

一种轻量化尺度感知调制 Swin Transformer 模型的轴箱轴承故障诊断方法

邓飞跃^{1,2}, 郑守禧¹, 郝如江^{1,2}

(1. 石家庄铁道大学机械工程学院, 050043, 石家庄;

2. 河北省工程机械动力与传动控制重点实验室, 050043, 石家庄)

摘要: 针对当前 Transformer 网络模型运算效率偏低, 且难以用于复杂工况条件下高速列车轴箱轴承故障诊断的问题, 提出了一种基于时频多域融合与轻量化结构的尺度感知调制 Swin Transformer (SMST) 模型的轴箱轴承故障诊断方法。首先, 采用格拉姆角场法、双谱法与 Chirplet 变换法, 将轴承振动信号转化为时域、频域与时频域内的二维图像, 基于多域特征融合思想集成新的特征图像; 然后, 设计了一种新的轻量化结构 SMST 模块, 在其内部实现了卷积运算与 Transformer 自注意力运算的进一步融合; 最后, 在层次化模型框架中引入特征金字塔模块 (FPB), 弥补不同层输出特征的不一致性, 实现了上下文信息的特征深度融合及复杂工况条件下轴箱轴承故障诊断。实验结果表明: 相比格拉姆角场法、双谱法、Chirplet 变换法、短时傅里叶变换法、连续小波变换法等单一领域图像生成方法, 时频多域融合方法生成的图像特征信息表征能力更好; 所提网络模型在 1 010、760、505 r/min 这 3 种转速变工况任务中的轴箱轴承故障识别准确率分别为 99.88%、99.92% 与 99.96%; 对比 ResNets、GoogleNet、ViT、Swin Transformer 和 SMT 这 5 种模型, 所提方法的故障识别准确率更高, 模型轻量化程度更好。所提方法可为实际工况中列车轴箱轴承故障诊断提供参考。

关键词: 故障诊断; 轴箱轴承; 多域融合; 轻量化; 尺度感知调制

中图分类号: TH17 **文献标志码:** A

DOI: 10.7652/xjtuxb202409009 **文章编号:** 0253-987X(2024)09-0083-11

A Lightweight Scale-Aware Modulation Swin Transformer Network for Axlebox Bearing Fault Diagnosis

DENG Feiyue^{1,2}, ZHENG Shouxi¹, HAO Rujiang^{1,2}

(1. School of Mechanical Engineering, Shijiazhuang Tiedao University, Shijiazhuang 050043, China;

2. Hebei Key Laboratory of Mechanical Power and Transmission Control, Shijiazhuang 050043, China)

Abstract: A scale-aware modulation swin transformer (SMST) model is proposed based on multi-domain fusion and lightweight structure, for axlebox bearing fault diagnosis, aiming to address the inefficiency of the current Transformer network model and its unsuitability for axlebox bearing fault diagnosis of high-speed trains under complex working conditions. Firstly, Gram angle field (GAF), bispectrum, and Chirplet transform techniques are applied to convert the bearing vibration signals into two-dimensional images in time, frequency, and time-frequency

domains, which are integrated into a new image based on the idea of multi-domain feature fusion. Then, a new lightweight structure SMST module is designed, combining the further fusion of convolution operation and Transformer self-attention operation. Finally, a feature pyramid block (FPB) is introduced into the SMST hierarchical model framework to address the inconsistency of the output features across layers and enable deep feature fusion with contextual information. The fault diagnosis of axlebox bearings under complex working conditions is realized. Results show that the time-frequency multi-domain fusion method offers a superior representation of image feature information compared to single-domain image generation methods such as Gram angle field, bispectrum, Chirplet transform, short-time Fourier transform, and continuous wavelet transform. The proposed network model achieves high accuracy rates of 99.88%, 99.92%, and 99.96% in identifying axlebox bearing faults in the three tasks under variable working conditions with different speeds of 1 010 r/min, 760 r/min, and 505 r/min, respectively. Comparative analysis against five models of ResNets, GoogleNet, ViT, Swin Transformer, and SMT showcases the higher fault identification accuracy and better model lightweight of the proposed method. The proposed method can serve as a reference for the diagnosis of axlebox bearing faults in trains under actual working conditions.

Keywords: fault diagnosis; axlebox bearing; multi-domain fusion; lightweight; scale-aware modulation

轴箱轴承是列车车轴与车架之间关键的旋转部件,起着承受车轮受力、传递车轴负荷、保证列车运行平稳舒适等重要作用。然而,在轮对踏面磨耗、轨道不平顺等复杂激励作用下,轴承故障信号中背景噪声干扰强烈,故障特征提取与识别非常困难^[1]。因此,探索更为有效的轴箱轴承故障诊断方法,对保障列车健康运行具有非常积极的意义。

伴随人工智能技术的快速发展,基于深度学习方法的滚动轴承故障诊断方法得到了蓬勃发展,自编码网络(AEN)、循环神经网络(RNN)、卷积神经网络(CNN)等模型不断涌现。向宙等^[2]提出了一种区分自编码网络模型用于轴承特征提取,通过在网络原始损失函数中添加交叉熵函数设计了一种新的复合损失函数;Jiang等^[3]构建了深度RNN模型,通过输入故障信号频谱对滚动轴承故障进行诊断。相比AEN、RNN模型,CNN模型采用局部连接和权值共享的设计,使得模型训练更为快捷,特征学习能力优越,应用更为广泛。近年来,AlexNet^[4]、ResNet^[5]、GoogLeNet^[6]和ShuffleNet^[7]等基于CNN模型的滚动轴承故障诊断方法不断涌现。周凌孟等^[8]利用主成分分析法对一维信号进行降维,将降维后的数据特征输入残差网络诊断轴承故障。Deng等^[9]将一维信号转换为二维图像格式后,输入构建的多尺度深度残差网络模型,实现了轴承不同健康状态的识别。Peng等^[10]利用短时傅里叶变换获取轴承振动信号的时频图,并输入深度残差网络

识别轴承故障类型。Yang等^[11]提出了一种基于小波和深度CNN模型的滚动轴承故障数据自动聚类方法,首先利用小波包分解获取特征信息,然后输入网络模型聚类分析,通过对信号进行预处理分析,可以获得更准确的特征信息,能有效增强网络模型的诊断效果。然而,CNN模型中感受野范围受卷积核大小所限,往往只能考虑特征图的局部信息而忽略全局信息,导致其长历程特征学习能力较差,全局特征感知能力不足。

基于Transformer的自注意力网络于2017年被首次提出^[12],后凭借具有并行计算、擅于捕捉长距离依赖关系、全局特征学习能力强等诸多优点,迅速在自然语言、机器视觉等领域得到了广泛应用。Vision Transformer(ViT)模型取得了媲美CNN模型的效果^[13],被视作CNN模型有力的替代者,Swin Transformer(Swin T)^[14]、Shuffle Transformer^[15]、Cswin^[16]等改进的Transformer模型更是不断涌现。与此同时,基于Transformer的故障诊断研究也成为国内外学者研究的热点。Yang等^[17]构建了基于Transformer的深度网络模型用于轴承故障诊断;Fang等^[18]提出了CLFormer自注意力模型用于轴承故障识别,在强噪声环境下具有较高的分类性能;Hao等^[19]提出了一种双通道视觉Transformer的轴承寿命预测模型,能够获取轴承信号中深层次的退化特征;Liang等^[20]将子域自适应与视觉Transformer相结合,基于融合后的Transformer网络实现了变工

况下轴承故障诊断。随着 Transformer 模型在故障诊断领域的广泛应用,其弊端也逐渐显现。传统 Transformer 模型多采用层次化框架,导致网络不同层之间特征信息集成较为困难,削弱了局部特征的学习能力。此外,Transformer 模型参数规模大,需要大量的计算资源,造成其工作效率偏低,计算成本高昂,在工业领域难以快速部署应用。因此,为更好应用于工程实际,有必要开发基于 Transformer 的轻量化网络模型。

基于上述分析,本文提出一种时频多域融合和尺度感知调制 Swin Transformer (scale-aware modulation Swin Transformer, SMST)模型用于高速列车轴箱轴承复杂变工况下故障诊断,并经过实验分析,验证了所提方法能够有效提取针对轴承故障状态的深度特征信息,克服复杂工况条件的干扰。本文主要内容包括以下3部分。

(1)分别采用格拉姆角场(GAF)、双谱与 Chirplet 变换(CT)方法,将轴承一维振动信号在时域、频域、时频域内转换为二维图像,通过时频多域融合集成为新的多通道特征图像。相比传统单一领域图像,二维特征图信息更为丰富,进一步增强了信号的特征表征能力。

(2)设计了一种轻量化结构的 SMST 模块,在模块内部实现了 CNN 和 Transformer 自注意力运算的深度融合:一方面利用尺度感知调制结构降低网络规模,捕捉不同尺度特征信息;另一方面利用 Swin T 模型更好地学习全局语义特征信息,提高模型检测性能。

(3)在 SMST 层次化模型框架中引入特征金字塔模块(FPB),弥补不同层输出特征的不一致性,实现上下文信息的特征融合,从而有效挖掘模型浅层和深层之间小目标语义特征,增强模型细粒度语义特征学习能力。

1 基于时频多域融合的图像生成方法

当前基于深度学习的轴承故障诊断研究中,通常采用图像生成方法,将轴承振动一维信号转换为二维图像输入网络模型。GAF 方法能将信号转换为时域图像,双谱方法能将信号转换为频域图像,短时傅里叶变换(STFT)、连续小波变换(CWT)方法等能将信号转换为时频域图像。上述方法虽被广泛采用,但仅能将信号转换为单一领域内的图像。基于此,本文提出基于时频多域融合的图像生成方法。

1.1 格拉姆角场法

格拉姆角场(GAF)法是一种基于格拉姆矩阵的图像转换方法^[21],通过格拉姆矩阵计算一维时间向量的线性相关关系,从而将一维时域信号转换为二维图像。GAF 保留了原始数据的全局信息,具有不受频域分辨率限制且不依赖时间窗口等优点。采用 GAF 对一维时间序列(用矩阵形式表示) $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$ 进行归一化处理,然后将归一化的时间序列使用极坐标重新进行放缩,表达式为

$$\begin{cases} \varphi = \arccos \tilde{x}_i, -1 \leq \tilde{x}_i \leq 1 \\ r = t_i/N \end{cases} \quad (1)$$

式中: φ 为极角; $\tilde{x}_i$ 表示归一化处理后的时间序列点; t_i 为时间步; N 为极坐标下的常数因子。由于 $\cos \varphi$ 在区间 $[0, \pi]$ 上单调递减,因此时间序列映射到极坐标下的值是唯一且固定的。将一维时域信号用极坐标系表示后,通过计算每个点之间的角度和、差表达数据点之间的相关性。采用格拉姆角和场作为信号处理方法的计算过程如下

$$\mathbf{G} = \tilde{\mathbf{X}}' \cdot \tilde{\mathbf{X}} - \sqrt{\mathbf{I} - \tilde{\mathbf{X}}'^2} \sqrt{\mathbf{I} - \tilde{\mathbf{X}}^2} \quad (2)$$

式中: $\mathbf{G}$ 为格拉姆角和场; $\mathbf{I}$ 为单位行向量; $\tilde{\mathbf{X}}$ 为时间序列; $\tilde{\mathbf{X}}'$ 为 $\tilde{\mathbf{X}}$ 的转置。通过上述处理方式, GAF 实现了将一维时间序列信号转化为二维图像,并保留了时间序列中的时间相关特性。

1.2 双谱分析方法

相比传统傅里叶变换方法,双谱分析是一种高阶谱处理方法,能够在频域范围内表征信号的非线性特征^[22]。由于信号构成不同,其频谱组成也不同,对应的双谱图分布也会存在明显差异。对于一维信号 $\mathbf{X}$, 其双谱 $B_x(\omega_1, \omega_2)$ 的表达式为

$$B_x(\omega_1, \omega_2) = \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} [c_{3x}(\tau_1, \tau_2) \exp[-j(\omega_1 \tau_1 + \omega_2 \tau_2)]] \quad (3)$$

式中: $c_{3x}(\tau_1, \tau_2)$ 为 $\mathbf{X}$ 的三阶累积量; ω_1, ω_2 表示双谱中傅里叶变换到频域得到的两个频率分量; τ_1, τ_2 为求取三阶累积量的时延。双谱分析能够有效抑制时域信号中的高斯噪声,保留信号频域信息,并以频域图像的形式进行显示。

1.3 Chirplet 变换法

Chirplet 变换(CT)法是在短时傅里叶变换与小波变换基础上的进一步衍变,通过引入平移算子与频率旋转算子的理念,利用线性调频参数表示核函数来处理信号^[23]。通过对轴承一维信号进行 CT

变换,可以在时频域内得到二维时频图。对于一维信号 $\mathbf{X}$, 其 CT 变换表达式为

$$S_x(f, t_c) = \int_{-\infty}^{+\infty} z(t) \psi(t - t_c) \exp(-j2\pi\varphi(f, t, t_c)) \cdot \exp(-j\omega(t - t_c)) dt \quad (4)$$

式中: S_x 为时间序列 $\mathbf{x}$ 的 Chirplet 变换; $z(t)$ 为 $\mathbf{X}$ 的解析信号; t 为时间; t_c 为时间中心; $\psi(t)$ 为窗函数, 一般使用高斯函数 $\psi(t) = \exp(-t^2/(2\sigma^2))$, 其中 σ 为窗口参数。

通过 CT 处理,可以得到信号的二维时频图像,其中蕴含时域与频域特征信息。

本文分别使用 GAF、双谱和 CT 3 种方法对轮对轴承一维信号进行处理,得到时域、频域和时频域二维图像,然后对 3 种图像进行灰度化处理,以减少图片内存,再将 3 类图像按照时域、频域与时频域的顺序叠加,构成 3 通道的时频多域融合图像,整个过程如图 1 所示。融合后的图像蕴含了轴承信号时域、频域和时频域的特征信息,集成后的多通道特征图像包含更为丰富的信号特征信息,故障特征表征能力进一步增强。

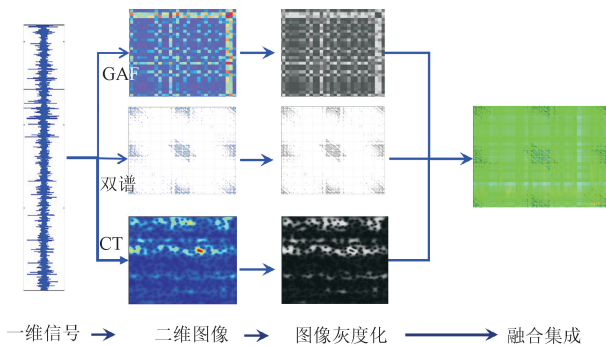


图 1 时频多域融合过程

Fig. 1 Time frequency multi domain fusion process

2 SMST 网络模型

2.1 尺度感知调制

Scale-aware modulation transformer(SMT)模型是基于 CNN 与 Transformer 的一种混合轻量化网络模型^[24]。SMT 模型基于卷积运算设计了一个新颖的轻量尺度感知调制单元 scale-aware modulation (SAM),由多头混合卷积模块和多尺度感知聚合模块两部分构成,具体结构如图 2 所示。多头混合卷积模块中使用多个卷积核尺寸不同的深度可分离卷积(DSC),一方面能够捕捉特征图多个尺度上的空间特征信息,另一方面能有效减少传统卷积的运算

量。通过构建多个卷积核尺寸不同的 DSC 层,以自适应方式提取不同尺度的图像特征信息,伴随着网络层数增加,能够更好地捕捉局部细节特征,提取更丰富的语义信息。为了增强多头混合卷积模块的信息交互,在多尺度感知聚合模块中对多头混合卷积模块提取到的不同尺度特征信息进行重组,使组内与组间的信息融合交互,可以实现轻量、高效的聚合效果。多尺度感知聚合模块增强了特征图的局部信息,更为明确地关注到特征图中的重要部分,具备更好的特征图细节特征捕捉与表征能力。

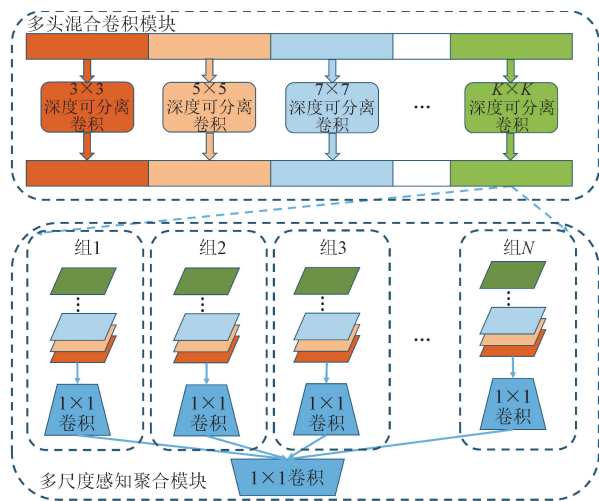


图 2 SAM 结构示意图

Fig. 2 Structure diagram of SAM

Swin T 模型基本模块包含 2 个串接的残差连接结构,第一个由归一化层(LN)和窗口注意力(W-MSA)或滑动窗口注意力(SW-MSA)构成;第二个由 LN 与多层感知机(MLP)构成。W-MSA 与 SW-MSA 是基于自注意力机制提出的改进算法,自注意力机制的表达式为

$$\text{attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}}\right)\mathbf{V} \quad (5)$$

式中: $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 分别为查询矩阵、键矩阵与值矩阵; $\mathbf{Q}\mathbf{K}^T$ 为注意力权重; $\sqrt{d}$ 为缩放因子; softmax 表示对权重进行归一化。W-MSA 模块中特征传递的计算过程为

$$\hat{z}^l = f_{\text{W-MSA}}[f_{\text{LN}}(z^{l-1})] + z^{l-1} \quad (6)$$

$$z^l = f_{\text{MLP}}[f_{\text{LN}}(\hat{z}^l)] + \hat{z}^l \quad (7)$$

式中: $\hat{z}^l$ 和 z^l 表示第 l 块中 W-MSA 和 MLP 的输出特征; $f_{\text{W-MSA}}(\cdot)$ 、 $f_{\text{MLP}}(\cdot)$ 、 $f_{\text{LN}}(\cdot)$ 分别表示 W-MSA 模块、MLP 模块和归一化层的输出函数。

W-MSA 模块中特征传递的计算过程为

$$\hat{z}^{l+1} = f_{\text{SW-MSA}}[f_{\text{LN}}(z^l)] + z^l \quad (8)$$

$$z^{l+1} = f_{\text{MLP}}[f_{\text{LN}}(\hat{z}^{l+1})] + \hat{z}^{l+1} \quad (9)$$

式中: $\hat{z}^{l+1}$ 和 z^{l+1} 分别表示第 $(l+1)$ 块 SW-MSA 和 MLP 的输出特征; $f_{\text{SW-MSA}}(\cdot)$ 表示 SW-MSA 模块的输出函数。

通过将 SAM 引入 Swin T 基本模块,代替原有的 W-MSA 或 SW-MSA,构建出更具轻量化结构的 SAM 模块。SMT 模型基于 SAM 模块,采用与 Swin T 模型相同的层次化网络结构,通过 SAM 模块聚合特征图的局部语义信息并对其进行调制,增强了模型的特征提取能力。在层次化网络架构中,特征图的分辨率随着层次的加深不断减小,使用 W-MSA 与 SW-MSA 对分辨率较大的浅层特征图进行提取时会带来庞大的计算成本,而 SAM 使用卷积对浅层特征图进行特征提取,避免了自注意力机制带来的庞大计算量,使得模型更加轻量化。

2.2 SMST 模块

SAM 模块是基于卷积运算提出的。单一的 SAM 模块无法提取图像特征的长距离语义信息,导致模型在全局特征提取上依旧存在不足。为此,本文在 SAM 模块的基础上,构建了一种新的 SMST 模块,结构如图 3 所示。SMST 模块依次堆叠 1 个 SAM 模块和 2 个 Swin T 基本模块。在所提的 SMST 模块中,同时包含 SAM、W-MSA 与 SW-MSA 模块。SAM 模块使用卷积进行特征提取,避免了自注意力机制带来的庞大计算量,使得模型更加轻量化。W-MSA 模块先对输入的特征进行窗口划分,后在窗口内进行特征注意力运算。SW-MSA 模块则采用滑动窗口的形式,弥补 W-MSA 模块窗口与窗口无法进行信息交流的不足。所提出的 SMST 模块由 SAM、W-MSA 与 SW-MSA 串接构

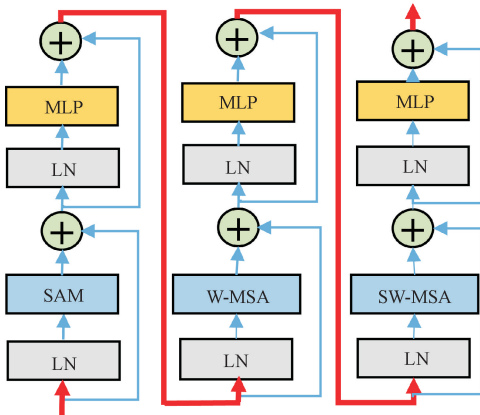


图 3 SMST 模块结构

Fig. 3 Architecture of the SMST block

成,同时兼容了卷积运算与自注意力运算,使得每部分都能够更好地匹配其计算特性,在网络模型由浅层到深层时,实现从局部特征到全局特征的综合提取。此外,所提 SMST 模块保持了轻量级结构,进一步减少了模型的计算成本。

2.3 FPB 模块

当前, Swin T 模型、SMT 模型及诸多变种 Transformer 网络模型均采用了层次化的网络结构,虽然有利于模型构建图像分类、目标检测、语义分割等任务,但随着网络层数增加,不同层之间特征信息的集成较为困难,会丢失一部分图像信息并影响模型精度。为此,本文构建了 FPB 模块,结构如图 4 所示,包含横向传递和自上而下两个部分,左侧为模型 4 个阶段后输出的不同尺度特征图,右侧自上而下为融合路径。在 FPB 模块中,首先对尺寸较小的特征图使用双线性插值法进行 2 倍上采样,将特征图的尺寸放大 2 倍,然后使用尺寸为 1×1 卷积改变尺寸较大的特征图的通道数,从而实现不同通道数、不同尺寸大小特征图的维度匹配,最后通过特征图逐元素相加操作实现融合。通过 FPB 模块可以实现层次化网络模型中不同深度特征、不同规模特征的细粒度融合,有效弥补不同尺度间特征的不一致性。

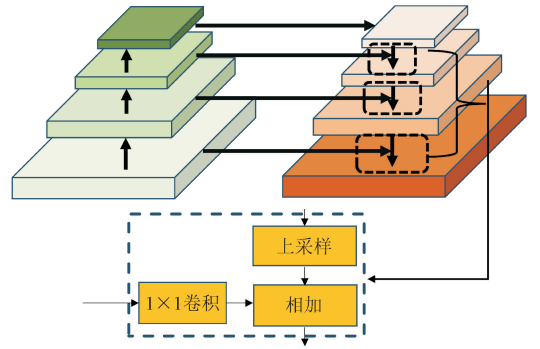


图 4 FPB 模块结构示意图

Fig. 4 Structure diagram of FPB block

2.4 所提模型框架

本文所提出的 SMST 模型整体框架及相应特征图参数如图 5 所示,主要由 4 个阶段(stage)的层次化结构与 FPB 模块构成。Stage 1 通过使用尺寸为 4×4 卷积的下采样层将 $224 \times 224 \times 3$ 像素的输入图像进行分块并将通道数扩大到 64,然后对分块的图像使用 SAM 模块进行特征提取。Stage 2 将 stage 1 的输出特征图通过使用尺寸为 2×2 卷积的

下采样层将特征图进行 2 倍下采样并将通道数扩大 2 倍。Stage 3 与 stage 4 下采样与 stage 2 相同,之后使用 SAM 模块进行特征提取,在 stage 3 中使用本文所提的 SMST 模块对下采样的特征图进行特征提取,从而实现从局部特征提取到全局特征提取

的转变。Stage 4 中仅使用 Swin T 对特征图进行特征提取,从而有效获取全局特征信息。将 FPB 结构嵌入到 4 个 stage 之后,使用 FPB 将不同深度特征、不同规模特征的细粒度融合,然后分类网络完成模式识别。

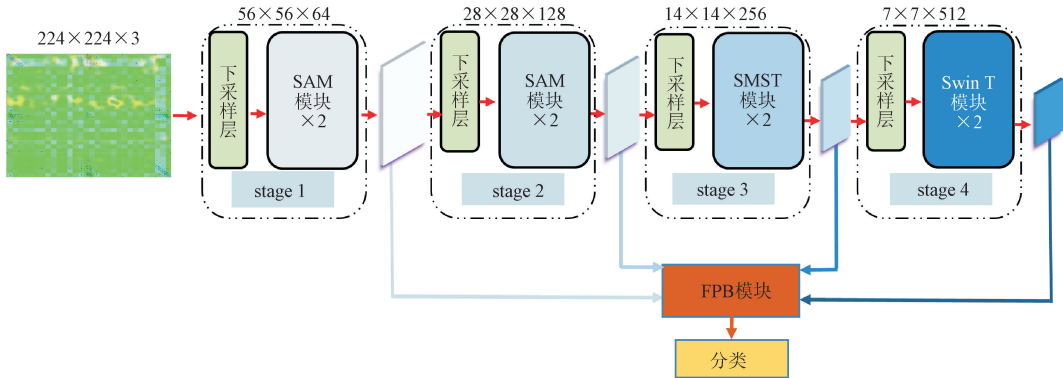


图 5 SMST 模型架构

Fig. 5 A model architecture of the SMST

3 实验结果与分析

3.1 实验介绍

使用交通工程结构力学行为与系统安全国家重点实验室高速列车轴承综合试验台测试的轴承数据验证本文所提方法的有效性,试验台如图 6 所示。轴承型号为 197726 双列圆锥滚子轴承,轴承故障类型分为 8 类,分别是正常、外圈故障(严重/轻微)、内圈故障(严重/轻微)、滚子故障(严重/轻微)、外圈滚子复合故障,如图 7 所示。测试轴承选自中车石家庄车辆有限公司轮对轴承检测车间,故障均为列车运行过程中产生的实际故障。轮对轴承故障类型与对应的类别标签如表 1 所示。

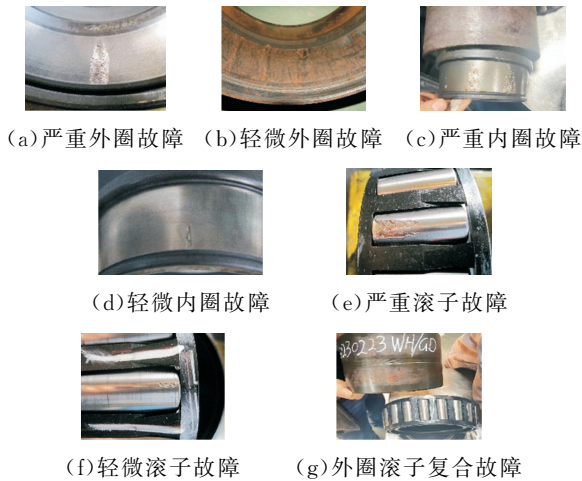


图 7 不同类型轮对轴承故障

Fig. 7 Different types wheelset bearing fault



图 6 高速列车轴承综合试验台

Fig. 6 Wheelset bearing comprehensive test rig of high-speed train

表 1 轮对轴承故障类型		
Table 1 Wheelset bearing fault categories		
类别	状态描述	标签
C1	正常状态	1
C2	严重外圈故障	2
C3	轻微外圈故障	3
C4	严重内圈故障	4
C5	轻微内圈故障	5
C6	严重滚子故障	6
C7	轻微滚子故障	7
C8	外圈滚子复合故障	8

实验测试中,在测试轴承端盖位置安装加速度传感器,采样频率为 25 600 Hz。为了验证模型在变工况条件下测试的准确性,选取 1 010、760、505 r/min 共 3 种转速进行测试,各自采集 1 000 组信号样本,共计 $1\,000 \times 3 \times 8 = 24\,000$ 组样本信号。将样本转化为图像之后,构建 3 个变转速轴承故障诊断任务,分别定义为任务 A1、任务 A2、任务 A3,具体如表 2 所示。实验过程为:以任务 A1 为例,将转速为 505 r/min 和 760 r/min 的样本作为训练集,将转速为 1 010 r/min 的数据样本作为测试集,训练得到最优模型后,输入测试集样本进行测试,模型训练和测试过程分别进行 10 次,测试结果评价指标取平均值。

表 2 变转速轴承故障诊断任务

任务	转速/(r · min ⁻¹)	
	训练数据集	测试数据集
A1	505, 760	1 010
A2	505, 1 010	760
A3	760, 1 010	505

3.2 模型参数设置

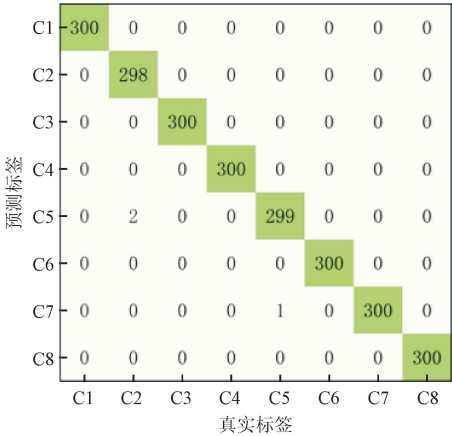
本文方法基于 python 3.8 语言,程序开发环境为 PyCharm,深度学习框架为 Pytorch,所用电脑硬件配置为 Intel(R)Core(TM) i9-13900K, 64 GB RAM,GPU 为 NVIDIA GeForce RTX 4090,输入图像为 224×224×3 像素,注意力头数为 8,学习率为 0.001,迭代次数为 50,批量大小为 16。

3.3 结果分析

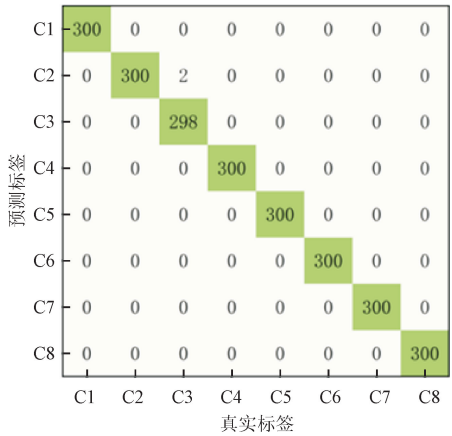
使用本文所提模型对 3 类任务分别进行实验验证。为验证 3 类任务测试实验的精确度,引入混淆矩阵分别对 3 次测试结果进行可视化,3 类任务模型的混淆矩阵如图 8 所示,识别结果准确率分别为 99.88%、99.92%与 99.96%,平均准确率为 99.92%。从图 8 可知,本文所提模型在 3 种变工况任务中都实现了较高的准确率,验证了模型在变工况条件下对高速列车轴箱轴承健康状态的识别精度较好。

为了进一步体现所提网络模型在变工况条件下对轴承健康状态的识别能力,利用 t 分布-随机近邻嵌入(t -SNE)方法,分析 SMST 模型各层提取的特征。选择模型中 6 个阶段进行 t -SNE 可视化,分别为模型输入端、stage 1、stage 2、stage 3、stage 4 和 FPB 输出端,特征可视化结果如图 9 所示。可以看出,在输入端,8 种健康状态的轴承图像的特征杂乱

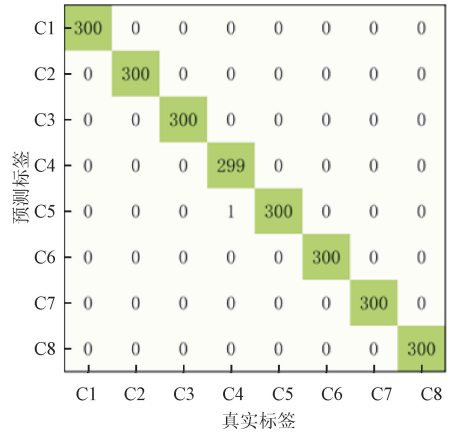
无章;随着模型层次的递进,8 种健康状态的轴承图像特征开始逐渐分离;在 stage 4 后特征基本分离,仍有一部分特征混合在一起;在经过 FPB 模块后,8 种特征完成了聚类。相比原有 SMT 模型,FPB 模块通过融合模型上下文特征信息,提高了 SMT 模型识别性能,聚类效果更好,这进一步验证了 SMST 模型的有效性。



(a) A1



(b) A2



(c) A3

图 8 混淆矩阵结果

Fig. 8 Confusion matrix results

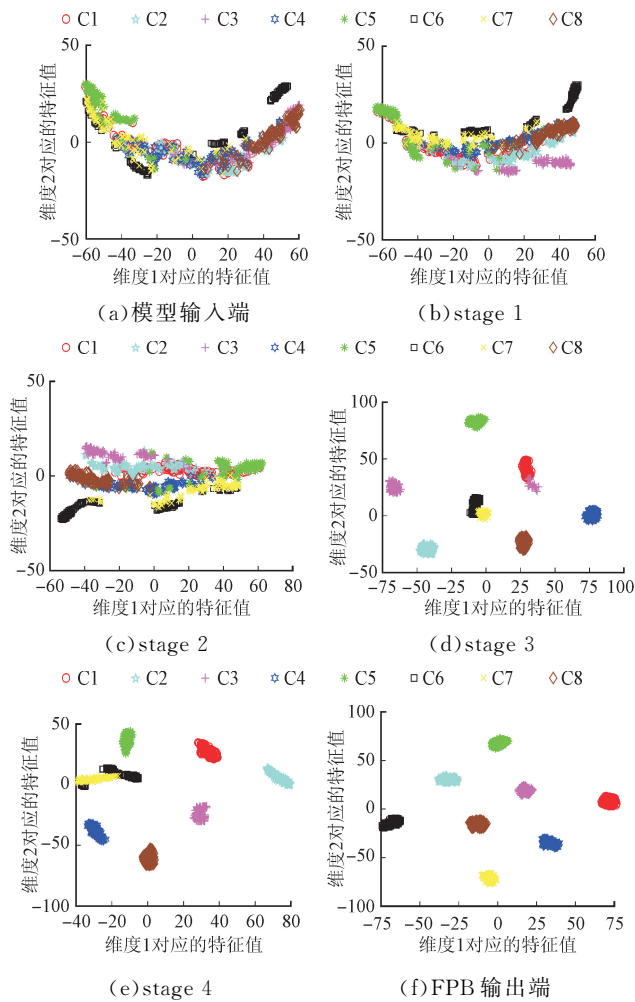


图 9 所提模型不同阶段的可视化结果
Fig. 9 Visualization results of different stages for the proposed model

3.4 消融实验

为进一步分析时频多领域图像融合及所提模型中 Swin T 及 FPB 模块对 SMST 模型性能的影响,通过两次消融实验分别进行验证分析,评价指标取 3 类任务的平均值。本文选取准确率、精准率、召回率和 F_1 值作为分析结果评价指标,上述 4 种指标定义及计算方式参见文献[25]。此外,为评估网络模型轻量化结构剪枝水平,选择运算速度与网络模型参数量两个指标对模型的计算成本与参数量进行评价。分别将 GAF、双谱、CT 法与本文方法获得的时频多领域融合图像作为信号样本输入到消融实验 1,采用 SMST 网络模型开展计算,得到的结果如表 3 所示。由表可知,GAF 与双谱法包含了信号的时域、频域信息,蕴含轴承故障信息相对单一,CT 法则包含了振动信号的时频域信息,因此 GAF 与双谱法的输入模型结果要低于 CT 法。时频多领域图像融合同时包含了时域、频域和时频域信息,模型测试结果最好,说明时频

多领域表达增强了信号中的特征信息表征能力。

表 3 消融实验 1 的结果
Table 3 Results of ablation experiment 1

评价指标	数值			
	GAF	双谱	CT	GAF+双谱+CT
准确率/%	94.61	95.03	98.79	99.92
精确率/%	94.61	95.03	95.88	99.92
召回率/%	95.71	98.79	98.78.	99.93
F_1 值/%	96.29	96.51	98.45	99.87

注:加粗数据为最优值。

消融实验 2 在 SMT 框架基础上,分别引入 Swin T 与 FPB 网络,连同 SMST 模型分别进行消融实验,结果如表 4 所示。可以看出,引入 Swin T 后,SMT 模型的准确率得到提升,参数量和运算速度显著减少;而 FPB 网络的引入则以增加少量参数量为代价,通过融合浅层和深层特征信息提升了 SMT 模型的性能。

表 4 消融实验 2 的结果
Table 4 Results of ablation experiment 2

评价指标	数值			
	SMT	SMT+ Swin T	SMT+ FPB	SMT+ Swin T +FPB
准确率/%	97.92	98.32	98.43	99.92
参数量/ 10^6	14.35	13.22	14.52	13.22
运算速度/ (10^9 s^{-1})	2.07	1.94	2.26	2.12

注:加粗数据为最优值。

3.5 对比实验

为验证本文构建模型的合理性,将 SAM、SMST 与 Swin T 按照不同顺序进行排列以获取最佳模型组合。根据文献[24]研究结果,SAM 模块位于层次化网络模型前端,模型的轻量化程度更高,运算成本更低。因此,模型的 stage 1 与 stage 2 均选择 SAM 模块,仅对模型 stage 3 与 stage 4 的构成进行分析。分别选取 4 种组合进行实验分析,结果如表 5 所示。从中可知,stage 3 与 stage 4 均选择 SMST 模块时,模型准备率较高,但轻量化结果较差;均选择 Swin T 基本模块时,模型准确率较差,但轻量化结果较好。这是因为 SMST 模块包含了 Swin T 基本模块还有 SMT 模块,性能较好,但网络结构较为复杂。本文所提模型的构成在 4 种组合中测试准确率最高,轻量化结果要稍差于采用 Swin T 模块的。

表 5 模块组合验证结果

Table 5 Module combination verification results

对比指标	数值			
	SAM+SAM+ SMST+SMST	SAM+SAM+ Swin T+Swin T	SAM+SAM+ Swin T+SMST	SAM+SAM+ SMS T+Swin T
准确率/%	99.72	99.40	99.60	99.92
参数量/ 10^6	15.65	12.17	15.03	13.22
运算速度/ (10^9 s^{-1})	2.19	2.04	2.15	2.08

为验证本文所提时频多域融合方法的优越性,选择 STFT、CWT 及 CT 这 3 种一维信号图像生成方法与本文所提方法进行对比,使用本文构建的 SMST 模型进行测试,结果如图 10 所示。由图可知,CT 的测试结果要好于 STFT、CWT 方法,这是因为 CT 是一种具有较高分辨率的参数化时频分析方法,采用变换核函数能更好地匹配待分析信号,相比 STFT、CWT 方法能达到更优的时频分析效果;本文所提方法在所有对比方法中的效果最好,所提方法生成的图像融合了信号在时域、频域、时频域内的特征信息,包含更为丰富的特征信息,信号特征信息表征能力更强。

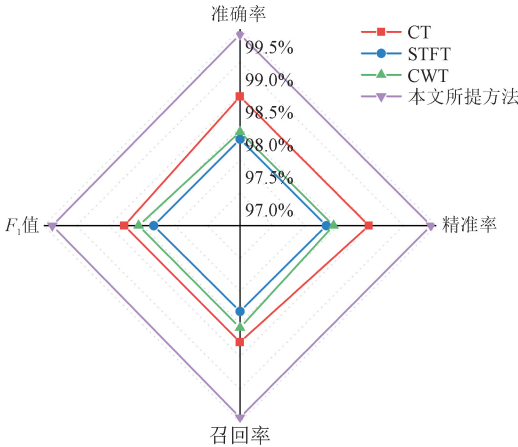


图 10 不同图像生成方法的结果对比

Fig. 10 Comparison of different image generation methods

本文选择基于 CNN 的 ResNet50^[5]、GooleNet^[6]模型,基于自注意力运算的 ViT^[13]、Swin Transformer^[14]、SMT^[24]模型共 5 种方法进行对比分析,得到的模型性能对比结果与轻量化剪枝结果分别如图 11 和表 6 所示。由图 11 可知,所有模型中,本文所提模型的识别准确率最高,为 99.9%,基于传统 CNN 框架的 ResNet50 和 GooleNet 模型的 GooleNet 网络测试结果较差;ResNet50 由于采用了残差结构,因而有效缓解了在深度神经网络中

增加深度带来的梯度消失问题,相比 GoogleNet 模型,网络深度增加模型测试准确率相对更高。根据文献[13],ViT 模型与 ResNet 模型的性能相近,但若建立在大量数据样本基础上,本次测试中 ViT 模型的准确率要低于 ResNet50;相比 ViT,Swin T 与 SMT 模型的性能虽然得到增强,但测试准确率仍低于本文所提方法。上述对比分析进一步证实了本文所提方法的优越性。

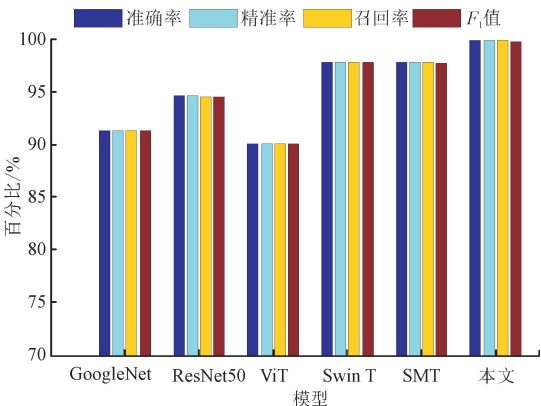


图 11 不同模型性能对比

Fig. 11 Model performance comparison

表 6 不同模型网络剪枝结果对比

Table 6 Comparison of method logical network pruning results

模型	参数量/ 10^6	运算速度/ (10^9 s^{-1})
ResNet50 ^[5]	25.56	4.12
GooleNet ^[6]	13.38	1.59
ViT ^[13]	86.57	15.28
Swin T ^[14]	28.29	4.36
SMT ^[24]	14.35	2.07
SMST	13.22	2.12

由表 6 和图 11 可知,GooleNet 网络结构简单参数量少,计算成本最低,但准确率要低于 ResNew50;相比 GooleNet 模型,ResNet50 网络深度增加,计算成本与参数量都有所提升,但模型测试

准确率相对更高。虽然 ViT 模型与 ResNet 模型性能相近,但计算成本与参数量要远远高于 ResNet 模型。相比 ViT, Swin T 模型的性能较好,计算成本与参数量均较少,这得益于 Swin T 基本模块中采用了 W-MSA 与 SW-MSA,有效减少了运算量,但测试准确率仍低于本文所提方法。SMT 模型结合 CNN 与 Transformer 两者的优势,虽然在计算成本与参数量上优于 Swin T 模型,但在准确率上要低于 Swin T 模型。本文所提方法在参数量与计算成本上与 SMT 相近,但准确率最高。上述对比分析进一步证实了本文所提方法的优越性。

4 结 论

本文提出了一种时频多域融合增强和 SMST 模型用于高速列车轮对轴承故障诊断,实现了复杂工况下轴箱轴承不同故障模式的准确识别,主要结论如下。

(1)分别采用 GAF、双谱与 CT 方法,在时域、频域与时频域内将轴承信号先编码为二维图像,再融合为新图像,利用时频多领域融合增强了信号特征信息表达能力。

(2)设计了一种轻量化结构的 SMST 模块,实现了卷积与 Transformer 的融合运算,并通过在层次化模型框架中引入 FPB 模块,实现了网络模型浅层信息和深层信息的细粒度融合,提升了模型的泛化性能。

(3)实验结果表明:本文所提模型能够在复杂工况下准确地识别轴承不同的故障类型及故障损伤程度,在提升识别准确率的同时降低模型的参数量以及计算成本,使得模型更加高效,轻量化程度更好。

参考文献:

- [1] SUN Runtao, YANG Jianwei, YAO Dechen, et al. A new method of wheelset bearing fault diagnosis [J]. Entropy, 2022, 24(10): 1381.
- [2] 向宙, 张西宁, 张雯雯, 等. 区分自编码网络及其在滚动轴承故障特征提取中的应用 [J]. 西安交通大学学报, 2019, 53(8): 47-55.
- XIANG Zhou, ZHANG Xining, ZHANG Wenwen, et al. Discriminative auto-encoding network with applications in fault feature extraction of rolling bearing [J]. Journal of Xi'an Jiaotong University, 2019, 53(8): 47-55.
- [3] JIANG Hongkai, LI Xingqiu, SHAO Haidong, et al. Intelligent fault diagnosis of rolling bearings using an

- improved deep recurrent neural network [J]. Measurement Science and Technology, 2018, 29(6): 065107.
- [4] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.
- [5] AL-QIZWINI M, BARJASTEH I, AL-QASSAB H, et al. Deep learning algorithm for autonomous driving using GoogLeNet [C]//2017 IEEE Intelligent Vehicles Symposium (IV). Piscataway, NJ, USA: IEEE, 2017: 89-96.
- [6] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 770-778.
- [7] ZHANG Xiangyu, ZHOU Xinyu, LIN Mengxiao, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 6848-6856.
- [8] 周凌孟, 邓飞其, 张清华, 等. 基于主成分分析的多域特征融合轴承故障诊断 [J]. 机床与液压, 2024, 52(6): 167-176.
- ZHOU Lingmeng, DENG Feiqi, ZHANG Qinghua, et al. Bearing fault diagnosis based on principal component analysis and multi-domain feature fusion [J]. Machine Tool & Hydraulics, 2024, 52(6): 167-176.
- [9] DENG Feiyue, DING Hao, YANG Shaopu, et al. An improved deep residual network with multiscale feature fusion for rotating machinery fault diagnosis [J]. Measurement Science and Technology, 2021, 32(2): 024002.
- [10] PENG Binsen, XIA Hong, LÜ Xinzhi, et al. An intelligent fault diagnosis method for rotating machinery based on data fusion and deep residual neural network [J]. Applied Intelligence, 2022, 52(3): 3051-3065.
- [11] YANG Yanli, FU Peiying. Rolling-element bearing fault data automatic clustering based on wavelet and deep neural network [J]. Shock and Vibration, 2018, 2018: 3047830.
- [12] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2017: 6000-6010.
- [13] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et

- al. An image is worth 16x16 words; transformers for image recognition at scale [EB/OL]. (2021-06-03) [2024-02-01]. <https://arxiv.org/abs/2010.11929>.
- [14] LIU Ze, LIN Yutong, CAO Yue, et al. Swin transformer; hierarchical vision transformer using shifted windows [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA; IEEE, 2021; 9992-10002.
- [15] HUANG Zilong, BEN Youcheng, LUO Guozhong, et al. Shuffle transformer; rethinking spatial shuffle for vision transformer [EB/OL]. (2021-06-07) [2024-02-01]. <https://arxiv.org/abs/2106.03650>.
- [16] CHU Xiangxiang, TIAN Zhi, WANG Yuqing, et al. Twins: revisiting the design of spatial attention in vision transformers [C]//Advances in Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2021; 9355-9366.
- [17] YANG Zhuohong, CEN Jian, LIU Xi, et al. Research on bearing fault diagnosis method based on transformer neural network [J]. Measurement Science and Technology, 2022, 33(8): 085111.
- [18] FANG Hairui, DENG Jin, BAI Yaouxu, 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, 2022, 71: 1-8.
- [19] HAO Wei, LI Zhixuan, QIN Guohao, et al. A novel prediction method based on bi-channel hierarchical vision transformer for rolling bearings' remaining useful life [J]. Processes, 2023, 11(4): 1153.
- [20] LIANG Pengfei, YU Zhuoze, WANG Bin, et al. Fault transfer diagnosis of rolling bearings across multiple working conditions via subdomain adaptation and improved vision transformer network [J]. Advanced Engineering Informatics, 2023, 57: 102075.
- [21] 张淑清, 杜灵韵, 王册浩, 等. 基于格拉姆角场与改进 CNN-ResNet 的风电功率预测方法 [J]. 电网技术, 2023, 47(4): 1540-1547.
- ZHANG Shuqing, DU Lingyun, WANG Cehao, et al. Windpower forecasting method based on GAF and improved CNN-ResNet [J]. Power System Technology, 2023, 47(4): 1540-1547.
- [22] 谢跃雷, 邓涵方. 基于改进 ResNet 的射频指纹识别方法 [J]. 电讯技术, 2022, 62(4): 416-423.
- XIE Yuele, DENG Hanfang. A radio frequency fingerprinting identification method based on improved ResNet [J]. Telecommunication Engineering, 2022, 62(4): 416-423.
- [23] 张力新, 贾义红, 许敏鹏, 等. 基于 Chirplet 变换的变频视觉诱发电位脑-机接口研究 [J]. 纳米技术与精密工程, 2014, 12(3): 157-161.
- ZHANG Lixin, JIA Yihong, XU Minpeng. Chirp stimuli visual evoked potential based brain-computer interface by Chirplet transform algorithm [J]. Nanotechnology and Precision Engineering, 2014, 12(3): 157-161.
- [24] LIN Weifeng, WU Ziheng, CHEN Jiayu, et al. Scale-aware modulation meet transformer [C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA; IEEE, 2023; 5992-6003.
- [25] LIU Yanting, ZHAO Junjuan, LUO Quanyong, et al. Automated classification of cervical lymph-node-level from ultrasound using depthwise separable convolutional Swin transformer [J]. Computers in Biology and Medicine, 2022, 148: 105821.

(编辑 亢列梅)