Conditional Neural Architecture Search Tailored for Reactor 3D Power Distribution Prediction

Minxiao Zhong
College of Computer Science
Sichuan University
Chengdu, China
zhongminxiao@stu.scu.edu.cn

Qing Li
State Key Lab. of Advanced Nuclear Energy Technology
Nuclear Power Institute of China
Chengdu, China
liqing@npic.ac.cn

Yanan Sun*
College of Computer Science
Sichuan University
Chengdu, China
ysun@scu.edu.cn

***Abstract*—The accurate prediction of three-dimensional power distribution is critical for ensuring the safety of nuclear reactors. Traditional prediction methods struggle in computational efficiency and prediction accuracy. Recent deep learning approaches offer a promising alternative but frequently employ generic architectures due to a lack of domain-specific knowledge guiding the architecture design. They fail to capture the inherent anisotropy, geometric symmetry, and state-dependent nonlinearity of reactor behavior. To overcome each limitations, this paper proposes a conditional neural architecture search method. It comprises three essential components: (1) designing an anisotropic convolution block to decouple radial and axial feature learning, thereby matching the anisotropy of physical principles; (2) proposing a geometry-aware attention block to embed rotational symmetry priors and adaptively focus on critical regions; and (3) developing a lightweight conditioner to dynamically activate the architecture for specific reactor states, effectively modeling state-dependent nonlinearities. Experiments show that our automatically searched architecture significantly outperforms manual baselines. For instance, it achieves a mean square error of 0.0109 with only 873K parameters, reducing error by 44.1% compared to a V-Net baseline. The results demonstrate that our method enables highly accurate, efficient, and adaptive prediction suitable for reactor safety evaluation.**

***Index Terms*—Nuclear reactor, 3D power distribution, neural architecture search.**

I. Introduction

The safety of nuclear reactors is the most critical element of nuclear industry, whose evaluation largely depends on accurate prediction of the three-dimensional (3D) power distribution [1], [2]. The 3D power distribution is geometrically discretized into a 3D voxel array corresponding to the fuel lattice and axial layers. In general, it exhibits three inter-related physical characteristics [3]: strong spatial anisotropy, geometric symmetry, and state-dependent nonlinearity. These characteristics collectively define the prediction as a nonlinear mapping task with strong structural priors.

Traditional prediction methods face a fundamental trade-off between efficiency and accuracy: Monte Carlo methods provide high accuracy but require millions of core-hours due to extensive particle tracking [4]. They are unsuitable for real-time monitoring [5]. Deterministic methods offer computational efficiency but suffer from accuracy limitations [6]. This is because they utilize physical approximations like homogenization and energy group reduction [7].

In recent years, deep learning has provided a new paradigm for learning complex nonlinear mappings of power distribution directly from data. However, existing studies have not deeply aligned their architecture design with the features of the 3D power distribution [8]. Specifically, current research primarily focuses on using 2D convolutional neural networks (2D-CNNs) to process 2D slice data of reactor [9]. It basically treats each axial layer independently. While straightforward, this approach has a fundamental flaw of dimensional separation. It completely ignores physical coupling and anisotropy between the axial and radial directions. Notably, 3D convolutional neural networks (3D-CNNs) can effectively capture 3D spatial features, and have become standard tools for processing volumetric data in fields like medical image analysis (e.g., CT, MRI segmentation) [10]. However, their direct application to reactor power distribution still faces challenges. This is because the isotropic mechanism of standard 3D convolution kernels does not match the inherent anisotropy of radial and axial physical effects. This mismatch inevitably leads to inefficient feature extraction. Furthermore, standard 3D-CNNs do not explicitly incorporate geometric symmetry priors. Therefore, there is an urgent need to design a new method addressing spatial anisotropy, geometric symmetry awareness, and architectural static rigidity.

In this paper, we propose a neural architecture search (NAS) method that integrates physical priors and reactor states. Specifically, to address the spatial anisotropy problem, we design an anisotropic 3D convolution block (ACB). To address geometric symmetry awareness, we propose a geometry-aware attention block (GAB). To tackle the architectural static rigidity problem, we develop a conditional NAS, and simultaneously incorporate the first two blocks into the search space to improve prediction performance. To achieve this goal, the following aims have been specified:

- We design an ACB. By decoupling the standard 3D convolution into a combination of radial 2D convolution and axial 1D convolution, the network can efficiently learn radial and axial power distribution patterns separately. This approach better matches the spatial heterogeneity of power distribution with fewer parameters.

*Corresponding Author: Yanan Sun.

- We propose a GAB. This block consists of a learnable symmetry constraint module and an adaptive multi-scale focal module. It explicitly embeds the geometric symmetry prior of the reactor and dynamically focuses on regions where the power distribution changes significantly. Based on this, it thereby enhancing overall prediction reliability.
- We develop a conditional NAS method that adapts to reactor states. Guided by a lightweight conditioner, a single supernet activates distinct sub-architectures in response to different reactor state inputs. This allows the architecture to specialize its feature extraction process according to each operational states.

II. Related Works

A. Feature Extraction for Anisotropy

To avoid the complexity of 3D power distribution prediction, earlier studies commonly adopted 2D-CNNs to process radial-plane prediction. For example, Jang *et al.* directly applied 2D-CNNs to learn the mapping between radial power distribution and other neutronics parameters [11]. However, such approach artificially separates the physical continuity along the axial direction. They cannot capture axial physical processes, such as power diffusion and moderation, leading to predictions that lack self-consistency axially. The standard 3D-CNN is a natural choice for volumetric data and has been widely used in medical image analysis, e.g., 3D ResNet-18 for segmentation [12]. It performs convolution with isotropic kernels across all three spatial dimensions to capture isotropic features. However, reactor power distribution exhibits physics-driven anisotropy. The 3D power distribution varies sharply in the radial plane, while varying gradually along the axial direction.

In this paper, we design a block based on anisotropic 3D convolution. It decouples a standard 3D convolution into an independent radial 2D convolution followed by an axial 1D convolution sequence.

B. Attention Mechanism

Attention mechanisms can be primarily categorized into three types: channel attention, spatial attention, and self-attention [13]. Channel attention, such as in SENet, generates a channel-wise weight vector via global average pooling to emphasize important feature channels [14]. However, it completely ignores spatial location information and cannot identify the spatial coordinates of locally critical regions in the reactor. Spatial attention, like the spatial attention module in CBAM, generates a 2D spatial weight map through convolutional operations to focus on important regions of an image [15]. Meanwhile, when directly applied to 3D volumetric data, it typically requires compression or slicing, which disrupts the integrity of the 3D space. Moreover, it struggles to adjust to geometric constraints such as reactor symmetry.

In this paper, we propose a block based on geometry-aware 3D attention. Its motivation lies in combining differentiable geometric transformation layers with 3D sparse attention.

C. Neural Architecture Search (NAS)

Current mainstream NAS technologies can be categorized, based on their search and deployment paradigms, into: 1) search based on reinforcement learning (RL) or evolutionary algorithm (EA), 2) differentiable architecture search using supernets. Methods based on RL/EA search a discrete architecture space via an agent or population iteration. Their ultimate goal is to find a single, statically optimal architecture for the entire task [16]. Consequently, they cannot generate specialized structures tailored to different, discrete reactor states. The differentiable supernet paradigm, exemplified by DARTS [17], significantly improves search efficiency by constructing a weight-sharing supergraph and optimizing architecture parameters via gradient descent [18]. However, its core output is similarly a fixed, final architecture representing a compromise across the data distribution. It lacks the capability for differentiated design targeting specific reactor states.

In this paper, we propose a conditional NAS method capable of automatic adaptation based on explicit reactor states. It dynamically activates the state-corresponded sub-network pathways within the supernet.

III. The Proposed Algorithm

A. Preliminary

The objective investigated in this work is a typical pressurized water nuclear reactor [3]. Specifically, the research examines a typical reactor with 177 fuels arranged in a 15×15 grid matrix with 16 axial layers. The targeting 3D power distribution can be written as $Y = \mathcal{F}(X)$ to explain complex physical effect. Training establishes a mapping $\mathcal{F}(\cdot)$ between initial physical attributes and power distribution formulated by $Y = \mathcal{F}(A, B, T, c)$, with the following meanings:

- A: The reactor scheme arrangement, recording the position information of fuels.
- B: Initial burnup value of fuel in matrix numerical form.
- T: The fuel types in semantic form, including t_1, t_2, t_3.
- c: The reactor states in semantic form, including the beginning of life-cycle (BOC), middle of life-cycle (MOC) and end of life-cycle (EOC).

B. Algorithm Overview

Algorithm 1 outlines our conditional NAS. It encodes reactor state c into embedding $\mathbf{e}_c$ (line 1), initializes supernet weights θ, architecture parameters α, and bias module Φ (line 2). Then, it defines candidate operations O including ACB and GAB (line 3) for anisotropic fields.

The search procedure follows a warm-up and alternating optimization strategy. Initially, network weights θ are trained with frozen architecture parameters α and Φ to stabilize the supernet (lines 5-6). Thereafter, for each data batch with state label c, a state-conditioned bias vector $\mathbf{b}_c = \Phi(c)$ is computed (line 8). This bias is added to the original architecture parameters α to form adjusted parameters $\tilde{\alpha} = \alpha + \mathbf{b}_c$ (line 9), effectively steering the architecture distribution based on reactor state. Architecture parameters and bias module are updated

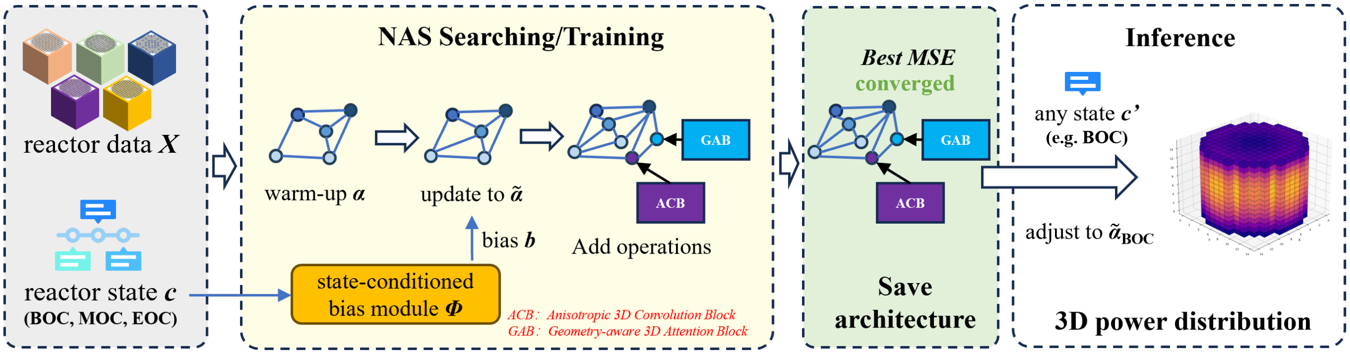


Fig. 1: The flowchart of the proposed conditional NAS algorithm.

using validation loss $\mathcal{L}_{val}$, while θ is updated using the training loss $\mathcal{L}_{train}$ in an alternating fashion (lines 10-11). The conditional architecture is evaluated each epoch (line 13), and the best parameter tuple (θ, α, Φ) yielding the lowest validation MSE is saved (line 14).

Upon completion of the search, the algorithm returns the optimized state-conditioned architecture parameters $(\theta^*, \alpha^*, \Phi^*)$ (line 16). It enables a single supernet to dynamically specialize its effective sub-architecture for different reactor states c, achieving adaptive accuracy without maintaining multiple discrete models. Fig. 1 shows the flowchart of the proposed algorithm for the better understanding.

Algorithm 1 Algorithm Overview

Input: 3D reactor data tensor $\mathbf{X} \in \mathbb{R}^{B\times5\times15\times15\times16}$, reactor state labels $c \in \{0, 1, 2\}$, and target power distribution $\mathbf{Y}$.

Output: Optimized network weights θ^*, architecture parameters α^*, and state-conditioned bias module parameters Φ^*.

1: Encode the discrete state c to obtain a condition vector $\mathbf{e}_c$;
2: Initialize the supernet weights θ, architecture parameters α; and the condition bias module Φ;
3: Define the candidate operation set O including *ACB* and *GAB*;
4: **for** each epoch e in total epochs E **do**
5:     **if** $e < E_{warm-up}$ **then**
6:         Train θ with fixed α and Φ on the training set;
7:     **else**
8:         Compute the state-conditioned bias: $\mathbf{b}_c = \Phi(c)$;
9:         Adjust architecture parameters: $\tilde{\alpha} = \alpha + \mathbf{b}_c$;
10:         Update α and Φ using the validation loss $\mathcal{L}_{val}$;
11:         Update θ using the training loss $\mathcal{L}_{train}$;
12:     **end if**
13:     Evaluate the current conditional architecture on the validation set;
14:     Save the tuple (θ, α, Φ) if validation MSE improves;
15: **end for**
16: **Return** the best-performing state-conditioned architecture parameters $(\theta^*, \alpha^*, \Phi^*)$.

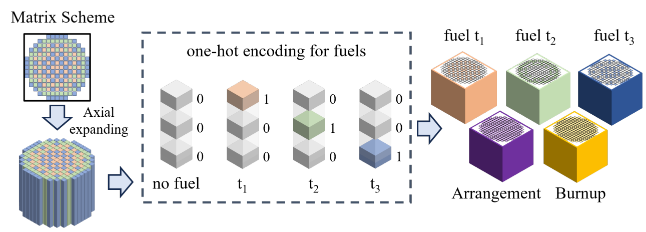


Fig. 2: Example of encoding fuel types to three separate tensors.

1) Encoding Strategy: To avoid confusion and unintended numerical relations, we use one-hot encoding for categorical attributes (Fig. 2). For example: fuel types $t_1 = [1, 0, 0]$, $t_2 = [0, 1, 0]$, absent fuel $[0, 0, 0]$. Non-categorical attributes use separate tensors. The 2D encoding is replicated axially across 16 layers. For discrete reactor states BOC, MOC, EOC, we assign labels 0,1,2 to preserve ordinality, then map to dense embedding e_c.

2) ACB: To handle anisotropy in reactor power distribution, ACB decomposes standard 3D convolution into sequential radial 2D and axial 1D convolutions.

First, radial 2D convolution applies 2D kernels K_{2D} to each axial slice of input $\mathbf{X}_{\text{in}}$ as $\mathbf{U} = \text{Conv}_{2D}(\mathbf{X}_{\text{in}}; \mathsf{K}_{2D}, \mathbf{b}_{2D})$.

Then, axial 1D convolution uses kernel K_{1D} along depth for each radial position, as $\mathbf{V} = \text{Conv}_{1D}(\mathbf{U}; \mathsf{K}_{1D}, \mathbf{b}_{1D})$.

Finally, pointwise convolution fuses channels. This provides a physically-grounded prior for anisotropic fields.

3) GAB: Power distribution is rotationally symmetric at steady state but asymmetric under control rod insertion. GAB dynamically assesses symmetry and applies data-dependent modulation.

Given input $\mathbf{F}_{\text{in}}$, GAB first estimates a symmetry confidence scalar γ from reactor state identifier c (embedded as $\mathbf{e}_c$) and global context vector $\mathbf{z}$ (summarized from $\mathbf{F}_{\text{in}}$), as $\gamma = \sigma\left(\mathbf{w}_2 \cdot \text{ReLU}\left(\mathbf{W}_1 [\mathbf{e}_c \| \mathbf{z}] + \mathbf{b}_1\right) + d\right)$, where σ is sigmoid, $\mathbf{W}_1, \mathbf{b}_1$ are first-layer weights/bias, $\mathbf{w}_2, d$ are output-layer weights/bias. $\gamma \approx 1$ indicates high symmetry.

Then soft symmetry modulation is performed by leveraging an N-fold rotational symmetry prior (N = number of azimuthal sectors, e.g., 4 or 8), compute modulation mask $\mathbf{M}_{\text{sym}}$ via

TABLE I: Comparison with various architectures on BOC, MOC and EOC reactor states.

Method \ State	*BOC*		*MOC*		*EOC*		**Parameter**	**GPU-day**	**Formation**
Loss type	MSE	MAE	MSE	MAE	MSE	MAE			
VGG-Net [19]	0.2364	0.4294	0.2184	0.4265	0.2061	0.4271	740K	-	2D Manually
ResNet [28]	0.0506	0.1776	0.0486	0.1784	0.0398	0.1631	414K	-	2D Manually
Inception-ResNet [20]	0.0577	0.2005	0.0520	0.1959	0.0481	0.1944	2334K	-	2D Manually
U-Net [8]	0.0678	0.1952	0.0839	0.2281	0.0622	0.1963	1481K	-	2D Manually
3D-CNN [10]	0.0342	0.1432	0.0311	0.1454	0.0247	0.1225	1542K	-	3D Manually
3D U-Net [21]	0.0223	0.11179	0.0198	0.1160	0.0168	0.1008	6543K	-	3D Manually
3D ResNet-18 [22]	0.0414	0.1604	0.0379	0.1613	0.0293	0.1348	1789K	-	3D Manually
3D ResNet-34 [23]	0.0556	0.1916	0.0490	0.1838	0.0359	0.1506	2185K	-	3D Manually
3D ResNet-50 [24]	0.0649	0.2063	0.0574	0.1969	0.0450	0.1704	3245K	-	3D Manually
V-Net [25]	0.0195	0.11178	0.0169	0.1089	0.0129	0.0902	798K	-	3D Manually
3D DenseNet [26]	0.0724	0.2330	0.0637	0.2240	0.0585	0.2209	1233K	-	3D Manually
3D EfficientNet [27]	0.6927	0.7815	0.7156	0.8131	0.7163	0.8153	348K	-	3D Manually
Ours	**0.0109**	**0.0676**	**0.0110**	**0.0706**	**0.0107**	**0.0645**	873K	1.2	3D NAS

feature similarities across the N sectors. This process is written as $\mathbf{F}' = \mathbf{F}_{\text{in}} \circ \left(\mathbf{1} + \gamma \cdot \mathbf{M}_{\text{sym}}\right)$.

Finally, a 3D attention mechanism is applied. From $\mathbf{F}'$, generate channel attention vector $\mathbf{a}_c$ and spatial-axial attention map $\mathbf{a}_{sa}$. Their broadcast product $\mathbf{\Pi} = \mathbf{a}_{sa} \circ \mathbf{a}_c$ is applied to $\mathbf{F}_{\text{in}}$ as $\mathbf{F}_{\text{out}} = \mathbf{\Pi} \circ \mathbf{F}_{\text{in}}$. This emphasizes both symmetric regions and other critical local features.

4) Conditional NAS: Conventional NAS learns a single static architecture while training per-state NAS is costly. We propose conditional NAS with a bias module, enabling a supernet to generate state-dependent architectures.

Based on gradient-NAS, a mixed operation on a cell edge is a weighted sum over candidates $\mathcal{O}$ (e.g., convolution, ACB, GAB). Here, architecture parameters $\boldsymbol{\alpha}$ depend on reactor state c. The bias module Φ produces a state-specific bias $\mathbf{b}_c = \Phi(c) = \lambda \cdot (\mathbf{W}_m \mathbf{e}_c + \mathbf{b}_m)$, with a scaling factor λ. For each operation o on an edge, the biased architecture parameter is then defined as $\tilde{\alpha}_o(c) = \alpha_o + b_{c,o}$. After this, softmax yields state-dependent selection probabilities, making the mixed output a probability-weighted sum conditioned on c.

IV. Experimental Design

A. Data Generation

Following the common approach in existing research, this study employs a traditional high-precision prediction software to simulate the reactor [7]. We randomly simulate 3,000 non-repetitive reactor schemes at three discrete states.

B. Peer Competitors

The peer competitors are two categories: (1) The manually designed 2D-CNNs from recent nuclear prediction works, including VGG-Net [19], ResNet [9], Inception-ResNet [20], and U-Net [8]. (2) State-of-the-art (SOTA) 3D-CNNs from related domains, including standard 3D-CNN [10], 3D U-Net [21], 3D ResNet-18 [22], 3D ResNet-34 [23], 3D ResNet-50 [24], V-Net [25], 3D DenseNet [26] and 3D EfficientNet [27].

C. Experiment Setup

During the architecture search phase, we warm up the supernet for 15 epoch and train the network for 50 epochs. The initial learning rate is set to 1.0. The architecture learning rate η is set to 6e-4 and the weight decay rate is set to 1e-3. Experiments are performed on a single NVIDIA RTX-4070 GPU with i7-14700KF CPU.

V. Result and Discussion

A. Overall Results

Our proposed method demonstrates significant and consistent performance advantages across all reactor states under both MSE and MAE evaluation metrics. As illustrated in Table I, our method achieves the lowest MSE values among all compared architectures, reaching 0.0109 (BOC), 0.0110 (MOC), and 0.0107 (EOC). This represents a substantial improvement over existing SOTA architectures.

Compared to classic 2D manually-designed architectures, our method exhibits a decisive advantage. It reduces the BOC MSE by 95.4% against VGG-Net (0.2364) and 78.5% against ResNet. This confirms the fundamental limitation of 2D representations for this 3D power distribution prediction on. When evaluated against advanced 3D manually-designed networks, our method maintains superior predictive accuracy. Specifically, our MSE is 68.1% lower than that of 3D-CNN and 51.1% lower than 3D U-Net in the BOC state. Notably, we outperform the strong V-Net baseline by 44.1% in BOC MSE. Even compared to deeper 3D ResNet variants, our method shows strong improvement. The performance gains are achieved alongside high model efficiency. Our method contains only 873K parameters, which is fewer than most 3D competitors and demonstrates superior parameter efficiency.

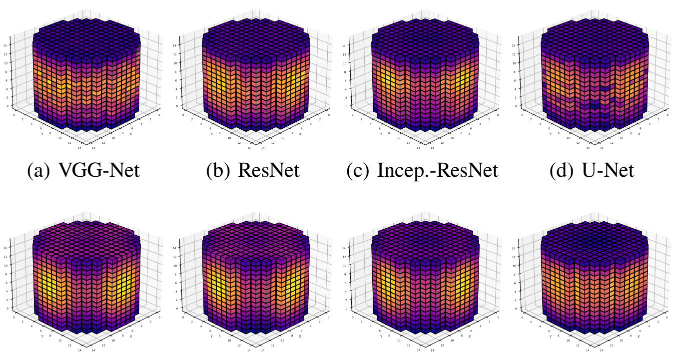
Furthermore, the training cost of 1.2 GPU-days remains practical, indicating the feasibility of our method in computationally constrained environments.

The superior performance stems from our physics-prioritized conditional NAS. Unlike manual general-purpose architectures, it automatically discovers optimal topologies tailored to reactor physics, enhancing generalization across BOC, MOC, and EOC with consistently low error rates.

B. Visual Inspection Over Architectures

To evaluate the performance of predicting 3D power distribution, we present the visual inspection at BOC and EOC state, as shown from Fig. 3 to Fig. 6.

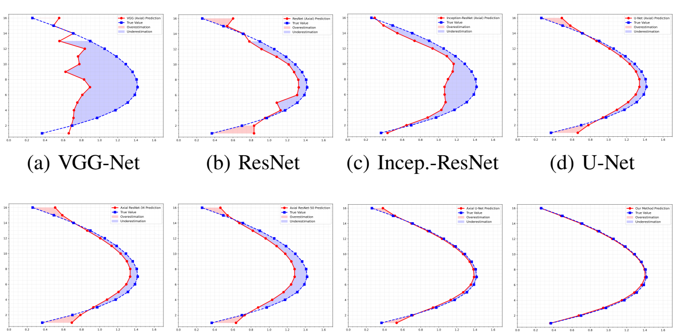
At BOC state, power distribution is smooth. Our method reconstructs a uniform radial pattern (Fig. 3(h)), contrasting with irregular ones from standard 2D architectures (Fig. 3(a), Fig. 3(d)). Axially, it also outperforms conventional 3D architectures (Fig. 4(e), Fig. 4(f)) in capturing physical effects.



(a) VGG-Net (b) ResNet (c) Incep.-ResNet (d) U-Net

(e) 3D ResNet-34 (f) 3D ResNet-50 (g) 3D U-Net (h) Ours

Fig. 3: Visual inspection of 3D power distribution prediction at BOC. The bright voxels represent higher power and the dark voxels represent lower power.

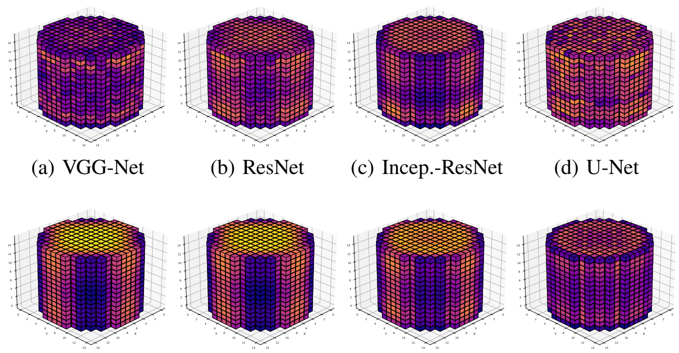


(a) VGG-Net (b) ResNet (c) Incep.-ResNet (d) U-Net

(e) 3D ResNet-34 (f) 3D ResNet-50 (g) 3D U-Net (h) Ours

Fig. 4: Visual inspection of axial power distribution prediction at BOC. The X-axis represents power distribution value and the Y-axis represents the reactor height. Blue line means the true value and the red line means the predicted value.

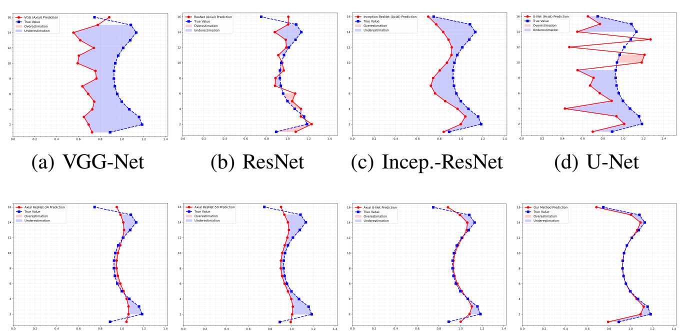
At EOC state, our method outperforms standard 2D (Fig. 5(a), Fig. 5(d)) and other 3D architectures (Fig. 5(e), Fig. 5(f)), clearly capturing non-uniform axial variations and pronounced fluctuations. The axial results (Fig. 6(a), Fig. 6(d), Fig. 6(e), Fig. 6(f)) further confirm its superiority, demonstrating the effectiveness of the geometry-aware attention mechanism.



(a) VGG-Net (b) ResNet (c) Incep.-ResNet (d) U-Net

(e) 3D ResNet-34 (f) 3D ResNet-50 (g) 3D U-Net (h) Ours

Fig. 5: Visual inspection of 3D power distribution prediction at EOC.



(a) VGG-Net (b) ResNet (c) Incep.-ResNet (d) U-Net

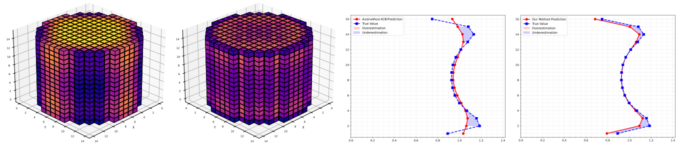
(e) 3D ResNet-34 (f) 3D ResNet-50 (g) 3D U-Net (h) Ours

Fig. 6: Visual inspection of axial power distribution prediction at EOC.

C. Ablation Studies

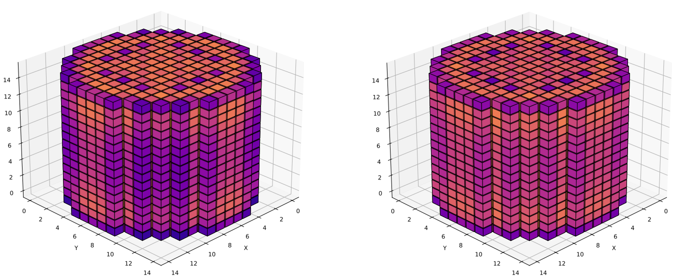
1) Effectiveness of ACB: We remove ACB from the NAS search space (Fig. 7) and replace it with a standard 3D convolution as baseline. Without ACB (Fig. 7(a)), radial and axial power distributions become less uniform than with ACB (Fig. 7(b)). Fig. 7(c) also shows reduced ability to capture axial physical effects. These results confirm ACB's effectiveness in modeling anisotropic 3D power distributions.

2) Effectiveness of GAB: We remove GAB from the NAS search space and retrain the resulting architecture on transient data with control rod insertion. Without GAB (Fig. 8(a)), the model detects the local power reduction but misses its symmetric increase, breaking geometric symmetry. With GAB (Fig. 8(b)), the symmetric phenomenon is captured. This shows superiority of GAB for geometry-dependent physical effects.



(a) Without ACB (b) Ours (c) Without ACB (d) Ours

Fig. 7: Ablation studies on ACB at EOC.



(a) Without GAB (b) Ours

Fig. 8: Ablation studies on GAB at transient state of control rod insertion.

VI. Conclusion

This paper presents a conditional NAS method tailored for 3D reactor power distribution prediction. Specifically, we design a search space incorporating an anisotropic 3D convolution block for directional feature extraction and a geometry-aware attention block for symmetry-aware feature refinement. Furthermore, a state-conditioned bias module allows a single supernet to specialize its architecture dynamically according to discrete reactor states. Experimental evaluations confirm that our method achieves superior prediction accuracy across all states compared to SOTA 2D and 3D manually-designed architectures. Ablation studies validate the individual contributions of the anisotropic 3D convolution block and geometry-aware attention block in modeling anisotropic patterns and geometric dependencies. The proposed architecture offers a practical and accurate data-driven solution for reactor safety evaluation.

References

[1] A. Adamantiades and I. Kessides, "Nuclear power for sustainable development: current status and future prospects," *Energy Policy*, vol. 37, no. 12, pp. 5149–5166, 2009.

[2] W. Lin, X. Miao, J. Chen, M. Ye, L. Zhang, Y. Xu, X. Liu, H. Jiang, and Y. Lu, "St-sam: Spatial-temporal state adaptation model for neutron detector fault detection and isolation in nuclear power plants," *IEEE Transactions on Industrial Informatics*, pp. 1–10, 2024.

[3] W. M. Stacey, *Nuclear reactor physics*. John Wiley & Sons, 2018.

[4] P. K. Romano, N. E. Horelik, B. R. Herman, A. G. Nelson, B. Forget, and K. Smith, "Openmc: A state-of-the-art monte carlo code for research and development," *Annals of Nuclear Energy*, vol. 82, pp. 90–97, 2015.

[5] H. O. Tekin, G. ALMisned, S. A. Issa, and H. M. Zakaly, "A rapid and direct method for half value layer calculations for nuclear safety studies using mcnpx monte carlo code," *Nuclear Engineering and Technology*, vol. 54, no. 9, pp. 3317–3323, 2022.

[6] T. M. Pandya, F. Bostelmann, M. Jessee, and J. Ortensi, "Two-step neutronics calculations with shift and griffin for advanced reactor systems," *Annals of Nuclear Energy*, vol. 173, p. 109131, 2022.

[7] H. Zhang, C. Zhao *et al.*, "Development of digital reactor high-fidelity neutronics code shark," *Atomic Energy Science and Technology*, vol. 56, no. 2, pp. 334–342, 2022.

[8] Q. Yang, N. Gui, X. Yang, J. Tu, and S. Jiang, "Predicting neutron flux density distribution in htr-10 using u-net based on dem-mc coupled simulations," *Nuclear Engineering and Technology*, vol. 57, no. 6, p. 103425, 2025.

[9] K. Lei, L. Cao, C. Wan, and H. Cao, "Evaluation of core refueling loading pattern with deep convolutional neural network," *Atomic Energy Science and Technology*, vol. 55, no. 2, p. 279, 2021.

[10] R. K. Rasel, F. Wu, M. Chiariglione, S. S. Choi, N. Doble, and X. R. Gao, "Assessing the efficacy of 2d and 3d cnn algorithms in oct-based glaucoma detection," *Scientific Reports*, vol. 14, no. 1, p. 11758, 2024.

[11] H. Janga, H. C. Shinb, D.-Y. Kimb, and H. C. Leea, "Prediction of opr-1000 neutronic design parameters using convolutional neural network for fuel loading pattern optimization," *Methods*, vol. 2, no. 1.14, pp. 6–92, 2020.

[12] C. Wang, "A review on 3d convolutional neural network," in *2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA)*. IEEE, 2023, pp. 1204–1208.

[13] A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017.

[14] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[15] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[16] B. Zoph, "Neural architecture search with reinforcement learning," *arXiv preprint arXiv:1611.01578*, 2016.

[17] H. Liu, K. Simonyan, and Y. Yang, "Darts: Differentiable architecture search," *arXiv preprint arXiv:1806.09055*, 2018.

[18] Z. Li, H. Li, and L. Meng, "An efficient neural architecture search based on pcdarts," in *2024 International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI)*. IEEE, 2024, pp. 49–53.

[19] Z. Guo, Z. Wu, S. Liu, X. Ma, C. Wang, D. Yan, and F. Niu, "Defect detection of nuclear fuel assembly based on deep neural network," *Annals of Nuclear Energy*, vol. 137, p. 107078, 2020.

[20] Z. Liu, J. Liu, S. Shao, L. Cao, and H. Wu, "Deep learning-based prediction of transient power variation in pressurized water reactors," *Progress in Nuclear Energy*, vol. 185, p. 105756, 2025.

[21] S. Qamar, H. Jin, R. Zheng, P. Ahmad, and M. Usama, "A variant form of 3d-unet for infant brain segmentation," *Future Generation Computer Systems*, vol. 108, pp. 613–623, 2020.

[22] A. Ebrahimi, S. Luo, and R. Chiong, "Introducing transfer learning to 3d resnet-18 for alzheimer's disease detection on mri images," in *2020 35th international conference on image and vision computing New Zealand (IVCNZ)*. IEEE, 2020, pp. 1–6.

[23] X. Gao and Y. Pang, "Brain age prediction with 3d resnet34 model in healthy control, mild cognitive impairment, and alzheimer's disease," in *2022 3rd International Conference on Electronic Communication and Artificial Intelligence (IWECAI)*. IEEE, 2022, pp. 490–494.

[24] S. Serte and H. Demirel, "Deep learning for diagnosis of covid-19 using 3d ct scans," *Computers in biology and medicine*, vol. 132, p. 104306, 2021.

[25] C. Zhao, S. Xiang, Y. Wang, Z. Cai, J. Shen, S. Zhou, D. Zhao, W. Su, S. Guo, and S. Li, "Context-aware network fusing transformer and v-net for semi-supervised segmentation of 3d left atrium," *Expert Systems with Applications*, vol. 214, p. 119105, 2023.

[26] T. Uemura, J. J. Näppi, T. Hironaka, H. Kim, and H. Yoshida, "Comparative performance of 3d-densenet, 3d-resnet, and 3d-vgg models in polyp detection for ct colonography," in *Medical Imaging 2020: computer-aided diagnosis*, vol. 11314. SPIE, 2020, pp. 736–741.

[27] Z. Wang, C. Yao, J. Ren, M. Feng, and X. Jiang, "Human activity recognition using 3d orthogonally-projected efficientnet on radar time-range-doppler signature," in *2022 IEEE 2nd International Conference on Software Engineering and Artificial Intelligence (SEAI)*. IEEE, 2022, pp. 26–30.

[28] D. Wang, W. Wang, C. Pan, and D. Wang, "Prediction of key core parameter of pwr by adaptive bp neural network," *Atomic Energy Science and Technology*, vol. 54, no. 1, p. 112, 2020.