

# 嵌入位置编码的多头注意力机制在堆积脉冲幅度估计中的应用

唐琳<sup>1,2,3</sup> 周爽<sup>1</sup> 李勇<sup>1</sup> 廖先莉<sup>1</sup> 李跃鹏<sup>4,1</sup>

1(成都大学 电子信息与电气工程学院 成都 610106)

2(安徽大学农业生态大数据分析与应用国家工程研究中心 合肥 230039)

3(南洋理工大学 电气与电子工程学院 新加坡 639798)

4(成都理工大学 核技术与自动化工程学院 成都 610059)

**摘要** 传统脉冲成形方法对堆积脉冲幅度计算存在的误差将会导致X射线荧光光谱失真,因此,在高堆积率背景下测量得到的光谱很难对光谱进行准确分析。本文提出一种基于深度学习的Transformer模型,该模型采用编码器-解码器结构,通过嵌入位置编码的多头注意力来估计堆积脉冲的幅度,并将其应用于高性能硅漂移探测器的辐射测量,以及X射线荧光光谱学分析。为了训练该模型,通过预定义的数学模型模拟得到探测器输出的脉冲信号,为了模拟真实的核脉冲,在信号中加入了与热噪声和散粒噪声相对应的高斯噪声。训练后的模型通过截取粉末铁矿样品和粉末岩石样品的实测脉冲序列进行验证,使用相对误差作为模型性能评估指标,对应于脉冲幅度估计的准确度。在粉末铁矿样品和粉末岩石样品的实验测量8组离线脉冲序列中,得到的平均相对误差为0.89%。结果表明:该模型能够实现堆积脉冲幅度的准确估计。

**关键词** 堆积脉冲, 深度学习, Transformer模型, 位置编码

中图分类号 O657.34

DOI: 10.11889/j.0253-3219.2023.hjs.46.090505

## Application of multi-head attention mechanism with embedded positional encoding in amplitude estimation of stacked pulses

TANG Lin<sup>1,2,3</sup> ZHOU Shuang<sup>1</sup> LI Yong<sup>1</sup> LIAO Xianli<sup>1</sup> LI Yuepeng<sup>4,1</sup>

1(College of Electronic Information and Electrical Engineering, Chengdu University, Chengdu 610106, China)

2(National Engineering Research Center for Agro-Ecological Big Data Analysis & Application, Anhui University, Hefei 230039, China)

3(School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore)

4(School of Nuclear Technology and Automation Engineering, Chengdu University of Technology, Chengdu 610059, China)

国家自然科学基金(No.12305214)、四川省自然科学基金项目(No.2023NSFSC1366)、安徽大学农业生态大数据分析与应用国家工程研究中心开放研究基金(No.AE202209)、广西多源信息挖掘与安全重点实验室研究基金(No.MIMS22-04)、中国留学基金委(No.202110640002)资助

第一作者: 唐琳, 女, 1988年出生, 2019年于成都理工大学获博士学位, 副教授, 新加坡南洋理工大学访问学者, 研究领域为核辐射探测与电子学

通信作者: 廖先莉, E-mail: liaoxianli@cdu.edu.cn

收稿日期: 2023-03-23, 修回日期: 2023-04-05

Supported by National Natural Science Youth Foundation of China (No. 12305214), Sichuan Natural Science Youth Fund Project (No. 2023NSFSC1366), Open Research Fund of National Engineering Research Center for Agro-Ecological Big Data Analysis & Application, Anhui University (No. AE202209), Research Fund of Guangxi Key Lab of Multi-source Information Mining & Security (No. MIMS22-04), China Scholarship Council (No.202110640002)

First author: TANG Lin, female, born in 1988, graduated from Chengdu University of Technology with a doctoral degree in 2019, associate professor, visiting scholar of Nanyang Technological University, Singapore, focusing on nuclear radiation detection and electronics

Corresponding author: LIAO Xianli, E-mail: liaoxianli@cdu.edu.cn

Received date: 2023-03-23, revised date: 2023-04-05

**Abstract** [Background] The calculation error of the stacked pulse amplitude generated by traditional pulse shaping methods leads to distortion in the X-ray fluorescence spectrum; thus, it is difficult to accurately analyze the spectrum measured in a high-stacking rate background. [Purpose] This study aims to propose a transformer model based on deep learning for the pulse amplitude estimation of radiation measurements using high-performance silicon drift detectors. [Methods] Firstly, multi-head attention was applied to the transformer model, and an encoder-decoder structure with embedded positional encoding was employed to estimate the amplitude of stacked pulses. Then, a predefined mathematical model was used to simulate the pulse signal output by the detector for model training, and Gaussian noise corresponding to thermal noise and shot noise was added to the signal to simulate real nuclear pulses. Finally, experimental verifications were carried out on powdered iron ore samples and powdered rock samples, and relative error, corresponding to the accuracy of pulse amplitude estimation, was used as a model performance evaluation indicator. [Results & Conclusions] Experimental verification results show that the average relative error obtained for eight offline pulse sequences of powdered iron ore samples and powdered rock samples is 0.89%, which means that the model can accurately estimate the amplitude of stacked pulses.

**Key words** Stacking pulses, Deep learning, Transformer model, Positional encoding

在核辐射测量中,感兴趣区域(Region of Interest, ROI)的计数率是非常重要的测量指标,在进行元素含量分析时,计数率的精度决定了元素含量分析的准确度<sup>[1-2]</sup>。因此,要得到准确的元素含量,每一个核脉冲信号都是至关重要的。近年来,许多学者在脉冲处理领域发表了相应的研究成果,包括采用遗传算法合成具有任意噪声的最优数字成形器<sup>[3]</sup>、基于双极性尖顶成形的基线恢复<sup>[4]</sup>、对堆积脉冲的测量和补偿技术等<sup>[5-6]</sup>。在前期研究中,笔者针对脉冲宽度不够的畸形脉冲提出了脉冲甄别<sup>[7]</sup>和脉冲修复<sup>[8]</sup>两种算法,有效降低了测量结果中计数率的统计涨落。上述成果都是采用传统的分析手段对核脉冲进行了分析和处理,在某些领域中也取得了较好的应用效果。但在实际应用中,对于那些严重堆积的脉冲,采用传统的脉冲处理技术难以实现较好的脉冲分离以及脉冲幅度估计,从而导致测量结果中计数率的准确性和稳定性较差。

人工智能的发展为X射线荧光光谱(X-ray Fluorescence, XRF)分析带来了新的发展空间,国内外学者在该领域都发表了不少研究成果。文献<sup>[9]</sup>使用U-Net网络作为脉冲整形器,对脉冲进行滤波并使其展开,在一定程度上解决了脉冲堆积问题;文献<sup>[10]</sup>基于神经网络和集成学习对核探测器信号的广义不确定性进行了准确估计;笔者所在的团队在前期研究中也已提出基于深度学习的递归神经网络(Recursive Neural Network, RNN)模型<sup>[11]</sup>对堆积脉冲进行参数估计,但该模型对数据集的要求较高,需要大量的数据样本和较长的训练时间。考虑到深度学习技术已经被广泛应用到生物<sup>[12]</sup>、控制<sup>[13]</sup>、医学<sup>[14-16]</sup>、通信<sup>[17]</sup>等多个领域,在核脉冲参数识别领域,深度学习的相关应用仍处于初步探索阶段,本文

将Vaswani等<sup>[18]</sup>提出的Transformer模型结构进行优化,并将其应用于核辐射测量的堆积脉冲幅度识别领域,通过嵌入位置编码、执行多头注意力机制来对堆积脉冲的幅度进行准确估计。在堆积脉冲识别领域中,多头注意力机制可以帮助识别并分类不同的脉冲,从而提高堆积脉冲的分析准确性和鲁棒性。

## 1 原理及方法

X荧光能谱多采用数字多道谱进行分析,在能谱图中,每一个脉冲幅度估计值都对应着相应道址上的一次计数事件,因此,幅度估计值的准确性就决定了XRF分析的准确率。当脉冲发生堆积时,要得到精确的谱图,堆积分离以及准确的幅值估计就异常重要了。本文提出一种深度学习模型实现堆积脉冲分离及其幅度估计,其原理如图1所示。输入事件经积分器输出一个堆积上升的阶跃脉冲信号,然后再由CR微分整形器滤除直流成分并将阶跃脉冲整形为如图1所示的负指数脉冲。数字化的负指数脉冲序列作为深度学习模型的输入,经学习模型最终输出为与输入事件对应的脉冲幅度估计事件。

如果测量系统得到的堆积脉冲幅度没能被准确估计,那么在最终获取的X荧光能谱ROI中的有效计数就会降低。因此,堆积脉冲幅度的准确估计对X荧光能谱的精细分析是非常重要的。采用基于深度学习模型对堆积脉冲幅值进行估计的第一步就是数据集的制作。

### 1.1 数据集制作

数据集的来源包括成形后的脉冲幅度采样值以及每个脉冲样本的参数,为了对本文提出的嵌入位置编码的Transformer模型进行训练,定义了负指数

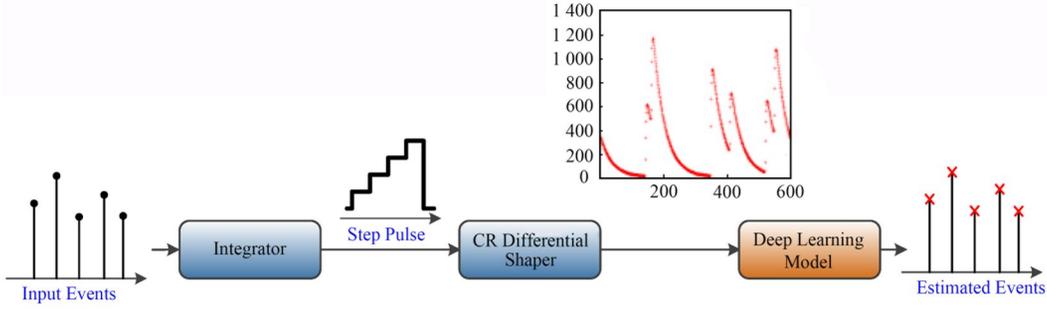


图1 堆积脉冲分离及脉冲幅度估计原理  
Fig.1 Principle of stacking pulse separation and pulse amplitude estimation

堆积脉冲的数学模型如式(1)所示：

$$V_e(t) = \sum_{i=1}^N \left[ u(t - T_i) A_i e^{-\frac{(t-T_i)}{\tau}} \right] + v(t) \quad (1)$$

式中： $u(t)$ 表示阶跃脉冲； $A_i$ 为第*i*个核脉冲的幅值系数； $T_i$ 表示第*i*个核脉冲的发生时间； $\tau$ 为时间常数； $v(t)$ 表示噪声。以采样周期 $T_{\text{clk}}$ 对其进行离散化且 $T_{\text{clk}}$ 为50 ns，则离散化后的脉冲序列如式(2)所示：

$$V_e(kT_{\text{clk}}) = \sum_{i=1}^N \left[ u(kT_{\text{clk}} - T_i) A_i e^{-\frac{(kT_{\text{clk}} - T_i)}{\tau}} \right] \quad (2)$$

负指数脉冲序列 $V_e(kT_{\text{clk}})$ 经过数字成形后将会得到一个新的脉冲幅度序列 $V_o(nT_{\text{clk}})$ 。常用的数字成形方法包括三角成形、梯形成形、高斯成形，这里以三角成形为例，其数学模型如式(3)所示：

$$V_o(nT_{\text{clk}}) = 2V_o[(n-1)T_{\text{clk}}] - V_o[(n-2)T_{\text{clk}}] + \frac{1}{n_a} \left\{ V_e[(n-1)T_{\text{clk}}] - V_e[(n-n_a-1)T_{\text{clk}}] \right\} - e^{-\frac{T_{\text{clk}}}{\tau}} \left\{ V_e[(n-2)T_{\text{clk}}] - V_e[(n-n_a-2)T_{\text{clk}}] \right\} \quad (3)$$

其中： $n_a = t_{\text{up}}/T_{\text{clk}}$ 。

本文建立的数据集由脉冲幅度采样值以及成形前负指数脉冲的真实幅值 $A$ 构成。根据需要，在对堆积脉冲的三角成形结果进行采样时，采样点数是可以调整的，为了实现快速谱分析，减小数据集规模，本文选择的是64点采样，其中包括63个成形结果的幅值采样点以及一个脉冲真实幅值的采样点。数据集的矩阵表示形式如式(4)所示：

$$\begin{bmatrix} [V_o(T_{\text{clk}})]_1 & [V_o(2T_{\text{clk}})]_1 & \cdots & [V_o(63T_{\text{clk}})]_1 & A_1 \\ [V_o(T_{\text{clk}})]_2 & [V_o(2T_{\text{clk}})]_2 & \cdots & [V_o(63T_{\text{clk}})]_2 & A_2 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ [V_o(T_{\text{clk}})]_m & [V_o(2T_{\text{clk}})]_m & \cdots & [V_o(63T_{\text{clk}})]_m & A_m \end{bmatrix} \quad (4)$$

式(4)中的数据集包含*m*个成形后的脉冲幅度

序列，每个脉冲幅度序列对应矩阵中的一行；每一行中包含的前63列代表堆积脉冲三角成形结果的幅度采样值，第64列代表该脉冲的真实幅值。本文按照7：2：1的比例把数据集分割为训练集、测试集和验证集。

## 1.2 Transformer建模

### 1.2.1 编码器-解码器框架

基于深度学习的Transformer模型其网络结构如图2所示，该模型是一个基于编码器(Encoder)-解码器(Decoder)框架的模型，不使用循环神经网络(Recurrent Neural Network, RNN)结构，完全基于注意力(Attention)机制<sup>[18]</sup>，使得并行性更好。在采用Transformer模型对堆积脉冲进行参数估计时，为了把位置编码信息传输到编码器，在编码器和解码器的输入中嵌入了位置编码(Positional Encoding)来对脉冲序列中样本的位置进行标记。

如图2所示，编码器的输入是一个嵌入了位置编码的脉冲幅度序列，内部由6个相同的编码器层组成，每个层的内部结构如图3所示，包括一个多头自注意力层(Multi-Head Self-Attention)和一个前馈层(Feed Forward)，每一个层产生的输出都传递给解码器对应的层。为了让不同的层之间便于连接，输入嵌入层、位置编码层以及编码器层、解码器层都保持着相同的维度64。

解码器将训练集中的脉冲参数集 $A$ 作为输入，其内部结构与编码器类似，也是由6个相同的层构成，如图3所示。它区别于编码器的地方在于每个解码器有三个子层，除了一个带掩码的多头自注意力层、一个前馈层，还有一个多头注意力层。在训练时将脉冲参数集 $A$ 输入到解码器中，对于一个序列，在时间为*t*的时刻，解码输出应该只能依赖于*t*时刻之前的输出，而不能依赖*t*之后的输出，因此需要一个掩码把*t*之后的输出隐藏起来。掩码的实质是一

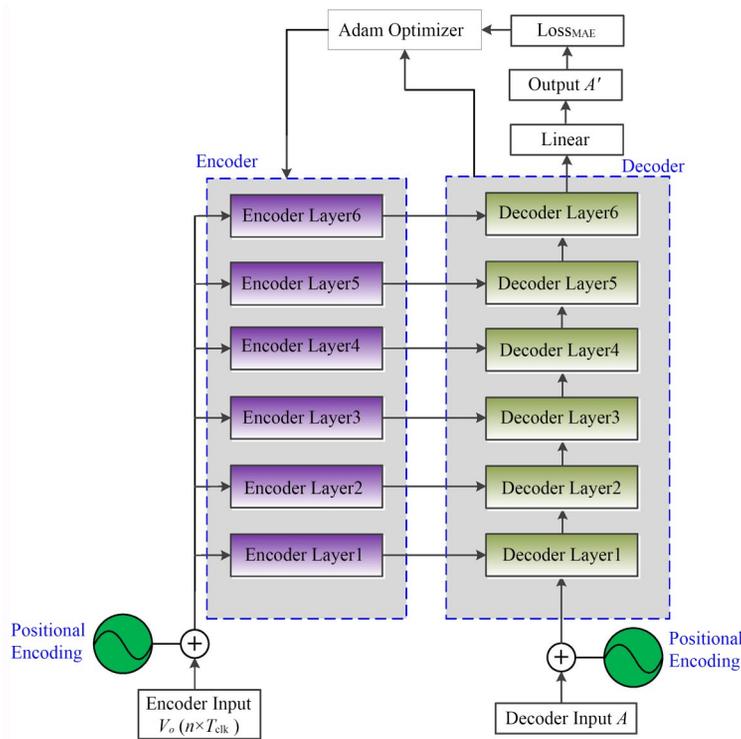


图2 Transformer模型结构图  
Fig.2 Structure diagram of the transformer model

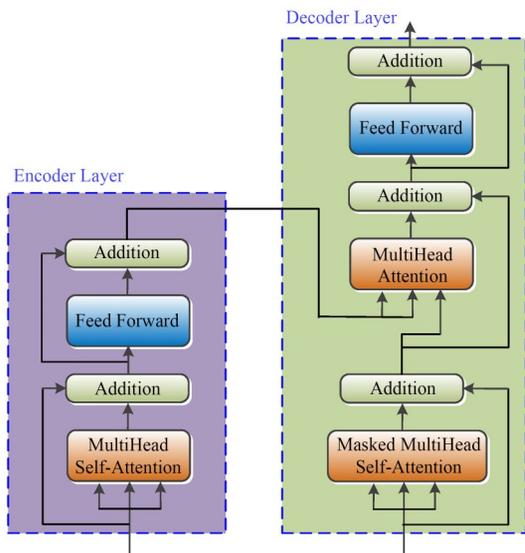


图3 编码器和解码器内部结构图  
Fig.3 Internal structures of the encoder and decoder

个下三角矩阵,即上三角的值全为0,把这个矩阵作用在每一个序列上,就可以达到把 $t$ 之后的输出隐藏起来的目的。这种掩码的设置,加上输出嵌入位置偏移的事实,确保了位置 $i$ 的预测只能依赖于小于 $i$ 的位置处的已知输出。

### 1.2.2 位置编码

常见的位置编码包括可学习的编码和固定位置

编码<sup>[19]</sup>。文献[18]证明了可学习编码与固定编码得到的结果相差不大,而最简单的固定编码莫过于二进制位置编码,其编码原理如图4(a)所示。为了对序列中的数据进行排序,给每个数据嵌入一个二进制的位位置编码,如果有16个数据需要编码,那么从0开始,编码到15结束,可以发现0~15的不同比特之间的变化率是不同的,最低位在每个数字上都会发生交替,次低位则在每两个数字上发生一次交替,依此类推。但是当所需的位置编码位数非常大时,使用二进制值就会浪费空间,占用内存。因此,本文考虑使用它们的浮点连续对应函数-正弦函数来进行位置编码。从本质上来讲,正弦函数也相当于一种交替变化的比特流, $y_2$ 所在的编码曲线中,1000个采样点内每个比特出现了6次,而通过调整频率后, $y_1$ 所在的编码曲线中,1000个采样点内每个比特仅出现了1次。从而可以得出,通过调整频率能够实现类似于二进制编码中从红色比特变为橙色比特的编码效果,如图4(b)所示。

本文使用正弦函数来进行位置编码,为了与输入数据进行匹配,编码维度为64,输入矩阵嵌入64维的位置编码后得到新的编码矩阵如式(5)所示:

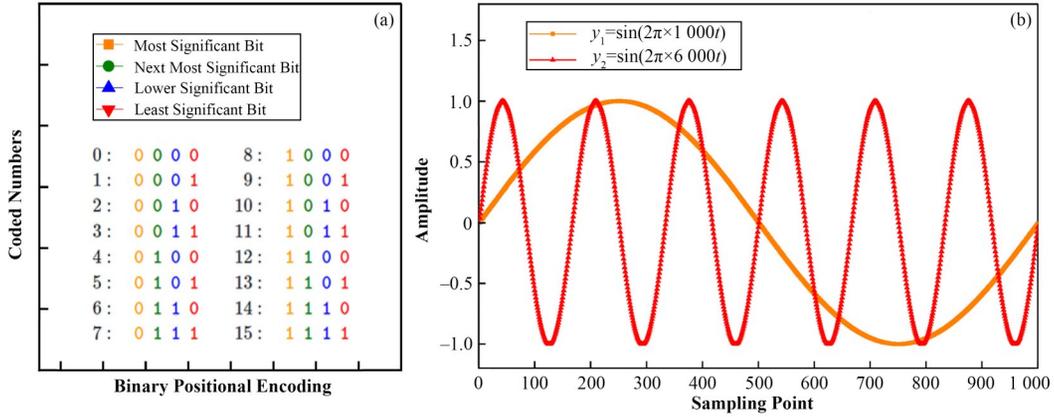


图4 位置编码原理(彩图见网络版)  
(a) 二进制编码, (b) 正弦函数编码

Fig.4 Principle of positional encoding (color online) (a) Binary encoding, (b) Sine function encoding

$$\begin{bmatrix} [V_o(T_{\text{clk}})]_1 + [\sin(\omega t)]_1 & \cdots & [V_o(63T_{\text{clk}})]_1 + [\sin(\omega t)]_{63} & A_1 + [\sin(\omega t)]_{64} \\ [V_o(T_{\text{clk}})]_2 + [\sin(\omega t)]_1 & \cdots & [V_o(63T_{\text{clk}})]_2 + [\sin(\omega t)]_{63} & A_2 + [\sin(\omega t)]_{64} \\ \vdots & \cdots & \vdots & \vdots \\ [V_o(T_{\text{clk}})]_m + [\sin(\omega t)]_1 & \cdots & [V_o(63T_{\text{clk}})]_m + [\sin(\omega t)]_{63} & A_m + [\sin(\omega t)]_{64} \end{bmatrix} \quad (5)$$

1.2.3 多头注意力机制

注意力函数本质上可以被描述为将查询向量(query)、键向量(key)、值向量(value)映射到一个输出向量的过程,输出向量是value的加权和,其中分配给每个value的权重由query与对应key的兼容性

函数计算,其计算公式如式(6)所示:

$$\text{Attention}_{(Q,K,V)} = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (6)$$

式中:  $d_k$  为键向量的维度。

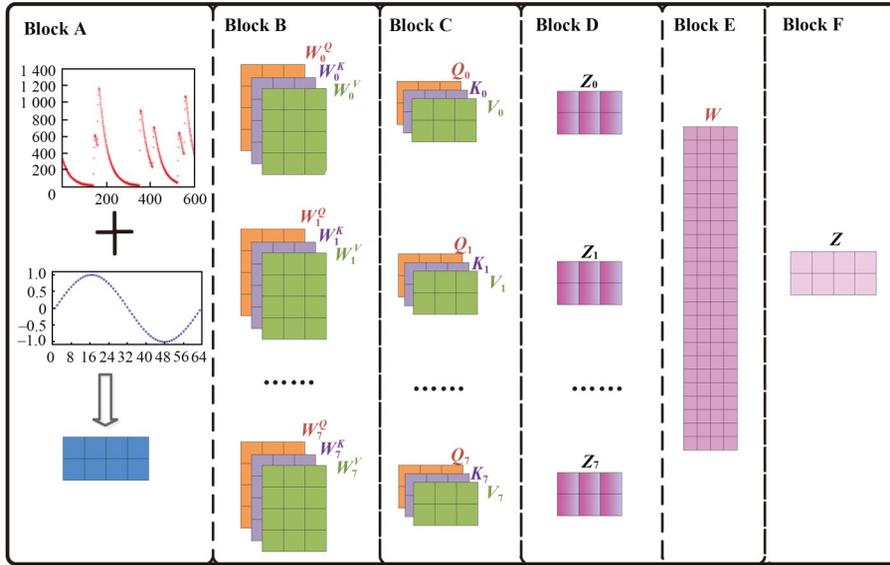


图5 多头注意力计算过程  
Fig.5 Calculation process of Multi-head attention

多头注意力的计算涉及多个模块,在图5中以6个不同的Block进行呈现。Block A中脉冲幅度形成的数据集嵌入64个维度的正弦位置编码后得到编码器的输入矩阵S,Block B中展示了8个权重矩阵

$W^Q, W^K, W^V$ 的组合,输入矩阵S与权重矩阵 $W^Q, W^K, W^V$ 进行点积得到Block C所示的8组Q、K、V矩阵,然后将8组变换后的Q、K、V进行注意力池化输出Block D中的8个注意力头 $Z_0 \sim Z_7$ ,将8个注意力头

拼接在一起并通过与 Block E 中的另一个可学习的权重矩阵进行变换,以产生 Block F 中的输出矩阵

$Z$ 。在图 5 所示的多头注意力计算过程示意图中每个 Block 的内容以及矩阵计算过程如表 1 所示。

表 1 Block 的具体特征  
Table 1 Specific characteristics of the blocks

模块名 Name	内容 Content	矩阵计算 Matrix calculation	矩阵维度 Matrix dimension
模块 A Block A	Input matrix $S$	$S \times W_i = Q$	2×4
模块 B Block B	Weight matrix $W_i$		4×3
模块 C Block C	Attention matrix $Q$		2×3
模块 D Block D	Attention head	$Z_i \times W = Z$	2×24
模块 E Block E	Weight matrix $W$		24×4
模块 F Block F	Output matrix $Z$		2×4

## 2 仿真与实验

### 2.1 模型训练

在定义 Transformer 时,三个矩阵  $W^Q$ 、 $W^K$ 、 $W^V$  的初值是随机的,为了得到最优的权重参数矩阵,验证模型对堆积脉冲进行参数估计的效果,以前文所述的堆积脉冲数学模型产生一系列堆积的负指数脉冲序列制作数据集。应用反向传播算法将  $L_{MSE}$  与损失函数的梯度一起反馈给网络来更新权重,实现减少后续迭代误差的目的。单次前向传播迭代输出的预测脉冲参数集  $A'_i$  与训练集中的实际脉冲参数集  $A_i$  的误差则可以通过损失函数来计算。对于有  $N$  个样本的训练集,将参数集  $A_i$  的平均绝对误差值作为损失函数的函数值  $L_{MSE}$ ,即损失函数的计算式为:

$$L_{MSE} = \frac{1}{N} \sum_{i=1}^N (A_i - A'_i)^2 \quad (7)$$

模型在训练集和验证集上得到的损失值的迭代变化过程如图 6 所示,不管是训练集还是验证集,损失值在第 25 个 epoch 之后就几乎趋近于零,并未出现过拟合现象,因此也无须再添加 drop out 层,在训练集和验证集上得到的准确率都趋近于 100%。

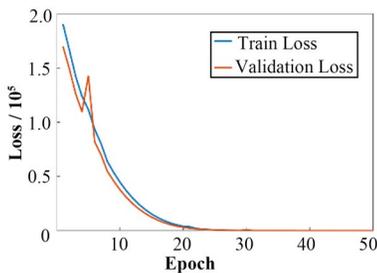


图 6 模型训练过程中在训练集和验证集上的损失值的迭代图

Fig.6 Iterative graph of the loss and accuracy for the training and validation sets during model training

### 2.2 实验验证

本文训练的深度学习模型在测试集和验证集都取得了较好的效果,为了进一步验证模型的泛化能力并对模型进行进一步评估,在测量过程中激发源采用的是上海科颐维 2000A 型 X 光管(额定管压 50 kV、额定管流为 0~1 mA),样品使用的是粉末铁矿样品(主要元素成为铁、锶、锡等),测量采用的探测器为 AMPTEK 公司的高性能硅漂移探测器 (FAST Silicon Drift Detector, FAST-SDD),其有效探测面积为 25 mm<sup>2</sup>,探测器厚度为 500 μm。取测量得到的堆积脉冲序列作为分析对象,验证模型对随机截取的脉冲序列幅度值的估计效果,如图 7 所示。

图 7(a)所示的随机脉冲序列中包含多个堆积脉冲,在采用三角成形对堆积脉冲序列进行脉冲分离及幅度估计时,堆积脉冲的分离效果较差,成形结果的幅度也存在较大的损失,如图 7(b)所示。对该脉冲序列调用深度学习模型进行参数估计,将模型输出结果与传统的成形方法得到的幅度估计值进行对比,得到如图 7(c)所示的随机脉冲幅度估计值对照图,实测脉冲幅度的真实值(黑色),采用三角成形得到的脉冲幅度(红色),采用深度学习模型获得的脉冲幅度(蓝色)。由图 7(c)所示的随机脉冲幅度估计值对照图中可以看出,对于那些没有堆积或者堆积不太严重的脉冲,不管是传统的三角成形方式还是深度学习模型都能够比较准确地估计出脉冲幅度。为了量化幅值估计效果,用  $\Delta$  表示绝对误差,  $\delta$  表示相对误差,  $A$  表示脉冲幅度,从而得出三角成形以及多头注意力估计模型对堆积脉冲幅值估计的绝对误差和相对误差的计算公式分别如式(8)和式(9)所示:

$$\begin{cases} \Delta_{Tri} = ABS(A_{real} - A_{Tri}) \\ \Delta_{Trans} = ABS(A_{real} - A_{Trans}) \end{cases} \quad (8)$$

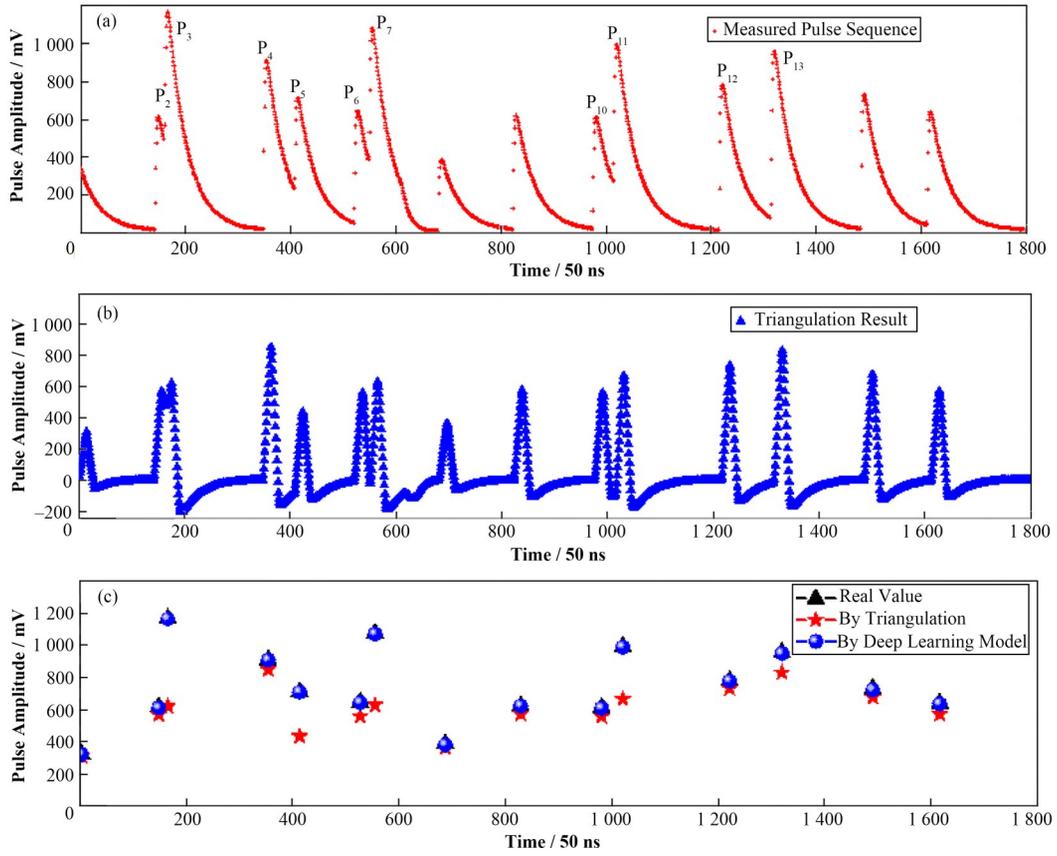


图7 (a) 实测脉冲序列图, (b) 三角成形结果, (c) 随机脉冲幅度估计值对照图(彩图见网络版)

Figure 7 (a) Measured pulse sequence diagram, (b) Triangulation results for the pulse sequence, (c) Comparison chart of the estimated random pulse amplitude (color online)

$$\begin{cases} \delta_{\text{Tri}} = \frac{\Delta_{\text{Tri}}}{A_{\text{real}}} \times 100\% \\ \delta_{\text{Trans}} = \frac{\Delta_{\text{Trans}}}{A_{\text{real}}} \times 100\% \end{cases} \quad (9)$$

实验采用的实测脉冲序列如图7(a)所示,该序列包含15个脉冲,其中堆积率为66%且堆积类型为双脉冲堆积。对图7(a)所示的15个脉冲采用不同方式计算得到的幅度值进行对比分析,结果如表2所示。

没有发生明显堆积的脉冲包括 $P_1$ 、 $P_8$ 、 $P_9$ 、 $P_{14}$ 以及 $P_{15}$ ,这些脉冲采用三角成形进行幅度计算得到的相对误差约在10%以内;而对于发生了堆积的脉冲,如图7中标注的 $P_2$ 和 $P_3$ ,类似的还有 $P_4$ 和 $P_5$ 、 $P_6$ 和 $P_7$ 、 $P_{10}$ 和 $P_{11}$ 以及 $P_{12}$ 和 $P_{13}$ ,对于这种发生了双脉冲堆积的堆积脉冲三角成形,通常第一个脉冲的幅度不会受到较大影响,而堆积上去的第二个脉冲则无法得出准确的成形结果,如 $P_3$ 、 $P_5$ 、 $P_7$ 、 $P_{11}$ 以及 $P_{13}$ ,这些脉冲采用三角成形进行幅度计算得到的相对误差成倍增长,最大相对误差高达47%,严重影响了脉冲幅度的估计。采用嵌入位置编码的多头注意力模型对幅值进行估计时,估计结果并不会受到脉冲堆积程

度的影响,平均相对误差约为0.55%。

为了进一步验证本文训练的模型在相同样品不同堆积程度的脉冲序列以及不同样品相同堆积程度的脉冲序列中的参数估计效果,在实验环节中截取堆积率为10%、30%、60%和90%,堆积类型为双脉冲堆积的粉末岩石样品和粉末铁矿样品的实测脉冲序列进行分析,分析结果如表3所示。从两种样品不同堆积程度得到的平均相对误差来看,深度学习模型对脉冲幅度估计的相对误差与样品类型并无直接关联,但与脉冲堆积程度却有一定的相关性,可以粗略估计随着堆积程度加深,模型对脉冲幅度估计的相对误差也会略有上升。表3所示的8组脉冲序列中,比较明显的一个“异常点”是第三组,该脉冲序列包含10个脉冲,其中双脉冲堆积的数量为6个,堆积程度为60%,按照理论分析深度学习模型对该序列进行幅度估计的相对误差应低于1%,但模型测试结果的相对误差高达3.42%。通过对堆积脉冲进一步分析发现,虽然该序列仅包含6个堆积脉冲,但其中两个脉冲发生了严重堆积,导致模型对堆积上去的第二个脉冲的幅度估计产生了较大误差。即便是这种极端情况下,本文训练的深度学习模型依然获

得了远优于传统脉冲估计方法的性能,在两种样品、8组离线脉冲序列的幅度估计中得到的平均相对误差为0.89%。

表2 深度学习模型对堆积脉冲的估计值与真实值的对比(堆积率为66%)  
Table 2 Comparison of estimated and real values of overlapping pulses in the deep learning model

脉冲 Pulse	是否堆积 Pile-up or not	真实值 Real value	三角成形值 Triangulation	成形相对误差 of Triangulation / %	模型估计值 Estimated value of model	模型相对误差 Relative error of model / %
P1	No	322.307 7	308.87	4.17	320.23	0.64
P2	Yes	615.384 6	568.21	7.67	610.48	0.80
P3		1 167.692	618.74	47.01	1 164.98	0.23
P4	Yes	910.769 3	849.85	6.69	908.56	0.24
P5		710.000 1	432.60	39.07	708.52	0.21
P6	Yes	644.615 4	557.88	13.46	643.19	0.22
P7		1 076.923	627.63	41.72	1 070.48	0.60
P8	No	386.153 9	363.08	5.98	380.86	1.37
P9	No	622.307 7	571.75	8.12	621.64	0.11
P10	Yes	611.538 5	556.17	9.05	607.77	0.62
P11		992.307 7	666.04	32.88	989.54	0.28
P12	Yes	780.769 3	732.13	6.23	777.57	0.41
P13		956.923 1	829.89	13.28	949.87	0.74
P14	No	730.769 3	675.82	7.52	722.75	1.10
P15	No	639.230 8	569.35	10.93	634.71	0.71

表3 深度学习模型的参数估计效果对照  
Table 3 Comparison table of parameter estimation effects of the deep learning model

序号 Number	样品类型 Sample type	堆积程度 Stacking degree	堆积类型 Stacking type	平均相对误差 Average relative error / %
1	粉末铁矿样品	10%	双脉冲 Double pulse	0.21
2	Powder iron ore sample	30%	双脉冲 Double pulse	0.32
3		60%	双脉冲 Double pulse	3.42
4		90%	双脉冲 Double pulse	0.85
5		粉末岩石样品	10%	双脉冲 Double pulse
6	Powder rock samples	30%	双脉冲 Double pulse	0.48
7		60%	双脉冲 Double pulse	0.65
8		90%	双脉冲 Double pulse	0.98

### 3 结语

本文提出了一种嵌入位置编码的多头注意力Transformer神经网络模型,并将其应用于堆积脉冲的幅度估计。该模型是一个基于Encoder-Decoder框架的模型,采用多头注意力机制,以预定义的数学模型产生负指数脉冲序列以及三角成形后的幅值序列来制作数据集进行训练,模拟和实验结果验证了所提模型的脉冲幅度估计性能。采用相对误差作为堆积脉冲幅度估计精度的指标,得出在粉末铁矿样品和粉末岩石样品的8组离线脉冲序列中得到的平均相对误差为0.89%。结果表明:采用嵌入位置编码的多头注意力模型成功地从具有堆积效应的脉冲序列中估计出了脉冲幅度,且估计结果并不受脉冲

堆积程度的影响。未来对深度学习模型在核辐射测量中的应用研究将旨在优化学习模型和数据集,此外,作者所在团队正在对HPGe和LaBr<sub>3</sub>(Ce)探测器应用类似方法进行研究,希望推广本模型在X射线荧光光谱学中的应用。

**作者贡献声明** 唐琳直接参与论文研究,负责神经网络模拟搭建并撰写论文;周爽负责实验测试结果分析;李勇负责模型训练;廖先莉负责资料的收集及整理;李跃鹏负责实验测试。

### 参考文献

- Ishikawa D, Baron A Q R. Practical measurement of the energy resolution for meV-resolved inelastic X-ray scattering[J]. Journal of Synchrotron Radiation, 2021, 28

- (Pt 3): 804 – 811. DOI: [10.1107/S1600577521003234](https://doi.org/10.1107/S1600577521003234).
- 2 Wu H X, Di R J, Liu Y J, *et al.* Application on straight-line shaping method for energy spectrum measurement in TXRF spectrometer based on SDD detector[J]. Spectroscopy and Spectral Analysis, 2021, **41**(7): 2148 – 2152.
  - 3 Regadio A, Sánchez-Prieto S, Tabero J, *et al.* Synthesis of optimal digital shapers with arbitrary noise using a genetic algorithm[J]. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2015, **795**: 115 – 121. DOI: [10.1016/j.nima.2015.05.059](https://doi.org/10.1016/j.nima.2015.05.059).
  - 4 Zhong X D, Chen L, Wang B C, *et al.* A spectrometer with baseline correction and fast pulse pile-up rejection for prompt gamma neutron activation analysis technology [J]. The Review of Scientific Instruments, 2018, **89**(12): 123504. DOI: [10.1063/1.5049517](https://doi.org/10.1063/1.5049517).
  - 5 Kafaee M, Moussavi-Zarandi A. Baseline restoration and pile-up correction based on bipolar cusp-like shaping for high-resolution radiation spectroscopy[J]. Journal of the Korean Physical Society, 2016, **68**(8): 960 – 964. DOI: [10.3938/jkps.68.960](https://doi.org/10.3938/jkps.68.960).
  - 6 Usman S, Patil A. Radiation detector deadtime and pile up: a review of the status of science[J]. Nuclear Engineering and Technology, 2018, **50**(7): 1006 – 1016. DOI: [10.1016/j.net.2018.06.014](https://doi.org/10.1016/j.net.2018.06.014).
  - 7 Tang L, Yu J, Zhou J B, *et al.* A new method for removing false peaks to obtain a precise X-ray spectrum[J]. Applied Radiation and Isotopes, 2018, **135**: 171 – 176. DOI: [10.1016/j.apradiso.2018.01.033](https://doi.org/10.1016/j.apradiso.2018.01.033).
  - 8 Tang L, Zhou J B, Fang F, *et al.* Counting-loss correction for X-ray spectra using the pulse-repairing method[J]. Journal of Synchrotron Radiation, 2018, **25**(Pt 6): 1760 – 1767. DOI: [10.1107/S160057751801411X](https://doi.org/10.1107/S160057751801411X).
  - 9 Regadio A, Esteban L, Sanchez-Prieto S. Unfolding using deep learning and its application on pulse height analysis and pile-up management[J]. Nuclear Instruments & Methods in Physics Research Section A: Accelerators Spectrometers Detectors and Associated Equipment, 2021, **1005**: 165403. DOI: [10.1016/j.nima.2021.165403](https://doi.org/10.1016/j.nima.2021.165403).
  - 10 Ai P, Deng Z, Wang Y, *et al.* Universal uncertainty estimation for nuclear detector signals with neural networks and ensemble learning[J]. Journal of Instrumentation, 2022, **17**(2): P02032. DOI: [10.1088/1748-0221/17/02/p02032](https://doi.org/10.1088/1748-0221/17/02/p02032).
  - 11 Ma X K, Huang H Q, Huang B R, *et al.* X-ray spectra correction based on deep learning CNN-LSTM model[J]. Measurement, 2022, **199**: 111510. DOI: [10.1016/j.measurement.2022.111510](https://doi.org/10.1016/j.measurement.2022.111510).
  - 12 Dong Y, Zhou S Z, Xing L, *et al.* Deep learning methods may not outperform other machine learning methods on analyzing genomic studies[J]. Frontiers in Genetics, 2022, **13**: 992070. DOI: [10.3389/fgene.2022.992070](https://doi.org/10.3389/fgene.2022.992070).
  - 13 Alwan N A S, Hussain Z M. Deep learning for robust adaptive inverse control of nonlinear dynamic systems: improved settling time with an autoencoder[J]. Sensors, 2022, **22**(16): 5935. DOI: [10.3390/s22165935](https://doi.org/10.3390/s22165935).
  - 14 Cheng N T, Fu J, Chen D J, *et al.* An antibody-free liver cancer screening approach based on nanoplasmonics biosensing chips via spectrum-based deep learning[J]. NanoImpact, 2021, **21**: 100296. DOI: [10.1016/j.impact.2021.100296](https://doi.org/10.1016/j.impact.2021.100296).
  - 15 Touch M, Clark D P, Barber W, *et al.* A neural network-based method for spectral distortion correction in photon counting X-ray CT[J]. Physics in Medicine and Biology, 2016, **61**(16): 6132 – 6153. DOI: [10.1088/0031-9155/61/16/6132](https://doi.org/10.1088/0031-9155/61/16/6132).
  - 16 Feng C Q, Kang K J, Xing Y X. Fully connected neural network for virtual monochromatic imaging in spectral computed tomography[J]. Journal of Medical Imaging, 2018, **6**(1): 011006. DOI: [10.1117/1.JMI.6.1.011006](https://doi.org/10.1117/1.JMI.6.1.011006).
  - 17 Kumar P, Kumar R, Gupta G P, *et al.* A blockchain-orchestrated deep learning approach for secure data transmission in IoT-enabled healthcare system[J]. Journal of Parallel and Distributed Computing, 2023, **172**: 69 – 83. DOI: [10.1016/j.jpdc.2022.10.002](https://doi.org/10.1016/j.jpdc.2022.10.002).
  - 18 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all You need[EB/OL]. 2017: arXiv: 1706.03762. <https://arxiv.org/abs/1706.03762>.
  - 19 Gehring J, Auli M, Grangier D, *et al.* Convolutional sequence to sequence learning[EB/OL]. 2017: arXiv: 1705.03122. <https://arxiv.org/abs/1705.03122>.
  - 20 唐琳, 赵卫东, 余松科, 等. X射线光谱数据处理平台的优化设计[J]. 光谱学与光谱分析, 2021, **41**(3): 763. DOI: [10.3964/j.issn.1000-0593\(2021\)03-0763-05](https://doi.org/10.3964/j.issn.1000-0593(2021)03-0763-05). TANG Lin, ZHAO Weidong, YU Songke, *et al.* Optimization design of X-ray spectrum data processing platform[J]. Spectrosc Spectr Anal, 2021, **41**(3): 763. DOI: [10.3964/j.issn.1000-0593\(2021\)03-0763-05](https://doi.org/10.3964/j.issn.1000-0593(2021)03-0763-05).