

基于ARGA-3D CNN的铅冷快堆三维中子 通量预测方法研究

杨子辉^{1,2} 莫紫雯^{1,2,3} 李中阳^{1,2} 孙国民^{1,2} 李兆东¹ 戈道川^{1,2} 郁杰^{1,2}

1(中国科学院合肥物质科学研究院 合肥 230031)

2(微小型反应堆技术安徽省重点实验室 合肥 230031)

3(淮师范大学 物理与电子信息学院 淮北 235000)

摘要 中子通量的三维预测对反应堆堆芯的设计、优化和安全分析至关重要,但由于微小型铅冷快堆空间紧凑且探测器布置困难,现有方法多集中在二维层面,较少关注三维通量的预测。本文提出了一种融合残差网络(Residual Network, ResNet)与多头自注意力机制(Multi-head Self Attention, MSA)的三维卷积神经网络(Genetic Algorithm-Enhanced 3D Convolutional Neural Network with Multi-Head Self-Attention and Residual Connections, ARGA-3D CNN)模型,该模型可以有效捕捉堆芯中子通量的空间分布特征,解决空间依赖性问题。通过ResNet缓解梯度消失与爆炸,增强训练稳定性,同时借助MSA强化关键区域识别。此外,采用遗传算法优化超参数,进一步提升堆芯中子通量预测精度。实验基于蒙特卡罗粒子输运模拟软件SuperMC计算结果构建数据集,并用该数据集训练与优化ARGA-3D CNN模型进行预测。结果显示,该模型预测值与SuperMC计算结果对比,在平均绝对误差(Mean Absolute Error, MAE)、均方误差(Mean Squared Error, MSE)和决定系数(R^2)指标上分别达到了 3.19×10^{-6} 、 2.14×10^{-11} 和0.973 5,计算效率有显著提升,单次预测仅耗时秒级,相比卷积神经网络(Convolutional Neural Network, CNN)、人工神经网络(Artificial Neural Network, ANN)、长短期记忆网络(Long Short-Term Memory, LSTM)以及Transformer等模型,预测效果更优。表明ARGA-3D CNN模型在三维中子通量预测中具有较高的精度和计算效率,为核反应堆堆芯参数的快速预测提供了新方法,具有一定的实用价值及意义。

关键词 铅冷快堆, 中子通量, 三维卷积神经网络, 多头自注意力机制, 残差网络, 遗传算法

中图分类号 TL99

DOI: 10.3724/j.0253-3219.2026.hjs.49.250171

CSTR: 32193.14.hjs.CN31-1342/TL.2026.49.250171

Prediction method of reactor core 3D neutron flux based on ARGA-3D CNN

YANG Zihui^{1,2} MO Ziwen^{1,2,3} LI Zhongyang^{1,2} SUN Guomin^{1,2} LI Zhaodong¹ GE Daochuan^{1,2} YU Jie^{1,2}

1(Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China)

2(Anhui Province Key Laboratory of Small Reactor and Micro-reactor Technology, Hefei 230031, China)

中国科学院战略性先导科技专项(No.XDA0410000)、中国科学院合肥研究院院长基金(No.YZJJ202403-CX)、安徽省自然科学基金面上项目(No.2008085MA23)资助

第一作者: 杨子辉, 男, 1988年出生, 2016年毕业于中国科学技术大学, 研究领域为蒙特卡罗粒子输运、多物理耦合、数字孪生等,

E-mail: zihui.yang@inest.cas.cn

共同第一作者: 莫紫雯, 女, 2000年出生, 2023年毕业于淮师范大学, 现为硕士研究生, 研究领域为先进核能系统设计与分析

收稿日期: 2025-04-25, 修回日期: 2025-06-20

Supported by the Strategic Priority Research Program of Chinese Academy of Sciences (No.XDA0410000), the President's Fund of Hefei Institutes of Physical Science, Chinese Academy of Sciences (No.YZJJ202403-CX), and the Natural Science Foundation of Anhui Province (No.2008085MA23)

First author: YANG Zihui, male, born in 1988, graduated from University of Science and Technology of China in 2016, focusing on Monte Carlo particle transport, multi-physics coupling, digital twin, etc., E-mail: zihui.yang@inest.cas.cn

Co-first author: MO Ziwen, female, born in 2000, graduated from Huaibei Normal University in 2023, master student, focusing on advanced nuclear energy system design and analysis

Received date: 2025-04-25, revised date: 2025-06-20

Abstract **[Background]** The three-dimensional prediction of neutron flux is crucial for the design, optimization, and safety analysis of reactor cores. However, due to the compact spatial configuration of small lead-cooled fast reactors and the difficulty in placing detectors, existing methods mainly focus on two-dimensional analysis, with limited attention to three-dimensional flux prediction. **[Purpose]** This study aims to propose a model called the Genetic Algorithm-Enhanced 3D Convolutional Neural Network with Multi-Head Self-Attention and Residual Connections (ARGA-3D CNN), which integrates Residual Networks (ResNet) and Multi-Head Self-Attention (MSA) mechanisms, to efficiently capture the spatial distribution features of neutron flux in reactor cores and address the issue of spatial dependencies. **[Methods]** The 3D convolutional neural networks (3D CNN) within this ARGA-3D CNN model were employed to effectively capture the spatial distribution of neutron flux in the reactor core, and the ResNet was used to alleviate the vanishing and exploding gradient problems, improving training stability, whilst the MSA was applied to enhance the identification of key regions. In addition, the Genetic Algorithm (GA) was used to optimize hyperparameters for further improving the accuracy of neutron flux prediction. Finally, this model was trained and optimized using a dataset constructed from results of Monte Carlo simulation software SuperMC, and thereafter applied to predicting 3D neutron flux of a small lead-cooled fast reactor. **[Results]** The prediction results show that the predicted values of the ARGA-3D CNN model, compared to the SuperMC simulation results, achieve Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2 values of 3.19×10^{-6} , 2.14×10^{-11} , and 0.973 5, respectively. The computational efficiency is significantly improved, with each prediction taking only seconds. Compared to traditional models such as Convolutional Neural Networks (CNN), Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM) networks and Transformer, the prediction performance is superior. **[Conclusions]** The ARGA-3D CNN model proposed in this study provides a high-precision and computationally efficient method for predicting three-dimensional neutron flux in a reactor core. It offers a new approach for the rapid prediction of nuclear reactor core parameters, which has significant practical value and implications for reactor design, operation, and safety analysis.

Key words Lead-cooled fast reactor, Neutron flux, 3D convolutional neural network, Multi-head self-attention, Residual network, Genetic algorithm

在核能技术的发展历程中,铅冷快堆作为一种具有独特优势的先进堆型,近年来受到了广泛关注^[1]。铅冷快堆以其优良的中子物理特性、较高的燃料利用率以及良好的安全性能,在核能领域展现出广阔的应用前景^[2]。然而,在堆芯设计、运行优化和安全分析中,堆芯中子通量分布作为评估堆芯性能的重要参数,其精确预测尤为关键^[3]。中子作为核反应的主要能量载体,其通量分布不仅直接影响反应堆的功率输出和燃料燃耗,还关系到反应性控制、热工水力特性以及辐照行为等关键因素^[4]。传统的中子通量预测方法,如基于确定性输运理论的数值计算和经验模型^[5],虽然能够在一定程度上提供有效结果,但在面对复杂堆芯结构和多物理场耦合问题时,常常存在计算成本高昂、计算速度慢以及对复杂工况适应性不足等问题^[6]。

近年来,人工智能技术,尤其是深度学习算法,因其在处理大规模数据和高效计算方面的优势,已被广泛应用于核反应堆物理参数的预测中^[7-10]。如

王端等^[11]提出的基于自适应反向传播神经网络(Backpropagation Neural Network, BPNN)的堆芯参数预测方法有效提高了预测精度,并减少了计算资源的消耗。Palmi等^[12]基于人工神经网络(Artificial Neural Network, ANN)的堆芯装载方案反应性及燃料循环周期预测方法,通过神经网络模型准确预测了不同堆芯装载配置下的反应性变化及燃料循环周期的长度。此外,Bei等^[13]基于BPNN开发了三维代理模型,显著提高了计算效率,尤其是在熔盐堆的中子通量预测中展现了其优势。陈镜宇等^[14]通过优化极限学习机(Extreme Learning Machine, ELM)模型并结合鲸鱼优化算法,进一步提升了中子通量和 k_{eff} 的预测精度,在提高训练效率方面取得了显著成效。尽管以上研究在反应堆中子通量和关键参数的预测上取得了显著进展,但现有方法在面对堆芯空间紧凑且结构复杂时,常面临计算精度与效率之间的平衡挑战,尤其在处理三维中子通量分布时仍显不足。

为提高堆芯三维中子通量分布的预测精度,解决现有方法在处理复杂空间结构和非线性关系时的局限性,本文提出了一种结合残差网络和自注意力机制的三维卷积神经网络(Genetic Algorithm-Enhanced 3D Convolutional Neural Network with Multi-Head Self-Attention and Residual Connections, ARGA-3D CNN)模型。该模型通过引入残差结构有效避免了梯度消失问题,同时利用自注意力机制自动识别并聚焦输入数据中的关键特征,从而提升了对堆芯三维中子通量分布的预测精度。通过遗传算法优化网络超参数,显著提升了模型性能。实验结果表明,ARGA-3D CNN模型能够高效处理三维复杂数据,尤其在三维中子通量预测中展现了显著的优势。

1 模型与方法

1.1 三维卷积模块

三维卷积神经网络(3D Convolutional Neural Network, 3D CNN)是卷积神经网络(Convolutional Neural Network, CNN)在三维数据处理中的扩展,通过引入三维卷积核,能够同时提取空间和时间特征,突破了二维卷积神经网络(2D Convolutional Neural Network, 2D CNN)仅能捕捉二维空间特征的局限^[15]。在图1(a)中,2D CNN应用于单张图像时,仅在空间维度滑动,处理后的结果为二维图像,而在图1(c)中,3D CNN则在空间和时间维度上都进行滑动,输出保留时间信息的三维体积(volume)。这种3D卷积操作特别适用于视频分析、动作识别^[16]和医学影像处理^[17]等任务,相较于2D卷积,展现出显著优势。

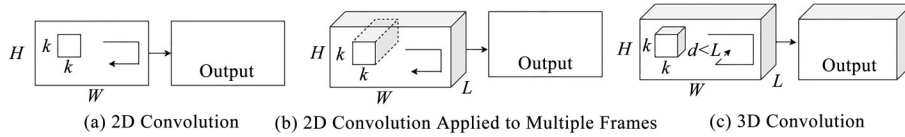


图1 3D CNN操作示意图 (a) 2D卷积, (b) 作用于多个帧的2D卷积, (c) 3D卷积

Fig.1 Diagram of 3D CNN operations (a) 2D convolution, (b) 2D convolution applied to multiple frames, (c) 3D convolution

在三维卷积中,卷积核的深度通常小于输入数据的深度(即核大小小于通道大小),如图2所示,使得过滤器可以在图像的高度、宽度和通道三个方向上滑动。每当卷积核移动到新的位置时,都会进行逐元素的乘法和加法操作,产生一个输出值。由于过滤器在整个三维空间中滑动,输出数据也按三维结构排列。三维卷积公式如下:

$$Y(t, i, j) = \sum_{d=0}^{D-1} \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} x(t+d, i+m, j+n) \times W(d, m, n) + b \quad (1)$$

式中: $Y(t, i, j)$ 表示输出特征图在时间步 t 的值; $x(t+d, i+m, j+n)$ 表示输入数据在时间步 $t+d$ 、位置 $(i+m, j+n)$ 的值; $W(d, m, n)$ 表示三维卷积核在位置 (d, m, n) 的权重; b 表示偏置项,且卷积核大小为 $D \times k \times k$ (时间深度 $\times$ 高度 $\times$ 宽度)。

在本文中,三维卷积操作的三个维度分别对应

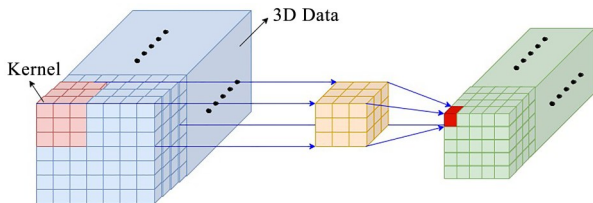


图2 三维卷积核

Fig.2 Diagram of 3D convolution kernel

于反应堆堆芯的空间坐标轴,卷积核对输入数据进行多维滑动操作,相当于在堆芯不同位置提取中子通量分布的局部空间特征。每次卷积的输出可视为在特定空间邻域内的中子通量特征响应,从而帮助网络捕捉三维通量在堆芯结构内的空间分布规律。

1.2 多头自注意力模块

多头自注意力机制(Multi-head Self Attention, MSA)通过并行计算多个注意力头,来提升模型对不同信息的处理能力^[18]。每个注意力头通过对输入的查询(Q)、键(K)和值(V)进行线性变换,计算出各自的注意力得分,并对结果进行加权求和,从而捕捉到输入的不同子空间特征。多个注意力头并行工作,使得模型能够在不同子空间上同时关注信息,进而提升了模型的表达能力和捕捉长程依赖的能力。图3中展示了多头注意力机制的实现过程:每个头通过线性变换分别从输入的 Q, K, V 中提取特征,然后对每个头的输出进行拼接(Concat),最后通过一个线性变换得到最终的输出。

具体计算过程如下:

对于每个注意力头,首先计算注意力得分,

$$\text{Attention}(K, V, Q) = \text{Softmax}(\text{score}(Q, K)) \times V \quad (2)$$

其中, $\text{score}(Q, K)$ 通常是通过点积计算得出。然后,多个头的输出通过拼接操作(Concat)结合起来,

最后通过线性变换得到最终结果:

$\text{MultiHeadSelfAttention}(K, V, Q) =$

$$\text{Concat}(\text{head}_1, \text{head}_2, \text{head}_3) \times W^0 \quad (3)$$

式中: W^0 是最终的权重矩阵。这种机制使得模型能够同时在不同的子空间中捕捉到多层次的特征信息,从而提升了整体的表示能力和泛化能力。

本文采用多头自注意力机制,其基本原理为通过并行计算多个注意力头,模型能够自动识别堆芯中子通量分布中的关键区域,并根据这些区域的中子通量对整体堆芯通量进行加权调整,从而能够更加精准地捕捉全局的空间特征,提升对复杂空间结构的建模能力。

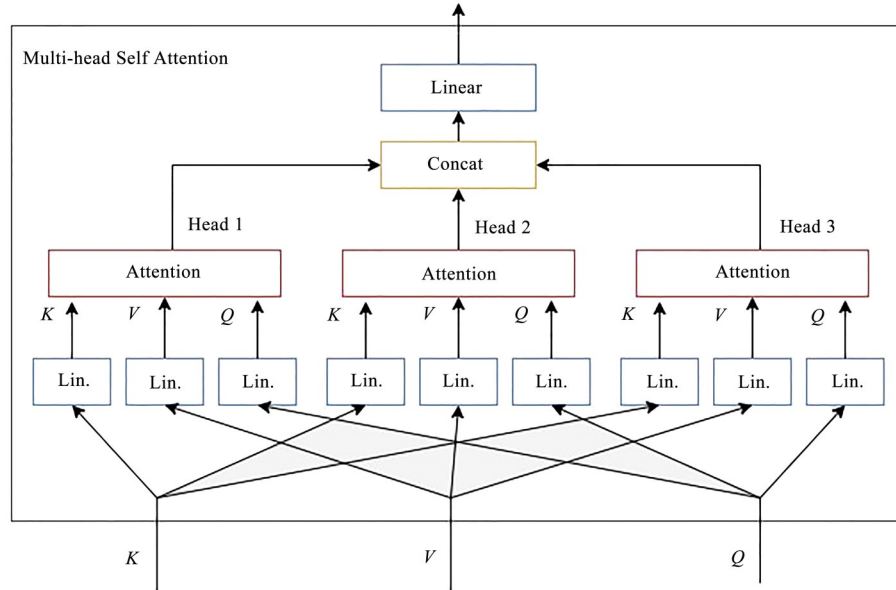


图3 多头自注意力模块
Fig.3 Diagram of multi-head self attention module

1.3 残差连接模块

残差连接(Residual Connections)是解决深层神经网络训练中梯度消失问题的有效技术^[19]。它通过直接将输入数据传递到网络的更深层,避免了梯度在反向传播过程中的衰减,从而加速了梯度的传播。He等^[20]在2015年提出的网络结构ResNet(Residual Network)的核心是引入了残差连接。图4描述了神经网络中的两种模块结构,正常块是神经网络中的基本构建单元,通常由权重层和激活函数组成,输入 x 经线性变换和非线性激活后输出 $f(x)$ 。残差连接块是ResNet中的关键组件,它通过引入残差学习来解决深度网络中的退化问题。残差连接块包括权重层、激活函数和残差连接。输入 x 首先通过权重层进行线性变换,然后通过激活函数处理,得到 $f(x)$ 。接着,将 $f(x)$ 与原始输入 x 相加,得到最终输出。这种结构可以表示为 $y=f(x)+x$,其中 y 是输出, $f(x)$ 是经过权重层和激活函数处理的结果, x 是输入。

本文采用残差连接的物理意义在于通过保留输入的原始信息,避免了深层神经网络训练中的梯度消失问题,从而帮助网络更好地传递堆芯中子通量的空间分布特征。特别是在处理堆芯中非线性、复杂的通量分布时,残差连接确保了模型能够更准确

地表达各区域间的依赖关系,从而提高了模型对中子通量分布的学习能力。

1.4 AR-3D CNN模型

AR-3D CNN模型的创新在于结合了残差网络、

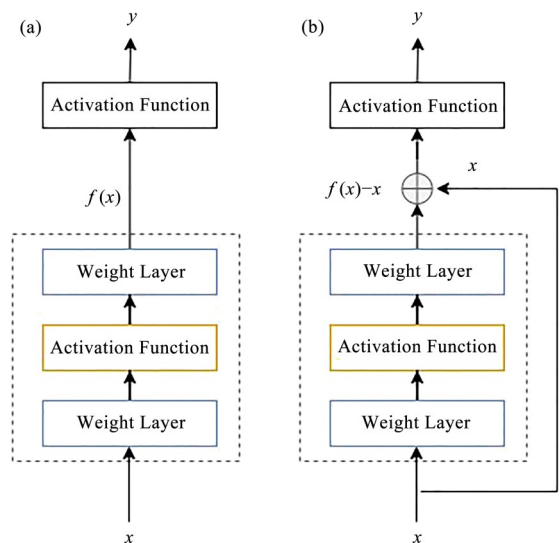


图4 残差连接模块 (a) 一个正常块, (b) 一个残差连接块
Fig.4 Residual connection module (a) A normal block, (b) A residual connection block

多头自注意力机制和三维卷积神经网络等结构,针对堆芯中子通量分布预测任务的复杂需求,提升了模型的预测精度、计算效率和鲁棒性,同时增强了对三维空间数据的处理能力,模型的整体流程如图5所示。模型包括以下几个主要步骤:1)在处理输入层数据时,通过3D CNN提取堆芯中子通量数据的三维空间特征,结合批归一化(BatchNorm)和ReLU激活函数,增强非线性表达能力;2)使用多个残差3D卷积块(Residual 3D ConvBlock)提取深层特征,避免梯度问题并学习复杂空间依赖;3)通过多头注意力层(Multi-Head Attention Layer)动态关注关键区域,提升特征学习和噪声鲁棒性;4)全连接层(Fully Connected Layer)进一步整合特征,完成非线性映射和最终预测;5)通过平均池化层(Average Pooling Layer)聚合特征图,优化输出结果。

(Fully Connected Layer)进一步整合特征,完成非线性映射和最终预测;5)通过平均池化层(Average Pooling Layer)聚合特征图,优化输出结果。

1.5 ARGA-3D CNN模型

本文结合深度学习与遗传算法的优化方法,采用基于验证集损失最小化的超参数优化策略^[21]。在该方法中,首先利用深度学习模型对输入数据进行特征提取和多任务预测建模,网络结构包括3D卷积层、残差连接层以及多头注意力机制等模块,旨在捕捉输入特征的时空特征和全局依赖关系。接着,遗传算法以验证集上的损失函数(Huber损失)^[22]的负值作为适应度函数进行定义:

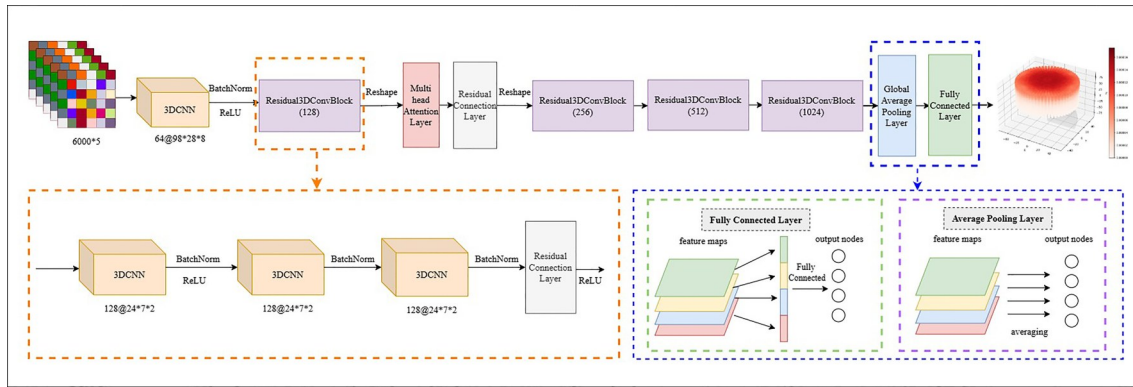


图5 AR-3D CNN网络结构图
Fig.5 AR-3D CNN network architecture diagram

$$\text{Fitness} = -\text{Loss} \quad (4)$$

$$\text{Loss} = \frac{1}{N} \sum_{i=1}^N \text{Huber}(Y_{\text{true}}^{(i)}, Y_{\text{pred}}^{(i)}) \quad (5)$$

式中: N 表示验证集样本数量; $Y_{\text{true}}^{(i)}$ 表示第 i 个样本的真实值; $Y_{\text{pred}}^{(i)}$ 表示第 i 个样本的预测值。Huber损失函数通过结合均方误差(Mean Squared Error, MSE)和平均绝对误差(Mean Absolute Error, MAE)的优点,能够在保证平滑性的同时,提高对异常值的鲁棒性,优化模型的稳定性和收敛性。

ARGA-3D CNN模型通过3D CNN提取三维特征,并利用遗传算法优化模型参数,提高预测精度和泛化能力。模型的整体流程如图6所示,包括两个主要部分:预测过程和遗传算法优化过程。首先,输入数据经过预处理后,初始化3D CNN模型并进行优化,以提取堆芯中子通量的空间特征;然后,构建最佳模型并进行训练,输出预测结果;接着,遗传算法优化超参数,通过选择、交叉和变异操作迭代搜索最佳超参数组合,进一步提高模型的预测精度和泛化能力;最终,输出最佳超参数,确保模型能在复杂的三维数据场景中获得最优的表现。

2 仿真实验

2.1 堆芯模型

本实验使用蒙特卡罗粒子输运模拟软件SuperMC构建铅冷快堆模型,堆芯组件的数量与材料汇总于表1。如图7所示,堆芯组件采用快堆常用的六棱柱形组件,堆芯由内到外分为三个区域:活性区、反射区和屏蔽区。活性区由101个组件组成,包括92个燃料组件(主要材料为19.75%的 UO_2)和9个控制棒组件;反射层组件共有98个,采用铅铋共晶材料;屏蔽组件为54个,材料为碳化硼。此外,安全棒、控制棒与补偿棒的数量分别为4根、3根和2根,材料分别为90% ^{10}B 的碳化硼、碳化硼和90% ^{10}B 的碳化硼。

SuperMC是一款由中国科学院合肥物质科学研究院核能安全技术研究所(Institute of Nuclear Energy Safety Technology, INEST)自主研发的大型核能中子学设计与辐射安全评价软件系统。该软件基于蒙特卡罗方法,能够对复杂的核系统进行精确的中子输运计算,并广泛应用于核反应堆设计、辐射安全分析及核能工程的各个方面。SuperMC能够处

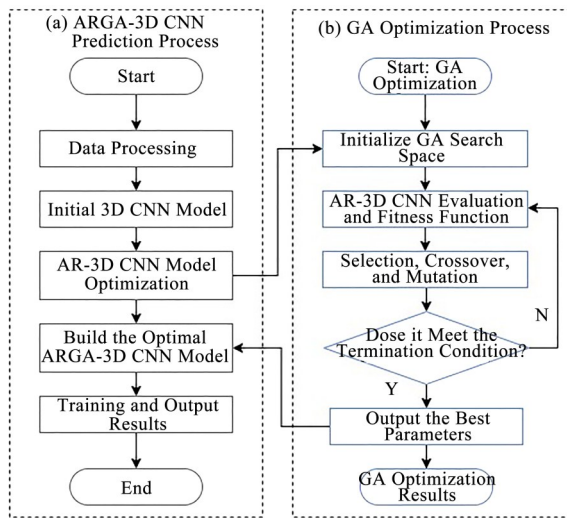


图6 ARGAs-3D CNN模型整体流程图
Fig.6 Overall flowchart of the ARGAs-3D CNN model

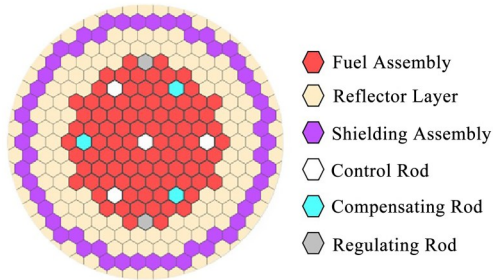


图7 铅冷快堆堆芯布局(彩图见网络版)
Fig.7 Layout of lead-cooled fast reactor core (color online)

理复杂几何形状的反应堆模型,并对多物质界面进行有效地模拟,提供高精度的中子流计算。其主要功能涵盖了核能系统中的各类重要分析,包括辐射输运、燃料消耗、辐射活化、剂量评估等全过程的中子安全分析。图8为使用SuperMC构建的铅冷快堆堆芯整体结构图。

2.2 实验数据集及评价指标

数据集通过堆芯模拟程序在考虑稳态工况的情况下进行堆芯中子通量密度的计算得到。输入数据包括三根控制棒的位置、两根补偿棒的位置以及30 000个空间坐标信息,这些参数共同描述了堆芯的物理状态及其对中子通量分布的影响。输出数据为与空间坐标对应的30 000个中子通量值,反映了在不同操作条件下,堆芯中各位置的中子通量强度,数据如表2所示。为了确保模型训练的稳定性与高效性,所有输入和输出数据均进行了归一化处理。通过这一数据集,训练模型学习输入参数与中子通量分布之间的复杂非线性关系,从而实现对堆芯中子通量的准确预测。

本实验采用4个广泛使用的指标,即MAE、

表1 铅冷快堆堆芯主要设计参数
Table 1 Main design parameters of the critical core of the lead-cooled fast reactor

参数 Parameter	值 Value
燃料富集度 Fuel enrichment	UO ₂ (19.75%)
冷却剂 Coolant	LBE
包壳 Cladding	316Ti
活性区高度 Activate height / mm	800
燃料组件数 Number of fuel assemblies	92
燃料组件内燃料棒数 Fuel pins per assembly	61
燃料组件间距 Fuel assemblies pitch / mm	120.7
燃料棒直径 Fuel rod diameter / mm	12
燃料芯块直径 Pellet diameter / mm	10.9
燃料棒包壳厚度 Cladding thickness / mm	0.4
反射层组件数 Number of reflector layer assemblies	98
屏蔽组件数 Number of shielding assemblies	54

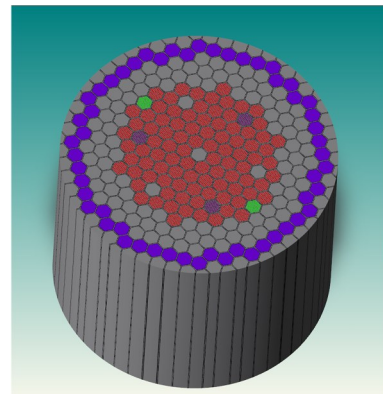


图8 铅冷快堆堆芯整体结构示意图(彩图见网络版)
Fig.8 Diagram of lead-cooled fast reactor core overall structure (color online)

表2 实验数据集
Table 2 Experimental dataset

模型参数 Model parameter	三维中子通量参数 Three-dimensional neutron flux parameter
输入参数 Input parameter	3根控制棒位置 3 control rod positions 2根补偿棒位置 2 compensating rod positions 空间坐标信息(6 000×5个) Spatial coordinate information (6 000×5 entries)
输出参数 Output parameter	中子通量分布(6 000×5个) Neutron flux distribution (6 000×5 entries)

MSE、平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)和决定系数(R^2),来衡量模型的准确性,评估预测模型的性能,反映了模型对数据的拟合效果。公式如下:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (7)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (8)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (9)$$

式中： y_i 是真实值； $\hat{y}_i$ 是预测值； n 是样本数量； $\bar{y}$ 是真实值的均值。 R^2 取值范围是 $[0, 1]$ ，值越接近1，表示模型拟合效果越好。

2.3 超参寻优

实验采用遗传算法来调整神经网络的超参数，以优化模型性能。遗传算法能够在高维度的搜索空间中进行全局优化，逐步找到最优的超参数组合^[23]。具体参数设置如下：种群大小为50，代数为100，交叉概率为0.8，变异概率为0.1，并采用轮盘赌选择方法进行个体选择。这种参数配置在计算效率和搜索

精度之间取得了较好的平衡，能够有效避免局部最优解，确保全局优化的效果。

由于数据集的维度和结构相对复杂且数据量较大，本实验采用混合策略进行超参数优化，其中学习率作为连续超参数使用对数尺度进行搜索，其他超参数（如批量大小、自注意力头数和全连接层节点数）则采用离散选择策略。这样的混合策略可以兼顾计算效率和搜索精度，优化过程既能细致地调整学习率，又能在离散超参数的有限集合中找到最优值。

经过遗传算法的多轮迭代与优化，最终确定了一组超参数配置，如表3中搜索结果所示，具体超参数为学习率 1.17×10^{-4} ，批量大小16，注意力头数4，全连接层节点数量1 024，代表了在当前搜索空间内模型性能的最优解。训练过程中，训练损失（Train Loss）和验证损失（Validation Loss）都显著下降，如图9所示，说明模型在学习过程中逐渐收敛并取得了较好的表现。从损失曲线来看，随着训练的推进，损失值趋于平稳，验证了优化过程的有效性。

表3 ARGA-3D CNN模型超参数搜索范围及搜索结果
Table 3 Hyperparameter search space and search results of the ARGA-3D CNN model

参数 Parameter	搜索空间范围 Search space range	搜索结果 Search result
学习率 Learning rate	$[1 \times 10^{-6}, 1 \times 10^{-2}]$	1.17×10^{-4}
批量大小 Batch size	[4, 16, 32, 128]	16
注意力头数 Number of attention heads	[2, 4, 8, 16]	4
全连接层节点数量 Number of fully connected layer nodes	[256, 512, 1 024, 2 048]	1 024

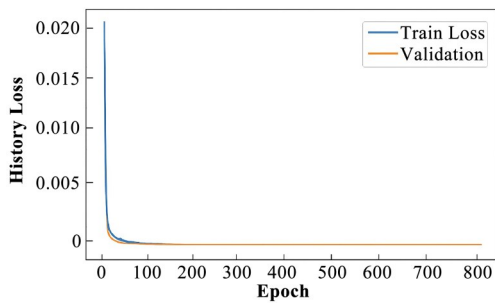


图9 训练和验证损失曲线(彩图见网络版)
Fig.9 Training and validation loss curve (color online)

2.4 对比实验

为了全面评估本文所提出 ARGA-3D CNN 模型的有效性，我们与多种先进的深度学习预测模型进行了对比实验。对比模型包括传统的CNN、ANN以及长短时记忆网络（Long Short-Term Memory, LSTM）以及Transformer，这些方法广泛应用于预测任务。所有模型均在铅冷快堆堆芯三维中子通量数据集上进行训练和测试。

对比实验的结果如表4所示，在MAE和MSE方

面，ARGA-3D CNN 的误差仅为 3.19×10^{-6} 和 2.14×10^{-11} ，相比于CNN、ANN、LSTM和Transformer，误差降低了一半以上，这表明该方法在预测中子通量值时具备极高的精度，能够有效避免传统方法常见的预测偏差问题。由于三维中子通量具有较复杂的空间分布特性，传统的CNN和ANN由于其局部感受野的限制，难以有效建模全局空间关系，导致预测误差较大。而ARGA-3D CNN结合了残差网络、多头自注意力机制和三维卷积神经网络，能够更好地学习全局空间特征。Transformer虽然也具备强大的全局建模能力，但在处理三维空间数据时，其计算复杂度较高，且对数据的局部特征捕捉能力相对较弱，因此在本实验中表现不如ARGA-3D CNN。

在MAPE方面，ARGA-3D CNN的相对误差仅为2.13%，相比CNN、ANN、LSTM和Transformer，减少了超过60%。在 R^2 方面，ARGA-3D CNN取得0.973 5的最佳拟合优度，远高于CNN的0.911 8、ANN的0.913 4、LSTM的0.921 2和Transformer的0.951 6。这表明，ARGA-3D CNN在三维空间特征

提取的优势,使得其能够有效学习中子通量的非线性特征,并构建出更符合物理规律的预测模型。为了进一步直观展示不同模型的表现,本文提供了可视化实验结果。

在测试数据集上对ARGA-3D CNN模型预测的中子通量分布与蒙特卡罗程序计算的分布进行了比较。考虑到模型预测值与计算值之间存在一定的相对误差,从测试数据集中随机抽取了一个实例以评估替代模型的性能。图10展示了预测与实际中子通量分布的二维对比图,而图11则提供了三维对比图。总体来看,预测的中子通量分布与实际分布吻

表4 ARGA-3D CNN模型对比实验结果
Table 4 Comparison of experimental results of the ARGA-3D CNN model

指标 Metrics	ARGA- 3D CNN	Transformer CNN	ANN	LSTM	
MAE	3.19×10^{-6}	1.75×10^{-2}	2.84×10^{-2}	2.82×10^{-2}	2.57×10^{-2}
MSE	2.14×10^{-11}	5.09×10^{-4}	1.48×10^{-3}	1.41×10^{-3}	1.29×10^{-3}
MAPE	2.13%	3.49%	8.06%	7.46%	6.59%
R^2	0.973 5	0.951 6	0.911 8	0.913 4	0.921 2

合良好,并且在反应堆堆芯中观察到了最高的中子通量,这与反应堆堆芯中子通量的预期行为完全一致。

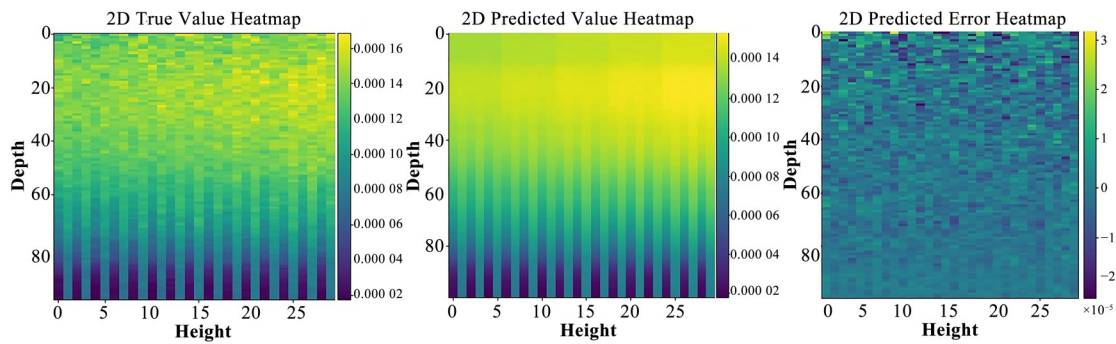


图10 真实中子通量分布、预测中子通量及其对应误差的二维图像(彩图见网络版)
Fig.10 2D image of actual neutron flux distribution, predicted neutron flux, and corresponding error (color online)

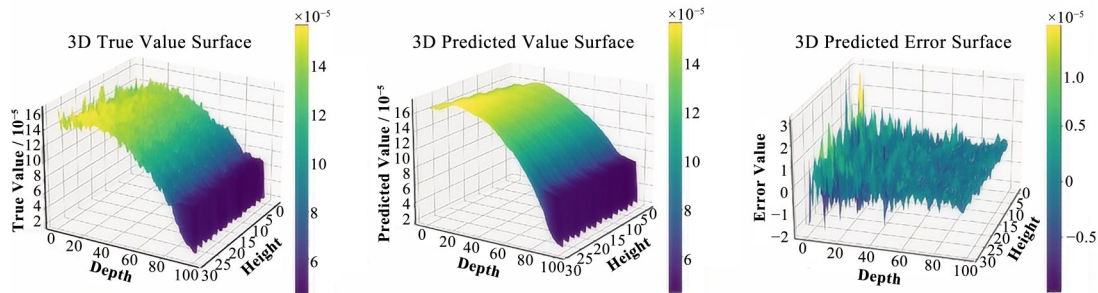


图11 真实中子通量分布、预测中子通量及其对应误差的三维图像(彩图见网络版)
Fig.11 3D image of actual neutron flux distribution, predicted neutron flux, and corresponding error (color online)

综上所述,ARGA-3D CNN通过深度三维卷积特征提取、残差网络的优化训练稳定性以及自注意力机制的全局信息建模,在中子通量预测任务中显著优于其他模型。相比于CNN、ANN、LSTM和Transformer,ARGA-3D CNN误差更低、预测更精准、模型更稳定,且在复杂数据环境下仍能保持良好的鲁棒性,为核反应堆设计和安全评估提供了一种新的高效、准确的预测方法,具有一定的应用前景。

2.5 消融实验

为了深入分析ARGA-3D CNN各个模块对模型整体性能的贡献,我们进行了消融实验,分别去除了遗传算法(AR-3D CNN)、残差网络(A-3D CNN)、多头自注意力机制(R-3D CNN)以及全局平均池化层

(AR-3D CNN(NAP)),并对比各版本模型的性能变化,以验证不同模块对最终预测结果的影响,实验结果如表5所示。

通过表5可见,提出的ARGA-3D CNN网络中

表5 ARGA-3D CNN模型消融实验结果
Table 5 Ablation experiment results of the ARGA-3D CNN model

指标 Metrics	ARGA- 3D CNN	AR-3D CNN	A-3D CNN	R-3D CNN	AR-3D CNN (NAP)
MAE	3.19×10^{-6}	3.62×10^{-6}	6.27×10^{-6}	4.45×10^{-6}	3.81×10^{-6}
MSE	2.14×10^{-11}	2.50×10^{-11}	6.40×10^{-11}	3.40×10^{-11}	2.70×10^{-11}
MAPE	2.13%	2.96%	5.53%	3.74%	3.20%
R^2	0.973 5	0.963 6	0.906 5	0.949 5	0.960 8

每个模块都能有效提升其在三维中子通量预测任务上的表现。通过引入遗传算法(Genetic Algorithm, GA)进行超参数寻优,显著增强了网络的全局搜索能力,降低了手动调参带来的局部最优风险,并进一步提高了模型的精度与稳定性。ResNet模块通过在深层结构中引入跳跃连接缓解了梯度消失和梯度爆炸问题,同时增强了特征传递效率,对网络性能的贡献尤为突出。MSA 则能精准捕捉输入数据在三维空间中的全局依赖关系,帮助模型关注关键区域并抑制干扰,使预测更为准确。全局平均池化层(Global Average Pooling layer, GAP)在压缩通道维度、减少冗余参数的同时,保持了主干特征的完整性,进而提升网络的泛化能力。当去除了GAP后,网络对细节特征的利用率有所下降,导致重建效果略微变差。整体而言,这些模块的协同作用使 ARG-3D CNN 对复杂的三维中子通量数据具有更强的适应性和更高的预测精度,为后续核反应堆堆芯关键参数的预测与分析奠定了坚实基础。

2.6 时效实验

本实验旨在评估所提出的 ARG-3D CNN 模型在时效性方面的优势,特别是在与核能仿真计算软件 SuperMC 的对比中。所有时效性实验均在相同硬件环境下进行,以确保结果的可比性。实验所用计算机配置为:CPU 为 AMD Ryzen 7 5800,内存 32 GB,显卡为 NVIDIA GeForce RTX 3060。通过比较两种方法在计算三维中子通量预测时所需的时间,来评估其时效性。在使用 SuperMC 进行 12 核并行计算时,完成一次完整预测所需的时间为分钟级别。相比之下,使用 ARG-3D CNN 模型进行相同预测时,计算时间仅为 0.401 1 s。由此可见,该模型在计算时效性方面具有显著优势,尤其是在处理复杂预测任务时,能够在极短时间内完成计算,大幅提升了处理效率。

3 结语

本文提出的基于 ARG-3D CNN 的三维中子通量预测模型,旨在解决铅冷快堆堆芯中子通量分布预测中的复杂空间特征提取问题,并显著提高了预测精度和计算效率。主要结论如下:

1)通过引入 ResNet 与 MSA,模型有效缓解了梯度消失问题,并增强了全局信息建模能力,从而在中子通量预测任务中提高了预测精度,减少了传统方法的预测误差。相比于 CNN、ANN、LSTM 和 Transformer, ARG-3D CNN 模型在 MAE、MSE、MAPE 和 R^2 等评价指标上均表现出优越性,且预测

30 000 个数据仅需 0.401 1 s,与传统计算软件相比,计算效率大幅提高。

2)结合遗传算法优化超参数,进一步提高了模型的鲁棒性与精度。经过优化后的超参数配置使得模型在大规模堆芯数据的预测任务中表现更为高效,同时具有更强的泛化能力。

综上所述,本文提出的模型为三维中子通量的高效预测提供了一种新的思路,尤其在核反应堆设计与安全评估中具有一定的应用前景。在后续工作中,将引入更加多样化的数据集,涵盖不同工况和操作条件,旨在进一步提升模型在实际应用中的适应性和鲁棒性。

作者贡献声明 杨子辉负责对文章作批评性审阅,研究经费支持;莫紫雯负责论文整体设计,神经网络程序设计,起草文章;李中阳、孙国民、戈道川、郁杰负责提供技术支持与指导;李兆东负责获取堆芯有效数据,数据分析。

参考文献

- Wei S Y, Wang C L, Tian W X, *et al.* Research progress in key thermal-hydraulic issue of lead-based fast reactor[J]. Atomic Energy Science and Technology, 2019, **53**(2): 326–336. DOI: 10.7538/yzk.2018.youxian.0335.
- Song T X, Zhang L, Gu Z X, *et al.* Transient analyses on overpower transients of small modular natural circulation Lead-cooled fast reactor SNCLFR-100[J]. Kerntechnik, 2025, **90**(2): 157–171. DOI: 10.1515/kern-2024-0076.
- Li J, Li W, Zhang X, *et al.* Research status and suggestions of nuclear reactor online core monitoring system[J]. Atomic Energy Science and Technology, 2024, **58**(7): 1393–1405. DOI: 10.7538/yzk.2024.youxian.0147.
- Yu Y, Park H, Novak A, *et al.* High fidelity multiphysics tightly coupled model for a lead cooled fast reactor concept and application to statistical calculation of hot channel factors[J]. Nuclear Engineering and Design, 2025, **435**: 113915. DOI: 10.1016/j. nucengdes. 2025. 113915.
- Zare S G, Ehsan Z. Neutronic investigation of ALFRED lead-cooled fast reactor core using Monte Carlo modeling [J]. Atomic Energy, 2024, **134**(3–4): 264–274. DOI: 10.1007/S10512-024-01054-2.
- 张大可, 唐智辉, 韦应靖, 等. 基于蒙特卡罗法的中子多球谱仪测量不确定度分析方法及应用[J]. 核技术, 2024, **47**(12): 120403. DOI: 10.11889/j.0253-3219.2024.hjs.47.120403. CSTR: 32193.14. hjs. CN31-1342/TL. 2024.47. 120403.

- ZHANG Dake, TANG Zhihui, WEI Yingjing, *et al.* Uncertainty analysis method and application of neutron multisphere spectrometer measurement based on Monte Carlo method[J]. Nuclear Techniques, 2024, **47**(12): 120403. DOI: 10.11889/j.0253-3219.2024.hjs.47.120403. CSTR: 32193.14.hjs.CN31-1342/TL.2024.47.120403.
- 7 王彤, 刘紫静, 赵鹏程, 等. 基于BP神经网络的高通量铅铋反应堆中子通量最大化设计方法研究[J]. 核技术, 2024, **47**(10): 100602. DOI: 10.11889/j.0253-3219.2024.hjs.47.100602. CSTR: 32193.14.hjs.CN31-1342/TL.2024.47.100602.

WANG Tong, LIU Zijing, ZHAO Pengcheng, *et al.* Design method of high-flux lead-bismuth cooled reactor neutron flux maximization based on BP neural network [J]. Nuclear Techniques, 2024, **47**(10): 100602. DOI: 10.11889/j.0253-3219.2024.hjs.47.100602. CSTR: 32193.14.hjs.CN31-1342/TL.2024.47.100602.

 - 8 张海明, 张昊春. 基于卷积神经网络模型的堆芯有效增殖因数预测方法研究[J]. 现代应用物理, 2022, **13**(2): 61–66. DOI: 10.12061/j.issn.2095-6223.2022.020209.

ZHANG Haiming, ZHANG Haochun. Prediction method of effective multiplication factor of reactor core based on convolution neural network model[J]. Modern Applied Physics, 2022, **13**(2): 61 – 66. DOI: 10.12061/j. issn. 20956223.2022.020209.

 - 9 周天泽, 虞凯程, 程懋松, 等. 基于KNN方法的熔盐堆系统瞬态识别模型开发及分析[J]. 核技术, 2023, **46**(11): 110604. DOI: 10.11889/j.0253-3219.2023.hjs.46.110604.

ZHOU Tianze, YU Kaicheng, CHENG Maosong, *et al.* Development and analysis of a transient identification model for molten salt reactor systems based on the KNN method[J]. Nuclear Techniques, 2023, **46**(11): 110604. DOI: 10.11889/j.0253-3219.2023.hjs.46.110604.

 - 10 冀南, 易金豪, 赵鹏程, 等. 基于自适应RBF神经网络预测堆芯热工水力参数的方法研究[J]. 核技术, 2022, **45**(9): 090601. DOI: 10.11889/j. 0253-3219.2022. hjs. 45. 090601.

JI Nan, YI Jinhao, ZHAO Pengcheng, *et al.* Research on adaptive RBF neural network prediction method for core thermal-hydraulic parameters of fast reactor[J]. Nuclear Techniques, 2022, **45**(9): 090601. DOI: 10.11889/j.0253-3219.2022.hjs.45.090601.

 - 11 王端, 王威策, 潘翠杰, 等. 基于自适应BP神经网络的压水堆堆芯换料关键参数的预测方法[J]. 原子能科学技术, 2020, **54**(1): 112 – 118. DOI: 10.7538/yzk. 2019. youxian.0016.

WANG Duan, WANG Weice, PAN Cuijie, *et al.* Prediction method for key parameters of core refueling in pressurized water reactors based on an adaptive BP neural network[J]. Atomic Energy Science and Technology, 2020, **54**(1): 112 – 118. DOI: 10.7538/yzk.2019.youxian.0016.

 - 12 Palmi K, Kubinski W, Darnowski P. Prediction of the evolution of the nuclear reactor core parameters using artificial neural network[J]. Annals of Nuclear Energy, 2025, **211**: 110891. DOI: 10.1016/j.anucene.2024.110891.
 - 13 Bei X Y, Dai Y Q, Yu K C, *et al.* Three-dimensional surrogate model based on back-propagation neural network for key neutronics parameters prediction in molten salt reactor[J]. Energies, 2023, **16**(10): 4044. DOI: 10.3390/en16104044.
 - 14 陈镜宇, 刘喜洋, 赵鹏程, 等. 基于优化极限学习机模型的反应堆中子通量与 k_{eff} 预测方法研究[J]. 核技术, 2024, **47**(10): 100604. DOI: 10.11889/j.0253-3219.2024.hjs.47.100604. CSTR: 32193.14.hjs.CN31-1342/TL.2024.47.100604.

CHEN Jingyu, LIU Xiyang, ZHAO Pengcheng, *et al.* Study on the prediction method of neutron flux and k_{eff} in reactors based on the optimized extreme learning machine model[J]. Nuclear Techniques, 2024, **47**(10): 100604. DOI: 10.11889/j.0253-3219.2024. hjs. 47.100604. CSTR: 32193.14.hjs.CN31-1342/TL.2024.47.100604.

 - 15 Liu S Q, Wang X, Zhao L, *et al.* 3DCANN: A spatio-temporal convolution attention neural network for EEG emotion recognition[J]. IEEE Journal of Biomedical and Health Informatics, 2022, **26**(11): 5321 – 5331. DOI: 10.1109/JBHI.2021.3083525.
 - 16 白静, 杨瞻源, 彭斌, 等. 三维卷积神经网络及其在视频理解领域中的应用研究[J]. 电子与信息学报, 2023, **45**(6): 2273–2283. DOI: 10.11999/JEIT220596.

BAI Jing, YANG Zhanyuan, PENG Bin, *et al.* Research on 3D convolutional neural network and its application to video understanding[J]. Journal of Electronics & Information Technology, 2023, **45**(6): 2273 – 2283. DOI: 10.11999/JEIT220596.

 - 17 李泽慧, 张琳, 山显英. 三维卷积神经网络方法改进及其应用综述[J]. 计算机工程与应用, 2025, **61**(3): 48–61. DOI: 10.3778/j.issn.1002-8331.2407-0031.

LI Zehui, ZHANG Lin, SHAN Xianying. Review of improvements in three-dimensional convolutional neural network methods and their applications[J]. Journal of Computer Engineering & Applications, 2025, **61**(3): 48–

61. DOI: 10.3778/j.issn.1002-8331.2407-0031.
- 18 Gao G S. Survey on attention mechanisms in deep learning recommendation models[J]. Computer Engineering and Applications, 2022, **58**(9): 9 – 18. DOI: 10.3778/j.issn.1002-8331.2112-0382.
- 19 Alaeddine H, Jihene M. Deep residual network in network [J]. Computational intelligence and neuroscience, 2021, **2021**(1): 6659083. DOI: 10.1155/2021/6659083.
- 20 He K M, Zhang X Y, Ren S Q, *et al.* Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). June 27-30, 2016, Las Vegas, NV, USA: IEEE, 2016: 770 – 778. DOI: 10.1109/CVPR.2016.90.
- 21 Bouktif S, Fiaz A, Ouni A L, *et al.* Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: comparison with machine learning approaches[J]. Energies, 2018, **11**(7): 1636. DOI: 10.3390/en11071636.
- 22 Ozturk N, Celik E. Application of genetic algorithms to core loss coefficient extraction[J]. Progress in Electromagnetics Research M, 2011, **19**: 133 – 146. DOI: 10.2528/pierm11051310.
- 23 Aszemi N M, Dominic P D D. Hyperparameter optimization in convolutional neural network using genetic algorithms[J]. International Journal of Advanced Computer Science and Applications, 2019, **10**(6): 269 – 278. DOI: 10.14569/ijacsa.2019.0100638.