Fast and Accurate Deep Neural Network (DNN) Model Extension Method for Signal Integrity (SI) Applications

Hyunwook Park, Junyong Park, Daehwan Lho, Seongguk Kim, Seungtaek Jeong, Gapyeol Park, Subin Kim, Hyungmin Kang, Boogyo Sim, Kyungjune Son, Taein Shin and Joungho Kim
Terabyte Interconnection and Pacakage Laboratory
Korea Advanced Institute of Science and Technology (KAIST)
Daejeon, South Korea
E-mail: hyunwookpark@kaist.ac.kr

***Abstract*— In this paper, we first propose a fast and accurate deep neural network (DNN) model extension method for signal integrity (SI) applications. Reusing pre-trained weights of DNN model, the model can be extended when new training data are given. Instead of updating whole weights of DNN in traditional machine learning (ML) approaches, fine-tuning of a part of weights can accelerate training. For verification, we applied the proposed method to regression model of peak time domain reflectometry (TDR) impedance of through hole via (THV) and classification model of through silicon via (TSV) void defects. Training time of the proposed method were 0.3 s and 2.3 s respectively, which are 99 % and 82.3 % reduction compared to the traditional approach. Moreover, test accuracy of the proposed method achieved 99.2 % and 100 %, respectively.**

***Keywords*— Deep neural network, fine tuning, signal integrity, transfer learning**

I. Introduction

Recently, various machine learning (ML) approaches have been introduced in signal integrity applications [1]. Especially, deep neural network (DNN) can solve complex and non-linear relationships between inputs and outputs as shown in Fig. 1. Hence, SI design metrics such as eye diagram and time domain reflectometry (TDR) impedance can be estimated, with input as channel design parameters. Also, high-speed channel modeling and through silicon via (TSV) defect classification can be performed via DNN model. These black box approaches of DNN do not require SI domain knowledge, and need only training data to find out mapping function between inputs and outputs. Since DNN directly map the input parameters to the SI design metrics, DNN can replace the whole process of EM simulations for extraction of channel characteristics and circuit simulations. Moreover, once DNN model is obtained via training, prediction of the outputs can be performed fast in a few milliseconds to seconds.

Various deep neural network (DNN) based approaches for SI applications have been studied [2]–[3]. Regression model of peak TDR impedance (Z_{TDR}) of differential via has been investigated [2]. Also, classification model of TSV void defects has been introduced [3]. However, these DNN models works only at specified range or conditions. In other words, the models can cover only the range of given training data sets. In conventional ML approaches, to extend the DNN models, whole weights of the DNN is re-trained with new training data sets, which is time-consuming. Therefore, time-efficient and accurate training methods to extend DNN models are essential.

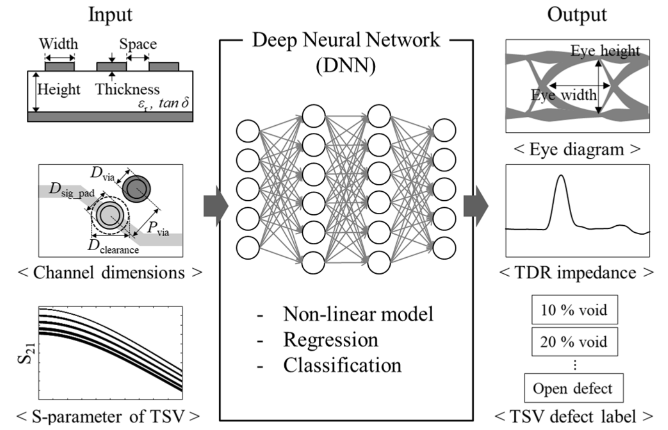

Fig. 1. Deep neural network (DNN) model for signal integrity (SI) applications.

In this paper, for the first time, we propose a fast and accurate DNN model extension method for SI applications. The DNN model can be extended by reusing weights of the pre-trained model, and fine-tuning a part of the weights with new training data sets. For verification, we applied the proposed method to regression model of peak Z_{TDR} of through hole via (THV) and classification model of TSV void defects. We compared training time and test accuracy, between the proposed method and the conventional ML method of re-training whole weights. The proposed method reduced 99 % and 82.3 % of the training time compared to the conventional method, respectively; the training time of the proposed method were 0.3 s and 2.3 s, respectively; those of the conventional method were 30.7 s and 13.0 s, respectively. Also, test accuracy of the proposed method were 99.2 % and 100 % respectively, which are almost equivalent to the conventional method of 99.3 % and 100 %, respectively. In other words, the proposed method can achieve accurate extended DNN models with significantly reduced training time.

II. Proposed DNN Model Extension Method for SI Applicaions

Fig. 2 shows a concept of the proposed DNN model extension method for SI applications. Two main ideas of the proposed method are as follows: reusing weights of the DNN model, pre-trained by original SI data sets; fine-tuning a part of the weights with the whole training data sets including new data sets.

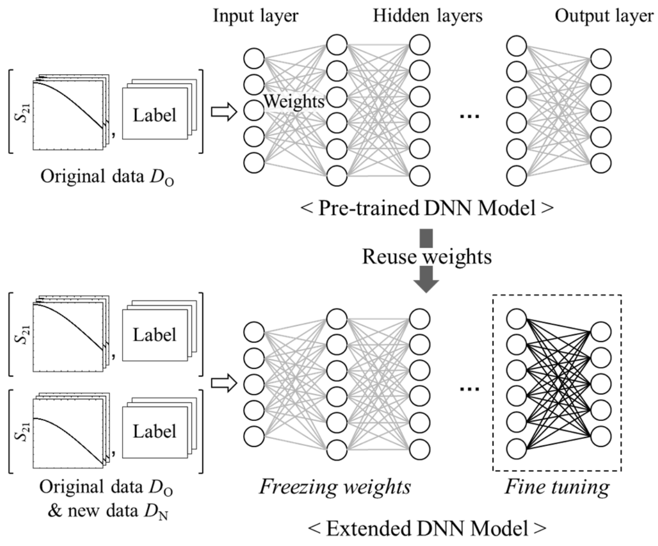

Fig. 2. Proposed DNN model extension method for SI applications.

Denote that original SI training data sets as D_O and new data sets as D_N, where $D_O = \{(x_1^O, y_1^O), \ldots, (x_n^O, y_n^O)\}$ consisting of n sets, and $D_N = \{(x_1^N, y_1^N), \ldots, (x_m^N, y_m^N)\}$ comprising m sets. The terms x and y are input feature vector and corresponding label vector, respectively. For SI applications, input features can be design parameters of high-speed channels, insertion loss S_{21}, return loss S_{11}, and etc. And corresponding label can be height/width of eye diagram, peak Z_{TDR}, TSV defect labels, and etc. The pre-trained DNN model indicates the weight optimized DNN by back propagation algorithms with the original data sets D_O [4].

To extend the DNN model to cover larger feature space or label space, weights of the pre-trained model should be updated with new additional training data sets D_N. This process is consisted of two procedure: cloning the weights of the pre-trained model and updating the weights of the last few layers via the back propagation algorithm, i.e. fine-tuning. When fine-tuning the weights of the last few layers, those of the rest layers keep frozen. Gradient descent algorithm is used to update the weights with whole data sets including the original D_O and the new data sets D_N. The gradient descent algorithm is a iterative optimization algorithm to find the minimum of loss function L, which is an error function between the estimated values by DNN and the given labels [4]. The equation for gradient updating the weights is as follows:

$$W = W - \alpha \frac{\partial L}{\partial W}. \tag{1}$$

where W is a weight matrix and α is learning rate. The DNN model extension can be achieved by iterating the gradient updating of the weights of the last few layers. Since only a part of the weights are updated, the training time can be further reduced compared to the conventional method.

III. Verification of the Proposed Method

For verification, we applied the proposed method to regression model of peak Z_{TDR} of THV and classification model of TSV void defect, and compared the training time and test accuracy with the conventional method [2], [3].

A. Regression Model of Peak TDR Impedance for THV

Fig. 3 shows geometry and design parameters of THV in 4-layer printed circuit board (PCB), for verification. We

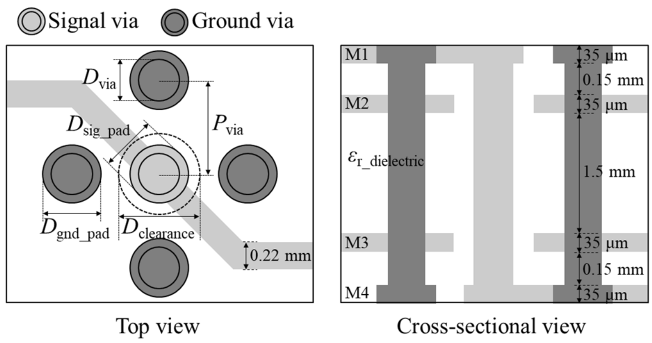

Fig. 3. Geometry and design parameters of THV in 4-layer PCB.

TABLE I. Range of Design Parameters of THV

Design Parameter	***Min***	***Max***
D_{via}	0.1 mm	0.35 mm
$D_{clearance}$	0.36 mm	0.53 mm
D_{sig_pad}	0.2 mm	0.45 mm
D_{gnd_pad}	0.2 mm	0.4 mm
P_{via}	0.45 mm	3.0 mm
$\varepsilon_{r\_dielectric}$	3.2	4.8

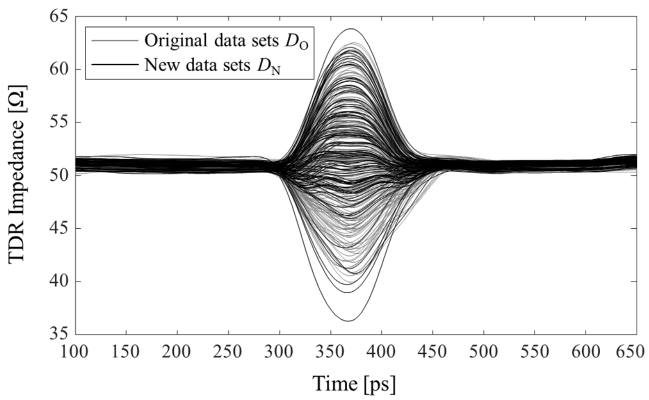

Fig. 4. TDR impedance of 198 original and 102 new training data sets.

assumed single signal via transition from top layer to bottom layer, with 4 ground vias. Input features are the design parameters of THV: diameter of signal/ground via D_{via}, diameter of clearance $D_{clearance}$, diameter of signal pad D_{sig_pad}, diameter of ground pad D_{gnd_pad}, signal via to ground via pitch P_{via}, and dielectric constant $\varepsilon_{r\_dielectric}$. And corresponding label is value of the peak Z_{TDR}. The range of values of the design parameters are summarized in Table I.

For data generation, we used latin hypercube sampling (LHS) to generate total 300 input feature vectors, and corresponding labels were obtained by ANSYS 3D EM simulation. We divided the whole data sets into original data sets D_O and new data sets D_N for verification: data sets which meets $3.74 < \varepsilon_{r\_dielectric} \leq 4.8$ configured D_O; otherwise D_N. Each D_O and D_N consisted of 198 and 102 data sets, respectively. TDR impedance of the original and new data sets, seen at the end of the signal line on the top layer, are depicted in Fig. 4. The rise time was set as 75 ps. The original data sets are indicated in gray lines; the new data sets are given in black lines. Depending on the values of the design parameters of THV, peak Z_{TDR} varies dynamically, as shown in the Fig. 4.

Fig. 5 shows structure of the DNN model of regression of peak Z_{TDR}. We used total of 5 fully connected layers including 3 hidden layers. Number of nodes in each layer are as follows:

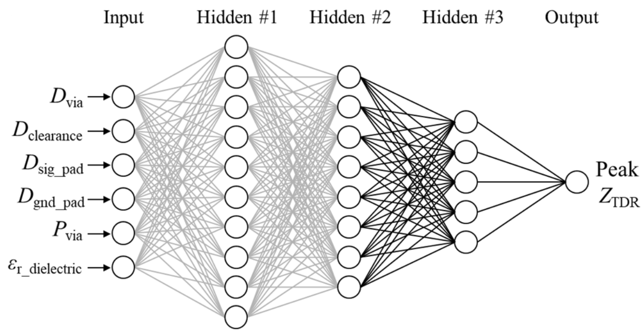

Fig. 5. Structure of DNN model for regression of peak Z_{TDR}.

6 for input layer; 10 for hidden layer #1; 8 for hidden layer #2; 5 for hidden layer #3; 1 for output layer. Linear function was used as activation function for the output layer, and the rest of the layers used sigmoid function.

The pre-trained DNN model was obtained with the 198 original data sets D_O. We extended the pre-trained model by reusing the weights and fine-tuning the weights connecting the last 3 layers, which are indicated in black lines in the Fig. 5. Total 300 data sets, consisting of D_O and D_N, were divided into training, validation and test sets of 70 %, 15 %, and 15 %, respectively. The fine-tuning was conducted with the training data sets until the training accuracy achieves 99 %, and simultaneously we validated the accuracy with validation sets. After the training, test accuracy was validated by the test sets. Fig. 6 shows a comparison of the peak Z_{TDR} between the proposed method and the given labels. The training, validation, and test accuracy were 99 %, 99.2 %, and 99.2 %, respectively. We also compared the training time of the proposed method and the conventional method of re-training the whole weights. The proposed method consumed 0.3 s, which is 99 % reduction compared to the conventional method of 30.7 s, as denoted as case A in Table II.

B. Classification Model of TSV Void Defect

We also applied the proposed method to classification model of TSV void defects. The task of the DNN model is to classify the void ratio for given insertion loss S_{21} [3]. We generated total 9 labels at 5 % intervals from 0 % to 40 % void ratio. The insertion loss S_{21} from 100 MHz to 20 GHz at 936 sampling points, which is input feature data, were obtained through 3D EM simulations. Total of 450 sets are generated: 50 sets for each label. For verification, data sets of 40 % void label were set as D_N; the rest were set as D_O.

The pre-trained DNN model comprised total of 5 layers. Number of nodes in each layer are as follows: 936 for input layer; 1000 for hidden layer #1; 800 for hidden layer #2; 15 for hidden layer #3; 8 for output layer. However, to classify the input data into 9 labels, the number of node in output layer was extended to 9 in the extended DNN model. In this case, we called it transfer learning, since the label space of the pre-trained and the extended DNN are different [5].

The weights connecting the last 3 layers were fine-tuned with data sets including additional sets of 40 % void label. And the results by the proposed method are summarized in Table II, denoted as case B. The training time of the proposed method was 2.3 s, reduced by 82.3 % compared to the conventional method. In addition, the test accuracy achieved 100 % equivalent to that of the conventional method. In other words, the proposed method achieved the accurate extended DNN model with significantly reduced training time.

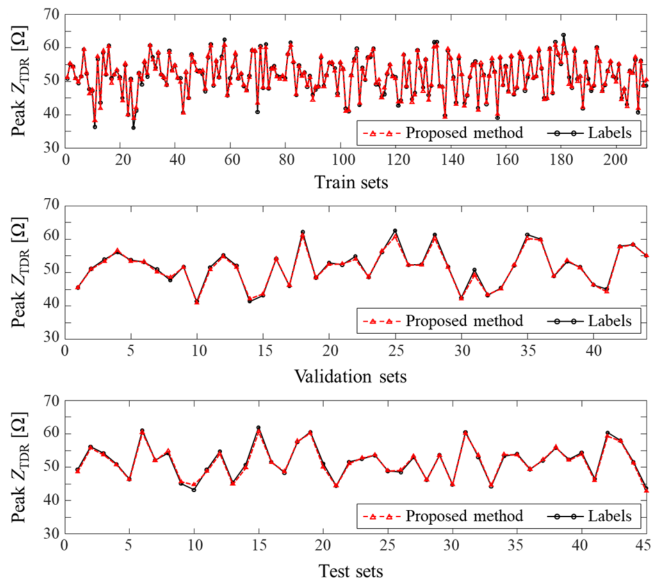

Fig. 6. Comparison of peak Z_{TDR} between the proposed method and labels.

TABLE II. COMPARISION OF TRAINING TIME AND TEST ACCURACY BETWEEN THE PROPOSED AND CONVENTIONAL METHOD

Method	*Training time*		*Test accuracy*	
	Case A	*Case B*	*Case A*	*Case B*
Proposed Method	0.3 s	2.3 s	99.2 %	100 %
Conventional Method	30.7 s	13.0 s	99.3 %	100 %

IV. CONCLUSION

In this paper, we proposed a fast and accurate DNN model extension method for SI applications. We applied the proposed method to the regression model of peak Z_{TDR} of THV and the classification model of TSV void defects. Reusing the pre-trained model and fine-tuning the weights between the last few layers, we successfully validated that the proposed method critically reduced the training time by 99 % and 82.3 %, respectively, compared to the conventional method. In addition, the test accuracy of the proposed method achieved 99.2 % and 100 %, respectively.

ACKNOWLEDGMENT

We would like to acknowledge the technical support from ANSYS Korea. This work was supported in part by the National Research Foundation of Korea Grant funded by the Korean Government (2019M3F3A1A03079612).

REFERENCES

[1] Tianjian Lu and Ken Wu, “Machine Learning Methods in High-Speed Channel Modeling”, in *Proc. DesignCon 2019*.

[2] J. Xu, L. Zhang, M. Sapozhnikov and J. Fan, "Application of Deep Learning for High-speed Differential Via TDR Impedance Fast Prediction," *2018 IEEE Symposium on Electromagnetic Compatibility, Signal Integrity and Power Integrity (EMC, SI & PI)*, Long Beach, CA, 2018, pp. 645-649.

[3] Y. Huang, C. Pan, S. Lin and M. Guo, "Machine-Learning Approach in Detection and Classification for Defects in TSV-Based 3-D IC," in *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 8, no. 4, pp. 699-706, April 2018.

[4] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA: MIT Press, 2016.

[5] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345-1359, Oct. 2010.