non-reciprocal bandpass filter with ultrawide high reverse isolation bandwidth," *Eng. Sci. Technol., Int. J.*, vol. 52, Apr. 2024, Art. no. 101674.
- [14] K. S. Kundert and A. Sangiovanni-Vincentelli, "Simulation of nonlinear circuits in the frequency domain," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. CADICS-5, no. 4, pp. 521–535, Oct. 1986.
- [15] D. Shiri, H. R. Nilsson, P. Telluri, A. F. Roudsari, V. Shumeiko, C. Fager, and P. Delsing, "Modeling and harmonic balance analysis of superconducting parametric amplifiers for qubit readout: A tutorial," *IEEE Microw. Mag.*, vol. 25, no. 11, pp. 54–73, Nov. 2024.
- [16] F. Feng, W. Na, J. Jin, J. Zhang, W. Zhang, and Q.-J. Zhang, "Artificial neural networks for microwave computer-aided design: The state of the art," *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 11, pp. 4597–4619, Nov. 2022.
- [17] H. Zhang, Y. Jing, and P. Zhou, "Machine learning-based device modeling and performance optimization for FinFETs," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 70, no. 4, pp. 1585–1589, Apr. 2023.
- [18] F. Feng, C. Zhang, J. Ma, and Q.-J. Zhang, "Parametric modeling of EM behavior of microwave components using combined neural networks and pole-residue-based transfer functions," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 1, pp. 60–77, Jan. 2016.
- [19] Q.-J. Zhang, K. C. Gupta, and V. K. 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, Apr. 2003.
- [20] 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, Oct. 2019.
- [21] A. Faraji, S. A. Sadrossadat, J. Jin, W. Na, F. Feng, and Q.-J. Zhang, "Hybrid batch-normalized deep feedforward neural network incorporating polynomial regression for high-dimensional microwave modeling," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 71, no. 3, pp. 1245–1258, Mar. 2024.
- [22] G. Pan, Y. Wu, M. Yu, L. Fu, and H. Li, "Inverse modeling for filters using a regularized deep neural network approach," *IEEE Microw. Wireless Compon. Lett.*, vol. 30, no. 5, pp. 457–460, May 2020.
- [23] C. Yu, Q. Li, F. Feng, and Q.-J. Zhang, "Convolutional neural network with adaptive batch-size training technique for high-dimensional inverse modeling of microwave filters," *IEEE Microw. Wireless Technol. Lett.*, vol. 33, no. 2, pp. 122–125, Feb. 2023.
- [24] Z. Wei and X. Chen, "Physics-inspired convolutional neural network for solving full-wave inverse scattering problems," *IEEE Trans. Antennas Propag.*, vol. 67, no. 9, pp. 6138–6148, Sep. 2019.
- [25] M. Noohi, A. Faraji, S. A. Sadrossadat, A. Mirvakili, and A. Moftakharzadeh, "Modeling and implementation of a novel active voltage balancing circuit using deep recurrent neural network with dropout regularization," *Int. J. Circuit Theory Appl.*, vol. 51, no. 5, pp. 2351–2374, Nov. 2022.
- [26] A. Faraji, M. Noohi, S. A. Sadrossadat, A. Mirvakili, W. Na, and F. Feng, "Batch-normalized deep recurrent neural network for high-speed nonlinear circuit macromodeling," *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 11, pp. 4857–4868, Nov. 2022.
- [27] A. Faraji, S. A. Sadrossadat, M. Yazdian-Dehkordi, M. Nabavi, and Y. Savaria, "A hybrid approach based on recurrent neural network for macromodeling of nonlinear electronic circuits," *IEEE Access*, vol. 10, pp. 127996–128006, 2022.
- [28] F. Charoosaei, M. Noohi, S. A. Sadrossadat, A. Mirvakili, W. Na, and F. Feng, "High-order deep recurrent neural network with hybrid layers for modeling dynamic behavior of nonlinear high-frequency circuits," *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 12, pp. 5340–5358, Dec. 2022.
- [29] M. Moradi A., S. A. Sadrossadat, and V. Derhami, "Long short-term memory neural networks for modeling nonlinear electronic components," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 11, no. 5, pp. 840–847, May 2021.
- [30] D. Jacoby, J. Ostrometzky, and H. Messer, "Short-term prediction of the attenuation in a commercial microwave link using LSTM-based RNN," in *Proc. 28th Eur. Signal Process. Conf. (EUSIPCO)*, Amsterdam, The Netherlands, Jan. 2021, pp. 1628–1632, doi: [10.23919/EUSIPCO47968.2020.9287835](https://doi.org/10.23919/EUSIPCO47968.2020.9287835).
- [31] K. P. Yalamarthy, S. Dhall, Mohd. T. Khan, and R. A. Shaik, "Low-complexity distributed-arithmetic-based pipelined architecture for an LSTM network," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 28, no. 2, pp. 329–338, Feb. 2020.
- [32] M. A. Alhartomi, M. T. Khan, S. Alzahrani, A. Alzahmi, R. A. Shaik, J. Hazarika, R. Alsulami, A. Alotaibi, and M. Al-Harthi, "Low-area and low-power VLSI architectures for long short-term memory networks," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 13, no. 4, pp. 1000–1014, Dec. 2023.
- [33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [34] S. Woo, J. Park, J. Lee, and I. Kweon, "CBAM: Convolutional block attention module," in *Proc. 15th Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 3–19, doi: [10.1007/978-3-030-01234-2_1](https://doi.org/10.1007/978-3-030-01234-2_1).



GIRDHARI CHAUDHARY (Member, IEEE) received the B.E. degree in electronics and communication engineering from Nepal Engineering College (NEC), Kathmandu, Nepal, in 2004, the M.Tech. degree in electronics and communication engineering from MNIT, Jaipur, India, in 2007, and the Ph.D. degree in electronics engineering from Jeonbuk National University, Republic of Korea, in 2013.

From March 2021 to August 2024, he was a Contract Professor with the JIANT-IT Human Resource Development Center, Jeonbuk National University, South Korea. Previously, he was a Principal Investigator (PI) of an Independent Project through the Basic Science Research Program of the National Research Foundation (NRF) funded by the Ministry of Education Korea. He is currently a Research Associate Professor with the Jeonbuk RICE Intelligence Innovation Research Center, Jeonbuk National University. His research interests include multi-band tunable passive circuits, in-band full duplex systems and high-efficiency power amplifiers, negative group delay circuits, and their applications.

Dr. Chaudhary was a recipient of the BK21 PLUS Research Excellence Award 2015 and Korean Research Fellowship (KRF) through the National Research Foundation (NRF) of Korea funded by the Ministry of Science and ICT. He has served as a Reviewer for IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES, IEEE MICROWAVE AND WIRELESS COMPONENT LETTERS, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I: REGULAR PAPERS, and IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS.



YAO MENG received the M.S. degree in information engineering from Capital Normal University, China, in 2018, and the Ph.D. degree in electronics engineering major with specialization in artificial intelligence and computer vision from the Division of Electronics and Information Engineering, Jeonbuk National University, in 2025. Her research interests include pattern recognition, computer vision, and machine learning.



DONG SUN PARK received the B.S. degree in electronic engineering from Korea University, South Korea, in 1979, and the M.S. and Ph.D. degrees in electronic engineering from the University of Missouri, USA, in 1984 and 1991, respectively. His research interests include deep neural networks, pattern recognition, image processing, and digital systems design.



YONGCHAE JEONG (Senior Member, IEEE) received the B.S.E.E., M.S.E.E., and Ph.D. degrees in electronic engineering from Sogang University, Seoul, Republic of Korea, in 1989, 1991, and 1996, respectively.

From 1991 to 1998, he was a Senior Engineer with Samsung Electronics, South Korea. In 1998, he joined the Division of Electronic Engineering, Jeonbuk National University, Jeonju-si, Republic of Korea. Since July 2024, he has been the Director of the Jeonbuk RICE Intelligence Innovation Research Center, Jeonbuk National University. He was the Vice-President of Plan Affairs and the Director of the HOPE-IT Human Resource Development Center of BK21 PLUS, Jeonbuk National University. From July 2006 to December 2007, he joined Georgia Institute of Technology as a Visiting Professor. He is currently teaching and conducting research in microwave passive and active circuits, mobile and satellite base-station RF systems, design of periodic defective transmission lines, negative group delay circuits and its applications, in-band full duplex radio, and RFIC design. He has authored and co-authored over 320 papers in international journals and conference proceedings. He is a member of Korea Institute of Electromagnetic Engineering and Science (KIEES).

...