

# 宽卷积局部特征扩展的 Transformer 网络故障诊断模型<sup>\*</sup>

张新良<sup>1</sup> 李 占<sup>1</sup> 周益天<sup>2</sup>

(1. 河南理工大学 电气工程与自动化学院 河南省智能装备直驱技术与控制国际联合实验室 焦作 454003;

2. 舟山洋旺纳新科技有限公司 舟山 316104)

**摘 要:**视觉 Transformer 网络的高精度诊断性能依赖于充分的训练数据,利用卷积网络在提取局部特征上的优势,构造能同时描述故障局部和全局特征的提取层,提高诊断模型的抗噪声干扰能力。首先,引入卷积网络模块将原始振动信号转换为 Transformer 网络可以直接接收的特征向量,提取故障局部特征,并通过增加卷积网络的感受野。然后,结合 Transformer 网络多头自注意力机制生成的全局信息,构建能同时描述故障局部和全局特征的特征向量。最后,在 Transformer 网络的预测层,利用高效通道注意力机制对特征向量的贡献度进行自动筛选。在西储大学(CWRU)轴承数据集上的故障诊断结果表明,在信噪比-4 dB 的噪声干扰下,改进后的 Transformer 网络轴承故障诊断模型的准确率达 90.21%,与原始 Transformer 模型相比,准确率提高了 13.2%,在噪声环境下表现出优异的诊断性能。

**关键词:**轴承故障诊断;视觉 Transformer;宽卷积核;自注意力机制;局部-全局特征;高效通道注意力

**中图分类号:** TP391 **文献标识码:** A **国家标准学科分类代码:** 460.99

## Fault diagnosis model of Transformer network with wide convolution local feature extension

Zhang Xinliang<sup>1</sup> Li Zhan<sup>1</sup> Zhou Yitian<sup>2</sup>

(1. Henan International Joint Laboratory of Direct Drive and Control of Intelligent Equipment, School of  
Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo 454003, China;

2. Zhoushan Yangwangnaxin Technology Co. Ltd., Zhoushan 316104, China)

**Abstract:** The high-precision diagnosis performance of the vision Transformer network depends on adequate training data. Using the advantage of convolutional network in extracting local features construct an extraction layer that can describe both local and global features of faults, and improve the anti-noise diagnosis capability of the diagnostic model. First, the convolutional network module is introduced to convert the original vibration signal into a feature vector that can be directly received by the Transformer network to extract the local features of the fault. Then, the global information generated by the multi-head self-attention mechanism of Transformer network is combined to construct the feature vector that can describe both local and global features of the fault. Finally, in the prediction layer of the Transformer network, the contribution of the feature vectors is automatically filtered using an efficient channel attention mechanism. The fault diagnosis results on the case western reserve university (CWRU) bearing dataset show that the improved Transformer network bearing fault diagnosis model achieves an accuracy of 90.21% under the noise interference with a signal-to-noise ratio of -4 dB, which is a 13.2% improvement in accuracy compared with the original Transformer model, and shows excellent diagnostic performance in a noisy environment.

**Keywords:** bearing fault diagnosis; vision Transformer; wide convolution kernels; self-attention mechanism; local-global feature; efficient channel attention

收稿日期:2023-08-15

<sup>\*</sup> 基金项目:国家自然科学基金(U1404612)、河南省高校基本科研业务费专项(NSFRF210305)、河南省科技攻关项目(222102210274)资助

## 0 引言

深度学习网络能够实现目标特征自动提取和分类,其摆脱了对信号预处理和专家知识的依赖,在海量数据处理方面具有显著的优势<sup>[1]</sup>。将深度学习网络用于机械系统的故障特征提取和诊断,具有重要的研究价值和广泛的应用前景。机械振动信号中包含了丰富的系统运行状态信息,基于振动信号的轴承故障诊断是目前常用和有效的方法<sup>[2]</sup>。作为旋转机械重要组件的滚动轴承实际工程中故障概率高达30%,及时高效地轴承故障诊断对于工业生产具有重要意义<sup>[3]</sup>。传统的深度学习网络轴承故障诊断模型通常是以二维图像数据作为输入,采用截取一重排、傅里叶变换、Hilbert-Huang变换等方式<sup>[4]</sup>,将原始一维时序信号转换为二维数据,如卷积神经网络(convolutional neural network, CNN)的模型<sup>[5]</sup>、深度残差网络(deep residual neural network, DResNet)模型<sup>[6]</sup>、长短期记忆网络(long short term memory, LSTM)模型<sup>[7]</sup>等。考虑到滚动轴承的周期性运动形式,时序信号中的故障信息往往体现为与距离相关的全局特征,上述截取一重排类的方法会因造成受距离约束的全局信息缺失,而傅里叶变换、Hilbert-Huang变换等则需要人工指定参数,对数据缺乏适应性<sup>[8]</sup>。

与上述一维振动信号转换二维图像的方法不同,Zhao等<sup>[9]</sup>从“端到端”的故障诊断模型需求出发,直接以轴承振动信号为模型输入,提出了一种深度残差收缩网络(deep residual shrinkage networks with channel-wise thresholds, DRSN-CW)故障诊断模型,并利用注意力机制提高模型的抗噪声能力。Yan等<sup>[10]</sup>利用深度置信网络对多传感器一维时序振动信号提取故障特征并进行轴承故障诊断。Wen等<sup>[11]</sup>采用一维卷积自编码进行时序信号的故障特征提取,实现了无监督状态下的故障诊断。受“端到端”卷积网络本身串行结构的限制,其提取到的轴承故障全局特征会减弱,需要通过增加额外的并行网络结构来增强全局信息的提取能力和模型的抗噪声能力<sup>[12]</sup>。

循环神经网络 LSTM,基于并行的网络结构,利用轴承振动信号的时间顺序性,实现受距离约束的全局信号特征提取<sup>[13]</sup>,但其样本训练无法并行化,网络的运行效率低。视觉 Transformer 网络<sup>[14]</sup>利用多头自注意力机制保证了对目标的全局特征提取能力,同时其训练过程可并行,对于构建“端到端”的轴承故障模型具有借鉴意义。Fang等<sup>[15]</sup>使用卷积操作来代替 Transformer 的特征提取和位置编码,将网络改进为 CLFormer,并使用裁剪和蒸馏方法对模型参数进行优化。黄星华等<sup>[16]</sup>使用 Transformer 的变体进行特征提取,结合域对抗策略提升跨域诊断的精度。Ding等<sup>[17]</sup>将时域振动信号转换为时-频图像,同样使用了 Transformer 编码器提取故障特征并选取分类标志,最终完成故障分类。刘俊锋等<sup>[18]</sup>基于 Trans-

former 编码器和 CNN 对轴承时序信号进行建模,并结合迁移学习的方法解决跨域下模型泛化能力差的问题。杜康宁等<sup>[19]</sup>将时域振动信号转换成时频图像,输入到视觉 Transformer 网络中捕获全局信息,提升了诊断精度。上述方法利用 Transformer 提取全局特征信息,虽然获得了较高的诊断精度,但是它们并没有充分考虑到局部特征提取对于网络诊断精度的影响,尤其是噪声干扰导致数据不充分的情况;另一方面,采用手动选取分类特征向量或初始化分类标志的模式存在步骤繁琐效率低下的问题,不利于提高复杂环境下诊断模型的运行效率。

考虑卷积网络在目标局部特征提取上的特性,将其引入至 Transformer 映射层,同时使用宽卷积核减少对序列之间长距离依赖关系的影响,构造能够同时描述故障局部-全局特征的提取层,为轴承故障生成特征向量。同时,在 Transformer 的预测层,为提高对故障特征向量的筛选能力,基于注意力机制设计一个高效特征筛选器,利用高效通道注意力(efficient channel attention, ECA)机制<sup>[20]</sup>对特征向量进行度量,自适应增强重要特征。

## 1 宽卷积核局部特征扩展映射

通过重新设计基于卷积网络的信号映射层以替代原来的嵌入层,直接处理振动信号,增加模型的局部特征提取能力,解决训练样本不足而引起的精度下降问题。

局部特征扩展卷积网络结构如图1所示,其采用一维宽核(wide kernel, WK)卷积作为主干网络,输入为一维振动信号时间序列  $\mathbf{X}=[x_1, x_2, \dots, x_n]^T$ ,经宽卷积变换和池化后,生成轴承故障的浅层特征向量。即:

$$\mathbf{Y}_i^l = \text{pooling}(\sigma(\sum_{j=1}^m \mathbf{Y}_j^{l-1} \otimes \mathbf{w}_i^l + \mathbf{b}_i^l)) \quad (1)$$

其中,  $\mathbf{Y}^l = \{\mathbf{Y}_i^l\}$ 、 $\mathbf{Y}^{l-1} = \{\mathbf{Y}_j^{l-1}\}$  为当前第  $i$  个卷积层及前一个卷积层的输出特征向量,  $n, m$  为其对应的输出通道个数。 $\mathbf{w}_i^l$  和  $\mathbf{b}_i^l$  分别为第  $i$  个通道的卷积核权重及偏置,  $\sigma(\cdot)$  为神经网络的激励函数。特别地,当  $l=1$  时,有:

$$\mathbf{Y}^0 = \mathbf{X} \quad (2)$$

卷积神经网络通过固定大小的卷积核与输入信号的局域做卷积运算,具有局部特征提取能力。同时,宽卷积核卷积神经网络(deep CNN with wide first-layer kernel, WDCNN)<sup>[21]</sup>,其卷积操作在功能上与短时傅里叶变换相似<sup>[22]</sup>,可以捕捉时序相关性特征和长距离依赖关系,在复杂的噪声环境中能保持特征的稳定性。考虑输入特征图像的大小如  $64 \times 64$  和  $80 \times 80$ ,其卷积核大小通常选择为  $5 \times 5$  或  $9 \times 9$ ,侧重于提取局部故障特征。本文中,输入故障样本为  $1024 \times 1$ ,在保证局部特征提取的同时减少对序列之间长距离依赖关系的影响,选择较宽的卷积核,图1中大小为  $120 \times 1$  的卷积核,步长为6,实现对时序相关性特征的提取。

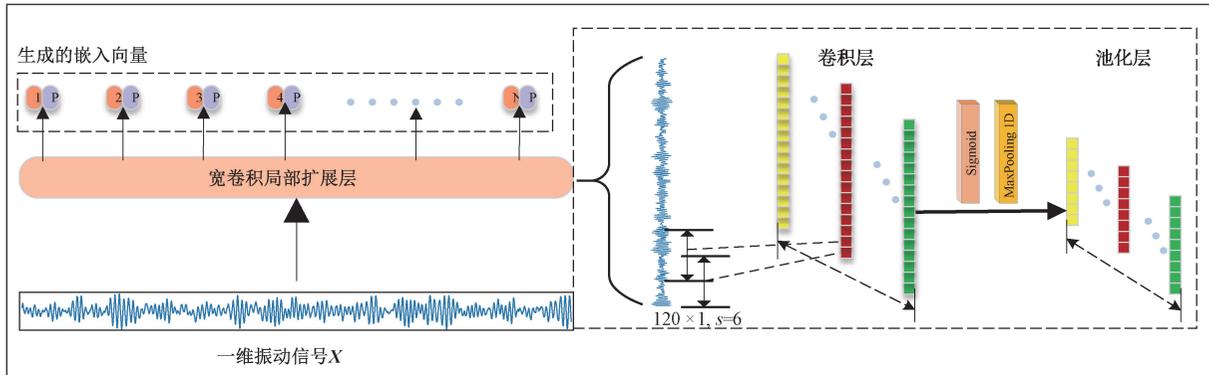


图1 宽卷积核局部特征扩展网络

考虑 Transformer 并行结构无法保留轴承振动信号时序数据的前后关联性,借鉴于 NLP 中的单词嵌入思想,对宽卷积核卷积网络输出的浅层特征进行位置编码,构造含有位置信息的嵌入向量  $\mathbf{Y}$ ,即:

$$\mathbf{Y} = \mathbf{Y}^l + \mathbf{P}, \mathbf{P} \in \mathbf{R}^{n \times L} \quad (3)$$

式中:  $L$  和  $n$  分别表示宽卷积核卷积网络输出特征向量的维度(长度)和通道数;  $\mathbf{P}$  为与特征向量同维度的自适应位置编码。

## 2 高效特征筛选器

当无差别地送入预测层进行轴承故障诊断时,将会因输入数据量过大而导致预测层的权重调整负担加重,针对这一问题本文创新性地多个全局-局部特征向量的信息看作为多个通道的信息,结合注意力机制在自动筛选高贡献度特征向量方面的作用引入 ECA 自动筛选用于分类的故障特征向量,提高故障诊断模型预测层的运行效率,网络结构如图 2 所示。

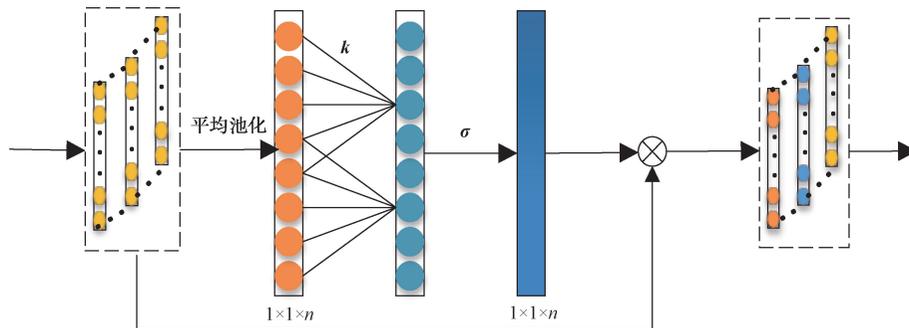


图2 高效通道注意力模块原理

Transformer 网络的编码层生成与嵌入向量通道数和维度大小相同的轴承故障多通道局部-全局特征向量,把这些向量全部送入预测层进行轴承故障诊断时,将会因输入数据量过大而导致预测层的权重调整负担加重。本文将多个全局-局部特征向量的信息类比为多个通道的信息,结合注意力机制筛选高贡献度特征向量,提高故障诊断模型预测层的运行效率,借助 ECA 构建一个特征向量组自动筛选器(图 2)。与传统注意力机制网络模型(squeeze-and-excitation networks, SENet)不同,ECA 将全连接层替换为一维卷积层,获取通道间约束关系。

对于 Transformer 网络编码层生成的特征向量  $\mathbf{X} \in \mathbf{R}^{n \times L}$ ,其在特征通道上的全局分布为:

$$X_i^{avg} = \frac{1}{L} \sum_{j=1}^L X_i(j) \quad (4)$$

各通道间的依赖关系,通过对输入特征进行加权操作实现,即:

$$\mathbf{G} = \sigma(\text{conv}(\mathbf{X}^{avg})) \quad (5)$$

$$\mathbf{X}^o = \mathbf{G} \cdot \mathbf{X} \quad (6)$$

式中:  $\mathbf{X}^{avg} = \{X_i^{avg}\}$ ,  $\mathbf{X}^o = \{X_i^o\}$ ,  $i = 1, 2, \dots, n$ , “ $\cdot$ ”表示点乘运算。卷积操作  $\text{conv}(\cdot)$  通过不同卷积核大小获取跨通道依赖关系,其卷积核大小  $k$  决定了当前通道特征与邻近的  $k$  个通道的交互方式,由特征向量的通道数  $n$  确定<sup>[21]</sup>,即:

$$k = \text{odd}\left(\frac{1 + \log_2 n}{2}\right) \quad (7)$$

式中:  $\text{odd}(\cdot)$  表示向上取奇数操作函数。

进一步,依据向量元素之和最大的原则选取最终用于故障分类的特征向量,即:

$$F = \underset{x}{\operatorname{argmax}}(|X_i^o|_1) \quad i = 1, 2, \dots, n \quad (8)$$

式中： $X_i^o$  为加权后的第  $i$  个通道所表示的特征向量； $|\cdot|_1$  为 1-范数； $F \in \mathbf{R}^{1 \times L}$  为从特征向量  $X^o$  中筛选出的分类特征向量。

### 3 改进的 Transformer 轴承故障诊断模型

利用设计的宽卷积核特征扩展网络和高效特征筛选

器对视觉 Transformer 网络进行改进,构造能够同时考虑特征稳定性、局部特征增强以及抗噪声能力的轴承故障诊断模型 WK-AVT (wide kernel attention vision transformer),如图 3 所示。其编码层主要由两个顺序连接的特征提取模块构成,在每一特征提取模块中,基于多头自注意力机制的残差连接完成多通道特征的提取,基于多层感知器的残差连接用于提升网络的特征描述能力。

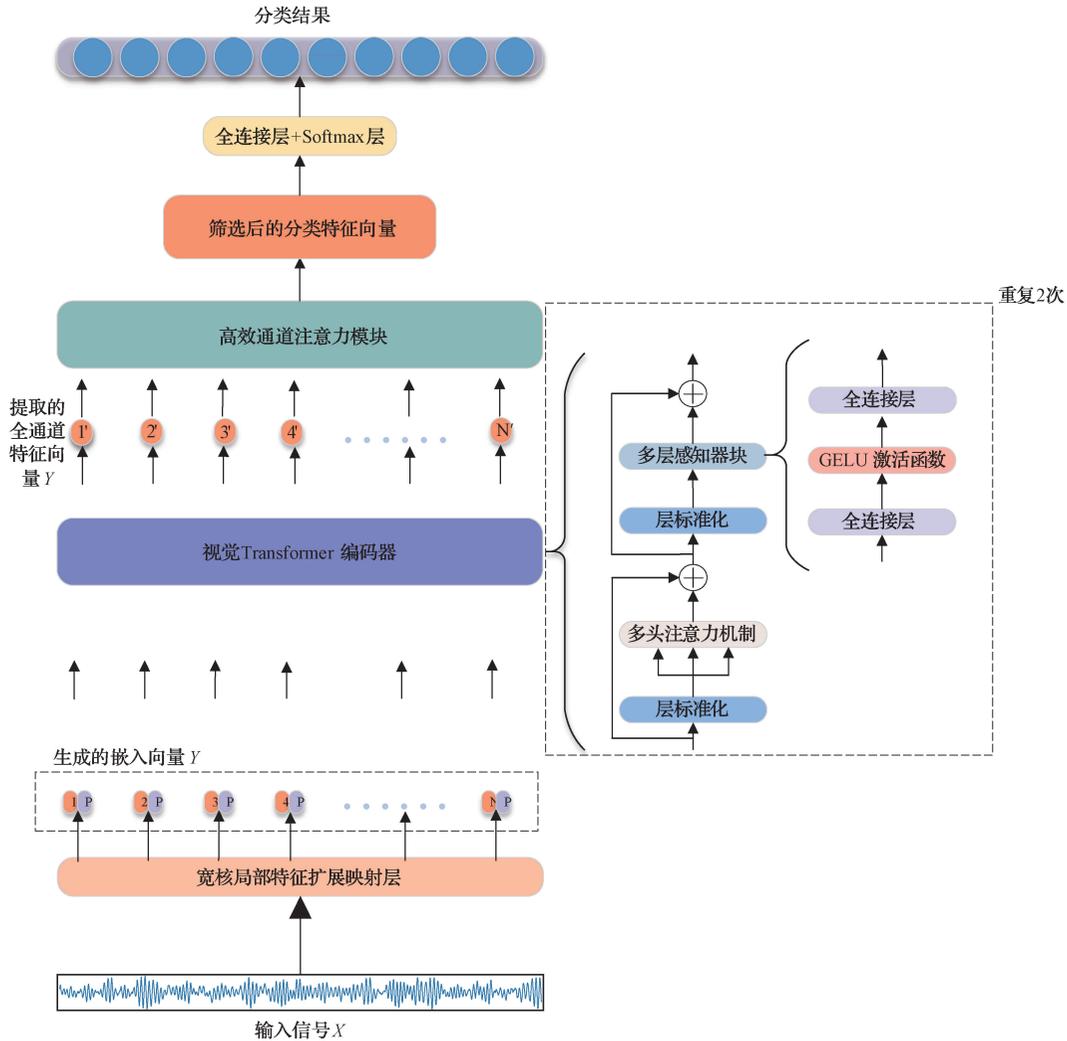


图 3 WK-AVT 模型结构框架

多头自注意力机制则由并行的  $h$  个自注意力模块构成,在每一个自注意力模块  $i$  中,通过度量各个向量之间的相似度来获得注意力权重,加权生成当前注意力模块的输出特征向量  $H_i$ ,相似度定义为:

$$S_i = \frac{Q_i K_i^T}{\sqrt{d}} \quad (9)$$

其中:

$$Q_i = \vec{Y}^T \times W_i^Q \quad (10)$$

$$K_i = \vec{Y}^T \times W_i^K \quad (11)$$

式中: $W_i^Q \in \mathbf{R}^{L \times d}$  和  $W_i^K \in \mathbf{R}^{L \times d}$  为查询向量矩阵  $Q_i$ 、关键字矩阵  $K_i$  的参数映射矩阵,由全连接网络构造,  $i = 1, 2, \dots, h$ ;  $d$  为映射后的向量维度; $\vec{Y}^T$  表示经过宽卷积核信号嵌入层生成的  $Y$  经过维度变换并归一化后的嵌入向量。

进一步,当前自注意力模块  $i$  的输出特征向量  $H_i$ ,可表示为:

$$G_i = \operatorname{Softmax}(S_i) \quad (12)$$

$$H_i = G_i \cdot V_i \quad (13)$$

其中:

$$V_i = \vec{Y}^T \times W_i^V \quad (14)$$

$\mathbf{W}_i^v \in \mathbf{R}^{L \times d}$  为值矩阵  $\mathbf{V}_i$  的参数映射矩阵,由全连接网络构造。

自注意力模块输出特征向量通过维度拼接和全连接网络,生成多头全通道特征向量,即:

$$\mathbf{X}_l = \text{concat}(\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_i, \dots, \mathbf{H}_h) \times \mathbf{W}^o \quad (15)$$

同时,引入残差连接解决模型训练过程中的梯度消失和梯度爆炸问题,残差网络模块输出为:

$$\mathbf{X}_{ii} = \mathbf{X}_l + \mathbf{Y} \quad (16)$$

进一步,通过前馈神经网络多层感知器进一步提升网络的特征描述能力,从而生成全通道轴承故障特征向量,即:

$$\mathbf{X} = \mathbf{X}_{ii} + \text{MLP}(\overline{\mathbf{X}}_{ii}) \quad (17)$$

式中:  $\text{MLP}(\cdot)$  为多层感知器。

$$\text{MLP}(x) = \sigma(\mathbf{W}_1 x + \mathbf{b}_1) \mathbf{W}_2 + \mathbf{b}_2 \quad (18)$$

式中:  $\mathbf{W}_1$  和  $\mathbf{W}_2$  为全连接的权重;  $\sigma$  为非线性激励函数;  $\mathbf{b}_1$  和  $\mathbf{b}_2$  为全连接的偏置。

最后,基于局部-全局信息的轴承故障特征向量,利用高效通道注意力网络(式(4)~(8)),自动筛选关键特征作为分类标志,完成故障类型的识别:

$$\mathbf{Pr} = \text{SoftMax}(\mathbf{F} \times \mathbf{W}) \quad (19)$$

式中:  $\mathbf{W}$  为预测层全连接网络权值矩阵;  $\mathbf{F}$  为高效通道注意力网络筛选输出的特征向量;  $\mathbf{Pr}$  为对应各轴承故障的预测输出值。

将 WK-AVT 故障诊断模型的损失函数定义为预测输出概率分布与目标概率分布之间的交叉熵,即:

$$H(p, q) = - \sum_x p(x) \log q(x) \quad (20)$$

式中:  $p(x)$  为目标分布;  $q(x)$  为预测分布。通过基于梯度下降算法的 SGD 优化器最小化损失函数来训练模型中的参数。

## 4 实验验证与结果分析

### 4.1 实验参数设置

利用西储大学(CWRU)轴承数据集验证所提出的

Transformer 故障诊断模型的有效性。该数据集包含了轴承内圈(inner race, IR)、外圈(outer race, OR)以及滚动体(ball)上的直径为 0.007、0.014 和 0.021 inch (1 inch=24.5 mm)的 9 类故障振动信号。实验中定义为 3 种损伤直径上的内圈故障、外圈故障和滚动体故障,结合正常工作(normal)状态,共 10 种故障类型,如表 1 所示。对振动信号序列进行无重叠连续分割采样,构造故障诊断模型的训练样本集和测试样本集,样本长度为 1 024,即每种故障类型 460 个样本。对应每一种轴承故障类型,选取 360 个样本用于模型训练,其他 100 个样本用于模型的测试。

表 1 CWRU 数据样本数量和标签

故障类型	训练样本数	测试样本数	标签
Ball_007	360	100	0
Ball_014	360	100	1
Ball_021	360	100	2
IR_007	360	100	3
IR_014	360	100	4
IR_021	360	100	5
Normal	360	100	6
OR_007	360	100	7
OR_014	360	100	8
OR_021	360	100	9

实验使用 Python 3.8 编程语言和 Pytorch 深度学习框架编写网络模型代码,并在 Ubuntu 20.04 操作系统、Torchvision 0.11.1 版本机器学习框架和带有 NVIDIA GeForce 3080Ti GPU 的 Intel i7-11700K CPU 上进行,Transformer 故障诊断模型的参数设置如表 2 所示。作为对比,相同实验条件下,利用 WDCNN 模型<sup>[20]</sup>、结合有循环神经网络(RNN)的 WDCNN 模型<sup>[23]</sup>以及深度残差收缩网络模型(DRSN-CW)<sup>[9]</sup>对上述轴承故障数据进行测试,其模型的参数设置同原文献中的保持一致。

表 2 WK-AVT 故障诊断模型结构参数

网络层数	结构名称	具体参数	输出尺寸
0	输入层	/	1 024 × 1
1	映射层	O_C:300;k:120;s:6;m_p:2;	75 × 300
2	编码层 1	h:3;dim:300;Act:GELU	75 × 300
3	编码层 2	h:3;dim:300;Act:GELU	75 × 300
4	ECA 层	k:3;s:1;p:1;Act:Sigmoid	1 × 300
5	全连接层	In:300;Out:10	1 × 10

注:O\_C 代表卷积核个数;k 代表卷积核大小;s 代表步长;m\_p 代表最大池化;h 表示多头自注意力网络的头数目;dim 表示映射的维度;Act 表示激励函数;p 表示 padding 大小;In 表示输入的维度;Out 表示输出的维度

为了测试诊断模型在噪声条件下的诊断性能,在 CWRU 数据集上叠加不同信噪比(SNR)的白噪声模拟其在噪声环境中的工作条件,其中信噪比定义为:

$$\text{SNR} = 10 \lg \frac{P_{\text{signal}}}{P_{\text{noise}}} \quad (21)$$

式中:  $P_{\text{signal}}$ 、 $P_{\text{noise}}$  分别表示信号和噪声的功率。

## 4.2 CWRU 数据集上的故障诊断结果

使用SGD优化器训练WK-AVT轴承故障诊断模型,训练过程中学习率设置为0.001,冲量为0.9,训练次数为50。作为对比,WDCNN诊断模型、RNN-WDCNN诊断模型以及DRSN-CW分别采用原文献中的Adam优化算法、Adam优化算法及SGD优化算法。

各个模型的训练准确率和测试准确率变化曲线如图4所示。由图4(a)可以看出,WK-AVT故障诊断模型以及WDCNN模型、RNN-WDCNN模型、DRSN-CW模型对轴承故障信号均具有较好的适应性,其对训练集的诊断准确率均达到99%以上,对测试集的预测准确率达到98%以上,

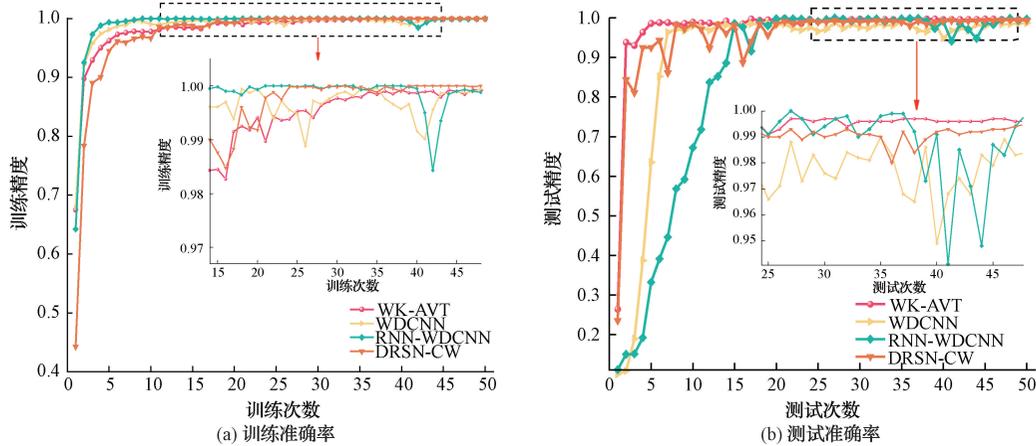


图4 故障诊断模型训练和测试准确率变化曲线

表3 不同模型的诊断准确率 (%)

模型	训练集	测试集
WK-AVT	99.93	99.7
WDCNN	99.80	98.70
RNN-WDCNN	99.98	99.60
DRSN-CW	100	99.4

对应图4(a)的训练数据集上的诊断准确率变化曲线,上述4个诊断模型在测试数据集上的准确率变化曲线如图4(b)所示。可以看出,相较于其他3个模型,RNN-WDCNN模型的测试准确率上升缓慢,与训练集上的准确率变化曲线不同,其在第20次训练周期才达到稳定状态,且模型的准确率在第42、43、44训练周期出现明显的波动。WDCNN模型的测试准确率最低,即使在训练准确率保持稳定的情况下,其准确率仍有明显的波动。与上述RNN-WDCNN模型和WDCNN模型相比,WK-AVT模型和DRSN-CW模型在其训练准确率达到稳定后,同样保持了较高的测试准确率,从第25个训练周期后,测试准确率均达到99.3%以上,且WK-AVT模型的测试准确率始终优于DRSN-CW模型,具有更好的预测性能。

## 4.3 噪声环境CWRU数据集故障诊断结果

在CWRU数据集上分别叠加倍噪比为10、8、4、2、0、

如表3所示。其中,模型训练过程中,RNN-WDCNN模型收敛速度最快,在第7个训练周期便基本达到稳定状态,准确率为99.5%,但其第42个周期出现了较大波动,精度下降至98.3%。WDCNN模型收敛速度最慢,且其收敛曲线一直存在较大的波动,最大波动幅值约1%。相比较而言,WK-AVT模型和DRSN-CW模型表现出了更稳定和平滑的收敛性能,虽然分别第33个和25个训练周期时才达到稳定状态,收敛速度低于RNN-WDCNN模型,但其显示出了更好的稳定性,WK-AVT模型在第33个训练周期后准确率保持在 $99.80\% \pm 0.13\%$ ,而DRSN-CW模型的准确率在第25个训练周期后保持在99.90%以上。

-2和-4 dB的噪声信号生成故障诊断测试数据。故障类别为“Ball\_021”数据集上不同噪声干扰下的振动数据如图5所示,表4为相应的欧氏距离和相关系数。可以看出,随着噪声影响程度的增加,欧氏距离的值在增加、相关系数的值在减小,说明噪声的引入对原始轴承振动信号产生了影响,对深度网络故障诊断模型的抗噪性能提出了挑战。诊断模型输入噪声数据后的训练准确率和损失变化曲线如图6所示,其中WK-AVT模型在训练后期的损失变化波动最小,结合图7所示的测试准确率可知其整体表现最优。

在生成的噪声数据测试样本上验证上述4种模型的故障预测能力,结果如图7所示。可以看出,4种模型的测试准确率随着噪声强度的增加都呈现下降趋势,在噪声干扰较小时,由于此时模型的诊断准确率主要取决于模型本身提取故障特征的能力,因此准确率下降比较缓慢,但当信噪比超过0 dB即噪声的功率大于原始信号功率时,准确率下降明显,4种诊断模型体现出对噪声的不同适应能力。其中,WDCNN模型由于网络中仅使用了宽卷积核,受噪声影响其对全局信息的提取能力减弱,准确率下降最明显,在-4 dB的噪声干扰下,其预测准确率仅为80.8%。DRSN-CW模型因其网络中跨连接增强了全局特征的提取能力,同时其内部阈值注意力模块能够自适应

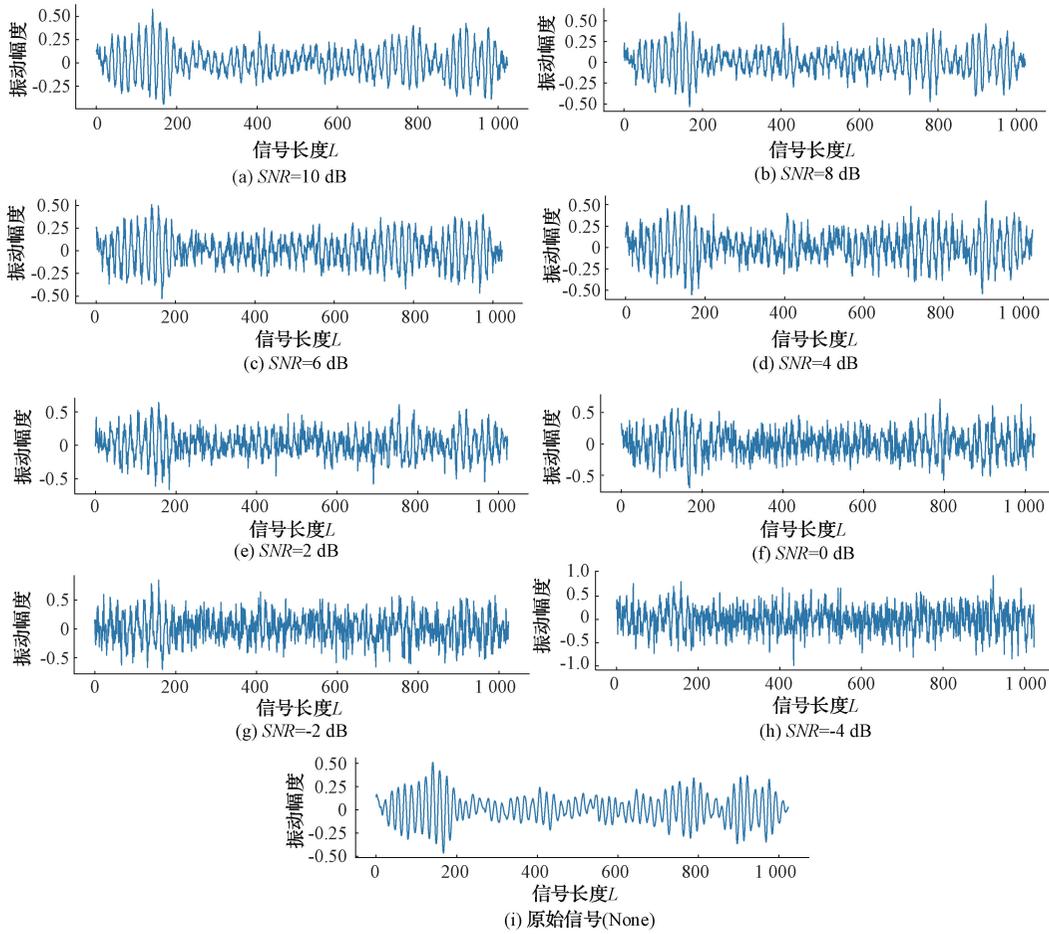


图5 不同信噪比条件下的振动信号时域波形

表4 在不同SNR下生成数据与原始信号之间的欧氏距离和相关系数

SNR/dB	10	8	6	4	2	0	-2	-4
欧氏距离	1.516 7	1.903 9	2.418 6	3.033 4	3.780 4	4.754 9	6.015 9	7.769 7
相关系数	0.957 2	0.927 3	0.892 8	0.835 1	0.794 8	0.695 2	0.630 2	0.542 1

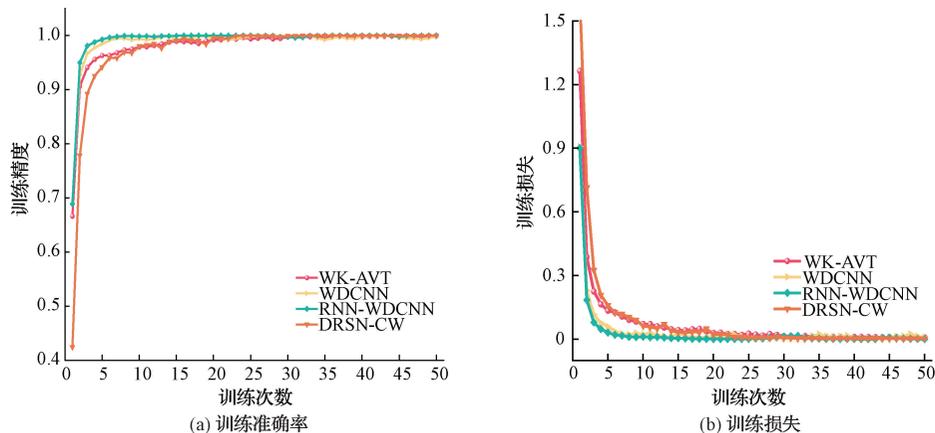


图6 加噪后诊断模型的训练准确率和损失变化曲线

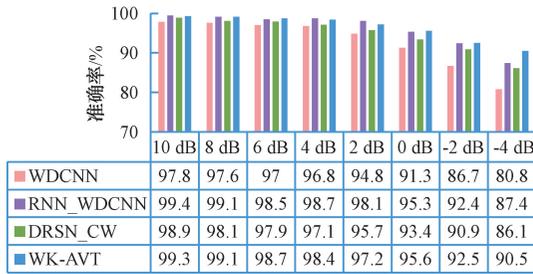


图7 不同信噪比条件下的故障诊断性能

筛选对故障分类有关的重要特征,对噪声干扰有一定的适应性,因此其预测准确率高于 WDCNN 模型,在-4 dB 的噪声干扰条件下的准确率比无噪声干扰时下降 13.3%,预测准确率为 86.1%。RNN-WDCNN 结合了 WDCNN 和时序循环神经网络的双路网络,其中循环神经网络具有提取振动信号中时序特征的能力,能够稳定提取轴承故障受距离约束特征的能力,且循环神经网络对时序信号的处理,功能上类似于时序信号的非线性滑动滤波器滤波,在时域上减小了噪声干扰的影响,其在信噪比为 10~2 dB 噪声干扰下,始终保持了 98% 以上的预测准确率;而在噪声与振动信号功率相当时,其预测准确率为 95.3%,明显高于 WDCNN 模型和 DRSN-CW 模型,在-4 dB 的噪声时,预测准确率为 87.4%,比无噪声干扰条件下的预测准确率下降 12.2%。对比其他 3 种 CNN 模型,WK-AVT 呈现出了更好的噪声适应性,在-4 dB 的噪声时,其预测准确率仍然达到 90.5%,呈现出高效的故障特征提取能力和良好的抗干扰能力。

#### 4.4 诊断结果分析

使用精确率(precision)、召回率(recall)、 $F_1$ -分数

( $F_1$ )和混淆矩阵评估各诊断模型对测试样本的分类能力,分别定义为:

$$Precision = \frac{TP}{TP + FP} \quad (22)$$

$$Recall = \frac{TP}{TP + FN} \quad (23)$$

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (24)$$

式中:TP、FN 和 FP 分别表示诊断模型对测试集中“正”类样本判定为“正”,“正”类样本判定为“负”,“负”类样本判定为“正”的数量。因此,精确率表示分类正确的正样本占判定为正样本总数的比例,召回率表示分类正确的正样本占实际正样本总数的比例,引入  $F_1$ -分数用以综合评估模型稳健性,避免单一评价精度率指标而引起的 FN 值升高、召回率值降低的影响。

由表 5~7 可知,上述 4 种模型的 4 项评价指标都呈现随着噪声强度的增加而呈现下降的趋势。在信噪比较小(不大于 0 dB)的严重噪声环境下,WK-AVT 在除精确率之外的其他两项指标中均表现最优,表明 WK-AVT 的抗噪能力较强;在信噪比增大的过程中,RNN-WDCNN 模型的召回率、 $F_1$ -分数分别在信噪比 2、4 和 10 dB 的情况下略高 WK-AVT,但是平均差值仅为 0.39%,其他信噪比情况均是本文网络的表现强于 RNN-WDCNN 模型。结合图 7 可知,在 4 种网络模型中,WK-AVT 的整体表现最好,这与利用宽卷积层替代传统视觉 Transformer 的原输入方式以及多头自注意力强大的全局信息特征提取能力密不可分,性能排名第 2 的是结合了宽卷积和循环网络的 RNN-WDCNN 模型,在所有对比网络中,仅包含宽卷积的 WDCNN 网络的各项评价指标最低,表现较差。

表 5 不同信噪比下诊断模型的精确率 (%)

模型	10 dB	8 dB	6 dB	4 dB	2 dB	0 dB	-2 dB	-4 dB
WK-AVT	99.33	<b>99.12</b>	<b>98.70</b>	98.45	97.32	95.73	<b>92.59</b>	<b>90.21</b>
WDCNN	97.95	97.84	97.07	97.07	95.10	92.12	87.27	81.56
RNN-WDCNN	<b>99.43</b>	99.10	98.53	<b>98.87</b>	<b>98.19</b>	<b>95.78</b>	92.49	87.90
DRSN-CW	98.91	98.12	97.97	97.20	95.95	93.63	91.18	86.75

表 6 不同信噪比下诊断模型的召回率 (%)

模型	10 dB	8 dB	6 dB	4 dB	2 dB	0 dB	-2 dB	-4 dB
WK-AVT	99.30	<b>99.10</b>	<b>98.70</b>	98.40	97.20	<b>95.60</b>	<b>92.50</b>	<b>90.50</b>
WDCNN	97.90	97.70	97.00	96.90	94.80	91.40	86.80	80.80
RNN-WDCNN	<b>99.40</b>	99.10	98.60	<b>98.80</b>	<b>98.10</b>	95.30	92.40	87.40
DRSN-CW	98.90	98.10	97.90	97.10	95.70	93.40	90.90	86.10

各模型分类混淆矩阵如图 8、9 所示,对不同故障类别正误判断情况。分别以信噪比为-4 和 6 dB 为例来进一步分析模型在各类别的预测情况,其中-4 dB 代

表了模型在强噪声时的情景,模型诊断结果可表明抗干扰能力的强弱,6 dB 情景下的诊断结果代表了模型自身提取故障特征的能力。将故障表述简单化,如

表7 不同信噪比下诊断模型的  $F_1$ -分数 (%)

模型	10 dB	8 dB	6 dB	4 dB	2 dB	0 dB	-2 dB	-4 dB
WK-AVT	99.32	<b>99.11</b>	<b>98.70</b>	98.42	97.26	<b>95.66</b>	<b>92.54</b>	<b>90.35</b>
WDCNN	97.92	97.77	97.03	96.98	94.95	91.76	87.03	81.18
RNN-WDCNN	<b>99.41</b>	99.10	98.56	<b>98.83</b>	<b>98.14</b>	95.54	92.44	87.65
DRSN-CW	98.91	98.11	97.93	97.15	95.82	93.51	91.04	86.42

注:加粗数据表示最高准确率

“Normal”表示正常状态,“Ball\_007”表示滚动体故障且故障尺寸为 0.007 inch,其他同理。由图 8 可知,WK-AVT、WDCNN、RNN-WDCNN 和 DRSN-CW 在“Ball\_007”故障上的准确率分别为 58%、59%、66%和 58%,在“OR\_014”故障上的准确率分别为 74%、54%、64%和 70%,其中在“Ball\_007”类别被识别错误的类型中,“OR\_014”类别的个数占的最多。同样,在“OR\_014”类别被识

别错误的类别中,“Ball\_007”类别的个数占的最多,因此“Ball\_007”和“OR\_014”类别存在很强的相关性,4种模型对“Ball\_007”、“OR\_014”的预测效果都较差。4种模型对“IR\_021”和“OR\_007”的识别准确率都比较高,但是只有本文所提算法对这两种故障的准确率都是 100%,4种模型中只有 WDCNN 对“OR\_021”故障的诊断率最低。

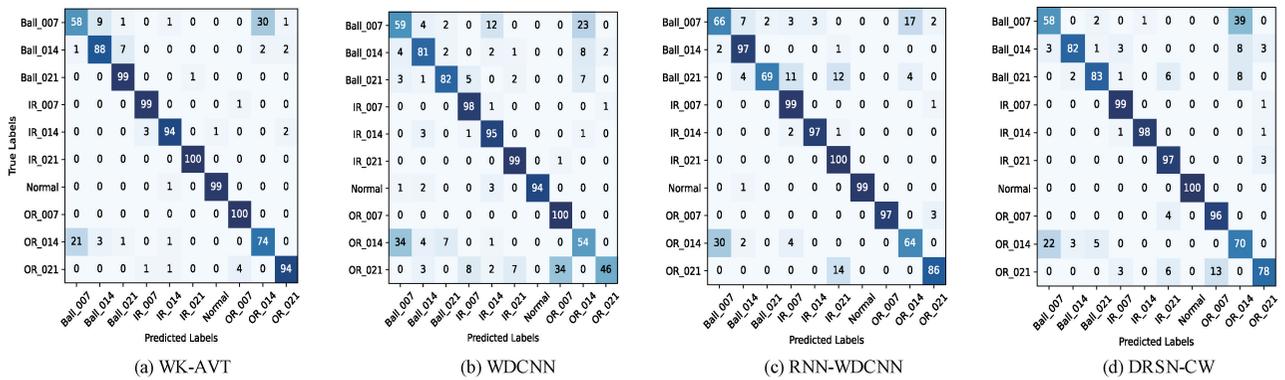


图8 SNR = -4 dB 各模型分类混淆矩阵

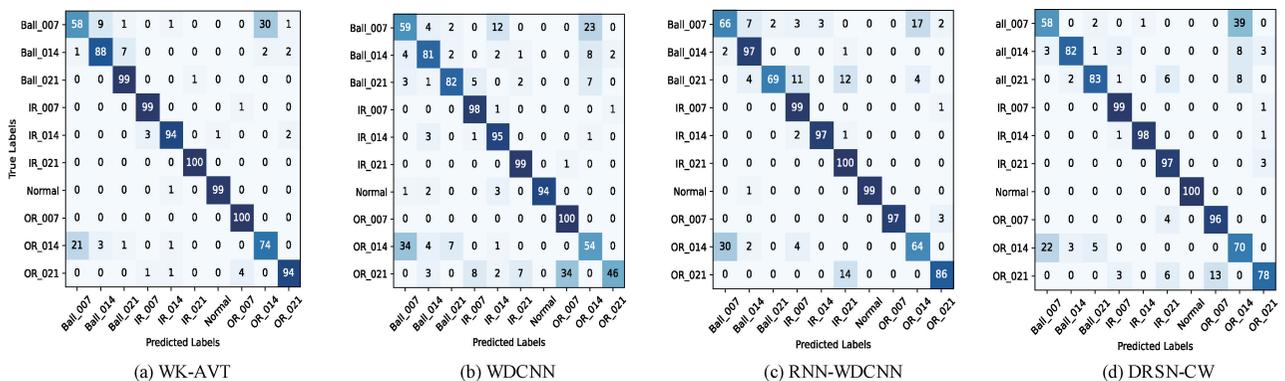


图9 SNR = 6 dB 各模型分类混淆矩阵

由图 9 可知,WDCNN 模型在预测“Ball\_007”、“Ball\_014”、“Ball\_021”、“OR\_014”和“OR\_021”时出现错误,其中对“OR\_014”的诊断准确率最差为 82%;RNN-WDCNN 模型在预测“Ball\_007”、“Ball\_014”、“OR\_007”、“OR\_014”、“OR\_021”时出现错误,其中对“OR\_014”的诊断准确率最差为 91%;DRSN-CW 模型在预测“Ball\_007”、“Ball\_014”、“Ball\_021”和“OR\_014”时出现错误,其中对

“OR\_014”的诊断准确率最差为 94%;本文所提算法在预测“Ball\_007”、“OR\_014”和“OR\_021”时出现错误,其中对“Ball\_007”和“OR\_014”的诊断准确率分别为 93%和 96%,因此 4 种模型中从低于 100%准确率的故障类型数量和出错类别准确率高低角度,本文所提算法均优于另外 3 种模型。总体来看,本文所提出的 WK-AVT 网络可以从轴承原始振动信号中有效提取到与故障相关的特征,并

进行高精度的识别和分类。

#### 4.5 消融实验

为了验证所设计的两个网络模块在噪声条件下对模型准确率的提升效果,在测试集上添加 SNR 从 10~-4 dB 的额外噪声,通过准确度变化反映不同模块对精度提升的贡献程度。根据组合顺序,分别为仅包含宽卷积层嵌入模块的 WKVT 模型、两个模块都不包含的 VT 模型、仅包含 ECA 模块的 E-VT 和两个模块都包含的 WK-AVT 模型。对比实验结果如表 8 所示, WK-AVT 模型即两个模块均使用时,其诊断结果的准

确率是最高的,VT 模型即两个模块均不使用时,其诊断效果是最差的。对比 WKVT 和 VT 模型,E-VT 和 WK-AVT 模型可以得出,使用宽卷积网络模块时,在噪声强度比较大的条件下(不高于 0 dB)对准确率的提高非常明显,比不使用此模块的情况平均准确率提升高达 10%。对比 WKVT 和 WK-AVT 模型,VT 和 E-VT 模型可以得出,使用基于 ECA 设计的高效筛选器比不使用此筛选模块的情况准确率也有所提升。因此,两个网络模块结合使用会大大提高网络的抗噪声能力,增强网络的鲁棒性。

表 8 不同信噪比下消融实验准确率 (%)

模型	10 dB	8 dB	6 dB	4 dB	2 dB	0 dB	-2 dB	-4 dB
WKVT	99.0	<b>99.3</b>	98.2	97.4	97.0	<b>95.6</b>	92.4	89.1
VT	96.1	94.8	96.0	94.0	93.1	88.7	84.1	77.3
E-VT	96.0	95.3	95.5	94.2	92.2	90.3	85.4	78.8
WK-AVT	<b>99.3</b>	99.1	<b>98.7</b>	<b>98.4</b>	<b>97.2</b>	<b>95.6</b>	<b>92.5</b>	<b>90.5</b>

## 5 结论

本文提出了一种深度网络模型用于解决滚动轴承在噪声条件下故障诊断精度低的问题,该模型结合了卷积神经网络和视觉 Transformer 的优点,同时设计高效特征筛选器有效选取最终用于分类的特征向量,增强模型的泛化能力。设计的宽卷积核局部特征扩展映射层直接处理原始振动信号,减少了对信号处理知识的依赖,有效增强了 Transformer 网络对轴承故障局部特征的提取能力,为轴承故障诊断模型提供了更全面的特征信息,提高了模型的鲁棒性。考虑到本文所提出的网络模型,可直接用于处理时序振动信号,因此,也可应用于齿轮、电动机和抽水机等旋转机械的故障诊断任务,下一步将研究基于设备采集的多通道时序信号,对输入层及特征选择进一步改进以适配噪声环境下多传感器采集到的多通道信号故障诊断任务。

#### 参考文献

- [1] VINCENT P, LAROCHELLE H, BENGIO Y, et al. Extracting and composing robust features with denoising autoencoders [C]. Proceedings of the 25th International Conference on Machine Learning, 2008: 1096-1103.
- [2] XU Y, YAN X, FENG K, et al. Attention-based multiscale denoising residual convolutional neural networks for fault diagnosis of rotating machinery [J]. Reliability Engineering and System Safety, 2022, 226: 108714.
- [3] WENG C, LU B, YAO J. A one-dimensional vision transformer with multiscale convolution fusion for bearing fault diagnosis [C]. 2021 Global Reliability

and Prognostics and Health Management (PHM-Nanjing). IEEE, 2021: 1-6.

- [4] HU T, TANG T, LIN R, et al. A simple data augmentation algorithm and a self-adaptive convolutional architecture for few-shot fault diagnosis under different working conditions [J]. Measurement, 2020, 156: 107539.
- [5] 杨小东,纪国宜.基于融合 CNN 的滚动轴承剩余寿命预测[J].国外电子测量技术,2021,40(3):62-67.
- [6] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [7] 范志伟,郭世伟,罗鑫,等.基于 Inception-DLSTM 双通道的滚动轴承故障诊断方法[J].电子测量技术,2023,46(7):53-59.
- [8] DING X, HE Q. Energy-fluctuated multiscale feature learning with deep ConvNet for intelligent spindle bearing fault diagnosis [J]. IEEE Transactions on Instrumentation and Measurement, 2017, 66(8): 1926-1935.
- [9] ZHAO M, ZHONG S, FU X, et al. Deep residual shrinkage networks for fault diagnosis [J]. IEEE Transactions on Industrial Informatics, 2019, 16(7): 4681-4690.
- [10] YAN L P, DONG X Z, WANG T, et al. A fault diagnosis method for gas turbines based on improved data preprocessing and an optimization deep belief network [J]. Measurement Science and Technology, 2019, 31(1): 015015.

- [11] WEN L, LI X, GAO L, et al. A new convolutional neural network-based data-driven fault diagnosis method[J]. IEEE Transactions on Industrial Electronics, 2017, 65(7): 5990-5998.
- [12] 温江涛,张鹏程,孙洁娣,等. 残差卷积自编码网络无监督迁移轴承故障诊断[J]. 中国机械工程, 2022, 33(14): 1707-1716.
- [13] 张昌凡,刘佳峰,何静,等. 基于改进卷积双向门控循环网络的轴承故障诊断[J]. 电子测量与仪器学报, 2021, 35(11): 61-67.
- [14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30.
- [15] FANG H, DENG J, BAI Y, et al. CLFormer: A lightweight transformer based on convolutional embedding and linear self-attention with strong robustness for bearing fault diagnosis under limited sample conditions [J]. IEEE Transactions on Instrumentation and Measurement, 2021, 71: 1-8.
- [16] 黄星华,吴天舒,杨龙玉,等. 一种面向旋转机械的基于 Transformer 特征提取的域自适应故障诊断[J]. 仪器仪表学报, 2022, 43(11): 210-218.
- [17] DING Y, JIA M, MIAO Q, et al. A novel time-frequency Transformer based on self-attention mechanism and its application in fault diagnosis of rolling bearings[J]. Mechanical Systems and Signal Processing, 2022, 168: 108616.
- [18] 刘俊锋,俞翔,万海波,等. 基于 MFMD 和 Transformer-CNN 的滚动轴承故障诊断方法[J]. 航空动力学报, 2023, 38(6): 1446-1456.
- [19] 杜康宁,宁少慧,邓功也. 基于视觉 Transformer 的滚动轴承智能故障诊断[J]. 组合机床与自动化加工技术, 2023(4): 96-99.
- [20] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.
- [21] ZHANG W, PENG G, LI C, et al. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals[J]. Sensors, 2017, 17(2): 425.
- [22] SONG X, CONG Y, SONG Y, et al. A bearing fault diagnosis model based on CNN with wide convolution kernels [J]. Journal of Ambient Intelligence and Humanized Computing, 2022, 13(8): 4041-4056.
- [23] SHENFIELD A, HOWARTH M. A novel deep learning model for the detection and identification of rolling element-bearing faults [J]. Sensors, 2020, 20(18): 5112.

#### 作者简介

张新良(通信作者),博士,教授,主要研究方向为智能控制,检测技术与自动化装置。

E-mail:zxldq@hpu.edu.cn

李占,硕士研究生,主要研究方向为故障诊断,图像处理,检测技术与自动化装置。

E-mail:332515392@qq.com

周益天,工程师,主要研究方向为工业自动化装置。

E-mail:tianyisjtu@163.com