# UnetPro: Combining Attention with Skip Connection in Unet for Efficient IR Drop Prediction


Zhengfei Qi*
*Super Scientific Software Laboratory*
*China University of Petroleum, Beijing*
Beijing, China
zhengfei.qi@student.cup.edu.cn

Wanchao Wang*
*Super Scientific Software Laboratory*
*China University of Petroleum, Beijing*
Beijing, China
wanchao.wang@student.cup.edu.cn

Chengxuan Yu*
*Super Scientific Software Laboratory*
*China University of Petroleum, Beijing*
Beijing, China
chengxuan.yu@student.cup.edu.cn

Dan Niu
*School of Automation*
*Southeast University*
Nanjing, China
101011786@seu.edu.cn

Xiao Wu
*Huada Empyrean Software Co. Ltd*
Beijing, China
wuxiao@mail.empyrean.com.cn

Zhou Jin
*Super Scientific Software Laboratory*
*China University of Petroleum, Beijing*
Beijing, China
jinzhou@cup.edu.cn



***Abstract*—IR drop analysis plays a key role in chip design. Unlike conventional time-consuming numerical analysis methods, which involve solving large-scale linear circuit equations, predicting the IR drop with machine learning shows great potential to significantly reduce computation time. However, applying machine learning for accurate prediction is non-trivial since achieving effective feature extraction poses significant challenges. Furthermore, as the number of layers in the neural network model increases, the loss of information in the transmission process gradually increases, leading to inaccurate prediction results. In this paper, we propose UnetPro, an innovative machine learning model to resolve these challenges. We leverage an attention mechanism that combines both global and local information and a multi-scale convolution module to make the model sufficiently perceive the various regions of the feature map, enhancing the feature extraction ability of the model. Moreover, we ensure the coherence of information by introducing skip connection. We also introduce the dropout mechanism to ensure the stability of model with information transfer. Compared with conventional Unet model, the error and correlation of our proposed algorithm are lower than it by 2.5e-4 and higher by 10.34%, respectively.**



***Index Terms*—IR Drop, Machine Learning, Attention Mechanism, Skip Connection, Multi-Scale Convolution**


## I. Introduction

Along with the rapid development of integrated circuits, the process nodes shrink, and the metal interconnects line widths become narrower, causing an increase in the resistance per unit length [1]. At the same time, the unit voltage on the chip, located far from the power supply, experiences a significant drop. This drop can lead directly to errors or failures in the unit's function, which is fatal to the design. Therefore, the signoff phase is critical for analyzing the IR drop.

There are two main types of IR drop, static IR drop and dynamic IR drop. Due to its significant impact on power distribution and overall chip reliability, static IR drop analysis attracts a lot of researches [2]. The static IR drop primarily arises from the division of voltage caused by the inherent resistance of the metal connecting lines [3]. So the static IR drop is related to the structure of the power supply network, mainly considering resistive effects. Generally, the IR drop problem ends up solving a system of sparse linear equations using the Modified Nodal Analysis, i.e., Gx=b [4], where G is the conductance matrix, x is the node voltage vector, and b is the excitation vector. However, with increasing system on chip (SoC) integration, the size of the system of linear equations grows to millions or more, incurring significant memory and computational overhead.

In order to reduce the computational overhead, multi-grid method [5] and random walk algorithms [6] are proposed. But there is still a certain computational overhead, especially when the power grids are uneven and irregular. On the other hand, the success of machine learning (ML) in computer vision tasks generates widespread interest in applying artificial intelligence (AI) to prediction problems. The application of ML provides new methods for the IR drop analysis [7] [8]. For example, some previous works extract localized properties of Power Delivery Network (PDN) into XGBoost [9] to predict the IR drop. There are also some works that consider the power transfer noise map as an image and employ a convolutional neural network (CNN)-based [10] [11] [12] strategy for IR drop prediction. Although the above work outperform traditional linear solver in terms of solution speed, their prediction accuracy and stability are issues of concern. Specifically, convolutional kernel selection is never easy, resulting in significant challenges in feature extraction. Smaller windows may violate the principle of locality, and larger windows may require a large amount of training time. Moreover, as the number of layers of the model network increases, the loss of information in the transmission process will gradually increase, which has an impact on the prediction accuracy.

To overcome the above challenges, we propose UnetPro, a new effective static voltage drop machine learning prediction

*Both authors contributed equally to this research.

model for SoC power networks. Our main contributions are as follows:

- We introduce multi-scale convolution and attention mechanism, which acquires multi-scale information about the target in different layers of the network. This enhances the feature extraction function of the model.
- Our model integrates skip connection and dropout mechanism to reduce loss of information transmission and provide improved prediction performance.
- Our proposed method improves the prediction performance with an average absolute error of about 1e-3 and a correlation coefficient higher than 85%. Compared with Unet, the error and correlation of our model are lower than it by 2.5e-4 and higher by 10.34% respectively.

The rest of this paper is organised as follows. Section II presents the background of IR drop analysis. Section III presents the details of our model. Section IV shows the experimental results. Finally, Section V concludes the paper.

## II. BACKGROUND

### A. IR Drop

IR drop is the voltage reduction in an integrated circuit caused by current flowing through wires and resistive elements. This IR drop can cause different parts of the circuit to operate at different voltage levels, thus affecting the performance of the circuit.

For full-chip IR drop analysis, commercial software typically abstracts the physical design into a mathematical model and then solves for a large system of linear sparse equations. However this process often requires a significant amount of computational time. In recent years, the application of machine learning in circuit design has gradually emerged to provide innovative solutions for IR Drop prediction. By learning a large amount of circuit design data and actual performance information [13], machine learning models are able to tap into the underlying complex relationships to achieve accurate prediction of circuit IR Drop.

An example of this is the conversion of a power network layout into a meaningful representation of image features, as shown in Figure 1. With a convolutional neural network, a static IR drop distribution map can be generated to provide designers with an intuitive reference.

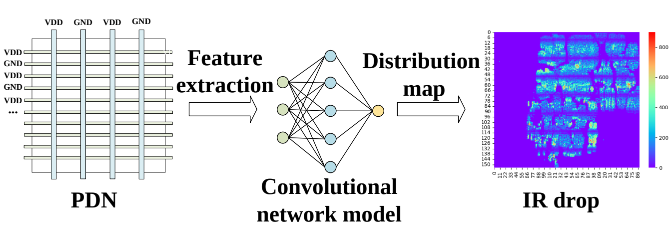



Fig. 1. Features are extracted from the circuit layout and IR Drop distribution maps are generated by convolutional network.

### B. Unet Model

The Unet model [14] is proposed by Olaf Ronneberger, Philipp Fischer, and Thomas Brox in 2015. It is initially designed to address medical image segmentation problems, such as segmenting organs like the lungs and liver. The introduction of Unet aims to improve the performance of image segmentation through a specific encoder-decoder structure and skip connection.

The Unet model uses an encoding-decoding structure. The encoder gradually extracts features and reduces the spatial dimensions of the input image through convolutional and pooling layers. The intermediate layer captures global information through deep convolutional layers. The decoder consists of convolutional and upsampling layers that gradually increase the spatial dimension of the feature map to reconstruct the image. The key skip connection mechanism connects the convolutional processing result of a layer in the encoder directly to the corresponding layer in the decoder, which helps to retain more detailed information. The model structure of Unet is shown in Figure 2.

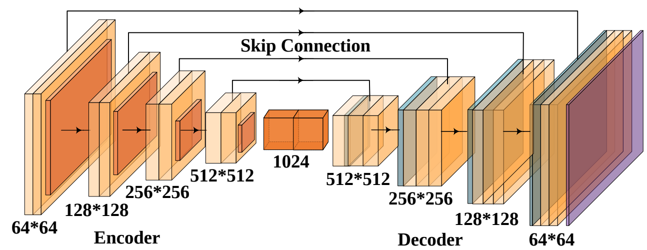



Fig. 2. The Unet model consists of an encoder and a decoder that passes information through skip connections. The encoder extracts high-level features, the decoder restores the image segmentation results, and the skip connections help to preserve detailed information.

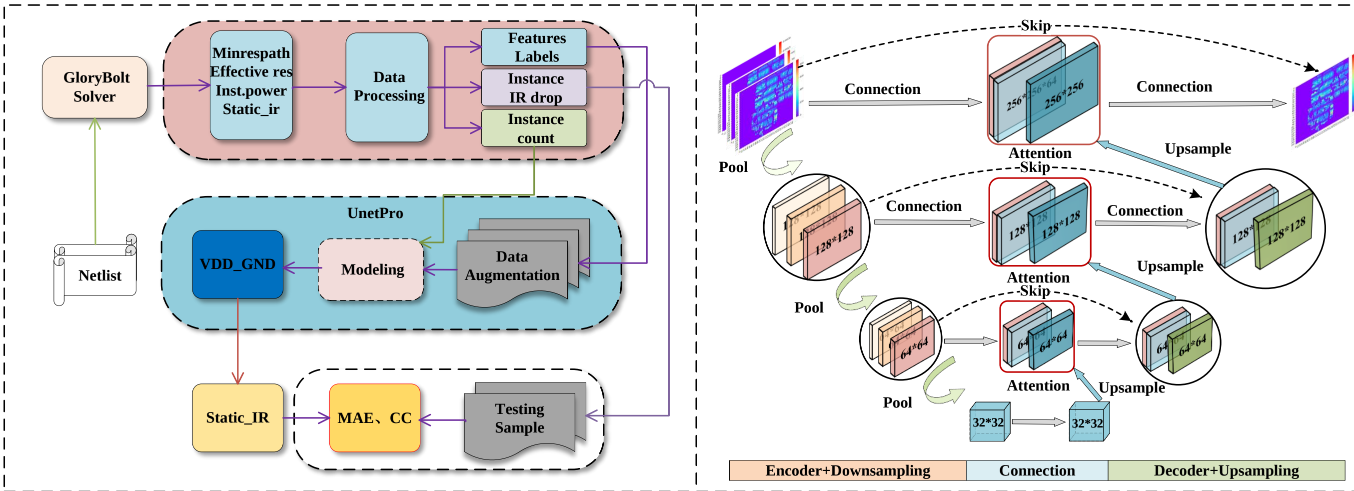
The strength of the Unet model lies in its unique encoding-decoding structure and skip connection mechanism, which helps to capture both local and global information efficiently. Its relatively simple implementation and superior performance make it one of the classic models in the field of image segmentation. However, Unet also has some shortcomings, including relatively low efficiency in processing large-size images, which may lead to information loss, as well as a possible overfitting problem for small-size datasets.

### C. Attention Mechanism

Attention mechanism is a widely used mechanism in deep learning, the core idea of which is to enable the model to focus more on specific spatial regions or channels when processing input data by learning dynamic weights [15]. It usually includes channel attention, which strengthens the model's ability to perceive important channels by weighting feature channels, and spatial attention, which enables the model to focus more on key regions in the image by weighting the spatial location of the input.

### D. Skip Connection

Skip connection is a common technique in neural network architectures designed to facilitate information flow and gra-




**(a) Overall flow chart** **(b) UnetPro model chart**

Fig. 3. (a) Shows the entire program design flowchart. (b) shows the model structure of UnetPro, which is based on the Unet structure and adopts the overall encoder-decoder architecture, incorporating structures such as the attention mechanism, skip connections and multi-scale convolution.

dient propagation [16]. Bypassing a number of intermediate layers by using the output of one of the upper layers of the network as the input to one of the lower layers, this form of upper and lower layer connectivity allows for better information communication to be established between the different layers, thus preserving and utilizing the detailed information in the original inputs more efficiently.

## III. Proposed Algorithm

### A. UnetPro Framework

In this paper, we propose a new machine learning model, UnetPro, to resolve the feature extraction and prediciton accuracy challenges in the IR drop prediction. The entire framework is shown in Figure 3(a). UnetPro is an image segmentation algorithm based on the Unet structure, which utilizes the overall encoder-decoder architecture. The Encoder succeeds in reducing the image dimensionality by stacking convolutional and maximal pooling layers, while the Decoder progressively reduces the resolution through the inverse convolutional layers and skip connection, and introduces an attention module to achieve accurate capture of global and local information. The UnetPro model has three input features: total power (the total power consumption of the instance), vdd-r (the equivalent resistance value of the instance connected to the powernet), and gnd-r (the equivalent resistance value of the instance connected to the ground net). The model structure of UnetPro is shown in Figure 3(b).

Specifically, to further enhance the model performance, we introduce an attention mechanism in the decoder part to highlight the key feature regions. At the same time, a multi-scale convolution module is introduced in the encoder to capture image features at different scales more comprehensively. In the decoder, we introduce a more complex skip connection mechanism due to the limitations of information transfer that may result from the simple splicing operation of the skip connection mechanism in Unet. Our skip connection incorporates an attention mechanism that allows the model to dynamically adjust the joining weights according to the importance of features at different levels, thus capturing the key information in the image more efficiently. This overall design aims to address the challenges of more fully and comprehensively utilizing the semantic information of the feature graph, thereby significantly improving the model performance.

### B. Data Pre-processing

Due to the differences in the number of samples of different categories or eigenvalues in the dataset, it may lead to the bias of the model towards certain categories or eigenvalues, which in turn affects the model performance and stability. In order to solve this problem, we adopt the method of coordinate mapping and region categorization for data preprocessing. Specifically, by mapping the component information to the corresponding regions, we record the maximum VDD drop and GND bounce in each region in detail, along with comprehensive statistics, including the number of components, total power consumption, and effective resistance data. In this way, we are able to better deal with the data imbalance problem, while extracting key features and preserving spatial information, thus effectively improving the generalization ability and performance of the model.

In order to improve the model performance and generalization ability, we introduce data augmentation techniques, which use three main data augmentation operations, i.e., Flip, Rotation, and Crop, to increase the diversity of the training data, so that the model can better adapt to different input conditions and improve its robustness. In the Flip operation, a decision is made on whether to flip the image horizontally or vertically based on a set random probability (default is 0.5). The Rotation operation also determines whether to rotate or not with a random probability, randomly selecting the rotation angle, which can be between 0, 90, 180 and 270 degrees. Finally, the Random Crop operation determines whether to crop by a set probability (default 0.5), first scaling the image to a specified size (default 384), and then randomly selecting a region in the image for cropping to obtain an image of the target size (default 256).

To address the performance challenges of small sample data types and to fully utilize all available data, we take an approach similar to rolling learning. In this approach, the entire dataset is divided into multiple training rounds, and the test set of each round is incorporated into the training set of the next round. The details of how this works are shown in Figure 4.

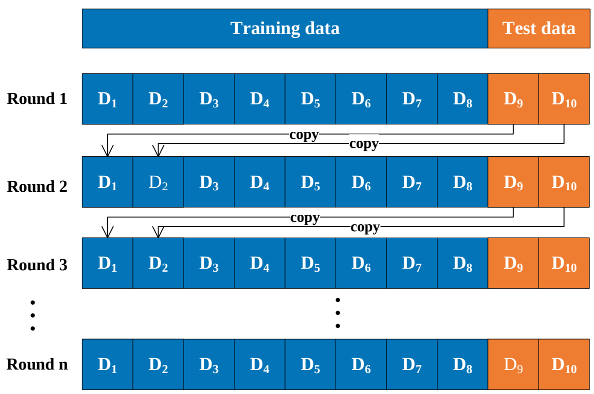



Fig. 4. The entire dataset is divided into multiple training rounds, and the test set from each round is incorporated into the training set for the next round.

Finally, we note that label dimensionality may have a potential impact on model training speed and prediction performance. Therefore, we adopt a label dimensionality reduction strategy to downscale the final prediction to one dimension by summing the values of VDD and GND, thus improving the training speed and reducing the computational burden while maintaining the prediction performance.

### C. Attention Mechanism

In order to focus more efficiently on important regions in an image, we introduce an attention module that combines global and local information fusion mechanisms. The module is designed to compute a set of attention weights to weight the feature maps of the encoder and decoder inputs through a combination of global and local feature transforms and ReLU and Sigmoid activation functions. This process allows the model to focus more on task-relevant information, improving the perception of critical regions in the image.

In particular, in our model, first, the module receives the feature map output from the encoder and the feature map input from the decoder, including global and local feature information. Next, the input feature maps are transformed by global and local feature transformations to extract global and local association information. Subsequently, a set of attentional weights are generated by the computation of ReLU and Sigmoid activation functions. These weights are used to weight the feature maps of the encoder and decoder. Finally, the computed attention weights are applied to the feature maps of the encoder and decoder to obtain a feature representation that pays more attention to the important regions.

### D. Skip Connection

The traditional Unet model employs a skip connection mechanism, the core idea of which is to connect the feature maps of the encoder with the feature maps of the corresponding layers of the decoder in order to provide more high-level semantic information. However, this mechanism may lead to specific problems in practice, such as information bottleneck and gradient vanishing. To address these issues, we propose an improved skip connection mechanism.

In our model, an improved skip connection mechanism is employed to solve the problem of information transfer limitations that may result from simple splicing operations in traditional Unet. In the process of skip connection, we first utilize the attention mechanism to regulate the features between the encoder and decoder. Specifically, we use the attention mechanism to dynamically adjust the connection weights so that the model can pay more attention to important feature information while suppressing irrelevant information. This skip connection mechanism combined with the attention mechanism enables the model to utilize the feature interactions between the encoder and decoder more flexibly, thus improving the accuracy and robustness of the model for the image segmentation task.

### E. Multi-scale Convolution Module

In order to further enhance the feature extraction performance of the model, we use multi-scale convolution module with different sizes of convolution kernels on the data for extraction. It consists of multiple parallel convolutional layers, each of which captures different scale features of the input data. After the capture is completed, the output results of these different scale convolutional layers are fused and output.

The principle of multi-scale convolution module can be explained by the multi-scale information in the visual scene. For the feature maps processed by our model, different scales of the extracted field of view are important for understanding the size, shape, and texture of the feature map content. So by using convolution kernels of different sizes, multi-scale convolution module capture local details while preserving global contextual information to a large extent. Its structure is shown in Figure 5.

## IV. Numerical Examples

### A. Experimental Setup

We use PyCharm to implement our work in Python. All experiments are conducted on an Intel Xeon dual-CPU server with a main frequency of 2.6GHz. The server is equipped with 256GB of RAM, and the graphics card model is NVIDIA 3060.

Our dataset covers five different categories, namely, Nvdla-small, RISCY, RISCY-FPU, Zero-riscy, and Vortex-small. Each of these categories contains four key data files. In the Min Path Res file, the feature information includes the minimum resistance value, winding name, and cell instance name for each instance. In the Effective Res file, the feature information includes the equivalent resistance values connected to the

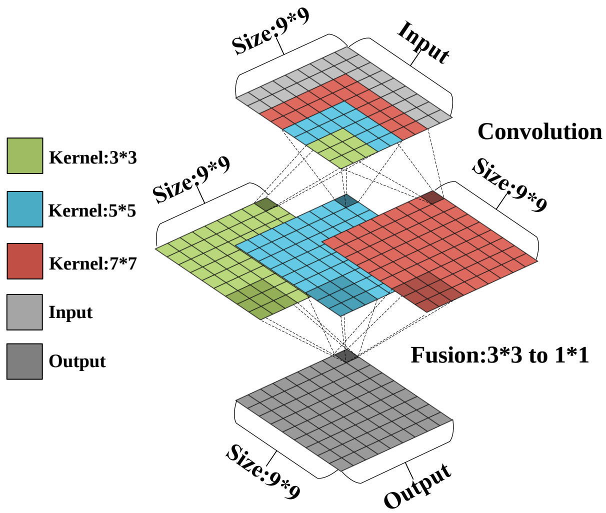



Fig. 5. Three different scales of features are captured for the input data and the results are fused for output.

TABLE I
PARAMETER SETTINGS IN UNETPRO.

| *Parameters* | | | Settings |
|---|---|---|---|
| Model Parameters | Conv1 | filter size | 128*128 |
| | | filters | 64 |
| | Conv2 | filter size | 64*64 |
| | | filters | 128 |
| | Conv3 | filter size | 32*32 |
| | | filters | 256 |
| Training Parameters | Batch-size | | 10 |
| | Max-iters | | 5000 |
| | In-channels | | 3 |
| | Out-channels | | 1 |
| | Learning rate | | 0.005 |
| | Weight-decay | | 0 |
| | Loss function | | L1+L2 |
| | GPU | | 1 |

power net and ground net, the coordinates and names of the instances on the power net and ground net, and so on. The power consumption report file contains features such as operating frequency, flip-flop rate, leakage power consumption, flip-flop power consumption, etc. of the instance. And the static IR drop report file contains feature information such as VDD drop and ground bounce, and so on.

Table 1 illustrates the hyperparameter settings of the model.In the encoder, the number of filters in each layer starts at 64 and gradually increases by a factor of 2. In the decoder, the highest layer starts with 256 filters and the number of filters is reduced by half in each subsequent layer. Additionally, in the decoder, the kernel size for each layer is set to 3. If the number of filters in a layer exceeds 128, dropout is applied with a dropout rate of 0.2. To ensure effective learning of the data, the model is trained in batches with a batch size of 10. The training process consists of 500 epochs, with a learning rate of 0.005 and a weight decay of 0.

In order to be able to better focus on the prediction accuracy of our model, we use the correlation coefficient (CC), the mean absolute error (MAE), and the mean absolute percentage error (MAPE) as measures of accuracy.

The correlation coefficient is defined as follows, $y_i$ is the predicted value of pressure drop, $\hat{y}_i$ is the Golden pressure drop, and $n$ is the number of data points.

$$CC = \frac{\sum_{i=1}^{n} [y_i - \text{mean}(y)] [\hat{y}_i - \text{mean}(\hat{y})]}{\sqrt{\sum_{i=1}^{n} [y_i - \text{mean}(y)]^2 \sum_{i=1}^{n} [\hat{y}_i - \text{mean}(\hat{y})]^2}} \quad (1)$$

The mean absolute error is a metric used to measure the prediction error of the model and is calculated as follows.

$$MAE = \frac{1}{n} \sum_{1}^{n} |\hat{y}i - yi| \quad (2)$$

The value of MAPE is expressed as a percentage and represents the magnitude of the mean relative error. The calculation formula is shown below.

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}i - yi}{\hat{y}i} \right| \times 100 \quad (3)$$

## B. Model Comparison

We conduct sufficient performance tests on our model and datasets, including testing the same model on different types of data and testing the same data on different types of models, using performance metrics such as mean absolute error (MAE), correlation coefficient (CC), and mean absolute percentage error (MAPE). Table 2 shows the performance test results of our final UnetPro model and the ordinary Unet model on four different datasets, where MAE and CC are used as the performance metrics, and the comparison clearly shows that UnetPro has a 10% improvement in CC compared with Unet and keeps the MAE around 0.0015.

The UnetPro split can be divided into three parts or phases, which are also the three main breakthrough points in our overall experiment, representing the new modules being added to our model. For the performance impact of this series of changes, we conduct comparative tests on each of the four datasets using the same amount of data. Figure 6 and Figure 7 show the performance of our model at each of the three phases and compare it with Unet, it can be seen that our incremental improvements reduce the model's prediction error and improve the prediction accuracy. Each modular addition to the model makes significant improvements to CC while ensuring that MAE is at a lower level.

Through the comparison, we can see that the skip connection is the core optimization point of the model. Furthermore, the prediction performance of the model is increased by about 10% through the introduction of the attention mechanism and multi-scale convolution module.

In addition, we also take an additional 10 cases from the data and make further comparison tests between the Unet

TABLE II
UNET AND UNETPRO TEST RESULTS FOR EACH METRIC UNDER THE FOUR CATEGORY DATASETS.

| Model | Zero-riscy | | RISCY | | RISCY-FPU | | Nvdla-small | |
|---|---|---|---|---|---|---|---|---|
| | MAE | CC | MAE | CC | MAE | CC | MAE | CC |
| Unet [14] | 0.0016 | 0.7953 | 0.0013 | 0.7369 | 0.0014 | 0.7459 | 0.0028 | 0.7219 |
| UnetPro | 0.0017 | 0.8786 | 0.0011 | 0.8830 | 0.0012 | 0.8345 | 0.0021 | 0.8178 |

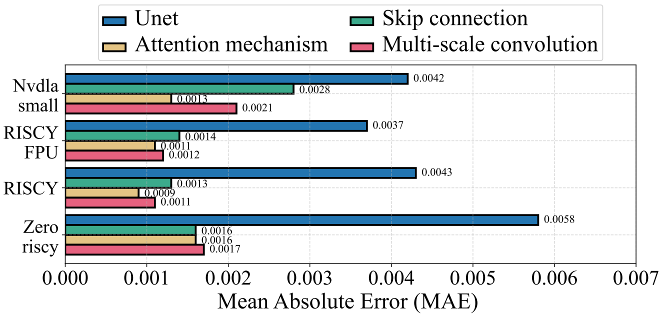



Fig. 6. A demonstration of the performance of the four-stage model on the MAE metric.

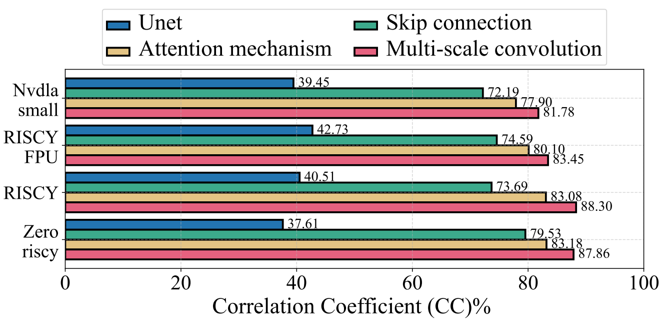



Fig. 7. A demonstration of the performance of the four stages of the model on the CC metrics.

model and the UnetPro model using the MAPE metric on them, as shown in Figure 8. The figure shows that the UnetPro generally decreases in MAPE by about 20% compared to the Unet, which indicates that the UnetPro is significantly smaller than the Unet in terms of the prediction error and keeps the MAPE below 40%.

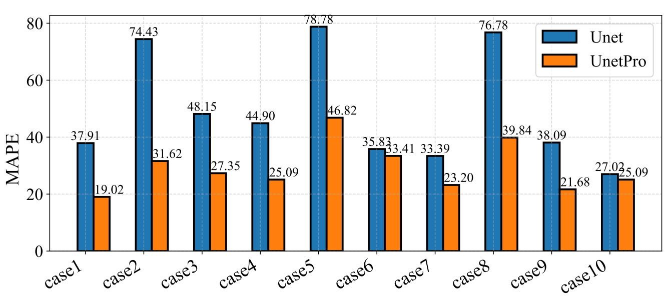



Fig. 8. Performance of Unet and UnetPro under the MAPE metric.

## V. CONCLUSION

In this paper, we introduce UnetPro as an IR drop prediction model. The prediction accuracy and stability of the model are significantly improved by combining structures such as attention mechanism, skip connection, and multi-scale convolution. Data enhancement further optimizes the model performance. Experiments demonstrate that the average absolute error of UnetPro is lower than 1e-3 and the correlation coefficient is higher than 85%. Compared with traditional simulation methods, UnetPro improves the prediction performance with more efficient speed and provides a reliable advanced solution for circuit design.


## ACKNOWLEDGMENT

We are deeply grateful to the reviewers for their valuable comments. Zhou Jin is the corresponding author of this paper. This work was supported by the Natural Science Foundation of China (Grant Nos. 62204265, 62234010, 62374031) and the National Natural Regional Key Fund (Grant No. U23A20301).


## REFERENCES


[1] C. K. Sarkar. *Technology computer aided design: simulation for VLSI MOSFET*. CRC Press, 2013.
[2] H. He and J.-Q. Lu. Compact models of voltage drops in power delivery network for tsv-based three-dimensional integration. *IEEE EDL*, 2013.
[3] Z. Xie, H. Li, X. Xu, J. Hu, and Y. Chen. Fast ir drop estimation with machine learning. In *ICCAD '20*, 2020.
[4] C. Zhuo, J. Hu, M. Zhao, and K. Chen. Power grid analysis and optimization using algebraic multigrid. *IEEE TC*, 2008.
[5] A. Todri, S.-C. Chang, and M. Marek-Sadowska. Electromigration and voltage drop aware power grid optimization for power gated ics. In ISLPED '07,2007.
[6] Y. Liang, W. Yu, and H. Qian. A hybrid random walk algorithm for 3-d thermal analysis of integrated circuits. In *ASP-DAC '14*, 2014.
[7] Y. Chen, H. Pei, X. Dong, Z. Jin, and C. Zhuo. Application of deep learning in back-end simulation: Challenges and opportunities. In *ASP-DAC '22*, 2022.
[8] V. A. Chhabria, V. Ahuja, A. Prabhu, N. Patil, P. Jain, and S. S. Sapatnekar. Thermal and ir drop analysis using convolutional encoder-decoder networks. In *ASPDAC '21*, 2021.
[9] C.-H. Pao, A.-Y. Su, and Y.-M. Lee. Xgbir: An xgboost-based ir drop predictor for power delivery network. In *DATE '20*, 2020.
[10] C.-T. Ho and A. B. Kahng. Incpird: Fast learning-based prediction of incremental ir drop. In *ICCAD '19*, 2019.
[11] Z. Xie, H. Ren, B. Khailany, Y. Sheng, S. Santosh, J. Hu, and Y. Chen. Powernet: Transferable dynamic ir drop estimation via maximum convolutional neural network. In *ASP-DAC '20*, 2020.
[12] X. X. Huang, H. C. Chen, S. W. Wang, I. H. R. Jiang, Y. C. Chou, and C. H. Tsai. Dynamic ir-drop eco optimization by cell movement with current waveform staggering and machine learning guidance. In *ICCAD '20*, 2020.
[13] S. Kundu, M. Prasad, S. Nishad, S. Nachireddy, and H. K. MLIR: machine learning based IR drop prediction on ECO revised design for faster convergence. IEEE VLSID, 2022.
[14] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI '15*, 2015.
[15] H. Hu, Z. Li, and W. Gao. Mpcnet: Improved meshsegnet based on position encoding and channel attention. *IEEE Access*, 2023.
[16] Z. Yu, L. Yu, W. Zheng, and S. Wang. Eiu-net: Enhanced feature extraction and improved skip connections in u-net for skin lesion segmentation. *CBM*, 2023.