

DOI: 10.12158/j.2096-3203.2025.01.008

基于 CNN 和 Transformer 的轻量化电能质量扰动识别模型

张彼德¹, 邱杰¹, 姜广鑫², 周灿¹, 罗靖清¹, 李天倩¹

(1. 西华大学电气与电子信息学院, 四川 成都 610039;

2. 国网四川省电力公司成都供电公司, 四川 成都 610000)

摘要:针对目前基于深度学习的电能质量扰动(power quality disturbances, PQDs)识别模型参数量多和计算复杂度较高的问题,文中提出了一种卷积神经网络(convolutional neural networks, CNN)融合 Transformer(CNN and Transformer, CaT)的轻量化 PQDs 识别模型。首先,利用深度可分离卷积初步提取扰动信号的局部特征;其次,提出一种高效的软阈值模块,在不显著增加模型参数量与计算复杂度的同时减少特征中的噪声与冗余特征;然后,利用 Transformer 模型挖掘 PQDs 信号的全局特征;最后,通过池化层、线性层和 Softmax 层完成 PQDs 识别。仿真实验表明,文中所提 CaT 模型在参数量和浮点运算数较少的情况下能够有效完成 PQDs 识别,对 PQDs 信号识别准确率高,具有良好的噪声鲁棒性。同时,得益于轻量化和端到端的模型设计,CaT 模型相对于其他深度学习模型的推理时间更短。

关键词:电能质量扰动(PQDs);轻量化;参数量;高效软阈值模块;深度可分离卷积;Transformer 模型

中图分类号:TM743

文献标志码:A

文章编号:2096-3203(2025)01-0069-10

0 引言

为了实现碳减排,电力系统正逐步增加风能和太阳能等新能源的并网设备;同时非线性设备,如新能源汽车等,接入电网的规模也在迅速扩大。这些都导致电压、电流或频率突然偏离额定值的电能质量扰动(power quality disturbances, PQDs)问题日益突出^[1-7]。如果能够迅速、准确地识别 PQDs,及时采取措施,可以减少电量损失。因此,PQDs 的准确识别与分类是电能质量治理的关键^[8-10]。

大量学者对 PQDs 信号识别进行了深入的研究。传统的 PQDs 信号识别可以分为 2 个步骤:提取 PQDs 信号的特征;构建分类器完成 PQDs 信号识别。在提取信号特征时,通常使用信号分析方法,将时域信号转换为频域信号后,在时域和频域上进行特征提取。常见的信号分析方法包括傅里叶变换^[11]、S 变换^[12]、小波变换^[13-14]和希尔伯特-黄变换^[15]等;然后利用人工先验知识选择重要的特征组;最后将选择的特征组送入分类器学习各特征与标签之间的非线性关系。其中常用的分类器有支持向量机^[16-17]、决策树^[18]和神经网络^[19-20]等。但上述方法需要人工确定特征组,检验人工经验的同时无法捕捉扰动信号的深层次特征,在一定的情况下适用性不强,造成分类准确率不高。

近年来,深度学习方法逐渐应用于 PQDs 信号

识别。相比传统的人工选取特征的 PQDs 信号识别方法,深度学习方法可以自适应学习扰动信号特征,极大程度上解决了传统 PQDs 信号识别提取特征值困难的问题。目前深度学习方法应用到 PQDs 信号识别领域主要可以分为 2 类。第一类为端到端的深度学习模型,使用一维卷积神经网络(convolutional neural network, CNN)或循环神经网络对原始 PQDs 信号进行特征提取,并完成 PQDs 信号的识别。文献[21]提出了一种基于深度卷积神经网络(Deep CNN)的 PQDs 识别方法,设计了一种由一维卷积层、池化层和批归一化层组成的识别模型,但上述模型缺少全局性的时序特征。文献[22]提出了一种 CNN 和长短期记忆(long short-term memory, LSTM)网络相结合的 PQDs 信号识别方法,使用 LSTM 提取扰动信号的时序特征,克服单一 CNN 模型学习特征后缺少全局时序特征的问题,但 LSTM 存在参数过大且无法并行加速的问题^[4]。第二类为将 PQDs 信号识别问题转换为图片分类问题,利用计算机视觉领域一些较为成熟的分类模型完成 PQDs 信号识别。文献[23]运用距离矩阵挖掘 PQDs 信号的递归特性,结合 CNN 与残差网络(residual network, Resnet)提取深度特征并完成信号识别。文献[24]提出一种结合格拉姆角场(Gramian angular fields, GAF)和 Resnet 的 PQDs 信号识别方法,首先利用 GAF 将时序信号映射为二维图像,然后利用 Resnet 完成特征提取和识别。上述方法将时序信号转化为图片数据,不仅增加数据处理时

收稿日期:2024-09-07;修回日期:2024-11-24

基金项目:四川省科技计划资助项目(2023YFG0191)

间,且需要更复杂的模型完成识别,导致模型参数量、计算复杂度和推理时间增加,不利于将模型部署到资源受限的环境下完成 PQDs 信号实时检测。

针对现有模型的参数量和计算复杂度较高的问题,文中提出一种 CNN 融合 Transformer(CNN and Transformer, CaT)的轻量化 PQDs 识别模型,在模型各层中充分利用 CNN 和 Transformer 模型特征提取的优势。首先设计轻量化的 CNN 模块和 Transformer 模块;然后通过堆叠 CNN 模块和 Transformer 模块,使得模型既具有 CNN 的高效和轻量化,又具有 Transformer 模块强大的全局建模能力,提升模型的识别能力。

1 CaT 模型

CaT 模型的具体结构如图 1 所示,其核心为文中提出的 CNN 模块和 Transformer 模块,其中 C1—C31 为模型输出的各个标签的预测概率。首先利用 CNN 模块提取局部特征并对扰动信号进行下采样,减少信号序列的长度。同时由于 CNN 模块中加入了高效软阈值模块,可在不显著增加模块整体参数的同时,减少模型中的冗余特征与噪声。然后通过级联的 Transformer 模块捕获信号各个时间点的相关性以获取扰动信号的全局特征。最后利用池化层完成特征的最终融合与降维,并由线性层与 Softmax 层完成 PQDs 的识别。

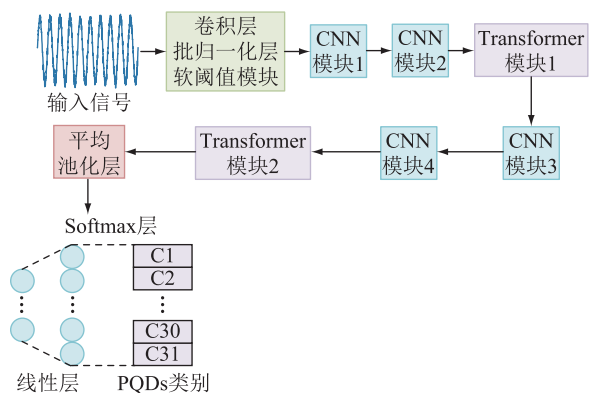


图 1 CaT 模型结构

Fig.1 Structure of CaT model

1.1 CNN 模块

CNN 模块由深度可分离卷积^[25-26]和文中提出的高效软阈值模块构成,其具体结构如图 2 所示。

CNN 模块利用深度可分离卷积提取信号的局部特征,相比传统卷积可大大减少参数量与计算复杂度,实现 CNN 模块的轻量化。在提取特征时嵌入文中提出的高效软阈值模块,进一步减少特征中的噪声与冗余特征,提升模型的抗噪能力。

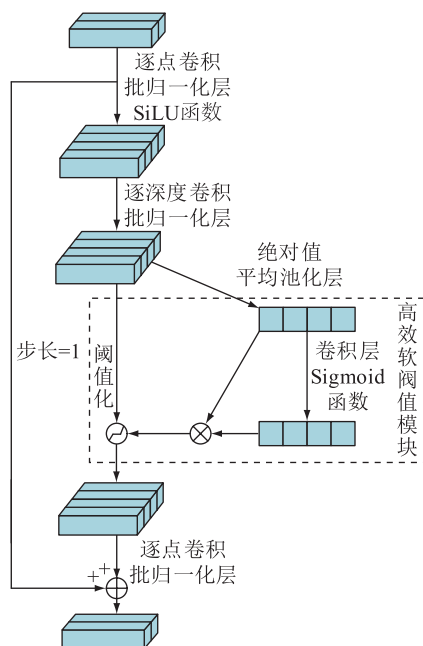


图 2 CNN 模块

Fig.2 CNN block

1.1.1 深度可分离卷积

深度可分离卷积能够在不显著降低性能的前提下,有效减少模型的参数量和计算复杂度,实现模型的轻量化。深度可分离卷积可拆分为 2 种卷积过程:逐深度卷积和逐点卷积,其运算过程如图 3 和图 4 所示。

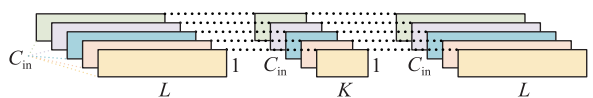


图 3 逐深度卷积

Fig.3 Depthwise convolution

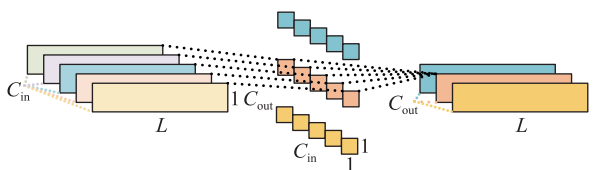


图 4 逐点卷积

Fig.4 Pointwise convolution

图 3 中,逐深度卷积的计算次数为:

$$P_{dw} = LKC_{in} \quad (1)$$

式中: L 为输入张量的长度; K 为卷积核的大小; C_{in} 为输入张量的通道数。

图 4 中,逐点卷积的计算次数为:

$$P_{pw} = LC_{in}C_{out} \quad (2)$$

式中: C_{out} 为输出张量的通道数。

普通卷积的计算次数为:

$$P = LKC_{in}C_{out} \quad (3)$$

则深度可分离卷积与普通卷积的计算次数之比为:

$$\frac{P_{dw} + P_{pw}}{P} = \frac{1}{C_{out}} + \frac{1}{K} \quad (4)$$

因此,当输出通道数或者卷积核大小较大时深度可分离卷积能够显著减少计算次数。同时深度可分离卷积能显著减少参数量,图 3 中逐深度卷积的参数量 N_{dw} 如式(5)所示,图 4 中逐点卷积的参数量 N_{pw} 如式(6)所示,普通卷积的参数量 N 如式(7)所示,参数量之比如式(8)所示,可见在通道数或卷积核大小较大时深度可分离卷积能够显著减少模型的参数量。

$$N_{dw} = KC_{in} \quad (5)$$

$$N_{pw} = C_{in}C_{out} \quad (6)$$

$$N = KC_{in}C_{out} \quad (7)$$

$$\frac{N_{pw} + N_{dw}}{N} = \frac{1}{C_{out}} + \frac{1}{K} \quad (8)$$

1.1.2 高效软阈值模块

软阈值能够有效地抑制特征量中的噪声,进而提升模型在噪声条件下的性能表现。其中软阈值函数的表达式为:

$$y = \begin{cases} x - \tau & x > \tau \\ 0 & -\tau \leq x \leq \tau \\ x + \tau & x < -\tau \end{cases} \quad (9)$$

式中: x 为输入特征量; y 为阈值化后的特征量; τ 为阈值,是一个恒正数。式(9)可将阈值范围内的特征量置 0,并将其他的特征量也朝着 0“收缩”。

阈值的选择至关重要,传统的通道独立软阈值模块^[27-28]通过通道注意力机制确定阈值,但多层线性层的使用会导致在通道数过多时,模型的参数量与计算复杂度增大,不利于模型的轻量化。为此文中提出一种高效软阈值模块,使用高效通道注意力机制确定软阈值函数的阈值,减少模块的参数量与计算复杂度,2 种软阈值模块的对比如图 5 所示。

图 5 中传统通道软阈值模块首先对输入量取绝对值,然后通过平均池化层将二维特征张量转换成长度为通道数 C 的一维张量,具体表达式为:

$$v_m = \frac{1}{L} \sum_{i=1}^L |x_i| \quad (10)$$

式中: v_m 为一维张量中第 m 个通道; x_i 为 F_m 中的第 i 个元素, F_m 为二维张量中第 m 个通道。

通过 2 层线性层获得阈值,具体表达式为:

$$\tau = v\delta(\mathbf{W}_2\sigma(v\mathbf{W}_1)) \quad (11)$$

$$\sigma(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (12)$$

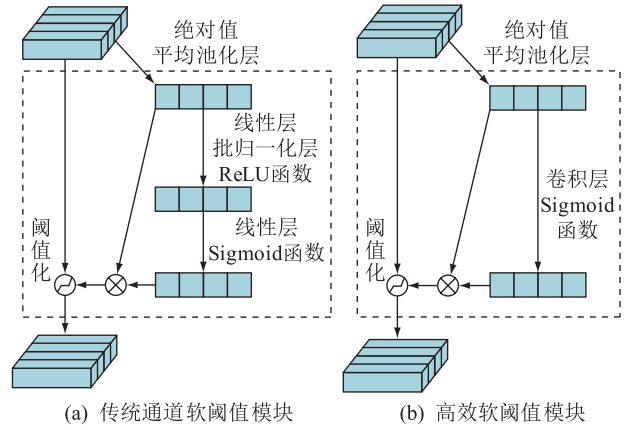


图 5 软阈值模块

Fig.5 Soft threshold block

$$\delta(x) = \frac{1}{1 + e^{-x}} \quad (13)$$

式中: v 为平均池化后的一维张量; σ 为 Sigmoid 函数; δ 为 ReLU 函数; x 为二维张量中元素; \mathbf{W}_1 、 \mathbf{W}_2 分别为第 1 个线性层和第 2 个线性层的权重参数,其参数量大小为 $C \times C$ 。当输入特征张量的通道数变大时,其参数量也将平方增大,不利于模块的轻量化。文中提出使用卷积层确定阈值,其表达式为:

$$\tau = v\delta(f_{conv}(v)) \quad (14)$$

式中: f_{conv} 为卷积运算。

高效软阈值模块中卷积层的卷积核大小为:

$$K = \left\lfloor \frac{\log_2 C}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}} \quad (15)$$

式中: $\gamma=2$; $b=1$; $\lfloor t \rfloor_{\text{odd}}$ 为最接近 t 的奇数。假设输入通道数为 C ,则图 5 中传统通道软阈值模块的线性层参数和计算次数均为 $2C^2$,而高效软阈值模块中卷积层的卷积核大小为 K ,计算次数为 $C \times K$,若计及 ReLU 激活函数,传统通道软阈值模块的计算复杂度将更大,因此高效软阈值模块相比传统通道软阈值模块在通道数较大时能够显著减少参数量与计算复杂度。

1.2 Transformer 模块

文中所提 CaT 模型为串行结构。这是因为文献[29]指出模型的分支结构会降低模型的并行度,造成推理时间的增加。同时 Transformer 模块的参数量和计算复杂度会随着时间序列长度的增加而大大增加,而串行结构可利用 CNN 模块对时间序列的编码及下采样,降低时间序列的长度,从而降低 Transformer 模块的参数量和计算复杂度。Transformer 模块的具体结构如图 6 所示。Transformer 模块首先对 CNN 模块的输出张量进行变换,以通道数作为时间序列每个时间点的维度,

满足 Transformer 模块对于输入数据的要求,然后将转置后的张量输入到 Transformer 模块中提取全局特征,并利用卷积层实现全局特征和原始局部特征的融合。其总体数学模型为:

$$\mathbf{x}_{\text{output}} = f_{\text{re}}(f_{\text{conv}}(\mathbf{x}_{\text{input}} + f_{\text{trans}}(f_{\text{re}}(\mathbf{x}_{\text{input}})))) \quad (16)$$

式中: $\mathbf{x}_{\text{output}}$ 为模块输出的张量; f_{re} 为张量转置操作; f_{trans} 为 Transformer 模型的一系列运算; $\mathbf{x}_{\text{input}}$ 为模块输入的张量;+表示连接操作。

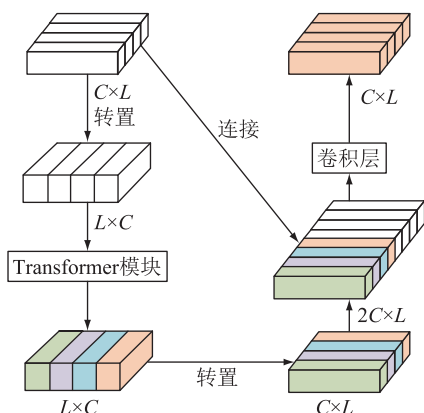


图6 Transformer 模块
Fig.6 Transformer block

Transformer 模型的具体结构如图7所示。其主要由自注意力机制、层归一化、残差连接和多层感知机构成。Transformer 模型的核心是自注意力机制,通过自注意力机制对输入特征进行全局编码,使得模型能够完成全局特征的建模;而多层感知机的引入可以使得 Transformer 模型具有更强大的非线性特征提取能力;残差连接使得映射函数在模型层数增多时更加容易被拟合,大大缓解模型层数增加时的梯度消失和梯度爆炸问题;层归一化对特征进行归一化,加快模型的收敛速度。

其中核心的自注意力机制的数学表达式为:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right) \mathbf{V}$$

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{d_k \sum_{i=1} e^{x_i}} \quad (17)$$

式中: $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 为输入张量的查询、键值、幅值线性映射矩阵; d_k 为缩放因子,用于稳定梯度,取 \mathbf{Q}, \mathbf{K} 的维度。首先 \mathbf{Q} 和 \mathbf{K} 进行点积操作,得到相似度;再进行缩放操作,除以 $\sqrt{d_k}$,保持梯度的稳定;然后使用 Softmax 操作进行归一化,得到权重矩阵;最后将权重矩阵与 \mathbf{V} 点积,进行加权求和,得到最终结果。

1.3 PQDs 识别流程

文中所提 CaT 模型为端到端的 PQDs 识别模型,将原始 PQDs 信号输入到模型后,模型可自适应

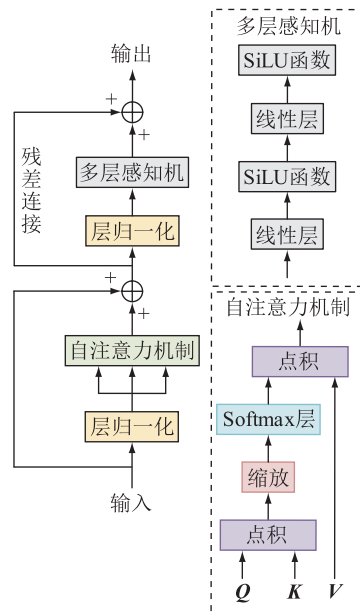


图7 Transformer 模型结构

Fig.7 Transformer model structure

的进行特征提取完成 PQDs 分类,其训练应用流程如图8所示,主要步骤如下。

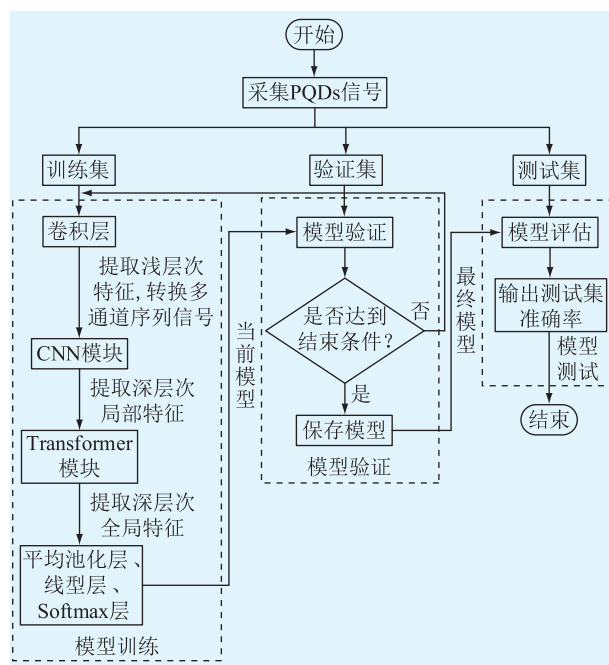


图8 PQDs 识别的 CaT 模型流程

Fig.8 The CaT model flow chart of PQDs recognition

(1) 采集或生成 PQDs 信号,划分训练集、验证集和测试集。

(2) 利用训练集和验证集协同训练 CaT 模型。其主要过程如下:

① 首先将训练集数据 PQDs 信号输入 CaT 模型进行前向传播,利用卷积层提取浅层次特征,并将 PQDs 信号转换为多元信号,再将其输入 CNN 模块,提取深层次局部特征;然后使用 Transformer 模块

对全局特征进行编码,并完成局部和全局特征的融合;最后将提取到的特征输入至平均池化层、线性层和 Softmax 层获取预测概率。

② 使用预测概率和真实标签计算损失值,并且进行反向传播,调整 CaT 模型的权重参数,使得模型能够自适应调整提取到的特征,提升模型的识别准确率。

③ 使用迭代得到的 CaT 模型计算验证集的准确率,若准确率高于上次保存的模型,则保存本次得到的模型,反之则使用训练集进行下一次迭代。若超过最大迭代次数或者验证集准确率超过 20 次没有上升,则停止训练,并保存得到的最终模型。

(3) 使用训练得到的最终模型计算测试集的准确率,评估最后得到的 CaT 模型。

2 算例分析

2.1 PQDs 数据集

参考文献[30-31]的 PQDs 模型,通过仿真生成 31 种扰动信号,其具体类型如表 1 所示。PQDs 信号的基波频率设置为 50 Hz,采样频率为 5 000 Hz。每种扰动生成了 1 200 条数据,每种扰动按照 8:2:2 的比例划分为训练集、验证集和测试集,训练集每类扰动 800 条数据,验证集和测试集每类扰动 200 条数据。为了测试模型对于噪声的鲁棒性,在数据集中分别增加 40、30、20 dB 的高斯白噪声,得到了不同噪声等级下的 PQDs 数据集,并对各个噪声等级的数据集都进行分类训练。

表 1 PQDs 类型

Table 1 PQDs types

编号	扰动类型	编号	扰动类型
C1	正常	C17	暂升+切口
C2	暂降	C18	暂升+闪变
C3	暂升	C19	暂升+谐波
C4	中断	C20	中断+振荡
C5	振荡	C21	中断+冲击
C6	冲击	C22	中断+切口
C7	切口	C23	中断+闪变
C8	闪变	C24	中断+谐波
C9	谐波	C25	振荡+闪变
C10	暂降+振荡	C26	振荡+谐波
C11	暂降+冲击	C27	冲击+闪变
C12	暂降+切口	C28	冲击+谐波
C13	暂降+闪变	C29	切口+闪变
C14	暂降+谐波	C30	切口+谐波
C15	暂升+振荡	C31	闪变+谐波
C16	暂升+冲击		

2.2 模型训练与网络参数

文中所提 CaT 模型与其他深度学习模型均使用 Pytorch 框架搭建,训练环境的具体参数如表 2 所示。在训练时设置最大迭代次数为 100 次,优化器选择 Adam,初始学习率为 0.01,设置学习率衰减策略和早停机制为当验证集的准确率迭代 5 次没有上升时,学习率按照 0.5 的比率衰减;当验证集的准确率迭代 20 次没有上升时停止训练,并保存最优模型,损失函数采用交叉熵损失函数。

表 2 硬件型号及软件版本

Table 2 Hardware model and software version

硬件/软件	型号/版本
Pytorch	2.0.0
Python	3.8
Cuda(compute unified device architecture)统一计算架构	11.8
图像处理器	RTX 2080Ti 11 GB
中央处理器	Xeon Platinum 8255C

为了简化模型的参数设置,CaT 模型中的 CNN 模块若无残差连接,输出通道数相比输入通道数增加 4;Transformer 模块中的嵌入序列维度则为输入通道数的 2 倍。因此,输入卷积层中的输出通道数和 CaT 模型卷积层中的卷积核大小将对整体模型的性能及轻量化程度产生巨大的影响。为了验证其影响,设置输入卷积层中的输入通道数为 8,对比卷积核大小分别为 3、5、7、9 时模型的参数量、浮点运算数(floating point operations, FLOPs)和 20 dB 环境下的 PQDs 识别准确率,如表 3 所示。

表 3 卷积核大小对参数量、FLOPs 和 PQDs 识别准确率的影响

Table 3 The Influence of convolutional kernel size on quantity, FLOPs, and PQDs recognition accuracy

卷积核	参数量	FLOPs	PQDs 识别准确率
3	10 516	2.380×10^6	0.925 8
5	12 292	2.652×10^6	0.957 1
7	14 068	2.924×10^6	0.960 8
9	15 844	3.192×10^6	0.954 7

从表 3 可以看出,当卷积核大小为 3 时,其准确率明显低于卷积核大小为 5、7 和 9 的情况。而在卷积核大小为 5、7 和 9 时,模型的准确率差异较小。然而,相较于卷积核为 7 和 9,卷积核为 5 时在模型参数量和 FLOPs 方面更具优势。综合考虑,将卷积核大小设置为 5。设置卷积核大小为 5 不变,不同模型通道数对参数量、FLOPs 和 PQDs 识别准确率的影响如表 4 所示。

表4 通道数对参数量、FLOPs 和 PQDs 识别准确率的影响

Table 4 The Influence of channel number on quantity, FLOPs, and PQDs recognition accuracy

通道数	参数量	FLOPs	PQDs 识别准确率
4	6 426	1.322×10^6	0.939 0
8	12 292	2.652×10^6	0.957 1
12	20 076	4.462×10^6	0.955 3
16	29 780	6.751×10^6	0.957 6

由表4可知,当模型的起始通道数为4时,其准确率比通道数为8、12和16时低较多。而在起始通道数为8、12和16时,模型的准确率差异较小。同时,起始通道数为8时,模型的参数量和FLOPs远低于起始通道数为12和16时。综上,模型的起始通道数选择为8。文中所提Cat模型的完整结构及相关参数设置如表5所示。

表5 CaT模型相关参数设置

Table 5 Parameter settings related to the CaT model

层名称	输出尺寸		参数设置
	通道数	序列长度	
输入卷积层	8	1 000	$C_{in} = 1, C_{out} = 8, K = 5, s = 1, p = 2$
CNN 模块1	8	1 000	$C_{in} = 8, C_{out} = 8, e = 2, K = 5, s = 1, p = 2$
CNN 模块2	12	250	$C_{in} = 8, C_{out} = 12, e = 2, K = 5, s = 4, p = 1$
Transformer 模块1	12	250	$C_{in} = 24, C_{out} = 12, k = 5, s = 1, p = 2, d = 24, z = 1$
CNN 模块3	12	250	$C_{in} = 12, C_{out} = 12, e = 2, K = 5, s = 1, p = 2$
CNN 模块4	16	62	$C_{in} = 12, C_{out} = 16, e = 2, K = 5, s = 4, p = 1$
Transformer 模块2	16	62	$C_{in} = 32, C_{out} = 16, K = 5, s = 1, p = 2, d = 32, z = 1$
平均池化层	16	1	
线性层	31	1	$h = 31$

表5中, s 为卷积核移动步长; p 为卷积层补零的个数; e 为CNN模块中的扩张率; d 为Transformer模块中的嵌入序列维度; z 为Transformer模块个数; h 为线性层神经元个数,其大小为PQDs类别数31。

2.3 消融实验

为了验证不同模块的积极作用,在其他模块参数不变的情况,建立了不同的模型进行对比实验。

(1) 模型1,无Transformer模块,无高效软阈值模块,高效软阈值模块由SiLU激活函数代替;

(2) 模型2,无Transformer模块,加入高效软阈值模块;

(3) 模型3,无Transformer模块,加入传统通道

软阈值模块;

(4) 模型4,加入Transformer模块,无高效软阈值模块,高效软阈值模块由SiLU激活函数代替;

(5) 模型5,加入Transformer模块,加入高效软阈值模块。

不同模块组合的模型中对PQDs识别准确率影响如表6所示。对比模型2和模型3可知高效软阈值模块和传统通道软阈值模块在不同噪声环境下互有胜负,但模型3可训练参数为6 824个,模型2可训练参数仅为3 192,其参数量仅为模型3的一半但模型性能相差不大,因此文中所提高效软阈值模块是有效的,能够大幅减少模型参数量,而不损失模型的性能。模型4在各个噪声等级下PQDs识别准确率均高于模型1,Transformer模块是有效的。同时文中所提模型5,在各个噪声等级下也取得了最高的识别准确率。以上分析可充分证明各模块对于PQDs识别的有效性,高效软阈值模块和Transformer模块的加入增加了模型在噪声环境中的鲁棒性,提高了模型在噪声环境中的识别准确率。

表6 不同模块组合的模型中PQDs识别准确率对比

Table 6 Comparison of PQDs recognition accuracy among models with different module combinations

模型	PQDs 识别准确率			
	无噪	40 dB	30 dB	20 dB
模型1	0.996 0	0.993 5	0.988 9	0.919 4
模型2	0.996 9	0.993 5	0.992 9	0.929 7
模型3	0.995 8	0.994 5	0.991 5	0.930 3
模型4	0.996 8	0.994 7	0.992 1	0.947 3
模型5	0.998 2	0.995 2	0.995 5	0.957 1

2.4 对比分析

为了验证文中所提方法相对于其他深度学习方法的优越性,对相同的数据集使用不同的模型进行训练和测试,模型包括Deep CNN^[21]、双向长短期记忆(bidirectional long short-term memory, BiLSTM)网络和GAF+Resnet^[24]。BiLSTM模型参考文献[32]对输入数据进行处理,LSTM层中隐藏层维度设置为128。为了验证各个模型的性能与噪声鲁棒性,在不同的噪声等级下均进行了训练、验证和测试,各个模型在不同噪声环境下的PQDs识别准确率如表7所示。

由表7可知,文中所提CaT模型在各个噪声环境下均取得了最高的准确率,特别是在20 dB噪声环境下,CaT模型准确率远高于其他3个模型。这是因为CaT模型加入了阈值化模块,使得模型对于噪声具有良好的鲁棒性。以上对比分析可充分表明

表 7 不同模型中 PQDs 识别准确率对比

Table 7 Comparison of PQDs recognition accuracy among different models

模型	PQDs 识别准确率			
	无噪	40 dB	30 dB	20 dB
CaT	0.998 2	0.995 2	0.995 5	0.957 1
Deep CNN	0.995 8	0.994 2	0.991 5	0.922 9
BiLSTM	0.984 2	0.981 1	0.957 3	0.932 1
GAF+Resnet	0.980 5	0.976 3	0.967 7	0.882 9

文中所提 CaT 模型识别的有效性及其在噪声环境下的优良鲁棒性。为了进一步表明文中所提模型的有效性,表 8 展示了 CaT 模型在不同噪声种类下,对于各类 PQDs 信号的识别准确率。

表 8 CaT 模型中不同种类 PQDs 识别准确率对比

Table 8 Comparison of PQDs recognition accuracy of different types in CaT model

扰动信号 编号	PQDs 识别准确率			
	无噪	40 dB	30 dB	20 dB
C1	1	1	1	0.970
C2	0.995	0.995	0.995	0.950
C3	1	1	0.995	0.930
C4	1	0.995	0.995	0.940
C5	1	1	1	0.965
C6	1	1	1	0.970
C7	1	1	1	0.955
C8	1	1	1	0.985
C9	1	1	1	0.960
C10	0.990	0.980	1	0.955
C11	1	0.995	0.990	0.960
C12	0.995	1	0.995	0.930
C13	0.985	0.985	0.980	0.965
C14	0.995	1	1	0.985
C15	1	0.995	0.995	0.960
C16	1	0.990	0.980	0.890
C17	1	1	1	0.915
C18	1	0.990	0.990	0.970
C19	1	1	1	1
C20	1	0.980	0.980	0.970
C21	0.995	0.990	1	0.955
C22	1	0.990	1	0.970
C23	0.990	0.990	0.990	0.980
C24	1	0.990	0.995	1
C25	1	1	1	0.965
C26	1	0.995	0.995	0.950
C27	1	0.995	0.995	0.940
C28	1	1	0.990	0.920
C29	1	1	1	0.960
C30	1	0.995	1	0.910
C31	1	1	1	0.995

从表 8 可知,随着噪声强度的加大,各类扰动信号的识别准确率均有所下降,但 C16(暂升+冲击)下降幅度较大。进一步分析混淆矩阵可知,模型将其错误识别为 C3(暂升),这可能是因为在 20 dB 噪声环境中暂升与冲击的特征极易混淆,最终导致模型识别错误。但除去 C16 外,其他扰动类型在 20 dB 噪声环境下仍能够保持识别准确率在 0.9 以上,以上分析可以表明 CaT 模型的 PQDs 识别有效性。

文中使用模型的参数量、FLOPs 和对 256 条 PQDs 信号的推理时间衡量各个模型的轻量化程度,各个模型的参数量、FLOPs 和推理时间如表 9 所示。

表 9 模型轻量化对比

Table 9 Comparison of lightweight models

模型	参数量	FLOPs	推理时间/ms
CaT	12 292	2.652×10^6	4.8
Deep CNN	166 207	94.5×10^6	7.8
BiLSTM	964 639	93.4×10^6	11.3
GAF+Resnet	11 186 112	$1 745 \times 10^6$	72.6

由表 9 可知,文中所提 CaT 模型的参数量和 FLOPs 远远小于其他模型,但是 CaT 模型在参数量和 FLOPs 最少的情况下,在不同的噪声等级下准确率却均高于其他模型,因此 CaT 模型更有利于部署到资源受限的环境中,降低 PQDs 识别的成本。同时 CaT 模型的推理时间更短,有利于工程的实际应用,完成 PQDs 的在线识别。图 9 展示了各个模型在 20 dB 的强噪声环境下的训练过程。

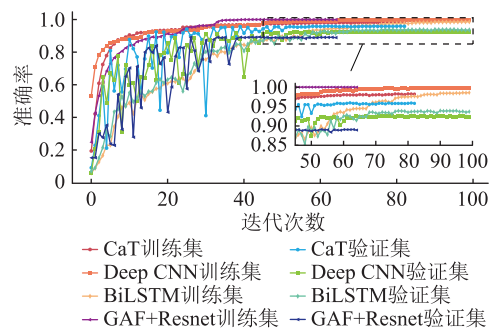


图 9 各个模型在 20 dB 噪声环境下的训练曲线

Fig.9 Training curves for each model in 20 dB noise environment

由图 9 可知,各个模型在训练前期都出现了较为严重的振荡,这是因为训练前期设置的学习率较大,随着迭代次数的增加各个模型都能够达到稳定,最终收敛。除 CaT 模型外,其他 3 个模型的训练集准确率和验证集的准确率出现了较为严重的偏差,训练集准确率已经近于 1,但验证集准确率在 0.9 左右,因此可以判断 Deep CNN、BiLSTM 和 GAF +

Resnet 均出现了较为严重的过拟合。这是由于对于 PQDs 识别来说 Deep CNN、BiLSTM 和 GAF+Resnet 的复杂度过高,参数过多。模型没有学习到深层次的规律,而仅是学到了训练数据中的每一个细节,最终导致模型的泛化性能不佳,在验证集上的准确率远低于训练集。而 CaT 模型的训练集准确率虽然低于其他模型,但验证集准确率均高于其他模型,同时 CaT 模型的训练集准确率和验证集准确率偏差较小,表明 CaT 模型并没有过拟合,具有良好的泛化性能。

2.5 特征可视化分析

为了增加模型的可解释性,理解模型各层特征的分布状态,文中采用 t 分布随机近邻嵌入 (t-distributed stochastic neighbor embedding, t-SNE) 将主要层输出的特征向量降维至二维空间进行可视化。20 dB 噪声环境下的测试集进行 t-SNE 可视化结果如图 10 所示。

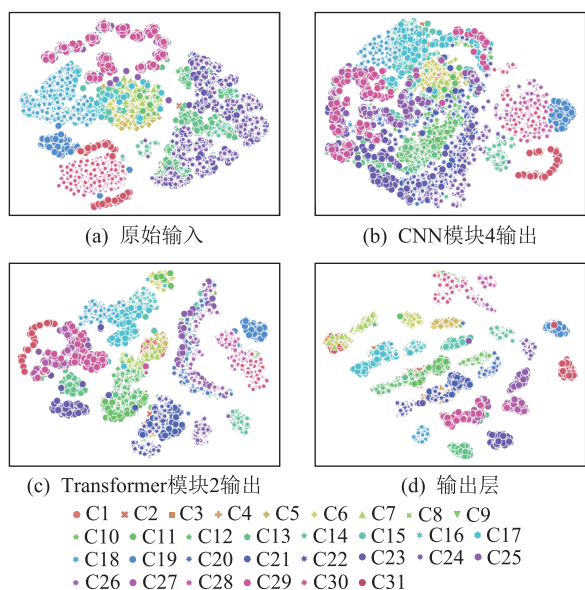


图 10 t-SNE 可视化

Fig.10 t-SNE visualization

由图 10 可知,原始 PQDs 信号具有较大的重叠部分,各类扰动信号相似度较高。经过 CNN 模块 4 后,各类扰动的特征重叠度减少,表明 CNN 模块的有效性。再经过 Transformer 模块 2 后各类扰动的特征扰动边界更为清晰。最后输出层的结果表明不同类的 PQDs 特征向量仅有少量的重叠,这进一步解释了 CaT 模型在 PQDs 识别上的有效性。

3 结论

为了将 PQDs 识别模型部署在资源受限环境下,文中提出了一种基于 CaT 的轻量化 PQDs 识别模型,使用深度可分离卷积提取局部特征,使用

Transformer 模块提取全局特征,使用高效软阈值模块提高模型对于噪声的鲁棒性。并经过仿真实验得到以下结论:

(1) 因 CaT 模型的轻量化和端到端设计,CaT 模型的参数量、FLOPs 和推理时间优于其他基于深度学习的 PQDs 识别模型,使得 CaT 模型更有利于部署在资源受限的环境下,更有利于工程实际应用。

(2) CaT 模型具有良好的泛化性能。CaT 模型在训练集上的准确率要低于其他深度学习模型,但在测试集和验证集上的准确率却高于其他模型。

(3) 引入高效软阈值模块,使得 CaT 模型具有良好的噪声鲁棒性。在 20 dB 的噪声环境下,CaT 模型的识别准确率能够达到 0.957 1,高于其他深度学习模型。

参考文献:

[1] 汪飞,全晓庆,任林涛. 电能质量扰动检测与识别方法研究综述[J]. 中国电机工程学报,2021,41(12):4104-4120.
 WANG Fei, QUAN Xiaoqing, REN Lintao. Summary of research on power quality disturbance detection and identification methods[J]. Proceedings of the CSEE, 2021, 41(12): 4104-4120.

[2] 黄雨涵,丁涛,李雨婷,等. 碳中和背景下能源低碳化技术综述及对新型电力系统发展的启示[J]. 中国电机工程学报, 2021, 41(S1): 28-51.
 HUANG Yuhuan, DING Tao, LI Yuting, et al. Decarbonization technologies and inspirations for the development of novel power systems in the context of carbon neutrality[J]. Proceedings of the CSEE, 2021, 41(S1): 28-51.

[3] KHOKHAR S, MOHD ZIN A A B, MOKHTAR A S B, et al. A comprehensive overview on signal processing and artificial intelligence techniques applications in classification of power quality disturbances[J]. Renewable and Sustainable Energy Reviews, 2015, 51: 1650-1663.

[4] 谈竹奎,文贤旭,杨涛,等. 面向新型电力系统的双馈风力发电机并网控制策略研究[J]. 电力系统保护与控制, 2023, 51(3): 181-187.
 TAN Zhukui, WEN Xiankui, YANG Tao, et al. A grid-connected control strategy for doubly-fed wind turbines for new power systems[J]. Power System Protection and Control, 2023, 51(3): 181-187.

[5] 丁坤,陈博洋,秦建茹,等. 大规模新能源集群接入弱电网的消纳能力评估方法[J]. 电力建设, 2023, 44(11): 86-94.
 DING Kun, CHEN Boyang, QIN Jianru, et al. Evaluation method of consumption ability of new large scale energy clusters connected to weak grids[J]. Electric Power Construction, 2023, 44(11): 86-94.

[6] 陈倩,朱婵霞,奚巍民,等. 基于国际对标的风光新能源发电市场政策机制思考[J]. 电力建设, 2024, 45(6): 120-129.
 CHEN Qian, ZHU Chanxia, XI Weimin, et al. Policy mechanism thinking for new energy power generation market based on international benchmarking[J]. Electric Power Construction, 2024,

- 45(6):120-129.
- [7] 吴雄,贺明康,何雯雯,等. 考虑储能寿命的风-光-火-储打捆外送系统容量优化配置[J]. 电力系统保护与控制,2023,51(15):66-75.
- WU Xiong, HE Mingkang, HE Wenwen, et al. Optimal capacity of a wind-solar-thermo-storage-bundled power transmission system considering battery life[J]. *Power System Protection and Control*, 2023, 51(15):66-75.
- [8] 钱倍奇,陈谦,李宗源,等. 基于马尔可夫转换场与多头注意力机制的电能质量扰动分类方法[J]. 电网技术,2024,48(2):721-729.
- QIAN Beiqi, CHEN Qian, LI Zongyuan, et al. Power quality disturbance classification method based on Markov transformation field and multi-head attention mechanism[J]. *Power System Technology*, 2024, 48(2):721-729.
- [9] 贺虎成,王承海,辛钟毓,等. 一种组合重构的电能质量扰动特征提取方法[J]. 电力系统保护与控制,2023,51(10):34-44.
- HE Hucheng, WANG Chenghai, XIN Zhongyu, et al. A feature extraction method for power quality disturbance based on reconstructed combination[J]. *Power System Protection and Control*, 2023, 51(10):34-44.
- [10] 王海东,程杉,徐其平,等. 基于深度学习融合网络的含噪电能质量扰动识别方法[J]. 电力系统保护与控制,2024,52(10):11-20.
- WANG Haidong, CHENG Shan, XU Qiping, et al. Identification of power quality disturbance with noises based on an integrated deep learning network[J]. *Power System Protection and Control*, 2024, 52(10):11-20.
- [11] 黄建明,瞿合祚,李晓明. 基于短时傅里叶变换及其谱峭度的电能质量混合扰动分类[J]. 电网技术,2016,40(10):3184-3191.
- HUANG Jianming, QU Hezuo, LI Xiaoming. Classification of power quality mixed disturbances based on short-time Fourier transform and its spectral kurtosis[J]. *Power System Technology*, 2016, 40(10):3184-3191.
- [12] 叶筱怡,刘海涛,吕干云,等. 基于BAS-BP分类器模型的电压暂降源识别[J]. 电力工程技术,2022,41(1):77-83.
- YE Xiaoyi, LIU Haitao, LYU Ganyun, et al. Identification of voltage sag source based on BAS-BP classifier model[J]. *Power System Technology*, 2022, 41(1):77-83.
- [13] ZHANG Y, JI T Y, LI M S, et al. Identification of power disturbances using generalized morphological open-closing and close-opening undecimated wavelet[J]. *IEEE Transactions on Industrial Electronics*, 2016, 63(4):2330-2339.
- [14] 赵立权,谢妮娜. 基于小波变换和改进的RVM的电能质量扰动分类[J]. 电工电能新技术,2013,32(4):74-78.
- ZHAO Liqun, XIE Nina. Classification of power quality disturbances based on wavelet transform and improved RVM[J]. *Advanced Technology of Electrical Engineering and Energy*, 2013, 32(4):74-78.
- [15] 田振果,傅成华,吴浩,等. 基于HHT的电能质量扰动定位与分类[J]. 电力系统保护与控制,2015,43(16):36-42.
- TIAN Zhenguo, FU Chenghua, WU Hao, et al. Power quality disturbance for location and classification based on HHT[J]. *Power System Protection and Control*, 2015, 43(16):36-42.
- [16] ZHAO L Q, GAI M J, WANG L. Classification of multiple power quality disturbances based on the improved SVM[C]//2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET). Chennai, India. IEEE, 2017:2625-2628.
- [17] 赵立权,龙艳. 基于改进的SVM的电能质量复合扰动分类[J]. 电工电能新技术,2016,35(10):63-68.
- ZHAO Liqun, LONG Yan. Classification of power quality compound disturbances based on improved SVM[J]. *Advanced Technology of Electrical Engineering and Energy*, 2016, 35(10):63-68.
- [18] 孔英会,车麟麟,苑津莎,等. 基于小波分解和数据挖掘中决策树算法的电能质量扰动识别方法[J]. 电网技术,2007,31(23):78-82.
- KONG Yinghui, CHE Linlin, YUAN Jinsha, et al. A power quality disturbance identification method based on wavelet decomposition and decision tree algorithm in data mining[J]. *Power System Technology*, 2007, 31(23):78-82.
- [19] KUMAR R, SINGH B, SHAHANI D T, et al. Recognition of power-quality disturbances using S-transform-based ANN classifier and rule-based decision tree[J]. *IEEE Transactions on Industry Applications*, 2015, 51(2):1249-1258.
- [20] 许立武,李开成,肖贤贵,等. 基于深度前馈网络的电能质量复合扰动识别[J]. 电测与仪表,2020,57(1):62-69,130.
- XU Liwu, LI Kaicheng, XIAO Xiangui, et al. Recognition of power quality complex disturbances based on deep feedforward network[J]. *Electrical Measurement & Instrumentation*. 2020, 57(1):62-69,130.
- [21] WANG S X, CHEN H W. A novel deep learning method for the classification of power quality disturbances using deep convolutional neural network[J]. *Applied Energy*, 2019, 235:1126-1140.
- [22] 曹梦舟,张艳. 基于卷积-长短期记忆网络的电能质量扰动分类[J]. 电力系统保护与控制,2020,48(2):86-92.
- CAO Mengzhou, ZHANG Yan. Classification for power quality disturbances based on CNN-LSTM network[J]. *Power System Protection and Control*, 2020, 48(2):86-92.
- [23] 王继东,王泽平,张迪. 基于递归图和预训练迁移学习的电能质量扰动分类[J/OL]. 南方电网技术:1-10[2024-11-10]. <http://kns.cnki.net/kcms/detail/44.1643.tk.20240517.1116.002.html>.
- WANG Jidong, WANG Zeping, ZHANG Di. Power quality disturbance classification based on recursive graph and pre-trained transfer learning[J/OL]. *Southern Power System Technology*:1-10[2024-11-10]. <http://kns.cnki.net/kcms/detail/44.1643.tk.20240517.1116.002.html>.
- [24] 贺才郡,李开成,杨王旺,等. 基于双通道GAF和深度残差网络的电能质量复合扰动识别[J]. 电网技术,2023,47

- (1):369-376.
- HE Caijun, LI Kaicheng, YANG Wangwang, et al. Power quality compound disturbance identification based on dual-channel GAF and depth residual network[J]. Power System Technology, 2023, 47(1):369-376.
- [25] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. IEEE, 2018;4510-4520.
- [26] 盖勤. 基于深度学习的电能质量检测方法研究[D]. 北京: 北京交通大学, 2021.
- GAI Qin. Research on power quality detection method based on deep learning[D]. Beijing: Beijing Jiaotong University, 2021.
- [27] 奚鑫泽, 邢超, 覃日升, 等. 基于深度卷积去噪网络的电能质量扰动识别方法[J]. 南方电网技术, 2022(12):118-125.
- XI Xinze, XING Chao, QIN Risheng, et al. Power quality disturbance identification method based on deep convolutional denoising network[J]. Southern Power System Technology, 2022(12):118-125.
- [28] 贺才郡, 李开成, 董宇飞, 等. 基于知识蒸馏与 RP-MobileNetV3 的电能质量复合扰动识别[J]. 电力系统保护与控制, 2023, 51(14):75-84.
- HE Caijun, LI Kaicheng, DONG Yufei, et al. Power quality compound disturbance identification based on knowledge distillation and RP-MobileNetV3[J]. Power System Protection and Control, 2023, 51(14):75-84.
- [29] MA N N, ZHANG X Y, ZHENG H T, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design [M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018:122-138.
- [30] 许立武. 基于深度学习的电能质量复合扰动识别研究[D]. 华中科技大学, 2020.
- XU Liwu. Study on recognition of complex power quality disturbances based on deep learning[D]. Huazhong University of Science and Technology, 2020.
- [31] 罗奕. 复杂电能质量扰动事件的分类方法研究[D]. 武汉: 华中科技大学, 2018.
- LUO Yi. Research on complex power quality disturbances classification[D]. Wuhan: Huazhong University of Science and Technology, 2018.
- [32] 龚正, 邹阳, 金涛, 等. 基于特征融合并行优化模型的电能质量扰动分类方法[J]. 中国电机工程学报, 2023, 43(3):1017-1026.
- GONG Zheng, ZOU Yang, JIN Tao, et al. Power quality disturbance classification method based on feature fusion parallel optimization model[J]. Proceedings of the CSEE, 2023, 43(3):1017-1026.

作者简介:



张彼德

张彼德(1977),男,博士,教授,研究方向为电气设备故障诊断及电能质量分析(E-mail:fyhzzx2015@sina.com);

邱杰(2000),男,硕士在读,研究方向为电能质量分析;

姜广鑫(1983),男,硕士,工程师,从事电能质量分析相关工作。

A lightweight power quality disturbance recognition model based on CNN and Transformer

ZHANG Bide¹, QIU Jie¹, LOU Guangxin², ZHOU Can¹, LUO Qingqing¹, LI Tianqian¹

(1. School of Electrical Engineering and Electronic Information, Xihua University, Chengdu 610039, China;

2. Chengdu Power Supply Company, State Grid Sichuan Electric Power Company, Chengdu 610000, China)

Abstract: A lightweight power quality disturbances (PQDs) recognition model that integrates convolutional neural network (CNN) and Transformer (CaT) is proposed to address the high number of parameters and computational complexity in existing deep learning-based models. Depthwise separable convolutions are first employed to extract local features from the disturbance signals. An efficient softthreshold block is then introduced to reduce noise and redundant features without significantly increasing the model's parameters or complexity. The Transformer model is used to capture global features of the disturbance signals. Finally, pooling layers, fully connected layers, and Softmax are applied to complete the recognition PQDs. Simulation experiments demonstrate that the CaT model effectively recognizes PQDs with fewer parameters and floating point operations, achieving high accuracy and strong noise robustness. Its lightweight, end-to-end design also results in shorter inference times compared to other deep learning models.

Keywords: power quality disturbances (PQDs); lightweight; number of parameters; efficient soft threshold block; depthwise separable convolution; Transformer model

(编辑 吴昊)