GLSVLSI'24
Great Lake Symposium on VLSI

PGAU: Static IR Drop Analysis for Power Grid using Attention U-Net Architecture and Label Distribution Smoothing

Feng Guo[+], Jiawei Liu[+], Jianwang Zhai*, Jingyu Jia, Kang Zhao, Chuan Shi*

School of Computer Science (National Pilot Software Engineering School),

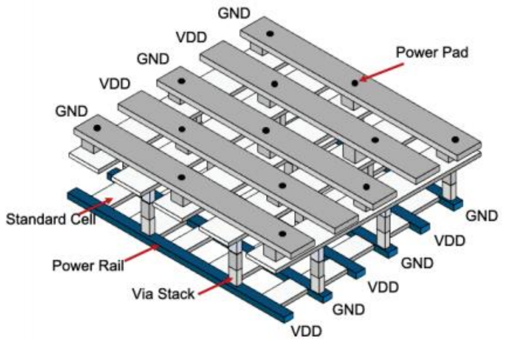
School of Integrated Circuits,

Beijing University of Posts and Telecommunications, China

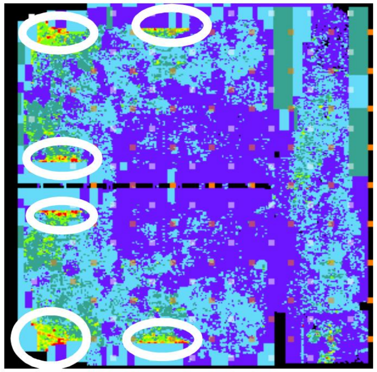
Contents

Introduction
IR drop problem

- **Reliability problems** due to the worst-case IR drop beyond limit
- **Time consuming problems** due to industrial-scale designs
- **IR drop distribution problems** due to the localization principle



An example of power grid (PG) [1].



IR drop violations cluster in local regions (red spots) [2].

[1] Dynamic IR-drop ECO optimization by cell movement with current waveform staggering and machine learning guidance[C]. ICCAD 2020.
[2] Machine-learning-based dynamic IR drop prediction for ECO[C]. ICCAD 2018.

Introduction
Related work

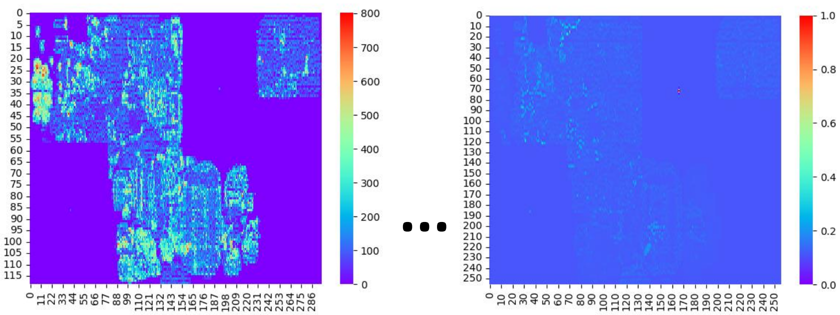
Previous methods

- **XGBIR [3]**: An instance-level predictor without spatial information
- **PowerNet [4]**: Not be suitable for designs with irregular grid densities
- **IREDGe [5]**: A U-Net-based network with limited accuracy
- **MAVIREC [6]**: Mainly for the dynamic IR drop problems

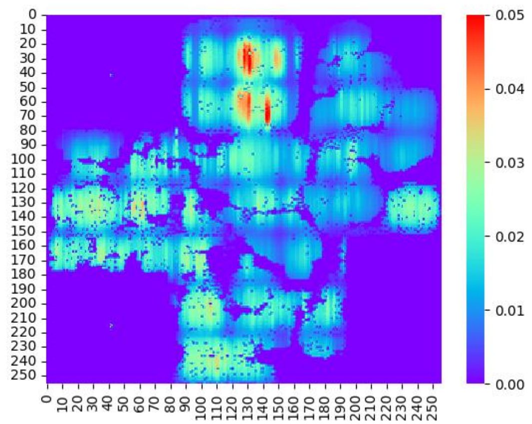
[3] XGBIR: An XGBoost-based IR Drop Predictor for Power Delivery Network. DATE 2020.
[4] PowerNet: Transferable dynamic IR drop estimation via maximum convolutional neural network. ASP-DAC 2020.
[5] Thermal and IR Drop Analysis Using Convolutional Encoder Decoder Networks. ASP-DAC 2021.
[6] MAVIREC: ML-aided vectored IR-drop estimation and classification. DATE 2021.

Introduction
Two key issues

- **Feature Confusion reflected in diversity and complexity of features**
- **Imbalanced Label Distribution showed large differences in IR drop distribution**



An example of feature confusion.



An example of IR drop distribution in PG, exhibiting locally consistent and overall imbalanced characteristics.

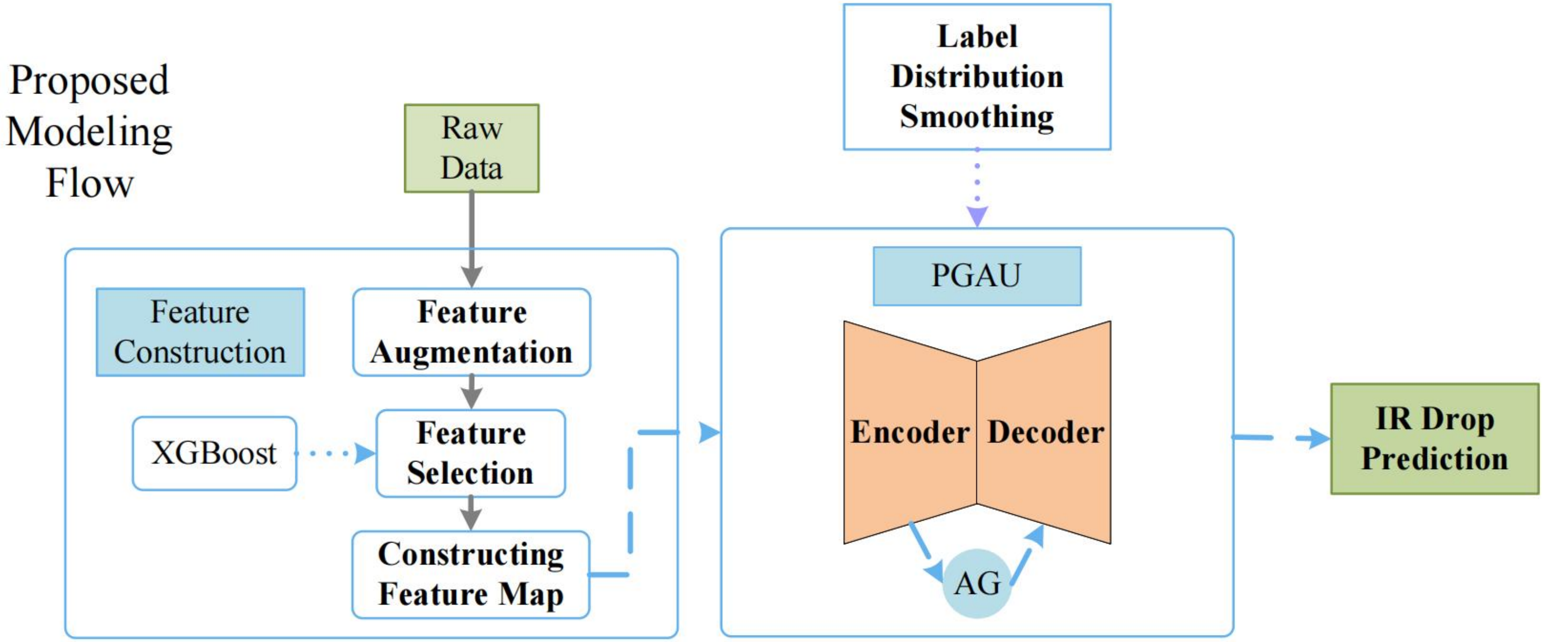
Preliminaries
Problem formulation

- Each feature map, denoted as P_{map_i} corresponds to a $w \times h$ data matrix
- The target ML model F tries to give the closest prediction F^* on y based on all n different feature maps $\{P_{map_i} \ldots\ldots P_{map_n}\}$

$$F : \left\{ P_{\mathrm{map}_1} \in \mathrm{R}^{w \times h}, \ldots, P_{\mathrm{map}_n} \in \mathrm{R}^{w \times h} \right\} \rightarrow \mathrm{R}^{w \times h}, \qquad (1)$$

$$F^* = \arg\min_{F} \mathrm{Loss} \left(F \left(\left\{ P_{\mathrm{map}_1}, \ldots, P_{\mathrm{map}_n} \right\} \right), y \right), \qquad (2)$$

Methodologies
Overall



The framework of our method.

Methodologies
Feature Augmentation

Table: Raw features of PG.

Feature	Symbol
The cell type to which the instance belongs.	C_{inst}
Instance's equivalent resistance to all power pads.	R_{vdd}
Instance's equivalent resistance to all ground pads.	R_{gnd}
Instance's loop resistance, i.e., $R_{vdd} + R_{gnd}$	R_{loop}
Instance's minimum resistance to nearest power pad.	mR_{vdd}
Instance's minimum resistance to nearest ground pad.	mR_{vss}
Instance's operating frequency.	$freq$
Instance's toggle rate per unit cycle.	r_{tog}
Instance's leakage power.	$P_{leakage}$
Instance's switching power.	$P_{switching}$
Instance's internal power.	$P_{internal}$
Instance's total power.	P_{all}
The ideal supply voltage of power net.	$P/G - volt$
The power net to which the instance is connected.	$P/G - domain$
Power pin.	$P/G - pin$

Methodologies
Feature Augmentation

- **Composite power:** $\boldsymbol{P_{composite}}$

$$P_{composite} = \left(P_{internal} + P_{switching}\right) r_{tog} + P_{leakage}, \qquad (3)$$

- **Nearby instance features [4]:** $\boldsymbol{Count_{neighbour}}$, $\boldsymbol{P_{neighbour}}$

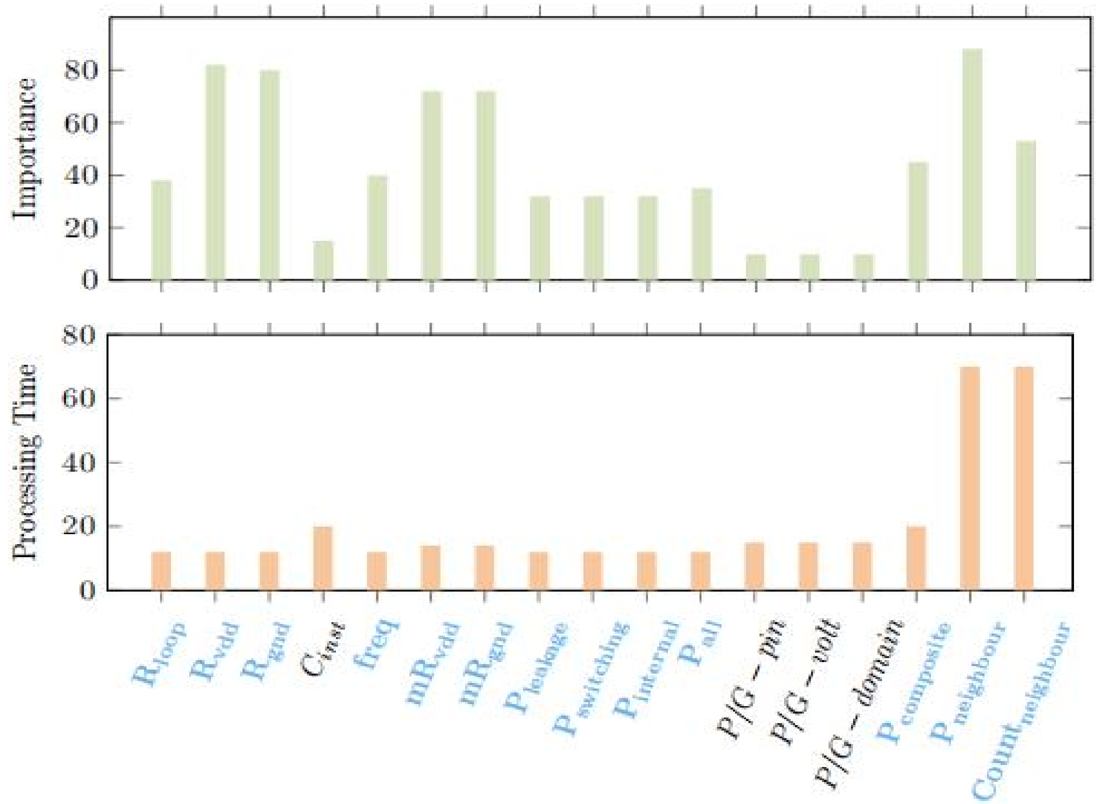
$$P_{neighbor} = \frac{1}{n}\sum_{i}^{n}(P_{all})_i, \qquad (4)$$

---

[4] PowerNet: Transferable dynamic IR drop estimation via maximum convolutional neural network. ASP-DAC 2020.

Methodologies
Feature Selection

- **Using XGBoost [7]**



Results of feature analysis and selection.

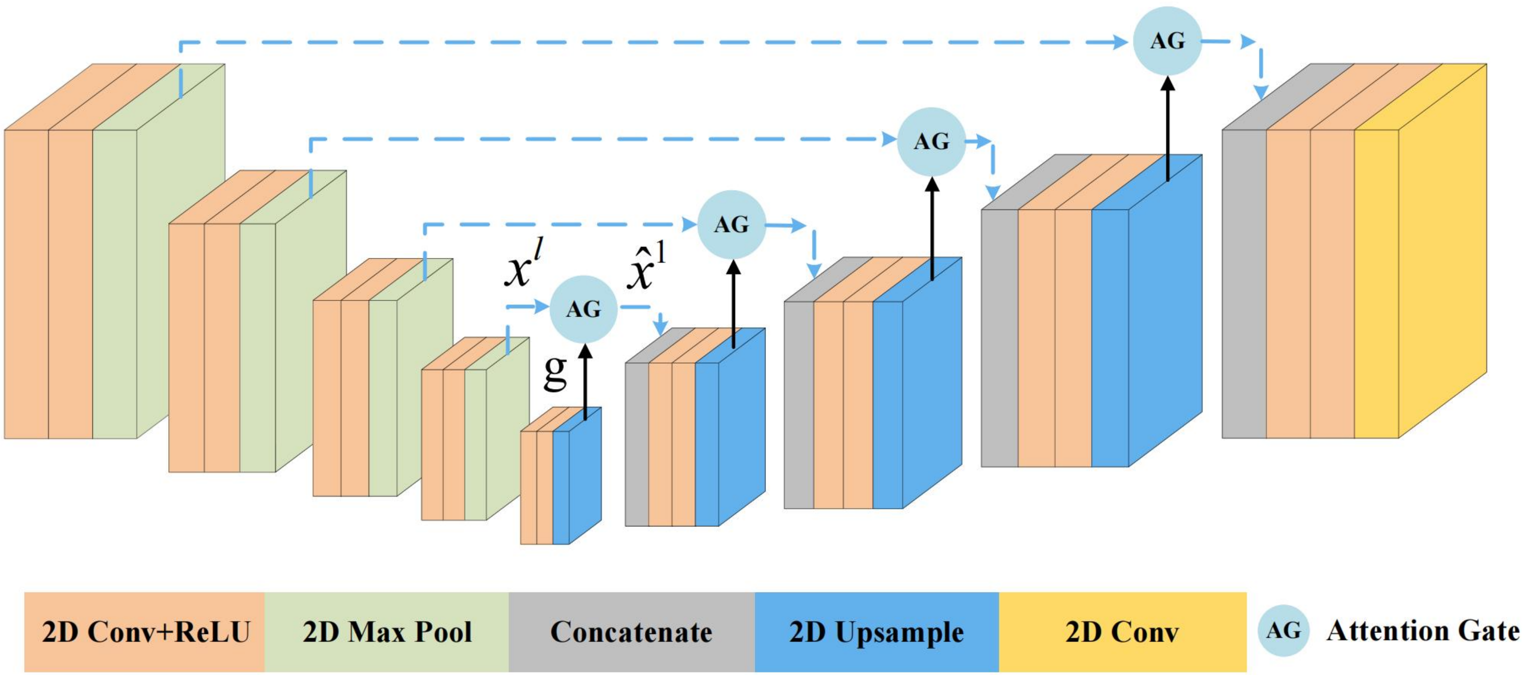
[7] XGBoost: A Scalable Tree Boosting System. KDD 2016.

Methodologies

Constructing 2D Spatial Feature Maps

- Based on the row height from Library Exchange Format (LEF), w and l, a design of size $W_c \times L_c$ translates to an image of W (= $W_c // w$) $\times L$ (= $L_c // l$) pixels.
- The coordinates of each node x_n and y_n will be translated to $x = x_n // w$ and $y = y_n // l$.

Methodologies
Encoder-Decoder Architecture with Attention Mechanism



Architecture of our PGAU model.

Methodologies

Encoder-Decoder Architecture with Attention Mechanism

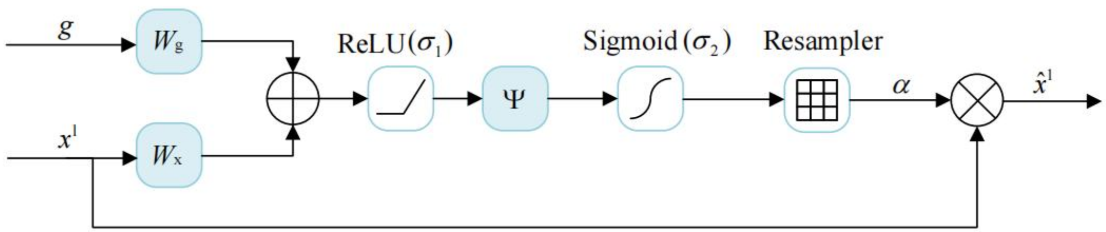


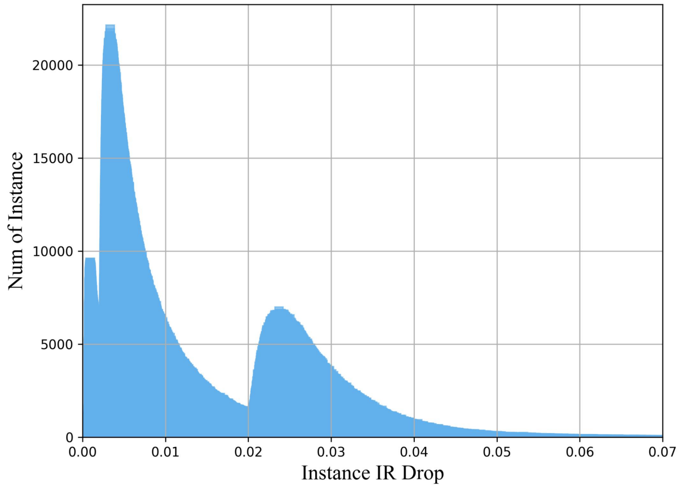
Illustration of the additive attention gate (AG).

- **AG is formalized as:** $$q_{att}^{l} = \psi^{T}\left(\sigma_1\left(W_x^T x^l + W_g^T g + b_g\right)\right) + b_{\psi}, \quad (5)$$

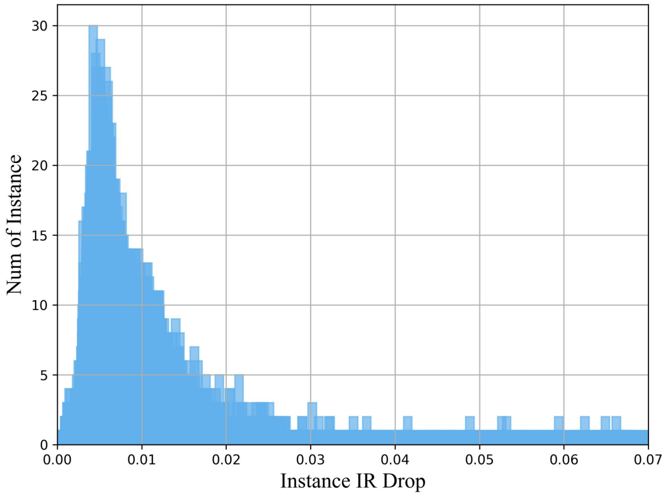
- **Attention coefficient is formalized as:** $$\alpha^{l} = \sigma_2\left(q_{\mathrm{att}}^{l}\left(x^{l}, g; \Theta_{\mathrm{att}}\right)\right), \quad (6)$$

- **Final feature map is formalized as:** $$\widehat{x}^{l} = \alpha^{l} x^{l}, \quad (7)$$

Methodologies
Loss Function with Label Distribution Smoothing



(a) Single PG



(a) Multiple PGs

The imbalanced distribution of IR drop.

Methodologies
Loss Function with Label Distribution Smoothing

- **The LDS calculation [8][9]:**

$$\tilde{p}\left(y'\right) = \int_{y} k\left(y, y'\right) p(y) dy, \tag{8}$$

- **Re-weighted loss function:**

$$\text{loss}\left(\hat{y}, y'\right) = \sum_{i=1}^{n} \frac{1}{\sqrt{\tilde{p}\left(y_i'\right)}} \left|\left(\hat{y}_i - y_i'\right)\right|. \tag{9}$$

[8] Yuzhe Yang, Kaiwen Zha, Yingcong Chen, Hao Wang, and Dina Katabi. Delving into Deep Imbalanced Regression. ICML 2021.
[9] Jialv Zou, Xinggang Wang, Jiahao Guo, Wenyu Liu, Qian Zhang, and Chang Huang. Circuit as Set of Points. NeurIPS 2023.

Evaluation
Experimental Settings

Datasets

- CircuitNet[10]: 7322 samples
- ICCAD2023[1]: 100 real samples; 20 fake samples

Table: Statistics of CircuitNet dataset.

Front-end Design	RISCY-FPU	zero-riscy	RISCY	nvdla-small	Vortex-small
# PGs per design	2400	2400	2400	56	66
# Nodes per PG	> 30000	> 50000	> 50000	> 600000	> 600000
Mean of IR drop	0.0028	0.0029	0.0034	0.0074	0.0076
Variance of IR drop	0.0003	0.0003	0.0004	0.0006	0.0006

[10] CircuitNet: an open-source dataset for machine learning applications in electronic design automation (EDA)[J]. Information Sciences, 2022, 65(227401): 1-227401.
1 https://iccad-contest.org/

Evaluation
Experimental Settings

Baselines

- Published IR Drop Prediction Baselines:
  - XGBIR(DATE 2020) [3]
  - IREDGe(ASPDAC 2021) [4]
  - PowerNet(ASPDAC2020) [5]
  - MAVIREC(DATE 2021) [6]
- Competitive Image Segmentation Models:
  - U-Net(CVPR2015) [11]
  - U-Net++(IEEE TMI 2019) [12]
  - SETR(CVPR 2021) [13]
  - MobileNetV3(ICCV 2019) [14]
  - LR-ASPP(ICCV 2019) [14]

---

[11] Fully convolutional networks for semantic segmentation. CVPR 2015.
[12] Unet++: Redesigning skip connections to exploit multiscale features in image segmentation[J]. IEEE TMI 2019.
[13] Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers[C]. CVPR 2021.
[14] Searching for mobilenetv3[C]. ICCV 2019.

Evaluation
Experimental Settings

Metrics

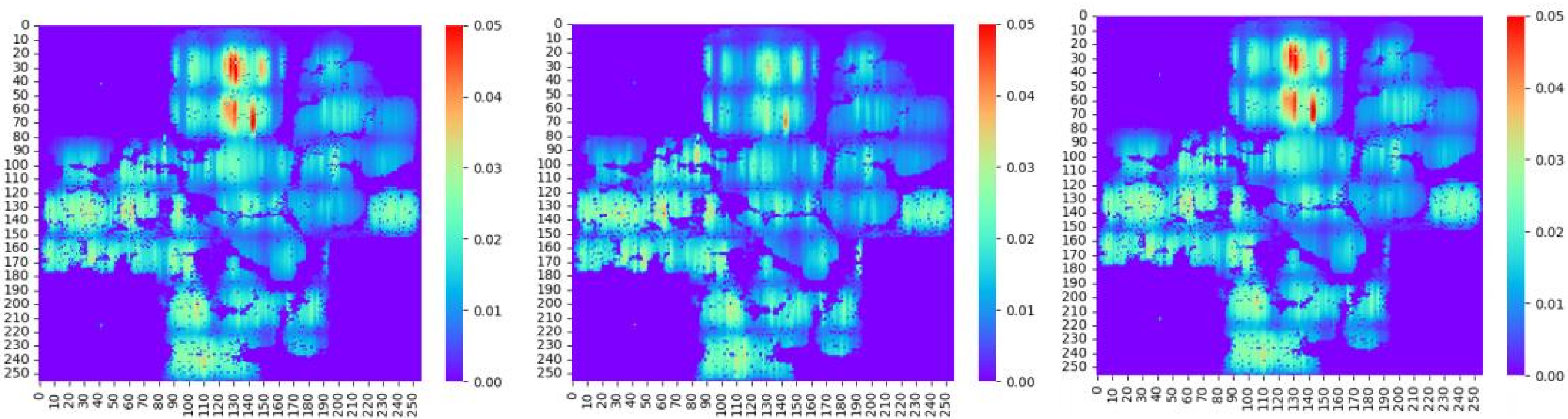
- Mean absolute error (MAE)
- Pearson correlation coefficient (CC)
- Maximum IR drop error (MIRDE)
- runtime

Evaluation
Main Results

Table: Main results on CircuitNet and ICCAD2023 datasets

Method	CircuitNet dataset				ICCAD2023 dataset			
	MAE	CC	MIRDE	Runtime	MAE	CC	MIRDE	Runtime
U-Net	0.0075	0.8704	0.0271	97s	0.3105	0.8413	0.8866	66s
U-Net++	0.0074	0.8736	0.0271	101s	0.3304	0.853	0.8195	71s
SETR	0.0275	0.7239	0.0829	120s	1.1526	0.7694	2.2441	87s
MobileNetV3	0.0266	0.7252	0.0734	117s	1.0786	0.7712	2.3087	87s
LR-ASPP	0.0258	0.7287	0.0717	117s	1.0025	0.7626	2.1358	85s
XGBIR	0.0016	0.7816	0.0296	**4s**	0.6735	0.7841	1.2554	**3s**
PowerNet	0.0087	0.8174	0.0308	140s	0.4214	0.7351	0.9671	79s
IREDGe	0.0035	0.8153	0.0256	128s	0.3907	0.7690	0.8203	70s
MAVIREC	0.0009	0.8854	0.0258	105s	0.3621	0.8651	0.8854	70s
PGAU (ours)	**0.0007**	**0.9085**	**0.0231**	97s	**0.2995**	**0.8692**	**0.7691**	68s

Evaluation
Main Results



(a) Golden (b) MAVIREC (c) PGAU (ours)

Visualization of IR Drop distribution of a PG.

Evaluation
Experimental Settings

Datasets

- CircuitNet[10]: 7322 samples
- ICCAD2023[1]: 100 real samples; 20 fake samples

Table: Statistics of CircuitNet dataset.

Front-end Design	RISCY-FPU	zero-riscy	RISCY	nvdla-small	Vortex-small
# PGs per design	2400	2400	2400	56	66
# Nodes per PG	> 30000	> 50000	> 50000	> 600000	> 600000
Mean of IR drop	0.0028	0.0029	0.0034	0.0074	0.0076
Variance of IR drop	0.0003	0.0003	0.0004	0.0006	0.0006

[10] CircuitNet: an open-source dataset for machine learning applications in electronic design automation (EDA)[J]. Information Sciences, 2022, 65(227401): 1-227401.
1 https://iccad-contest.org/

Evaluation
Transfer Results

Table: Transfer Results on CircuitNet dataset

Method	MAE	CC	MIRDE	runtime
UNet	0.0718	0.3525	0.2088	10s
UNet++	0.0676	0.3726	0.1849	12s
SETR	0.5356	0.1962	0.2775	20s
MobileNetV3	0.3387	0.1731	0.3845	19s
LR-ASPP	0.4673	0.2003	0.4803	18s
XGBIR	0.0793	0.3277	0.3638	**3s**
PowerNet	0.1996	0.2905	0.4472	20s
IREDGe	0.0883	0.2914	0.3873	16s
MAVIREC	0.0711	0.3374	0.2058	15s
PGAU (ours)	**0.0647**	**0.3589**	**0.1689**	12s

Evaluation
Ablation Study

Table: Ablation study results on CircuitNet dataset

Method	MAE	CC	MIRDE	Runtime
PGAU (w/o. $P_{composite}$)	0.0008	0.9048	0.0236	95s
PGAU (w/o. $P_{neighbour}$)	0.0010	0.8943	0.0244	95s
PGAU (w/o. $Count_{neighbour}$)	0.0010	0.8952	0.0239	95s
PGAU (w/o. LDS)	0.0007	0.9085	0.0231	97s
PGAU (w/o. AG)	0.0075	0.8704	0.0271	97s
PGAU (ours)	**0.0007**	**0.9094**	**0.0223**	97s

Conclusion

- This work focuses for the first time on the complex and redundant features of PG and imbalanced IR drop distribution.

- To solve the problem of feature confusion, this work extracts more effective combined features, analyzes the importance, and filters low-quality features.

- This work proposes a customized universal U-Net-based ML model for accurate static IR drop prediction, using an attention gate to capture the special area (e.g., hotspots) of IR drop in the PGs.

- To solve the problem of imbalanced IR drop distribution, this work adopts label distribution smoothing to re-weight the loss function and improve the model's performance.

THANK YOU!

Feng Guo/ Beijing University of Posts and
Telecommunications