MAUnet: Multiscale Attention U-Net for Effective IR Drop Prediction

Mingyue Wang[1,2], Yuanqing Cheng[1,2], Yage Lin[1], Kelin Peng[1], Shunchuan Yang[1], Zhou Jin[3], Wei W. Xing[4*]

[1] School of Integrated Circuit Science and Engineering, Beihang University, Beijing, China
[2]National Key Laboratory of Spintronics, Beihang University, Hangzhou, China
[3] Super Scientific Software Laboratory, China University of Petroleum, Beijing, China
[4] SoMas, The University of Sheffield, U.K.
{Mia_wang,yuanqing,linyage,colinpeng,scyang}@buaa.edu.cn,jinzhou@cup.edu.cn,wxingphd@gmail.com

ABSTRACT

The efficient analysis of power grids is a crucial yet computationally challenging task in integrated circuit (IC) design, given the shrinking power supply voltage of ultra deep-submicron VLSI design. Different from the conventional modified nodal analysis technique, this paper introduces MAUnet, an innovative machine-learning model that redefines state-of-the-art full-chip static IR drop prediction. MAUnet ingeniously integrates multi-scale convolutional blocks, attention mechanisms, and U-Net architecture to optimize prediction accuracy. The multi-scale convolutional blocks significantly enhance feature extraction from image-based data, while the attention mechanism precisely identifies hotspot regions. The U-Net architecture, on the other hand, enables scalable image-to-image prediction applicable to circuits of any size. Uniquely, MAUnet also incorporates a pioneering fusion method that synergies both power grids and image-based data. Additionally, we introduce a low-rank approximation transfer learning technique to extend MAUnet's applicability to unseen test cases. Benchmark tests validate MAUnet's superior performance, achieving an average error of less than 6% relative to the average IR drop on three benchmarks. The performance enhancements offered by our proposed method are substantial, outperforming the current state-of-the-art method, IREDGe, by considerable margins of 29%, 65%, and 68% in three canonical benchmarks. Transfer learning is validated to enable model to achieve effective improvement on real circuit test cases. Compared to commercial tools, which often require hours to deliver results, the proposed method provides orders of magnitude speed-up with negligible error in practice.

KEYWORDS

IR drop analysis, U-Net, Attention mechanism, Machine learning

** Corresponding author.
This work is supported in part by the Natural Science Foundation of China under Grant No. 92373205 and the Natural Science Foundation of Beijing, China under Grant No. Z230002.

DAC '24, June 23–27, 2024, San Francisco, CA, USA

ACM ISBN 979-8-4007-0601-1/24/06…$15.00
https://doi.org/10.1145/3649329.3658465

1 INTRODUCTION

As technology nodes continue to scale down, there is a concomitant reduction in supply voltage and an increase in current density. These phenomena collectively narrow the power supply noise margin, amplifying the significance of meticulous power supply noise analysis in electronic design automation (EDA). Two principal types of power supply noise emerge in this context: IR drop and transient voltage drop. The former arises from resistive behavior in interconnects, while the latter is a consequence of both inductive and capacitive interactions within interconnects. These noise forms can adversely affect gate performance and may even precipitate timing violations. More critically, voltage overshoot can lead to irreversible device damage and chip malfunction. Given these potential repercussions, this paper prioritizes the analysis of IR drop, which holds considerable sway over both the performance and the reliability of the chip [1].

Estimating power supply noise in contemporary semiconductor chips presents a formidable computational challenge, particularly with respect to IR drop. A prevailing approach to address this is Modified Nodal Analysis (MNA), which leverages Kirchhoff's Current Law to formulate a system of linear equations. The system is characterized by a conductance matrix, a node voltage vector, and a current source vector. Solving this extensive system of equations yields the node voltages across the entire power delivery network. However, the challenge exacerbates as integration density escalates; the linear equation system can swiftly balloon to encompass millions of voltage nodes and interconnect segments, thereby making the process both computationally intensive and memory-demanding. Various algorithms and techniques have been proposed to mitigate this computational burden, including the multigrid method [2], random walk algorithms [3], and the boundary element method [4]. Nevertheless, the computational overhead remains substantial, particularly for chips featuring non-uniform and irregular power grid designs.

The incorporation of Machine learning (ML) methodologies has heralded new avenues for expediting IR drop analysis via data-driven approaches. Despite the promise, extant ML-based solutions have largely been restricted to more abstract models of the power delivery network (PDN). In the work presented in [5], a voltage drop predictor was devised by extracting local attributes of the PDN and integrating them within an XGBoost framework. However, the generalizability of this approach to full-chip analysis—especially for larger chips with intricate floor-plan structures—remains an

open question. Alternatively, the study in [6] employed a convolutional neural network (CNN)-based strategy for full-chip IR drop prediction. This approach partitions the chip's power maps into discrete units and makes predictions based on the features of these units. Despite its merits, the method introduces complexities associated with the selection of the window size. Smaller windows risk violating the principle of locality, as cited in [7], whereas larger windows contribute to models that are both memory-intensive and time-consuming during the training and inference phases. Moreover, this study does not adequately address the influence of pad positions on IR drop, which could be a confounding factor affecting prediction accuracy. The work in [8] introduces the IREDGe model, designed to predict full-chip IR drop by employing the current maps, PDN density maps, and effective distance (eff-dist) maps. These maps serve to characterize the PDN topology and pad distribution respectively. Being an image-to-image model, IREDGe conducts the prediction in a single pass, which may compromise the model's prediction accuracy.

Informed by existing literature, we introduce MAUnet, a state-of-the-art ML model tailored for full-chip static IR drop prediction. Our model ingeniously integrates multi-scale convolutional blocks, the attention mechanism, and the U-Net architecture to deliver optimized predictive performance. The multi-scale convolutional blocks facilitate the extraction of a rich set of feature representations. Concurrently, the attention mechanism serves to isolate and suppress irrelevant features, thereby sharpening the model's focus on accurate IR drop prediction, especially for values exhibiting deviations from the mean. Through this harmonized combination, denoted as "M+A+U", our method sets a new standard by requiring fewer training data and yielding higher accuracy. Most notably, it also exhibits the ability to transfer learned knowledge across real circuit data. Furthermore, we introduce a feature extraction method that combines power grids and feature maps to improve the input features and enhance the model accuracy consistently. The contributions of this work are as follows:

(1) We introduce MAUnet, an advanced deep learning model for full-chip static IR drop prediction that synergistically combines multi-scale convolutional blocks, the attention mechanism, and the U-Net architecture.
(2) We present a general power delivery network (PDN) topology feature extraction, which substantially and uniformly boosts predictive capability.
(3) We propose an efficient fine-tuning strategy to enable pre-trained model transfer to real circuit test cases.
(4) Our method shows excellence in four indicators compared to SOTA, especially after adding additional extracted features. Overall, it achieves improvements of 35%, 66%, and 69% on three benchmarks.

2 PRELIMINARIES AND RELATED WORK

2.1 Conventional IR Drop Computation

MNA offers a systematical apparatus to analyze complex circuit networks, which can include both passive and active components, e.g., voltage sources and controlled current sources. When applied to PDN for the purpose of computing IR drop, the network can be modeled as a resistive network augmented with current and voltage sources.

MNA generalizes classical nodal analysis by incorporating additional equations that allow for the representation of a broader range of components like voltage sources. By exploiting Kirchhoff's current law, we can establish a system of linear equations capturing the network's behavior. Formally, for a network with N nodes, B branches, and V independent voltage sources, the MNA equations can be written as:

$$\begin{bmatrix} G & B \\ B^T & D \end{bmatrix} \begin{bmatrix} \mathbf{V} \\ \mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{ext} \\ \mathbf{E} \end{bmatrix} \tag{1}$$

where G is the conductance matrix, B and B^T are incidence matrices relating nodes to branches, D is a diagonal matrix often used for controlled sources, $\mathbf{V}$ is the node voltage vector, $\mathbf{I}$ represents the currents in the independent voltage sources, $\mathbf{I}_{ext}$ is the vector of external currents entering each node, and $\mathbf{E}$ is the vector of voltage sources in the network. Once this system of equations is assembled, solving it yields the nodal voltages, providing insights into the IR drop across the PDN. This methodology serves as a foundation for more sophisticated analyses and improvements, including those enabled by machine learning algorithms, as it presents an efficient, albeit classical, approach for IR drop computation.

For modern chip design, the power delivery network contains millions nodes and branches, which results in huge memory and time overhead when solving Eq.1. Although many techniques have been proposed, such as multi-grid technique [1], random walk method [2] and boundary element method [3]. Solving the large scale equation system above is still a challenging task especially for non-uniform and irregular grids.

2.2 Introduction to U-Net

U-Net [9] is a classic convolutional network architecture first proposed in the field of image segmentation. It consists of two paths: a contracting path on the left and an expansive path on the right. In each step of contracting path, the number of features is doubled by a union of two convolution kernels, ReLU and max pooling. In the expansive pathway, an upsampling is performed by a up-convolution operation, which reduces the number of features by half. After that, these upsampling features are concatenated with those copped downsampling features from the corresponding contracting steps, this process is so-called skip connection. By superimposing upsampling, the feature size is restored step by step. After restoring to the original size, the final layer undergoes a 1x1 convolution to output the result.

U-Net is inherently a fully convolutional network, which means it can accept input images of varying sizes, allowing it to handle images with diverse dimensions. Besides, U-Net's distinctive characteristic lies in its skip connection structure, which helps preserve features from initial layers. CNNs typically lose fine-grained information as they advance from layer to layer, but U-Net's skip connections enable it to maintain and exploit low-level details and local features throughout the network. These make U-Net a powerful choice for a variety of image processing tasks, especially where accommodating different input sizes and preserving detailed information are critical.

2.3 Introduction to Attention Mechanism

Attention is a technique used to emphasize relevant activation during training, thereby reducing the computational resources allocated to irrelevant activation and enhancing the network's generalization capabilities. In essence, it enables the network to selectively

focus on specific portions of an image, and broadly attention mechanisms can be categorized into two types: hard attention and soft attention. Hard attention operates by either cropping the image or iteratively proposing regions to emphasize relevant areas. Conversely, soft attention operates by assigning different weights to various parts of the image. High-correlation regions receive larger weight values, while low-correlation regions are assigned smaller weights. As the model undergoes training, greater emphasis is placed on regions with higher weight values. The model also learns to adjust these weights to improve its ability to discern which parts of the input require attention.

3 THE PROPOSED MAUNET IR DROP PREDICTION APPROACH

3.1 MAUnet: M For Multi-Scale

As is the classic network architecture in the field of image segmentation, U-Net has had many variants and has been used for regression prediction problems [10]. One common way of improvement is to enhance the receptive field of the convolution kernel. The initial U-Net employed a fixed receptive field for its convolution kernel, potentially resulting the performance of the network depends on the choose of convolution kernel size. Small receptive field of convolution kernel may lead to redundant features, while larger may ignore the target details. Thus, constructing multi-scale convolutional blocks with different receptive fields composed of convolution sequences and multiple convolution kernels is an effective way to deal with it. The work in [11] proposed multi-scale convolutional blocks based on several commonly used convolution kernels to improve the ability of U-Net receptive field, which help to extract more diverse information and better feature maps.

In the context of PDN analysis, the IR drop at each location relies on both local and global power information. Typically, the convolution process captures the spatially correlated distribution by sliding an appropriately sized averaging window over the input image. In this process, the influence of local, adjacent and global features on nodes is mainly extracted by the receptive field of convolution kernel. Consequently, crafting extensive receptive fields can aid prediction tasks that depend on intricate spatial information. In light of these considerations, this paper introduces the multi-scale convolutional blocks [11] in the down sampling process to enhance the feature extracting. Each block consists of two convolution kernels with different receptive fields achieved by employing multiple kernel sizes. Assuming the input of the multi-scale convolution block is x, and the output of block is T, the calculation process can be simplified as follows:

$$\begin{aligned} x_1 &= w_{32}(w_{31}x + b_{31}) + b_{32}, \\ x_2 &= w_{72}(w_{71}x + b_{71}) + b_{72}, \\ X &= Cat[x_1, x_2], \\ T &= w_f X + b_f. \end{aligned} \tag{2}$$

x_1 and x_2 are the output of multi-scale convolutional blocks with kernel size 3 and 7. The choice of block kernel sizes 3 and 7 stems from experiments conducted in [11], demonstrating suitability of this combination in prediction tasks.

3.2 MAUnet: A For Attention

In the standard U-Net architecture, skip connections are employed to combine spatial information from the down-sampling path with the up-sampling path to retain better spatial features. Nonetheless, it is commonly observed that the features from initial layers exhibit a poorer feature representation compared to the features from deeper network. To address this issue, the incorporation of a soft attention mechanism [12] can help suppress the activation of irrelevant regions within down sampling features. Specifically, the attention mechanism operates through an attention gate that takes two inputs: g, representing the features from deeper layers, and x, representing the features from the earlier layers. By employing a series of convolution kernel, followed by ReLU and Sigmoid activation functions, relevant portions of the initial features are assigned higher weights, while less relevant portions receive lower weights. The detailed structure of the attention gate is depicted in Fig. 1. Given the soft attention mechanism's capability to auto-

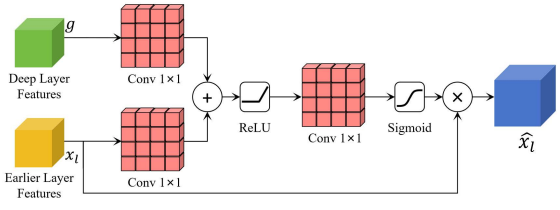

Figure 1: The attention gate proposed in [12].

matically prioritize more relevant features, we apply it to the skip connection part of downsampling and upsampling. This strategic integration can enhance the feature representation where it matters most, thereby allowing our model to more effectively concentrate on critical regions, such as hotspots.

3.3 MAUnet: U For U-Net

Adapting the multi-scale convolutional blocks and the attention mechanism, we name this model MAUnet, and the whole flowchart of MAUnet is shown as Fig. 2. MAUnet uses different numbers of filters for downsampling, specifically 16, 32, 64 and 128, and the corresponding numbers for upsampling are 64, 32, 16. In the downsampling pathway, these filters are essentially the multi-scale convolutional blocks, incorporating kernel sizes of 3 and 7. And the filters in upsampling process named Up-conv, which utilize kernels of sizes 4x4 and 3x3 for feature upsampling. Then the expanded features and down sampling features are directed to the attention gate. Following this, the expanded features are concatenated with the output of the gate, undergo convolution, and are then passed on to the next filter.

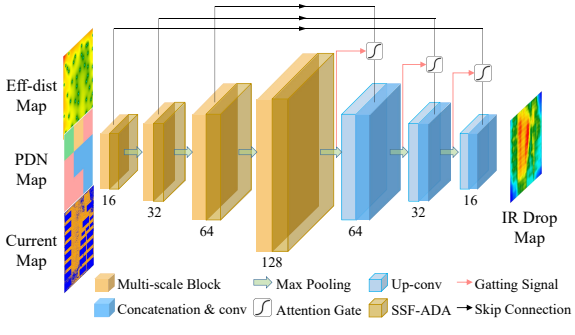

Figure 2: The flowchart of MAUnet proposed in this work, which integrates multi-scale convolutional blocks and attention mechanism.

3.4 Deep Features Extraction for PDN

In our architecture, we extend the feature extraction process beyond the conventional current maps, PDN density maps, and eff-dist maps methodologies described in [8]. We also incorporate insights from [5] to derive supplementary design information from the power grids, elevating the granularity of our analysis.

The power grid serves as a nuanced electrical model generated through SPICE simulations, encapsulating the intricacy of the PDN's topology. For each node p in this grid, we gather essential data, including spatial coordinates, adjacent resistors, and the interconnect topology. Furthermore, the model provides vital metrics such as via positions, pad locations, and current source attributes.

One of the innovative aspects of our approach lies in the mapping between the feature maps and the power grids. In this mapping, the coordinates of a pixel directly relate to a specific sector within the power grid. Using the Nangate 45nm technology as an example, the SPICE-defined units of the power grids (where 2000 units equate to 1μm) are converted into image-based coordinates by a factor of 2000, thereby facilitating a seamless transition from grid-based to image-based representations.

To streamline the extraction process for vias, resistances, and pitches, we propose the following structured methodology:

(1) **Vias**: Components characterized as resistors with connections spanning multiple layers undergo a coordinate transformation. After the transformation, vias are precisely mapped to specific pixels on the feature maps.
(2) **Resistance (R)**: For resistors connected within the same layer, a coordinate transformation maps them onto the feature maps. The pixel value is then set equal to the corresponding resistance. In the case of overlapping features, an additive approach is employed to approximate the combined resistance.
(3) **Pitch**: Utilizing the connection data of layer-specific resistors, we ascertain the precise coordinates of these connecting points. Points sharing the same x/y coordinates are deemed to be part of the same pitch, enabling us to derive pitch data layer-by-layer. For efficiency, our analysis can be restricted to key layers of interest.

This comprehensive feature extraction process equips us with a rich dataset that not only captures the fundamental characteristics of the PDN but also integrates supplementary information to enhance predictive modeling.

3.5 Transfer to Real Simulation

Due to intellectual property (IP) considerations within the field of IC design, there is a scarcity of publicly available benchmarks for real circuits, particularly for advanced technology nodes. However, a significant development in this regard is the BeGAN network introduced in the work referenced as [13]. This innovation enables the generation of a substantial number of benchmarks that closely resemble real circuit benchmarks, thus addressing the scarcity issue. The main difference between real benchmarks and BeGAN benchmarks is the source of current maps, the former uses current maps from physical design simulations, and the latter generates real-like current maps using the BeGAN network. Although both types have high similarity, their subtle difference can lead to variations in IR drop values.

Here, we introduce a transfer learning approach aimed at refining models initially trained on BeGAN benchmarks by incorporating a limited amount of real circuit data. The transfer learning strategy utilizes scale and shift feature adapter (SSF-ADA) proposed in [14], which mainly introduces scale factors and translation factors to modulate features for efficient fine-tuning of parameters. We introduce a SSF-ADA following each multi-scale convolution block in upsampling, which normally participates in the training process. When transferring a pre-trained model to a real circuit dataset, we selectively unfreeze parameters associated with SSF-ADA within the downsampling path. Simultaneously, we maintain the trainability of parameters in the deep layers within the upsampling layer. Parameters in all other layers remain frozen. The update of SSF-ADA is used to adjust the extraction method of down-sampling features, while deep layers in the up-sampling path can be updated with all parameters to adapt to the different IR drop distributions.

4 EXPERIMENT RESULTS

4.1 Experimental setup

To assess MAUnet, we use three open-source benchmarks [15] as our primary data. The benchmarks include the IR drop information of three technology nodes: Nangate 45nm, Asap 7nm, and Skywater 130nm PDK. Nangate 45nm has three kinds of feature maps we need for training, and the other two contain only the power grids and current maps. We supplement eff-dist maps and PDN density maps by applying the method proposed by [8]. The samples of these three datasets are 1000,1000 and 418.

We intend to compare to the SOTA IR drop prediction method, PowerNet [6], which is based on CNN. In their method, a full power map is partitioned into tiles, and a sliding window technique is employed to perform IR drop predictions for each tile. However, we found that we could not afford the time cost—we applied it to Nangate 45nm, where the map sizes change from 200 to 900. When we choose the window size=31 disclosed by this method, we find that a forward pass time of completing an overall prediction of PowerNet on a 204×204 map on NVIDIA GA100 is 372.3s. As we have 1000 samples, it at least takes 4 days to complete a round of training. Therefore, we can not implement the results of PowerNet here.

We partition each set of benchmarks into a training set and a test set, following an 8:2 ratio. Each dataset is max-min normalized and used to train the MAUnet. The details about parameters settings are shown in Tab. 1. To evaluate the performance of our MAUnet, we use the mean absolute error (MAE), F1, the correlation coefficient (CC) and the structural similarity index (SSIM) to measure our prediction accuracy. F1 is a binary classification metric utilized to assess the prediction performance of hotspot regions. This metric utilizes a threshold to differentiate pixels with voltage drop values exceeding a specific threshold observed in each benchmark. In this context, we employ 80% of the maximum IR drop as the threshold. That is, the node with the top 20% is defined as positive class. Then, We use TP, TN, FN standing for true positive, true negative, false positive and false negative. F1 can be calculated as: $F1 = 2{*}P/(P{+}R)$, where $P = TP/(TP + FP)$ and $R = TP/(TP + FN)$. The specific

calculation of CC and SSIM are as follows:

$$CC(A,B) = \frac{\sum_{i=1}^{n}(A_i,\mu_A)(B_i-\mu_B)}{\sqrt{\sum_{i=1}^{n}(A_i-\mu_A)^2\sum_{i=1}^{n}(B_i-\mu_B)^2}},$$
$$SSIM(A,B) = \frac{(2\mu_A\mu_B+C_1)(2\sigma_{AB}+C_2)}{(\mu_A^2+\mu_B^2+C_1)(\sigma_A^2+\sigma_B^2+C_2)}, \quad (3)$$

where the μ_A and μ_B are the means of matrix A an B, σ_A and σ_B are the corresponding variances. σ_{AB} is the covariance, and C_1, C_2 are the hyperparameters.

Table 1: Parameter settings in MAUnet.

Parameters		Settings
Model parameters	Filters	[64,32,12,64]
	Maxpooling filter size	2x2
	Multi-scale block filters	[3,7]
Training parameters	Epoch	400
	Optimizer	ADAM
	Loss function	MAE
	Learning rate	1e-3
	Decay rate	0.6
	Decay step	50

4.2 Compared to IREDGe

We mainly compare our work with the IREDGe proposed in [8]. The specific structure of IREDGe is exactly the same as that shown in the paper. We conducted three independent experiments on each of the three datasets. For each test sample in these nine experiments, we calculated the relative error, the distribution of which is illustrated in Fig. 3. It can be seen that the prediction accuracy of MAUnet is much better than IREDGe. Fig. 4 shows the F1 distribution comparison of three datasets, overall, MAUnet exhibits a more favorable F1 probability distribution compared to IREDGe. The predictive performance of IREDGe and MAUnet in regions characterized by significant IR drop is illustrated in the Fig. 5. Since the map sizes of Asap 7nm are typically smaller compared to the others, it appears to offer a finer level of granularity. The display in Fig. 5 reveals that MAUnet excels in accurately pinpointing hotspot areas in comparison to IREDGe.

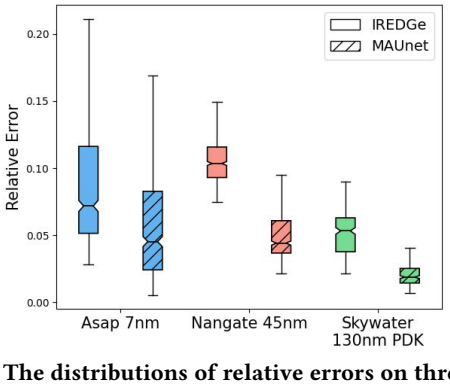

Figure 3: The distributions of relative errors on three benchmarks under three random experiments.

4.3 Ablation Study of Feature Extraction

In order to verify the effectiveness of feature fusion, we compare the performance of the model before and after adding features on both IREDGe and MAUnet. The specific results are shown in Tab 2. The experimental settings for adding extracted features only differ in the number of input feature maps. For each model, we ensure convergence to the desired range by conducting training for 400 epochs. It can be seen from the results that whether adding features to the IREDGe or to MAUnet, the effect is better than the original effect. By adding three additional extracted features, MAUnet can achieve 32%, and 24% improvements on Nangate 45nm and Skywater 130nm PDK respectively, which illustrates the feature extraction method we proposed has an important impact on supplementing the topological information of the chip. Our method obtains the best prediction error reaching 5.9%, 3.8%, and 1.6% of the average IR drop on each benchmark. Combining features extraction, it outperforms IREDGe by 29%, 65%, and 68% respectively.

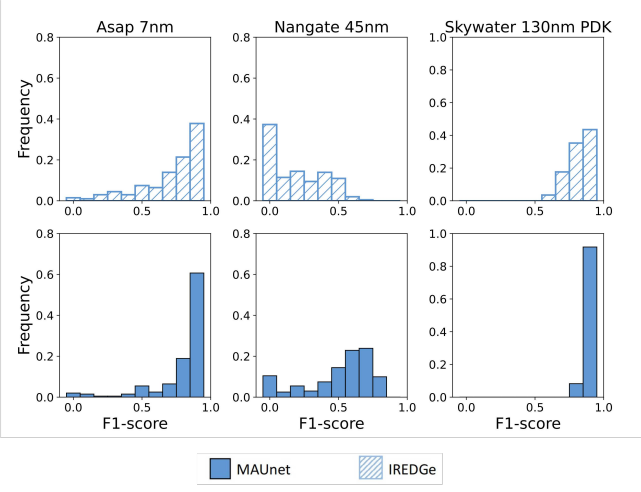

Figure 4: The F1 probability distribution of three samples with average MAE on three benchmarks. The higher the frequency of regions with high F1, the better ability the model's hotspot prediction achieves.

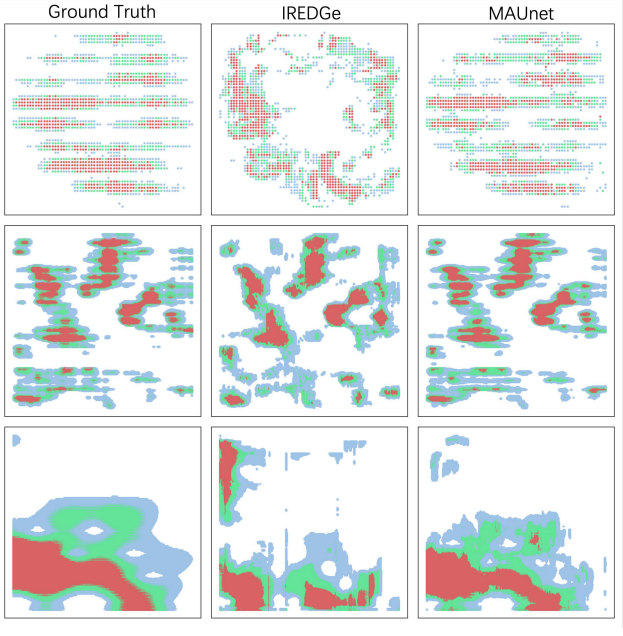

Figure 5: IREDGe and MAUnet performance in high IR regions. The red areas represent the top 10% IR drop, the green corresponds to the regions top 10%-20%, and the blue denotes 20%-30%. The three rows from top to bottom represents Asap 7nm, Nangate 45nm and Skywater 130nm PDK.

Table 2: Numerical Results of Three Nodes before and after Adding Extracted Features

	Asap 7nm				Nangate 45nm				Skywater 130nm PDK			
Method	MAE (V)	F1	CC	SSIM	MAE (V)	F1	CC	SSIM	MAE (V)	F1	CC	SSIM
IREDGe	1.87E-3	0.786	0.708	0.602	1.25E-4	0.256	0.881	0.891	8.22E-5	0.877	0.809	0.896
IREDGe with DF	1.64E-3	0.809	0.789	0.639	1.15E-4	0.318	0.8937	0.8716	5.95E-5	0.933	0.908	0.948
MAUnet	**1.23E-3**	**0.843**	0.948	0.841	6.35E-5	0.580	0.980	0.982	3.43E-5	0.959	0.982	0.967
MAUnet with DF	1.33E-3	0.833	**0.9721**	**0.941**	**4.33E-5**	**0.770**	**0.996**	**0.994**	**2.61E-5**	**0.976**	**0.994**	**0.991**

4.4 Transfer Learning

In the preceding phase of our experiment, we have trained three distinct models, each of them specialized in learning from their respective training datasets. To assess the validity of our transfer learning approach, we conducted the transferring experiments on Nangate 45nm. We utilize a additional dataset consisting of 20 real circuit data samples from the Nangate 45nm. Out of these, we separate 10 cases as fine-tuning cases, while the remaining 10 cases are kept hidden for evaluating the prediction. Before ten sets of real data participate in each set training, we perform upward, left-right and diagonal flips on them. Fine-tuning settings are consistent with training experiments mentioned above. Fig. 7 shows the MAE of the trained Nangate 45nm model using by 10 fine-tuning cases respectively. It's evident that by applying transferring learning, the predictive accuracy of the model has effectively improved on ten unseen 10 samples. The maximum improvement can be achieved by 60%, and the minimum improvement can be achieved by 4%. The higher RMSE values observed in cases 5, 6, 7, and 8 can be attributed to the fact that these cases were generated under extreme simulation conditions.

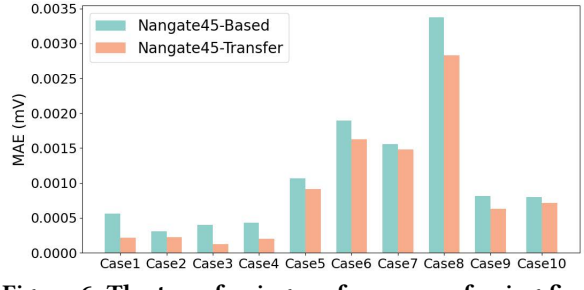

Figure 6: The transferring performance of using five fine-tuning cases on six hidden cases.

5 CONCLUSION

In this paper, a novel U-Net-based power supply noise prediction technique is proposed, which combines multi-scale convolutional blocks, attention mechanism, and U-Net to deliver SOTA performance in IR drop prediction. Furthermore, we introduce a versatile feature extraction method and an effective transfer learning scheme to consistently boost the model performance. The limitation of this work includes validation under more real-world testing cases and large-size datasets.

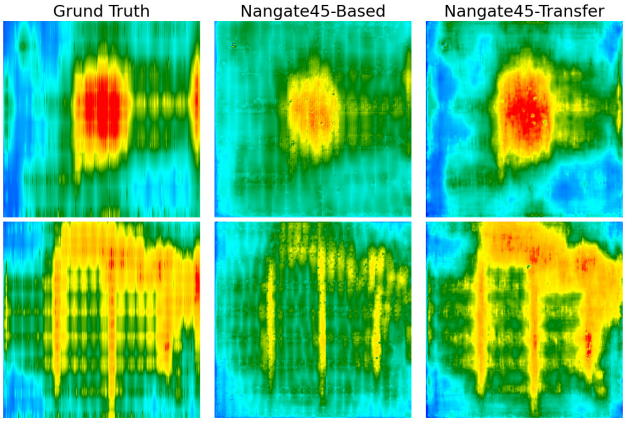

Figure 7: The performance of Nangate45 pre-trained model before and after fintuning. The top and bottom rows show results in Case 2 and 4, respectively.

REFERENCES

[1] H. He and J. Q. Lu. Compact models of voltage drops in power delivery network for tsv-based three-dimensional integration. *IEEE Electron Device Letters*, 34(3):438–440, 2013.
[2] A. Todri, Shih-Chieh Chang, and M. Marek-Sadowska. Electromigration and voltage drop aware power grid optimization for power gated ics. pages 391–394.
[3] Yuan Liang, Wenjian Yu, and Haifeng Qian. A hybrid random walk algorithm for 3-d thermal analysis of integrated circuits. In *2014 19th ASP-DAC*, pages 849–854. IEEE, 2014.
[4] Goong Chen and Jianxin Zhou. *Boundary element methods with applications to nonlinear problems*, volume 7. Springer Science & Business Media, 2010.
[5] Chi-Hsien Pao, An-Yu Su, and Yu-Min Lee. Xgbir: An xgboost-based ir drop predictor for power delivery network. In *2020 DATE*, pages 1307–1310, 2020.
[6] Zhiyao Xie, Haoxing Ren, Brucek Khailany, Ye Sheng, Santosh Santosh, Jiang Hu, and Yiran Chen. Powernet: Transferable dynamic ir drop estimation via maximum convolutional neural network. In *2020 25th ASP-DAC*, pages 13–18. IEEE, 2020.
[7] Eli Chiprout. Fast flip-chip power grid analysis via locality and grid shells. In *IEEE/ACM ICCAD, 2004.*, pages 485–488. IEEE, 2004.
[8] Vidya A Chhabria, Vipul Ahuja, Ashwath Prabhu, Nikhil Patil, Palkesh Jain, and Sachin S Sapatnekar. Thermal and ir drop analysis using convolutional encoder-decoder networks. In *Proceedings of the 26th ASP-DAC*, pages 690–696, 2021.
[9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.
[10] Deep coastal sea elements forecasting using unet-based models. *Knowledge-Based Systems*, 252:109445, 2022.
[11] Run Su, Deyun Zhang, Jinhuai Liu, and Chuandong Cheng. Msu-net: Multi-scale u-net for 2d medical image segmentation. *Frontiers in Genetics*, 12:639930, 2021.
[12] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.
[13] Vidya A. Chhabria, Kishor Kunal, Masoud Zabihi, and Sachin S. Sapatnekar. Began: Power grid benchmark generation using a process-portable gan-based methodology. In *2021 IEEE/ACM ICCAD*, pages 1–8, 2021.
[14] Dongze Lian, Daquan Zhou, Jiashi Feng, and Xinchao Wang. Scaling & shifting your features: A new baseline for efficient model tuning. *Advances in Neural Information Processing Systems*, 35:109–123, 2022.
[15] Vidya A. Chhabria, Kishor Kunal, Masoud Zabihi, and Sachin S. Sapatnekar. Began: Power grid benchmark generation using a process-portable gan-based methodology. In *2021 IEEE/ACM ICCAD*, pages 1–8, 2021.