UniCircuit: Multimodal Circuit Representation Learning with Anchor-Free Alignment

Jingxin Wang and Weikang Qian
Global College, Shanghai Jiao Tong University, Shanghai, China
Emails: {jingxin.wang, qianwk}@sjtu.edu.cn

***Abstract*—Electronic design automation requires unified circuit representations that jointly capture functionality, structure, and optimization behavior. Aligning heterogeneous circuit data (*e.g.*, Verilog codes and AND-inverter graphs (AIGs)) remains challenging due to substantial semantic gaps. Existing contrastive learning approaches rely on carefully defined cross-modal anchors, which are often difficult to construct and limit alignment quality. We propose UniCircuit, an anchor-free multimodal framework that integrates Verilog, AIG, and synthesis flow representations. UniCircuit employs a singular value decomposition-based alignment mechanism to capture shared semantics without predefined anchors. Experimental results show that UniCircuit achieves 52.94% higher quality-of-result (QoR) prediction accuracy and 79.35% better cross-modal retrieval recall@5 compared to state-of-the-art methods.**

I. Introduction

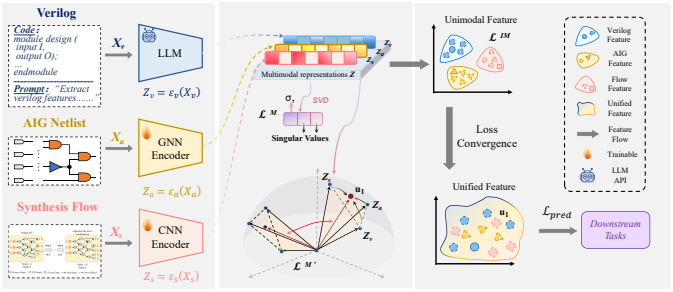
In electronic design automation (EDA), hardware description languages (HDLs), structural circuit representations, and logic synthesis flows form a sequential pipeline [1], [2], [3] (*e.g.*, Verilog code describes behavior, while an AND-inverter graph (AIG) implements the corresponding logic). However, synthesis is lossy and irreversible, often discarding functional context from HDLs during structural optimization [4]. Joint multimodal learning is therefore crucial for tasks like quality-of-result (QoR) prediction, where the target QoR is evaluated after executing a complete synthesis flow, and accuracy depends on the interplay between high-level intent and synthesis flows—the latter acting as dynamic trajectories that expose latent circuit properties absent in static netlists.

Existing methods [5], [6], [7] typically rely on *anchor points*, *i.e.*, reference instances from one modality, to guide others through contrastive learning [8] or reconstruction [9]. However, they face the **alignment difficulty**: defining reliable anchors across heterogeneous abstractions is challenging. To address this issue, we propose **UniCircuit**, a novel **anchor-free** multimodal framework. Leveraging a singular value decomposition (SVD)-based alignment [10], [11], [12], UniCircuit captures shared semantics without predefined anchors by maximizing dominant singular values. This ensures robust consistency across various modalities including Verilog code, AIG, and synthesis flow.

II. UniCircuit Methodology

This section presents the UniCircuit framework. As shown in Fig. 1, it first extracts multimodal representations from Verilog codes, AIGs, and synthesis flows, which will be detailed in Sec. II-A. It then applies a multimodal circuit anchor-free alignment, which will be detailed in Sec. II-B. Finally, it adopts the leading singular vector from the alignment process as the unified representation and optimizes a composite loss that balances alignment and QoR prediction, which will be detailed in Sec. II-C.

This work is supported by the National Key R&D Program of China under grant number 2021ZD0114701. Corresponding author: Weikang Qian.



(a) Multimodal Representation Extraction (b) Multimodal Anchor-free Alignment (c) Loss Convergence and Downstream Tasks

Fig. 1. Overview of UniCircuit: (a) First, multimodal representations are extracted from Verilog codes, AIGs, and synthesis flows; (b) then, they are aligned via SVD with an anchor-free objective; (c) finally, as the alignment loss converges, the leading singular vector $\mathbf{u}_1$ emerges as the unified representation for QoR prediction.

A. Multimodal Circuit Representation Extraction

We extract complementary representations from three modalities: Verilog code, AIG structure, and synthesis flow, capturing functional intent, structural topology, and optimization dynamics, respectively.

For Verilog, large designs are partitioned into modules, and DeepSeek-V3 [13] is used with customized prompts to generate concise functional summaries, which are aggregated and encoded into a dense semantic embedding using a lightweight sentence encoder all-MiniLM-L6-v2 [14], [15].

For AIGs, we adopt HOGA [16] to capture structural representations by fusing multi-hop neighborhood information through gated self-attention, followed by global pooling to obtain a graph-level embedding [17].

For synthesis flows, they are modeled as a discrete sequence of logic transformations (*e.g.*, rewrite, balance, *etc.* [3]), encoded using multi-scale one-dimensional convolutions to capture transformation motifs at different receptive fields within the flow [4]. Pooling and a lightweight multi-layer perceptron (MLP) are then used to aggregate these features into a representation of the synthesis flow [3].

B. Anchor-Free Multimodal Circuit Alignment

Following principled multimodal representation learning [12], we unify multimodal circuit embeddings without predefined anchors. Given a batch of B instances with $K = 3$ modalities, Verilog code, AIG, and synthesis flow, denote

the embedding for instance i ($1 \leq i \leq B$) and modality k ($1 \leq k \leq K$) as $\mathbf{z}_i^{m_k}$. The embeddings for instance i over all modalities, *i.e.*, $\mathbf{z}_i^{m_1}, \ldots, \mathbf{z}_i^{m_K}$, are L_2-normalized and concatenated into a joint matrix $\mathbf{Z}^{(i)}$. We perform singular value decomposition for $\mathbf{Z}^{(i)}$ as $\mathbf{Z}^{(i)} = \mathbf{U}^{(i)}\mathbf{\Sigma}^{(i)}{\mathbf{V}^{(i)}}^\top$, where the leading left singular vector $\mathbf{u}_1^{(i)}$ represents the dominant shared semantic direction across modalities.

To promote robust alignment, we optimize three complementary objectives. First, we encourage the shared semantic direction to dominate the representation space by maximizing the leading singular value, using the alignment loss $\mathcal{L}^M = -\frac{1}{B}\sum_{i=1}^{B} \log \frac{\exp(\sigma_1^{(i)}/\tau)}{\sum_{j=1}^{K}\exp(\sigma_j^{(i)}/\tau)}$, where $\sigma_j^{(i)}$ denotes the j-th largest singular value of $\mathbf{Z}^{(i)}$ and τ is a temperature parameter controlling the spectral concentration. This objective concentrates variance along the principal multimodal direction.

Second, to discriminate different instances, we introduce an instance-matching loss $\mathcal{L}^{\text{IM}} = \mathbb{E}_{(m_1,\ldots,m_K)\sim\mathcal{M}}[y \log \hat{y} + (1-y)\log(1-\hat{y})]$, where $\mathcal{M}$ represents the distribution of multimodal tuples consisting of both matched and mismatched modality pairs sampled from the dataset, y indicates whether a multimodal instance is correctly matched, and $\hat{y}$ is predicted by a lightweight MLP classifier [18]. This objective prevents degenerate alignment that collapses instance identities.

Finally, we regularize the shared semantic space by enforcing consistency among projections onto the leading singular direction using the following objective: $\mathcal{L}^{\mathcal{M}'} = -\frac{1}{B}\sum_{i=1}^{B} \log \frac{\exp\left[(\mathbf{u}_1^{(i)})^\top \mathbf{u}_1^{(i)}/\tau\right]}{\sum_{j=1}^{K}\exp\left[(\mathbf{u}_1^{(i)})^\top \mathbf{u}_1^{(j)}/\tau\right]}$, which encourages representations of different instances to remain coherent in the shared semantic space, where $\mathbf{u}_1^{(i)}$ denotes the leading singular vector corresponding to the i-th instance. Upon convergence, we adopt the leading singular vector as the unified circuit representation, $\mathbf{z}_i^f = \mathbf{u}_1^{(i)}$, enabling effective alignment despite large abstraction gaps among various modalities.

C. Composite Training Objective

We jointly optimize multimodal alignment and QoR prediction in a unified training objective. Given the fused representation $\mathbf{z}_i^f$ (the unified circuit embedding for instance i), QoR metrics are predicted using a lightweight regression head $\mathbf{f}_{\text{reg}}(\cdot)$ and supervised with a mean squared error loss, $\mathcal{L}_{\text{pred}} = \frac{1}{B}\sum_i \|\mathbf{f}_{\text{reg}}(\mathbf{z}_i^f) - \mathbf{g}_i\|_2^2$, where $\mathbf{g}_i$ denotes the ground truth. Training follows a two-stage scheme: we first optimize the alignment objectives, formulated as a weighted sum of the losses $\mathcal{L}^M, \mathcal{L}^{IM}$, and $\mathcal{L}^{\mathcal{M}'}$ to obtain a stable unified representation, and then fine-tune it for QoR prediction while retaining alignment regularization. This progressive optimization ensures both robust multimodal alignment and accurate QoR estimation.

III. Experimental Results

We evaluate UniCircuit on the OpenABC-D benchmark [3], which comprises 870,000 AIG samples. These samples are obtained from 29 hardware IPs [19], [20], [21]. We adopt an IP-level hold-out evaluation protocol, using 20 IPs for training and the remaining IPs for testing.

AIGs are processed as PyTorch Geometric graphs with 100–50,000 nodes, and synthesis flows are encoded as integer sequences of up to 20 logic transformations.

We train the model in two stages using Adam with learning rate 1×10^{-4} and batch size 64 on four 4090 GPUs for 500 epochs, first optimizing alignment objectives and then QoR prediction.

1) QoR Prediction: We consider QoR prediction for designs after applying the synthesis flow. As shown in Table I, UniCircuit substantially improves QoR prediction over state-of-the-art baselines. Using HOGA-5 as AIG encoder, average mean absolute percentage error (MAPE) decreases from 3.57% to 1.68%, while similar gains are observed for DeepGate2 and GCN. These results demonstrate the robustness of UniCircuit in leveraging complementary information from Verilog, AIGs, and synthesis flows.

TABLE I. QoR prediction performance (MAPE) using optimized circuit area as the metric. Each model shows results without (w/o) and with (w/) UniCircuit.

Design	**GCN** [22]		**DeepGate2** [17]		**HOGA-5** [16]	
	w/o	w/	w/o	w/	w/o	w/
fir	7.79%	3.66%	3.85%	1.84%	3.65%	**1.70%**
iir	5.25%	2.49%	2.93%	1.47%	2.78%	**1.31%**
jpeg	6.91%	3.30%	3.69%	1.71%	3.49%	**1.61%**
mem_ctrl	11.23%	5.81%	6.88%	3.58%	6.48%	**3.14%**
tv80	9.48%	5.21%	5.12%	2.61%	4.82%	**2.37%**
vga_lcd	4.54%	1.84%	2.35%	1.07%	2.25%	**1.02%**
wb_conmax	9.77%	4.81%	5.31%	2.50%	5.01%	**2.29%**
wb_dma	4.95%	2.26%	2.57%	1.31%	2.42%	**1.16%**
sasc	2.78%	1.08%	1.30%	0.58%	1.20%	**0.54%**
Average	6.97%	3.38%	3.78%	1.85%	3.57%	**1.68%**

2) Cross-modal Retrieval: As shown in Table II, we evaluate retrieval performance using Recall@K ($R@K$), which quantifies the percentage of queries where the ground-truth target is correctly identified within the top-K ranked results [23]. UniCircuit consistently outperforms baselines. For example, in the task of retrieving AIG from Verilog code, *i.e.*, Verilog→AIG, R@5 improves from 0.222 to 0.398. This validates the effectiveness of anchor-free SVD-based alignment combined with instance-level matching.

TABLE II. Multi-modal retrieval performance evaluation results on OpenABC-D dataset.

Retrieval Tasks	Model	R@1	R@5	R@10
Verilog → AIG	SynNetV3 [3]	0	0.042	0.012
	CircuitFusion [23]	0.037	0.2218	0.3698
	UniCircuit	**0.081**	**0.3978**	**0.4729**
AIG → Verilog	SynNetV3 [3]	0.01	0.032	0.017
	CircuitFusion [23]	0.0736	0.0739	0.2588
	UniCircuit	**0.0975**	**0.1247**	**0.3963**

IV. Conclusion

We present UniCircuit, an anchor-free multimodal circuit alignment framework that enables flexible integration of Verilog code, AIG, and synthesis flow for EDA tasks such as QoR prediction and cross-modal retrieval. By removing the reliance on predefined anchors, UniCircuit improves generalization and efficiency compared to traditional methods.

REFERENCES

[1] R. K. Brayton, R. Rudell, A. Sangiovanni-Vincentelli, and A. R. Wang, "MIS: A multiple-level logic optimization system," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 6, no. 6, pp. 1062–1081, 2004.

[2] A. Mishchenko, S. Chatterjee, R. Brayton, and N. Een, "Improvements to combinational equivalence checking," in *Proceedings of the 2006 IEEE/ACM international conference on Computer-aided design*, 2006, pp. 836–843.

[3] A. B. Chowdhury, B. Tan, R. Karri, and S. Garg, "OpenABC-D: A large-scale dataset for machine learning guided integrated circuit synthesis," *arXiv preprint arXiv:2110.11292*, 2021.

[4] A. B. Chowdhury, B. Tan, R. Carey, T. Jain, R. Karri, and S. Garg, "Bulls-Eye: Active few-shot learning guided logic synthesis," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 42, no. 8, pp. 2580–2590, 2022.

[5] W. Fang, W. Li, S. Liu, Y. Lu, H. Zhang, and Z. Xie, "NetTAG: A multimodal RTL-and-layout-aligned netlist foundation model via text-attributed graph," *arXiv preprint arXiv:2504.09260*, 2025.

[6] W. Fang, S. Liu, H. Zhang, and Z. Xie, "A self-supervised, pre-trained, and cross-stage-aligned circuit encoder provides a foundation for various design tasks," in *Proceedings of the 30th Asia and South Pacific Design Automation Conference*, 2025, pp. 505–512.

[7] H. Wu, H. Zheng, Y. Pu, and B. Yu, "Circuit representation learning with masked gate modeling and Verilog-AIG alignment," *arXiv preprint arXiv:2502.12732*, 2025.

[8] J. Wang et al., "Multi-level graph subspace contrastive learning for hyperspectral image clustering," in *2024 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2024, pp. 1–8.

[9] Z. Shi et al., "Alignment unlocks complementarity: A framework for multiview circuit representation learning," *arXiv preprint arXiv:2509.20968*, 2025.

[10] G. Cicchetti, E. Grassucci, L. Sigillo, and D. Comminiello, "Gramian multimodal representation learning and alignment," *arXiv preprint arXiv:2412.11959*, 2024.

[11] J. S. Chipman, "'Proofs' and proofs of the Eckart–Young theorem," in *Stochastic processes and functional analysis*, CRC Press, 2020, pp. 71–83.

[12] X. Liu, X. Xia, S.-K. Ng, and T.-S. Chua, "Principled multimodal representation learning," *arXiv preprint arXiv:2507.17343*, 2025.

[13] A. Liu et al., "Deepseek-v3 technical report," *arXiv preprint arXiv:2412.19437*, 2024.

[14] J. Pan et al., "Crop: Circuit retrieval and optimization with parameter guidance using llms," *arXiv preprint arXiv:2507.02128*, 2025.

[15] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using siamese BERT-networks," *arXiv preprint arXiv:1908.10084*, 2019.

[16] C. Deng et al., "Less is more: Hop-wise graph attention for scalable and generalizable learning on circuits," in *Proceedings of the 61st ACM/IEEE Design Automation Conference*, 2024, pp. 1–6.

[17] Z. Shi et al., "DeepGate2: Functionality-aware circuit representation learning," in *2023 IEEE/ACM International Conference on Computer Aided Design (ICCAD)*, IEEE, 2023, pp. 1–9.

[18] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*, PmLR, 2020, pp. 1597–1607.

[19] B. Chetwynd, K. Bush, and K. Ingols, *Common evaluation platform*, version v4.0, MIT Lincoln Laboratory, 2024.

[20] *Opencores hardware rtl designs*, https://opencores.org/.

[21] C. Albrecht, "IWLS 2005 benchmarks," in *International Workshop for Logic Synthesis (IWLS)*, vol. 9, 2005.

[22] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[23] W. Fang, S. Liu, J. Wang, and Z. Xie, "Circuitfusion: Multimodal circuit representation learning for agile chip design," in *The Thirteenth International Conference on Learning Representations*, 2025.