MBGA-FuseNet：双分支图网络与多域自适应融合的轴承故障诊断框架

姜佩贺[1] 刘银河[1] 王晨旭[2*]
1.烟台大学物理与电子信息学院，烟台，264005
2.哈尔滨工业大学(威海)信息科学与工程学院，威海，264209

摘要：针对复杂工况下滚动轴承故障信号存在多尺度特征提取不足、时空依赖能力弱及跨域特征融合不充分等问题，提出了双分支图网络与多域自适应融合(MBGA-FuseNet)的轴承故障诊断框架。构建双分支图神经网络架构：将改进位置编码的图级Transformer与自适应多尺度切比雪夫图卷积结合以增强节点拓扑性从而实现不同邻域信息的跨尺度自适应融合，并结合块间全局自注意力编码器提升特征的深度表达与层级关联。设计了非线性典型相关分析(NL-CCA)逐节点级跨通道特征融合模块，在节点层面实现跨分支特征的自适应关联与动态加权融合，并提出一种改进的通道融合注意力模块(CFAM)，通过引入快速卷积门控线性单元(Faster-CGLU)机制，细化通道门控信号的细粒度。最后在时序特征建模方面，采用双向门控递归单元(BiGRU)结构捕获时间依赖关系，结合提出的多域交互融合框架(HIFTFS)，将时域、频域与空间域特征进行深度交叉融合。将所提方法在多个公开轴承数据集上进行了实验验证，不同工况与信噪比条件下的准确率均优于其他对比模型。

关键词：故障诊断；图神经网络；自适应融合；注意力；交叉融合
中图分类号：TH133.3
DOI:10.3969/j.issn.1004-132X.2026.09.009 **开放科学(资源服务)标识码(OSID)**：

MBGA-FuseNet: Bearing Fault Diagnosis Framework Based on Dual-branch Graph Network and Multi-domain Adaptive Fusion

JIANG Peihe[1] LIU Yinhe[1] WANG Chenxu[2*]
1.School of Physics and Electronic Information, Yantai University, Yantai, Shandong, 264005
2.School of Information Science and Engineering, Harbin Institute of Technology, Weihai, Shandong, 264209

Abstract: To address the problems of insufficient multi-scale feature extraction, weak spatio-temporal dependency, and inadequate cross-domain feature fusion in rolling bearing fault signals under complex operating conditions, a multi branch graph attention fusion network(MBGA-FuseNet) was proposed for bearing fault diagnosis. First, a dual-branch graph neural network architecture was constructed: a graph-level Transformer with improved positional encoding was combined with an adaptive multi-scale Chebyshev graph convolutional layer to enhance node topology and achieve cross-scale adaptive fusion of information from different neighborhoods. Furthermore, an inter-block global self-attention encoder was incorporated to enhance the deep representation and hierarchical correlation of feature representation. Second, a nonlinear canonical correlation analysis(NL-CCA) node-level cross-channel feature fusion module was designed to realize adaptive correlation and dynamic weighted fusion across branches at the node level. Third, a improved channel fusion attention module(CFAM) was proposed by introducing a faster convolutional gated linear unit(Faster-CGLU) mechanism, which refines the granularity of channel gating signals. Finally, for temporal feature modeling, a bidirectional gated recurrent unit (BiGRU) was employed to capture temporal dependencies. Moreover, a hierarchical inter-domain frequency-temporal feature synthesis (HIFTFS) was developed to deeply cross-fuse features from the temporal, frequency and spatial domains. Experimental results on multiple publicly available bearing datasets demonstrate that the proposed method consistently outperforms other comparative models under various operating conditions and signal-to-noise ratios.

Key words: fault diagnosis; graph neural network; adaptive fusion; attention; interactive fusion

收稿日期：2025-11-07
基金项目：山东省重大科技创新工程(2022ZLGX04)

0　引言

在工业应用中,轴承往往工作在高速、重载以及复杂多变的环境,而在多种复杂工况下长期运行不可避免地会出现性能下降或故障[1],不仅会引起设备振动与噪声增加,还可能造成传动系统精度下降,甚至引发连锁损坏导致严重事故,因此对其健康状态检测并采取有效保护措施是保障工业系统连续运行的关键,开发一种先进的故障诊断方法具有重要的工程应用价值。

近年来深度学习在故障诊断领域得到广泛应用。REN等[2]创建了一个带有主成分分析的去噪自编码器,将原始KPCA手动确定内核的方式转换为端到端的通过训练自动获取网络权重参数的方式。LUO等[3]提出了一种改进的双向长短期记忆网络(LSTM)模型,结合双向时间序列对完整上下文信息的提取能力,提高故障分类的准确性。邹筱瑜等[4]提出一种融合时域与时频域特征的深度网络,通过嵌入差异化注意力机制实现多模态协同诊断。LEI等[5]针对故障精度不高和泛化能力不理想的问题,提出一种基于双流交互式卷积神经网络的诊断方法:将时域和频域信号作为神经网络的输入,经过动态加权信道注意机制和空间自注意机制对两个信息流充分交互。LIANG[6]等通过一种具有残差连接的一维扩张卷积网络来诊断不同噪声条件和负载下的滚动轴承问题。YANG等[7]提出了一种动态卷积胶囊网络的滚动轴承诊断方法,将快速傅里叶变换处理过的一维振动信号经过动态卷积层和池化层提取特征,并将特征信息传输到胶囊网络,在噪声条件下表现出有效的缺陷诊断性能。阴艳超等[8]通过将深度时间卷积神经网络、时序注意力机制与迁移学习的序列相结合,显著提高小样本场景下的预测精度与计算效率。

图卷积网络广泛应用于对图数据建模并有效识别样本之间的相关性[9]。宋李俊等[10]提出了改进图注意力双向门控循环单元网络,有效捕获轴承退化状态的深度时空特征。ZUO等[11]基于多层图卷积网络和多层图同构网络提出了一种多尺度加权图同构的滚动轴承故障诊断方法,将多尺度与加权规则相结合,充分避免了噪声点的干扰。FAN等[12]提出多尺度图引导动态增强对齐网络,通过多尺度图构建缓解域内类别不平衡,并结合动态增强对齐损失降低对比例失衡的敏感性。WEN等[13]针对故障诊断跨域偏移制约问题,提出多重对抗学习的聚类图卷积网络,实现面向多方位故障的无监督深度聚类诊断。

传感器获取的实时滚动轴承数据本身包含空间拓扑关系与时间动态特性,目前方法针对时空特征解耦存在一定的局限性,不能很好地对多维属性有效融合[14],提取的特征仍缺乏多维表示。为解决上述问题,提出了一种双分支图神经网络与多域混合交互融合框架下的多域自适应融合(multi branch graph attention fusion network,MBGA-FuseNet)轴承故障诊断框架。

1　原理及方法

1.1　图相关理论

传统切比雪夫图卷积网络(GCN)使用固定阶数K来定义卷积范围,为增强图的多尺度特征提取能力,根据图的结构动态选择合适的卷积核大小,本文提出自适应多尺度切比雪夫图卷积层:

$$\boldsymbol{h}_m^{(l+1)}=\sigma\left(\sum_{k=0}^{K_M-1}\boldsymbol{u}_{k,m}T_k(\bar{L})\boldsymbol{h}^{(l)}\right)$$

$$m=1,2,\cdots,M$$

式中:$\bar{L}$为归一化Laplacian矩阵;$T_k(\bar{L})$为切比雪夫系数的K阶多项式[15];$\boldsymbol{u}$为切比雪夫多项式向量参数;$\boldsymbol{h}^{(l)}$为第l层的特征输入;σ为非线性激活函数;k为节点邻接跳数;其中$\{K_1,K_2,\cdots,K_M\}$表示不同卷积尺度m下分别对应的切比雪夫阶数,$K_m=2m+1$。

在$\{K_1,K_2,\cdots,K_M\}$中,采用邻接矩阵的平均度数

$$d_{\text{avg}}=\frac{1}{N}\sum_i d(i)$$

作为最大卷积尺度M的值,其中N代表节点数,$d(i)$是节点i的度[16]。通过引入自注意系数α_m并通过softmax函数归一化处理:

$$\boldsymbol{\alpha}_m=\frac{\exp(\boldsymbol{e}_m)}{\sum_M\exp(\boldsymbol{e}_m)}$$

$$\boldsymbol{e}_m=softmax(\boldsymbol{q}^{\mathrm{T}}\boldsymbol{k}_m/\sqrt{d})$$

动态融合实现高效多尺度图卷积:

$$\boldsymbol{h}^{(l+1)}=\sigma\left(\sum_{m=1}^{M}\boldsymbol{\alpha}_m\boldsymbol{h}_m^{(l+1)}\right)$$

1.2　双向门控循环单元网络(BiGRU)

GRU作为循环神经网络(RNN)的变体,通过更简洁的门控机制和链式架构顺序捕捉和处理数据中的时序信息,同时避免RNN梯度消失和梯度爆炸的问题,标准GRU结构如图1所示。

$\boldsymbol{Z}_t$和$\boldsymbol{r}_t$分别表示更新门和重置门,$\tilde{\boldsymbol{h}}_t$表示时间t处的候选隐藏状态,$\boldsymbol{h}_{t-1}$是来自前一时间步的输入,$\boldsymbol{h}_t$是输出。其中更新门$\boldsymbol{Z}_t=\sigma(\boldsymbol{W}_{\text{xz}}\boldsymbol{x}_t+\boldsymbol{W}_{\text{hz}}\boldsymbol{h}_{t-1}+\boldsymbol{b}_z)$,重置门$\boldsymbol{r}_t=\sigma(\boldsymbol{W}_{\text{xr}}\boldsymbol{x}_t+\boldsymbol{W}_{\text{hr}}\boldsymbol{h}_{t-1}+\boldsymbol{b}_{\text{r}})$,候选隐藏状态$\tilde{\boldsymbol{h}}_t=\tanh(\boldsymbol{W}_{\text{xh}}\boldsymbol{x}_t+$

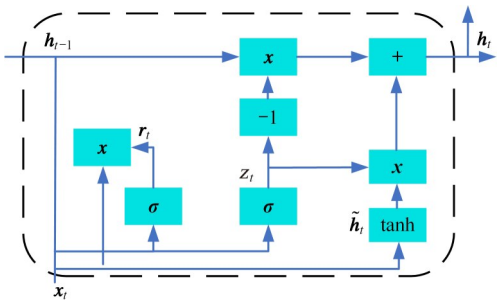


图1 GRU内部结构图

Fig.1 Internal structure diagram of GRU

$r_t \odot (h_{t-1}W_{hh}) + b_h)$，输出 $h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t$。在上述公式中，$W$用来表示当前输入$x_t$和前一时刻隐藏状态$h_{t-1}$对应于更新门和重置门的权重矩阵，$\sigma$表示Sigmoid激活函数，$b$表示偏置向量。更新门控制输出隐藏状态$h_t$中新候选隐藏状态$\tilde{h}_t$和上一隐藏状态$h_{t-1}$分别所占的比例：$Z_t = 0$时表示完全保留历史记忆信息，不对现有状态进行更新；重置门表示在生成新状态时，是否对过去时刻的隐藏状态进行考虑：$r_t = 0$时表示新的生成状态完全不依赖于过去时刻的状态信息，只与当下输入有关。

BiGRU通过并行化的双向扩展GRU信息流设计能够同时捕捉历史与未来上下文信息。如图2所示，BiGRU由两个独立且对称的GRU单元构成。

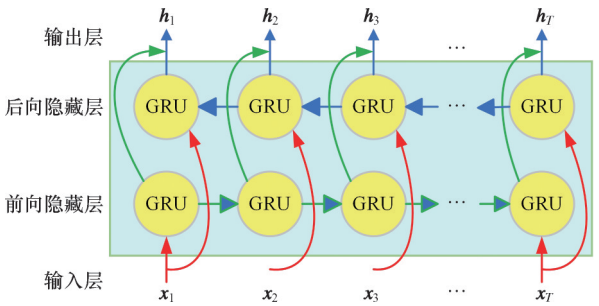


图2 BiGRU结构简图

Fig.2 BiGRU structure schematic

对于输入序列$\{x_1, x_2, \cdots x_T\}$，每个时间步经过BiGRU分别被馈送到前向和后向隐藏层。对于前向GRU层，从$t=1$到$t=T$，每个时间步t有

$$h_t' = GRU(x_t, h_{t-1})$$

对于后向GRU层，从$t=T$到$t=1$，每个时间步t有

$$h_t'' = GRU(x_t, h_{t+1})$$

对于序列中的每个时间步t，双向组合输出：

$$h_t = [h_t'; h_t'']$$

式中：$GRU(\cdot)$为门控循环单元网络，h_t'为前向GRU时间步t的隐藏层状态；h_t''为后向GRU时间步t的隐藏层状态；h_t为BiGRU时间步t的最终输出向量；$[a;b]$为向量a和b的纵向拼接。

2 MBGA-FuseNet方法

2.1 双分支图网络特征提取层

2.1.1 基于改进位置编码的图Transformer层（WGT）

如图3所示，在图G中节点i的特征表示$\alpha_i \in \mathbf{R}^{d_n \times 1}$经线性映射嵌入为$d$维隐藏表示：

$$\hat{h}_i^{(0)} = A^{(0)}\alpha_i + a^{(0)}$$

式中：$A^{(0)} \in \mathbf{R}^{d \times d_n}$为可学习的线性变换矩阵，$a^{(0)} \in \mathbf{R}^d$为线性映射层的可学习偏置向量。

用于图同构检测的WL算法作为一种嵌入位置编码方式，首先通过哈希函数$c_i^{(0)} = hash(\alpha_i)$为各节点分配初始标签，并在每轮迭代中聚合邻居标签形成多重集。随后将该多重集通过多层感知机（multi layer perception，MLP）压缩映射为新标签，并赋给节点完成一轮迭代。通过k轮迭代，节点标签更新为

$$c_i^{(k)} = MLP^{(k)}(c_i^{(k-1)} \| \sum_{u \in N(i)} c_u^{(k-1)})$$

在第k次迭代中通过聚合将$k-1$轮中节点i的标签$c_i^{(k-1)}$和邻居多重集$\{c_u^{(k-1)}: u \in N(i)\}$编码为新标签$c_i^{(k)}$。最终将各轮表示拼接并通过映射得到WL位置编码：

$$p_i^{(0)} = W_{pos}[c_i^{(1)}; c_i^{(2)}; \cdots; c_i^{(k)}] + b_{pos}$$

式中：W_{pos}为可学习的线性权重矩阵；b_{pos}为偏置项。

将得到的节点位置编码添加到节点特征中：

$$h_i^{(0)} = LayerNorm(W_x \hat{h}_i^{(0)} + W_p p_i^{(0)})$$

接下来将多头注意机制和初始化的可学习图邻接矩阵相结合，来加权节点彼此的重要信息，计算公式如下：

$$w_{ij}^{(k,l)} = softmax((1-\lambda)softmax(\frac{q_i^{(k,l)} \cdot k_j^{(k,l)}}{\sqrt{d_k}}) + \lambda \cdot softmax(ReLU(A)))$$

式中：$A \in \mathbf{R}^{N \times N}$为由邻接矩阵初始化的可学习邻接矩阵；$\lambda \in (0,1)$用于平衡由邻接矩阵确定的结构相关性与多头注意力机制下的动态相关性；$ReLU(\cdot)$为非线性激活函数。

节点更新为

$$h_i^{(l+1,k)} = \sum_{j \in N_i} w_{ij}^{(k,l)} V_j^{(k,l)} h_j^{(l)}$$

$$\hat{h}_i^{(l+1)} = O^{(l)}(Connect(h_i^{(l+1,1)}, \cdots, h_i^{(l+1,K)}))$$

式中：$h_j^{(l)} \in \mathbf{R}^{d \times 1}$为第$l$层的邻居节点信息；$V_j^{(k,l)} \in \mathbf{R}^{d_k \times d}$为邻居节点$j$在第$l$层经过线性变换后的值向量；$w_{ij}^{(k,l)}$为节点$i$与邻居节点$j$之间的自适应注意权重系数；$O^{(l)} \in \mathbf{R}^{d \times (Kd_k)}$为输出线性变换层。

得到第k头注意力下节点i在第$l+1$层的特征表示$h_i^{(l+1,k)} \in \mathbf{R}^{d_k \times 1}$。将多头注意力通过$Connect(\cdot)$拼接得到更新后的多头注意节点表示

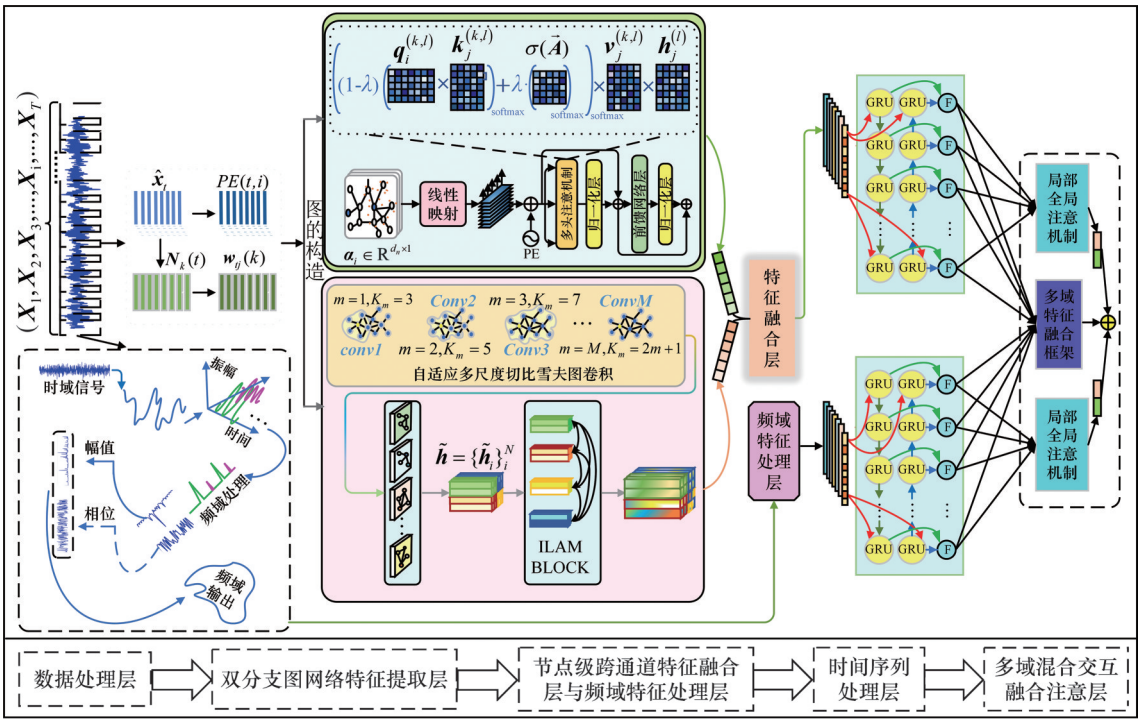


图3　MBGA-FuseNet轴承故障诊断总体框架

Fig.3　General framework for bearing fault diagnosis with MBGA-FuseNet

$\hat{\boldsymbol{h}}_i^{(l+1)}\in\mathbf{R}^{d\times 1}$。最后将输出结果依次经过归一化层(LN)、残差连接层、前馈网络层(FN)、归一化层、残差连接层得到完整编码器输出：

$$\boldsymbol{h}_{i,\mathrm{LN}}^{(l+1)}=LN(\hat{\boldsymbol{h}}_i^{(l+1)})+\boldsymbol{h}_i^{(l)}$$

$$\boldsymbol{h}_{i,\mathrm{FN}}^{(l+1)}=\boldsymbol{W}_2^{(l)}GELU(\boldsymbol{W}_1^{(l)}\boldsymbol{h}_{i,\mathrm{LN}}^{(l+1)})$$

$$\boldsymbol{h}_i^{(l+1)}=LN(\boldsymbol{h}_{i,\mathrm{FN}}^{(l+1)})+\boldsymbol{h}_{i,\mathrm{LN}}^{(l+1)}$$

式中：$\boldsymbol{W}_1^{(l)}$、$\boldsymbol{W}_2^{(l)}$为第l层前馈网络可学习权重矩阵；$GELU(\cdot)$为高斯误差单元非线性激活函数。

2.1.2　自适应多尺度图卷积融合层(MFCN)

在1.1节中介绍了自适应多尺度切比雪夫层，如图3所示，不同卷积层($Conv$)对应不同的切比雪夫多项式阶数，从而获取不同邻居的多层次结构信息。因此提取的特征具有显著的层次差异：浅层卷积更关注于局部邻域低阶信息，深层则聚合跨域高阶特征[17]。为将不同卷积层次下的多尺度特征充分融合，引入块间全局自注意力编码器(inter layer global attention module，ILAM)。首先将各层输出视为层级词元并对输出$\boldsymbol{h}=(\boldsymbol{h}_1,\boldsymbol{h}_2,\cdots,\boldsymbol{h}_N)\in\mathbf{R}^{N\times d}$执行块嵌入操作：

$$\tilde{\boldsymbol{h}}=\boldsymbol{h}\boldsymbol{W}_{\mathrm{emb}}+\boldsymbol{E}_{\mathrm{pos}}$$

式中：$\boldsymbol{W}_{\mathrm{emb}}\in\mathbf{R}^{d\times d_e}$为嵌入线性映射层；$\boldsymbol{E}_{\mathrm{pos}}\in\mathbf{R}^{N\times d_e}$为可学习位置编码，将第$l$层嵌入向量$\tilde{\boldsymbol{h}}^{(l)}\in\mathbf{R}^{N\times d_e}$送入ILAM单元。

如图4所示，ILAM由两个并行的全局平均池化(GAP)与全局最大池化层(GMP)、归一化层、基于门控机制多头自注意模块(layer gated multi self attention，LGMSA)、残差连接层和多层感知机构成，计算过程如下：

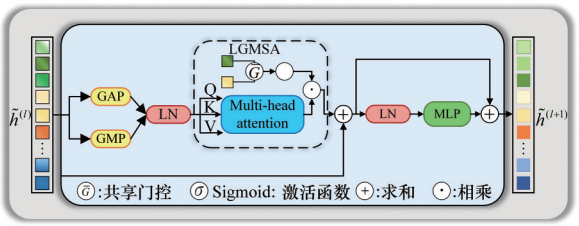


图4　ILAM示意图

Fig.4　Schematic diagram of ILAM

$$\tilde{\boldsymbol{h}}_{\mathrm{inr}}^{(l)}=LN(GAP(\tilde{\boldsymbol{h}}^{(l)})\oplus GMP(\tilde{\boldsymbol{h}}^{(l)}))$$

$$\tilde{\boldsymbol{h}}_{\mathrm{msa}}^{(l)}=LGMSA(\tilde{\boldsymbol{h}}_{\mathrm{inr}}^{(l)})+\tilde{\boldsymbol{h}}^{(l)}$$

$$\tilde{\boldsymbol{h}}^{(l+1)}=MLP\{\mathrm{LN}[\tilde{\boldsymbol{h}}_{\mathrm{msa}}^{(l)}]\}+\tilde{\boldsymbol{h}}_{\mathrm{msa}}^{(l)}$$

对于$\tilde{\boldsymbol{h}}_{\mathrm{inr}}^{(l)}$中任意两个层级$\boldsymbol{z}^{(l)}\in\mathbf{R}^{2d_e}$和$\boldsymbol{z}^{(m)}\in\mathbf{R}^{2d_e}$，通过一个共享门控判定它们的交互强度：

$$\bar{\boldsymbol{G}}_{\mathrm{lm}}=\boldsymbol{a}_g^{\mathrm{T}}\tanh(\boldsymbol{W}_g^{(l)}\boldsymbol{z}^{(l)}+\mathbf{W}_g^{(m)}\boldsymbol{z}^{(m)}+\boldsymbol{b}_g)$$

$$\boldsymbol{G}_{\mathrm{lm}}=\sigma(\bar{\boldsymbol{G}}_{\mathrm{lm}})$$

式中：$\boldsymbol{W}_g^{(l)}$和$\boldsymbol{W}_g^{(m)}\in\mathbf{R}^{d_a\times 2d_e}$为第$l$层和第$m$层的可学习权重矩阵；$\boldsymbol{b}_g$为可学习偏置向量；$\boldsymbol{a}_g\in\mathbf{R}^{d_a}$为线性投影向量；$\tanh(\cdot)$为非线性激活函数。

最后通过Sigmoid激活函数得到介于[0，1]之间的层交互关系$\boldsymbol{G}_{\mathrm{lm}}$。LGMSA可视作$H$个独立自注意力头的扩展，每个注意力头系数可表示为

$$\tilde{\boldsymbol{\alpha}}^h=softmax(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}})$$

$$\boldsymbol{Q}=\tilde{\boldsymbol{h}}_{inr}^{(l)}\boldsymbol{W}_q\qquad\boldsymbol{K}=\tilde{\boldsymbol{h}}_{inr}^{(l)}\boldsymbol{W}_k$$

式中：$\sqrt{d_k}$为特征维度缩放因子；$\boldsymbol{W}_q$，$\boldsymbol{W}_k\in\mathbf{R}^{2d_e\times D_e}$为对应的线性映射层。

将其与门控相乘并执行归一化操作：

$$\boldsymbol{\alpha}^h = Softmax(\tilde{\boldsymbol{\alpha}}^h \boldsymbol{G}_{\mathrm{lm}})$$

得到单个注意头表示：$\boldsymbol{S}_H(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \boldsymbol{\alpha}^h \boldsymbol{V}$，其中 $\boldsymbol{V} = \tilde{\boldsymbol{h}}_{\mathrm{inr}}^{(l)} \boldsymbol{W}_v$，$\boldsymbol{W}_v \in \mathrm{R}^{2d_e \times D_e}$。

通过拼接 H 个S头，并经由线性映射层 $\boldsymbol{W}_{\mathrm{lgmsa}} \in \mathrm{R}^{HD_e \times d_e}$ 重新将维度恢复到起始维度，LGMSA最终可以被定义为

$$\mathrm{LGMSA}(\tilde{\boldsymbol{h}}_{\mathrm{inr}}^{(l)}) = Connect(\boldsymbol{S}_1, \boldsymbol{S}_2, \cdots, \boldsymbol{S}_H)\boldsymbol{W}_{\mathrm{lgmsa}}$$

2.2 节点级跨通道特征融合层(NL-CCA)

为将双分支图网络的双通道输出特征 $\boldsymbol{H}_t \in \mathrm{R}^{N \times d_t}$、$\boldsymbol{H}_e \in \mathrm{R}^{N \times d_e}$ 更好地自适应融合，如图5所示，将来自每个通道的输出看作由 N 个特征子节点构成的节点集合，每个节点由 d 个特征值组成，构成了基于通道数 C 的 $N \times d$ 维特征矩阵 $\boldsymbol{M}_{\mathrm{ar}} = [\boldsymbol{M}_1 \quad \boldsymbol{M}_2 \quad \cdots \quad \boldsymbol{M}_c] \in \mathrm{R}^{C \times N \times d}$。将各个通道的每个特征矩阵基于节点数分成由多个特征块组成的节点级信道矩阵 $\boldsymbol{C} = [\boldsymbol{c}_1 \quad \boldsymbol{c}_2 \quad \cdots \quad \boldsymbol{c}_N]$，为了捕捉不同通道在节点级的交互，对双通道每个节点的特征进行线性变换：

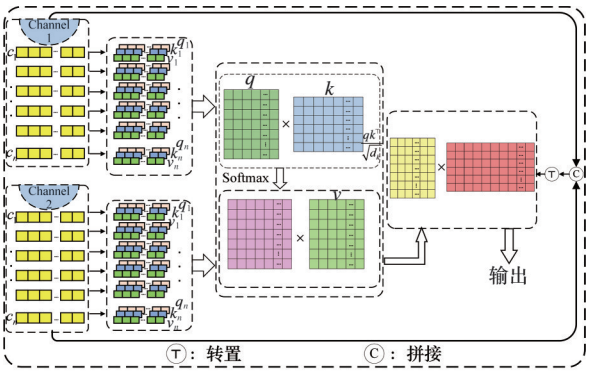


图5 节点级跨通道特征融合层

Fig.5 Node-level cross-channel feature fusion layer

$$\boldsymbol{q}_i^e = \boldsymbol{c}_i^e \boldsymbol{W}_q^e \qquad \boldsymbol{k}_i^e = \boldsymbol{c}_i^e \boldsymbol{W}_k^e \qquad \boldsymbol{v}_i^e = \boldsymbol{c}_i^e \boldsymbol{W}_v^e$$

$$\boldsymbol{q}_i^t = \boldsymbol{c}_i^t \boldsymbol{W}_q^t \qquad \boldsymbol{k}_i^t = \boldsymbol{c}_i^t \boldsymbol{W}_k^t \qquad \boldsymbol{v}_i^t = \boldsymbol{c}_i^t \boldsymbol{W}_v^t$$

式中：$\boldsymbol{c}_i^e$ 为单通道 e 中节点 i 的特征矩阵，并通过线性单元 $\boldsymbol{W}^e$ 分别映射到对应的 $\boldsymbol{q}_i^e$、$\boldsymbol{k}_i^e$ 和 $\boldsymbol{v}_i^e$ 向量中。

节点的跨通道特征融合得到单个节点通道融合后对应的 $\boldsymbol{q}_i$、$\boldsymbol{k}_i$、$\boldsymbol{v}_i$：

$$\boldsymbol{q}_i = [\boldsymbol{q}_i^e;\ \boldsymbol{q}_i^t] \qquad \boldsymbol{k}_i = [\boldsymbol{k}_i^e;\ \boldsymbol{k}_i^t] \qquad \boldsymbol{v}_i = [\boldsymbol{v}_i^e;\ \boldsymbol{v}_i^t]$$

再将所有节点拼接得到 N 个节点双通道融合后的 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 矩阵。由 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 计算特征融合权重矩阵，然后将权重矩阵与 $\boldsymbol{V}$ 相乘并和拼接后的初始双通道特征相乘，得到最终的融合特征结果 $\hat{\boldsymbol{C}}$。最后通过线性映射层和非线性激活函数恢复至最初维度，具体计算过程如下：

$$\boldsymbol{C}_{\mathrm{attention}} = softmax\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V}$$

$$\hat{\boldsymbol{C}} = \boldsymbol{C}_{\mathrm{attention}} \times (Connect(\boldsymbol{C}_e, \boldsymbol{C}_t))^{\mathrm{T}}$$

$$\hat{\boldsymbol{C}}_{\mathrm{out}} = \sigma(\hat{\boldsymbol{C}}\boldsymbol{W}_1 + \boldsymbol{b}_1)\boldsymbol{W}_2 + \boldsymbol{b}_2$$

式中：$\sqrt{d_k}$ 为特征维度缩放因子；$\boldsymbol{C}_e$、$\boldsymbol{C}_t$ 为初始双通道特征矩阵；$\boldsymbol{W}_1$、$\boldsymbol{W}_2$ 为两层前馈网络可学习权重矩阵；$\boldsymbol{b}_1$、$\boldsymbol{b}_2$ 为对应的偏置向量；$\hat{\boldsymbol{C}}_{\mathrm{out}} \in \mathrm{R}^{N \times d}$ 为最终融合输出结果。

2.3 频域特征处理层(CFAM)

如图3所示，原始振动信号经过快速傅里叶变换(FFT)转换为频域特征信号。对频域信号进行频域特征提取，见图6。首先将 d 个特征维度映射为 d 个通道表征，将全局池化与平均池化应用于每个通道，得到两个 $1 \times 1 \times C$ 的特征图，随后送入MLP(线性层神经元个数为 C/r)中，并将输出结果分别通过一维卷积层促进通道交互并恢复至 C 维后进行加权操作，通过Sigmoid函数为每个通道生成注意力权重。最后，依次计算其与原始输入特征之间的点积，得到通道信息增强后的特征表示。

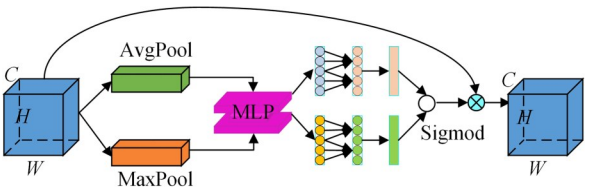


图6 CAM模块

Fig.6 CAM module

传统空间注意模块同一通道的特征图中所有词元共享相同的空间门控信号，导致作用粒度粗糙。为此在结构中加入Faster-CGLU模块[18]，通过卷积门控线性单元(CGLU)细致化通道门控信号，其结构如图7所示。

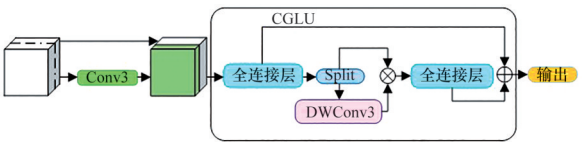


图7 Faster-CGLU模块

Fig.7 Faster-CGLU module

首先，对输入通道进行卷积操作提取空间特征，之后经过CGLU模块的全连接层分成两部分：其中一部分经过3×3深度可分离卷积生成门控信号，保证每个通道与特征图中的每个词元都能从局部邻域特征的上下文得到独立的门控信息。最后将基于最近邻特征的细粒度门控信号与另一分支的特征图逐元素相乘，再经过全连接层处理并与原始输入通过残差连接，得到最终输出。

2.4 多域混合交互融合注意层

将双分支特征融合层与频域特征层处理的空间频域特征，分别通过时间序列处理层(BiGRU)，从时间维度上捕获信号的双向依赖关系。

为将时域、频域和双分支图网络提取的空间域特征充分融合，实现各个域特征之间的融合互补，提出图8a所示的多域特征融合框架（hierarchical inter-domain frequency-temporal feature synthesis，HIFTFS）并协同基于门控的局部-全局自适应注意机制共同构成多域混合交互融合注意层。

2.4.1　频域特征增强（FSA）模块

对于第 i 层输入FSA模块的分量 $\boldsymbol{X}_{\text{freq},i}\in\mathbf{R}^{n\times d}$，如图8b所示，先经过频域自注意机制得到增强后的频域特征信息，再通过IFFT转换回到时域中，并与经过自适应权重校正的输入特征进行融合后得到最终模块输出 $\hat{\boldsymbol{X}}_{\text{freq},i}$。上述过程如下：

$$\boldsymbol{Q}=\boldsymbol{X}_{\text{freq},i}\boldsymbol{W}^{Q}\qquad \boldsymbol{K}=\boldsymbol{X}_{\text{freq},i}\boldsymbol{W}^{K}\qquad \boldsymbol{V}=\boldsymbol{X}_{\text{freq},i}\boldsymbol{W}^{V}$$

$$\boldsymbol{Q}_{\text{f}}=FFT(\boldsymbol{Q})\qquad \boldsymbol{K}_{\text{f}}=FFT(\boldsymbol{K})\qquad \boldsymbol{V}_{\text{f}}=FFT(\boldsymbol{V})$$

$$\boldsymbol{X}_{f}=softmax\left(\frac{\boldsymbol{Q}_{\text{f}}\boldsymbol{K}_{\text{f}}^{\text{T}}}{\sqrt{d_k}}\right)\boldsymbol{V}_{\text{f}}$$

$$\hat{\boldsymbol{X}}_{\text{freq},i}=IFFT(\boldsymbol{X}_{f})\oplus IFFT(\boldsymbol{W}_{k}\cdot\text{FFT}(\boldsymbol{X}_{\text{freq},i}))$$

式中：$FFT(\cdot)$为快速傅里叶变换算子；$IFFT(\cdot)$为逆快速傅里叶变换算子；$\boldsymbol{Q}_{\text{f}}$、$\boldsymbol{K}_{\text{f}}$、$\boldsymbol{V}_{\text{f}}$分别为 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 经过快速傅里叶变换在频域中的表示；$\boldsymbol{X}_{\text{f}}$为经过频域自注意机制增强后的频域特征信息；$\boldsymbol{W}_{k}$为自适应权重校正可学习矩阵。

2.4.2　时频双向交互注意（TFF）机制

首先通过具有线性映射的特征嵌入模块分别计算时域特征 T 的 T_Q、T_K、T_V 和频域特征 F 的 F_Q、F_K、F_V，并使用点积注意机制计算时域特征和频域特征之间的相关性度量。在双向交互的注意机制关系中，由时域主导的相关性度量

$$\boldsymbol{\alpha}_{i,j}^{\text{TF}}=softmax\left(\frac{(T_{Q,i},F_{K,j})}{\sqrt{d}}\right)$$

用来表示第 i 个时域特征对第 j 个频域特征的注意力权重系数，其中 i 和 j 分别对应时域特征和频域特征时间步长上的特征向量。

对另一域的value向量作加权求和，最终得到从频域特征导出的时域上下文增强向量：

$$\boldsymbol{C}_{\text{T}}=\left(\sum_{j}\boldsymbol{\alpha}_{i,j}^{\text{TF}}\right)\cdot\boldsymbol{F}_{V,j}$$

把频域中与 T_i 最相关的若干频率分量按权重聚合为一个多维表示，使时域特征能够注意到频域特征中的重要特征。相反也能得到从时域导出的频域上下文表示：

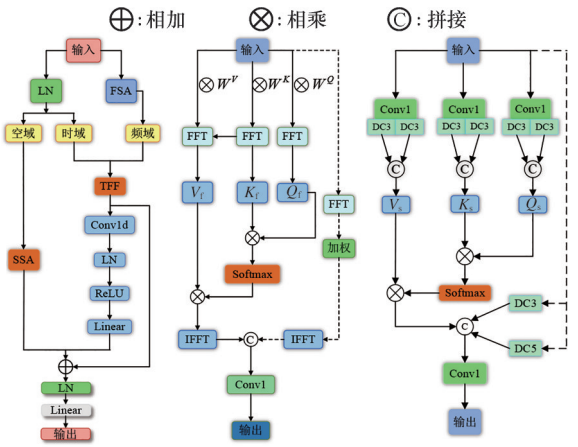
$$\boldsymbol{C}_{\text{F}}=\left(\sum_{i}\boldsymbol{\alpha}_{j,i}^{\text{FT}}\right)\cdot\boldsymbol{T}_{V,i}$$

频域的每个词元也可以借助时域的信息来增强频域表示。最终，时-频交互产生融合的上下文表示：

$$\boldsymbol{C}_{\text{T,F}}=\boldsymbol{C}_{\text{T}}+\boldsymbol{C}_{\text{F}}$$

2.4.3　空间增强自注意（SSA）机制

为增强空间特征表示，如图8c所示，首先通过1×1卷积对空间位置线性映射生成位置编码，随后在3个并行分支上分别采用拼接2个3×3的深度可分离卷积并结合自注意力机制提取空间注意力特征。将不同感受野的卷积输出与注意力结果拼接，并经卷积块投影回原通道维度，得到增强的空间特征。



(a)HIFTFS模块　(b)FSA模块　(c)SSA模块

图8　HIFTFS框架

Fig.8　HIFTFS framework

2.4.4　基于门控的局部-全局自适应注意机制

为更好地融合局部细节（捕捉短时突发故障）与全局结构（识别周期性故障），提出带门控的局部-全局自适应注意模块，通过可学习参数 γ 对局部和全局特征动态加权融合：

$$\gamma=\sigma(\boldsymbol{W}_{\gamma}\boldsymbol{H}_{\text{Input}}+\boldsymbol{b}_{\gamma})$$

式中：σ 为Sigmoid激活函数；$\boldsymbol{W}_{\gamma}$ 为可学习权重矩阵；$\boldsymbol{H}_{\text{Input}}$ 为不同特征模态的输入向量；$\boldsymbol{b}_{\gamma}$ 为偏置项。

最终融合输出由参数 γ 和 $1-\gamma$ 动态调整，实现模型对局部和全局特征的依赖关系：

$$\boldsymbol{C}_{\text{fusion}}=\gamma\sum_{i}\boldsymbol{\alpha}_{\text{local},i}\boldsymbol{H}_{i}+(1-\gamma)\sum_{i}\boldsymbol{\alpha}_{\text{global},i}\boldsymbol{H}_{i}$$

式中：$\boldsymbol{\alpha}_{\text{local},i}$、$\boldsymbol{\alpha}_{\text{global},i}$ 分别为局部注意力和全局注意力在位置 i 处的权重向量。

3　实验验证

3.1　实验设置与数据集建立

3.1.1　数据集介绍

分别利用凯斯西储大学（CWRU）数据集和西安交通大学转子数据集（XJTU-SY）验证所提出的故障诊断方法的有效性。CWRU数据集实验采用驱动端轴承在12 kHz采样频率，转速为1797 r/min、负载为0工况条件下运行采集的振动信号，故障类别及其标签编号见表1。

XJTU-SY数据集根据转速和径向载荷力的不同共设置3种工况，如表2所示。从3种工况中选取7组具有代表性的数据进行实验。

表1 CWRU滚动轴承数据集说明

Tab.1 Description of CWRU rolling bearing data set

故障位置	故障直径	标签	数据集
内圈	0.007	1	105
	0.014	2	169
	0.021	3	209
外圈	0.007	4	130
	0.014	5	197
	0.021	6	234
滚珠	0.007	7	118
	0.014	8	185
	0.021	9	222
正常	0	0	97

表2 西安交通大学滚动轴承数据集

Tab.2 Description of XJTU-SY rolling bearing data set

工况	标签	数据集	故障位置
35 Hz/12 kN	0	Bearing 1_1	外圈
	1	Bearing 1_4	保持架
37.5 Hz/11 kN	2	Bearing 2_1	内圈
	3	Bearing 2_2	外圈
	4	Bearing 2_3	保持架
40 Hz/10 kN	5	Bearing 3_1	外圈
	6	Bearing 3_3	内圈

3.1.2 实验数据处理

初始学习率设为0.003，批大小为32，训练100个轮次，采用交叉熵损失函数和Adam优化算法。在CWRU数据集中，先使用长度为2048的滑动窗口对序列进行切分，每个滑动窗口步长为256，分割后的样本按6∶2∶2的比例依序划分为训练集、验证集和测试集（XJTU-SY数据集在同工况实验中采用相同数据处理方式并按照8∶2的比例依序划分为训练集和测试集）并剔除跨边界的窗口样本。将每个样本窗口划分为58个子窗口，每个子窗口$\boldsymbol{x}_t$视作一个特征子节点并对原始信号归一化处理：

$$\hat{\boldsymbol{x}}_t=\frac{\boldsymbol{x}_t-\mu}{\sigma}$$

式中：μ为时间序列的均值；σ为标准差。

为捕获时间序列中的动态演化并增强模型对时间位置信息的感知能力，对每个时间步引入32维正弦位置编码：

$$PE(t,2i)=\sin\left(\frac{t}{10\,000^{2i/32}}\right)$$

$$PE(t,2i+1)=\cos\left(\frac{t}{10\,000^{2i/32}}\right)$$

式中：i为位置编码在频率上的离散化配置，$i=0,1,\cdots,15$，每个i对应一对不同频率尺度的正弦函数与余弦函数。

使用欧氏距离并基于K最近邻（KNN）来构造节点相似性矩阵：

$$\boldsymbol{N}_k(t)=Top_k\left(\operatorname{argmin}_{j\neq t}\left\|\boldsymbol{x}_t-\boldsymbol{x}_j\right\|_2\right)$$

基于节点t的K最近邻集$\boldsymbol{N}_k(t)$，计算节点t与其相邻节点j之间的初始边权重：

$$\boldsymbol{w}_{tj}(k)=\frac{1}{\left\|\boldsymbol{x}_t-\boldsymbol{x}_j\right\|_2+\varepsilon}$$

式中：ε为非零的最小常数。

最终基于邻接关系构造邻接矩阵，完成图拓扑的数据集构造：

$$\boldsymbol{A}(k^*)=\begin{cases}\dfrac{1}{\left\|\boldsymbol{x}_t-\boldsymbol{x}_j\right\|_2+\varepsilon} & j\in\boldsymbol{N}_k(t)\\ 0 & \text{其他}\end{cases}$$

3.2 CWRU数据集实验

3.2.1 消融实验

为验证各模块的有效性，设计了9组消融对比实验（模型A～I），每组实验组合方式以及实验结果见表3。以未改进的双分支卷积-图Transformer网络（模型A）为基准，与经过优化的双分支图网络（模型B）比较，精确率从88.68%提高到94.12%，表明模型B在捕获不同拓扑层次方面的有效性。模型C在B的基础上精确率提高

表3 消融实验结果

Tab.3 Results of the ablation experiments

模型	WGT	MFCN	NL-CCA	CFAM	BiGRU	HIFTFS	精确率/%	召回率/%	F1分数
A（基准）	√	√					88.68	88.50	0.8859
B	√	√					94.12	93.20	0.9367
C	√	√			√		94.36	93.45	0.9390
D	√	√	√		√		95.12	94.87	0.9491
E	√	√		√	√		94.52	94.00	0.9426
F	√			√	√		91.16	90.96	0.9106
G		√		√	√		89.09	89.02	0.8899
H			√	√	√		86.72	86.74	0.8673
I	√	√	√	√	√		97.32	97.29	0.9731
本文方法	√	√	√	√	√	√	**99.28**	**99.46**	**0.9937**

0.24%,有效地从时间维度上增强了信号双向依赖关系的捕捉,提高时序特征的提取能力。模型D和模型E分别在模型C的基础上增加了NL-CCA模块和CFAM模块,使模型能够更好地通道融合并聚焦关键时空特征,精确率分别进一步提高到了95.12%和94.52%。通过比较E和G发现:缺少WGT时,精确率下降约5%,这表明改进位置编码的图Transformer结构对挖掘图节点间的长程依赖关系起到了不可替代的作用。同理,在模型E和F的比较中,可以看出MFCN模块自适应充分融合图卷积各层之间的多尺度特征后对模型性能的提升。

由图9混淆矩阵可看出,基准模型A的分类效果最差,尤其是在0.014内圈故障的分类中将22个样本错误分类为滚珠类型的故障;而在所提方法中,所有样本仅有4个被误分类。

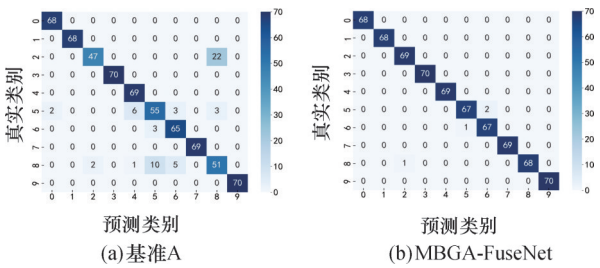


图9　不同消融模型的混淆矩阵

Fig.9　Confusion matrices of different ablation models

3.2.2　比较实验描述

本研究主要与多种主流网络模型进行比较,并在相同的数据集及工况(负载为0)下进行训练。首先将所提模型与3种典型的时间特征序列模型CNN-BiGRU、BILSTM、Informer进行比较,这些模型仅在时间域特征提取,难以同时考虑故障信号的空间结构关联性与多尺度特征分布。相比之下,MBGA-FuseNet采用双分支特征融合结构,充分将图结构的空间建模能力与时间序列特征提取的优势结合。相较于对比模型中的图类方法无法有效捕获空间结构信息与多通道关联特征以及长期依赖自注意进行长距离建模的缺点,通过改进KNN时间步长联合编码算法的加权图构造方法,使节点间的关联从线性时间依赖扩展到非欧空间的全局关系,并通过MFCN分支的自适应多尺度滤波机制提取多频率、多层次的特征信息,分别实现了准确率2.02%、11.92%、9.89%的提高。如表4所示,尽管所提方法增加了推理时间和计算资源消耗,但在对可靠性要求极高的轴承故障诊断上实现了更精确的诊断评估。

表4　对比实验结果

Tab.4　Results of the comparative experiments

模型	准确率/%	F1分数	召回率/%	参数量/10^6	浮点运算次数/10^9
CNN-BiGRU[19]	97.50	0.9775	98.00	0.780	1.23
BILSTM[3]	87.60	0.8815	88.65	0.672	1.08
Informer[20]	89.63	0.8998	90.34	8.96	18.6
SGCN[21]	94.99	0.9524	95.50	1.09	2.64
Graph-Transformer[15]	91.54	0.9216	92.79	1.46	2.14
SpGAT[22]	79.60	0.8005	80.45	0.830	2.91
本文方法	**99.52**	**0.9942**	**99.46**	**3.83**	**7.36**

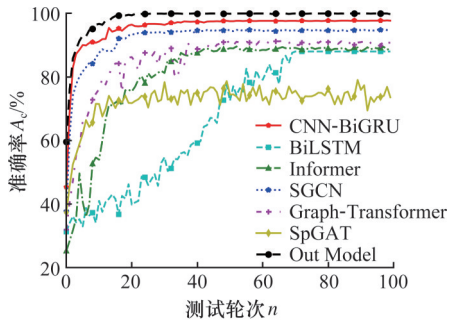
为验证所提方法的抗噪性及泛化性,又分别在负载为0和2237.1 W(3HP)的条件下进行了不同信噪比R的对比实验。如表5所示,即便在不同的负载环境下,所提方法对噪声干扰均表现出较强的抑制能力。

表5　不同负载和信噪比下的实验结果

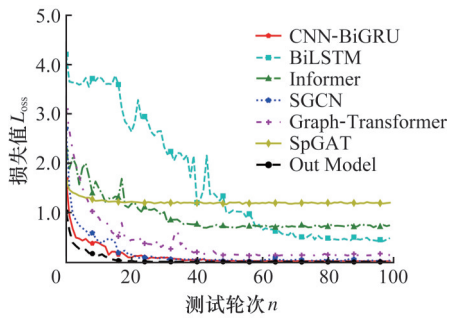
Tab.5　Experimental results with different loads and signal-to-noise ratios　%

模型	负载为0					负载为2237.1 W(3 HP)				
	R=2 dB	R=4 dB	R=6 dB	R=8 dB	R=10 dB	R=2 dB	R=4 dB	R=6 dB	R=8 dB	R=10 dB
CNN-BiGRU	85.23	88.30	92.52	94.70	96.43	82.26	86.27	91.13	93.50	96.13
BILSTM	69.83	75.10	79.88	83.50	85.59	67.96	74.10	78.47	82.68	84.80
Informer	76.83	82.25	85.47	87.14	88.21	74.71	80.94	84.24	86.83	87.94
SGCN	83.17	86.21	89.98	91.72	93.20	81.08	84.68	88.71	91.08	92.76
Graph-Transformer	78.52	84.40	87.97	89.50	90.27	78.04	84.27	86.94	89.32	90.32
SpGAT	67.43	72.07	75.56	77.17	78.64	65.83	70.80	74.34	76.67	78.40
本文方法	**93.41**	**94.09**	**95.58**	**96.95**	**98.71**	**91.32**	**93.82**	**95.34**	**96.81**	**98.58**

图10显示了不同对比模型在100轮次迭代后的测试准确率和损失曲线,所提模型可以快速平稳达到验证损失收敛水平,且准确率在短时间内提高到较高水平。引入t分布随机邻域嵌入算法对高维特征空间进行非线性降维和可视化分析。如图11所示,与未经过分类的原可视化数据相比,对比的图模型表现出类间重叠和模糊的决策边界,而图11c为所提模型的可视化结果:类间特征紧密性显著增强,类间特征可区分性增强,每种故障类别形成具有明确几何边缘的独立聚类。



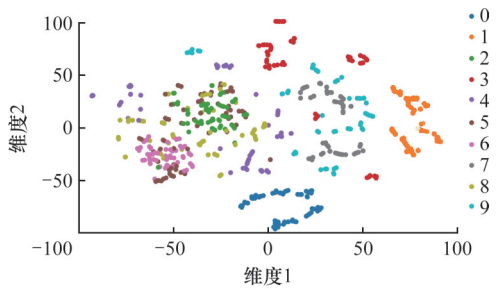
(a)不同模型比较的准确率曲线



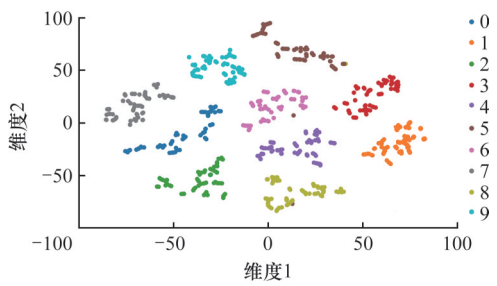
(b)不同模型比较的损失值曲线

图10 对比实验损失与准确率

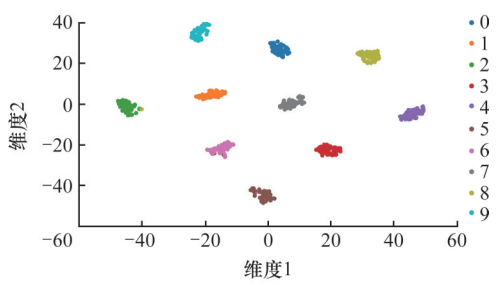
Fig.10 The loss and accuracy of comparison experiments



(a)原始数据



(b)Graph-Transformer特征提取部分



(c)所提方法特征提取

图11 t-SNE可视化对比结果

Fig.11 The t-SNE visualization comparison results

3.3 XJTU-SY数据集实验验证

3.3.1 不同模型的信噪比实验对比

将信噪比分别为:2、4、6、8、10 dB的高斯白噪声分别加入原始振动信号中。为了避免噪声信号带来的随机性,每组实验重复5次计算平均值。图12显示了不同噪声条件下7种模型对比的平均准确度,表6给出了不同信噪比下的平均准确度和标准差。所提方法在所有信噪比场景下都实现了最好的诊断性能,即使在低信噪比(2 dB)时所提出的方法仍保持了92.79%的平均准确率,优于其他对比模型;同时,标准差为0.19%,表明在噪声条件下依然具有较强稳定性。

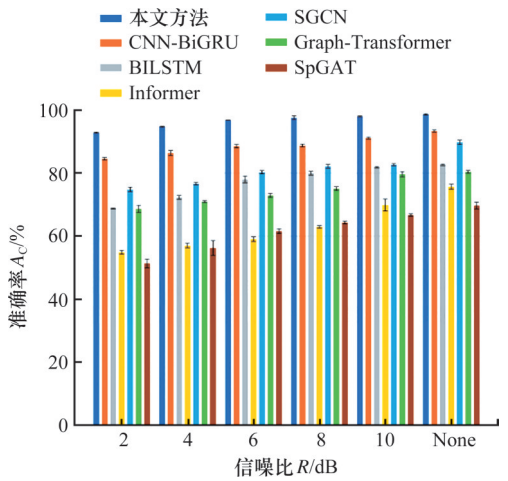


图12 不同信噪比下的结果对比

Fig.12 Comparison of results under different signal-to-noise ratios

3.3.2 变工况条件下轴承数据集的实验

为验证跨工况条件下的迁移泛化能力,设计了变工况迁移实验。数据集包括的3种不同工况:35 Hz/12 kN(记为工况A)、37.5 Hz/11 kN(记为工况B)以及40 Hz/10 kN(记为工况C),分别代表不同转速与载荷下的运行状态。为了模拟实际工程中因转速与负载变化导致的数据分布差异,构建了6组跨工况迁移实验,包括A-B、A-C、B-A、B-C、C-A、C-B 6种组合。实验过程中首先对各工况数据进行标准化与统一分段处理以保证输入特征维度一致,并在噪声为−4 dB的条件下进行验证,实验结果如图13所示。

在6种不同跨工况样本实验中,所提方法的平均识别准确率为98.21%,均高于其他对比模型的平均准确率。其中工况B-A的故障识别准确率最高,为98.63%,比其他对比模型在相同工况下提高了3.48%~23.43%。此外,从图中的误差线可以看出,所提模型的误差范围更小,具有更强的泛化能力,在噪声环境下也更稳定。

表 6　不同信噪比下的对比试验结果

Tab.6　Comparison results under different signal-to-noise ratios　%

模型	信噪比					
	2 dB	4 dB	6 dB	8 dB	10 dB	未添加噪声
CNN-BiGRU	84.53±0.42	86.38±0.78	88.55±0.50	88.74±0.41	91.10±0.32	93.32±0.35
BILSTM	68.76±0.12	72.26±0.58	77.98±1.08	79.93±0.62	81.85±0.25	82.63±0.26
Informer	54.88±0.61	57.00±0.72	59.02±0.74	62.90±0.41	69.84±1.89	75.66±0.82
SGCN	74.80±0.71	76.61±0.37	80.27±0.51	82.11±0.61	82.63±0.33	89.76±0.68
Graph-Transformer	68.62±1.12	70.93±0.28	72.91±0.62	75.07±0.59	79.61±0.84	80.39±0.45
SpGAT	51.34±1.38	56.23±2.34	61.55±0.76	64.31±0.45	66.67±0.38	69.70±1.08
本文所提方法	**92.79±0.19**	**94.67±0.15**	**96.77±0.03**	**97.56±0.06**	**97.93±0.02**	**98.59±0.02**

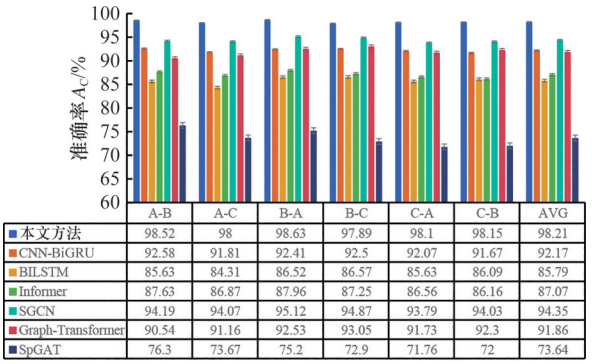


图 13　噪声环境下不同负载的结果对比

Fig.13　Comparison of results under different loads in a noisy environment

4　结论

1)针对传统图神经网络在处理非欧几里得结构时多维信息融合能力不足、动态特征提取能力弱的问题，通过引入WGT分支与MFCN分支实现了空间结构与局部多尺度特征的自适应融合、图结构局部与全局多尺度特征的协同提取。

2)改进双分支输出通道简单拼接融合，通过NL-CCA节点级跨通道特征融合模块实现了跨分支特征的自适应关联，进一步提高了多层级特征交互效率，减少时空特征流动过程中的信息损失。

3)引入卷积门控线性单元(CGLU)增强的频域特征处理层(CFAM)，实现细粒度特征选择与关键信息强化，弥补传统卷积网络对于FFT频域特征提取能力的不足。

4)构建多域混合交互融合注意层，并结合双向门控循环单元(BiGRU)捕获长距离时序依赖，实现多域特征的动态交互与自适应加权融合。

5)实验结果表明，在多种公开轴承数据集上均取得了显著优于传统深度学习模型的分类精度与稳定性，尤其在低信噪比和跨工况场景下展现出更强的抗干扰能力与迁移适应性。

未来工作将从以下方面展开：①对模型压缩、参数修建、量化等方法来降低计算开销，使其更适合在嵌入式终端上部署和运行。②探索基于图对比学习与跨域自监督机制的无标注特征迁移，推动该方法在实际工业环境中的落地应用。

参考文献：

[1] HUANG Delin, SU Xiangdong, YANG Jinghui, et al. An Improved Dual-channel CNN-BILSTM Fusion Attention Model for Fault Diagnosis of Aero-engine Bearings[J]. Measurement, 2025, 253: 117761.

[2] REN Zelin, JIANG Yuchen, YANG Xuebing, et al. Learnable Faster Kernel-PCA for Nonlinear Fault Detection: Deep Autoencoder-based Realization[J]. Journal of Industrial Information Integration, 2024, 40: 100622.

[3] LUO Yijing, FENG Chen, ZHANG Yuquan, et al. Vibration Fault Diagnosis of Hydroelectric Generator Sets Based on Multi-dimensional Feature Extraction and CNN-BiLSTM[J]. Measurement, 2026, 264: 120034.

[4] 邹筱瑜，孙国庆，王忠宾，等. 基于时频融合深度网络的矿用钻机轴承故障诊断[J]. 中国机械工程，2024，35(8)：1405-1413.
ZOU Xiaoyu, SUN Guoqing, WANG Zhongbin, et al. Bearing Fault Diagnosis of Mining Drilling Rig with Time-frequency-fused Deep Network[J]. China Mechanical Engineering, 2024, 35(8): 1405-1413.

[5] LEI Chunli, WANG Lu, ZHANG Qiyue, et al. Rolling Bearing Fault Diagnosis Method Based on DSICNN under Strong Noise Background[J]. Nondestructive Testing and Evaluation, 2025, 40(12): 5804-5820.

[6] LIANG Haopeng, ZHAO Xiaoqiang. Rolling Bearing Fault Diagnosis Based on One-dimensional Dilated Convolution Network with Residual Connection[J]. IEEE Access, 2021, 9: 31078-31091.

[7] YANG Aoyu, ZHONG Zhidan, ZHAO Yao, et al. Rolling Bearing Fault Diagnosis Method Based on Dynamic Convolution Capsule Network[J]. Journal of Physics: Conference Series, 2023, 2419(1): 012091.

[8] 阴艳超，施成娟，邹朝普，等. 基于深度时间卷积神经网络与迁移学习的流程制造工艺过程质量时序关联预测[J]. 中国机械工程，2023，34(14)：1659-1671.
YIN Yanchao, SHI Chengjuan, ZOU Chaopu, et al. Time-series Correlation Prediction of Quality in Process Production Processes Based on Deep TCN and Transfer Learning[J]. China Mechanical Engineering, 2023, 34(14): 1659-1671.

[9] LI Tianfu, ZHOU Zheng, LI Sinan, et al. The Emerging Graph Neural Networks for Intelligent Fault Diagnostics and Prognostics: a Guideline and a Benchmark Study[J].

Mechanical Systems and Signal Processing, 2022, 168: 108653.
[10] 宋李俊，刘松林，辛玉，等. 基于轴承退化状态评估和改进图注意力双向门控循环单元网络的轴承剩余寿命预测[J]. 中国机械工程，2025，36(7)：1562-1572.
SONG Lijun, LIU Songlin, XIN Yu, et al. Residual Life Prediction for Bearings Based on Bearing Degradation State Assessment and IGAT-BiGRU Network[J]. China Mechanical Engineering, 2025, 36(7): 1562-1572.
[11] ZUO Dong, TANG Tang, CHEN Ming. Rolling Bearing Fault Diagnosis Based on Multi-scale Weighted Visibility Graph and Multi-channel Graph Convolution Network[J]. Measurement Science and Technology, 2023, 34(11): 115019.
[12] FAN Xiaoxuan, DUAN Lixiang, ZHANG Na. A Multi-scale Graph-guided Dynamic Enhanced Alignment Network for Mechanical Fault Diagnosis Considering Domain Shift and Data Imbalance[J]. Neurocomputing, 2025, 625: 129546.
[13] WEN Haoran, GUO Wei, LI Xiang. A Novel Deep Clustering Network Using Multi-representation Autoencoder and Adversarial Learning for Large Cross-domain Fault Diagnosis of Rolling Bearings[J]. Expert Systems with Applications, 2023, 225: 120066.
[14] WANG Zhe, WU Zhiying, LI Xingqiu, et al. Attention-aware Temporal-Spatial Graph Neural Network with Multi-sensor Information Fusion for Fault Diagnosis[J]. Knowledge-Based Systems, 2023, 278: 110891.
[15] 吴百礼. 基于图卷积神经网络的轴承故障诊断与剩余寿命预测方法研究[D]. 广州：华南理工大学，2023.
WU Baili. Research on Bearing Fault Diagnosis and Remaining Life Prediction Method Based on Graph Convolutional Neural Network[D]. Guangzhou: South China University of Technology, 2023.
[16] WANG Jinhua, CAO Wenbao, CAO Jie, et al. Fault Diagnosis Method Based on Multi-scale Adaptive Supergraph Convolutional Neural Networks[J]. Applied Acoustics, 2025, 240: 110889.
[17] NI Peihao, ZHANG Yuanyuan, XIONG Xiaoyun, et al. MSGAFN: Multi-scale Graph Attention Fusion Network for Machine Fault Diagnosis[J]. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2024, 238(15): 7894-7907.
[18] 张永宏，许鑫豪，尹贺峰，等. MFSF-DETR：一种基于多尺度特征移位融合的PCB缺陷与元件检测算法[J]. 仪器仪表学报，2025，46(8)：266-285.
ZHANG Yonghong, XU Xinhao, YIN Hefeng, et al. MFSF-DETR: a PCB Defect and Component Detection Algorithm Based on Multi-scale Feature Shift Fusion[J]. Chinese Journal of Scientific Instrument, 2025, 46(8): 266-285.
[19] XU Ziwei, LI Yanfeng, HUANG Hongzhong, et al. A Novel Method Based on CNN-BiGRU and AM Model for Bearing Fault Diagnosis[J]. Journal of Mechanical Science and Technology, 2024, 38(7): 3361-3369.
[20] ZHANG Xin, YANG Kaiyue, ZHENG Liaomo. Transformer Fault Diagnosis Method Based on TimesNet and Informer[J]. Actuators, 2024, 13(2): 74.
[21] HUANG Zhenhua, ZHOU Wenhao, LI Kunhao, et al. SGCN: a Scalable Graph Convolutional Network with Graph-shaped Kernels and Multi-channels[J]. Knowledge-Based Systems, 2023, 279: 110923.
[22] CHANG Heng, RONG Yu, XU Tingyang, et al. Spectral Graph Attention Network with Fast Eigen-approximation[C]//Proceedings of the 30th ACM International Conference on Information & Knowledge Management. Online, 2021: 2905-2909.

（编辑　王旻玥）

作者简介：姜佩贺，男，1988年生，博士、副教授。研究方向为智能信号处理技术、工业设备预测性能维护及诊断。E-mail：jiangpeihe@ytu.edu.cn。王晨旭*（通信作者），男，1977年生，教授、博士研究生导师。研究方向为智能感知与无人系统。E-mail：wangchenxu@hit.edu.cn。

本文引用格式：
姜佩贺，刘银河，王晨旭. MBGA-FuseNet：双分支图网络与多域自适应融合的轴承故障诊断框架[J]. 中国机械工程，2026，37(9)：2166-2176.
JIANG Peihe, LIU Yinhe, WANG Chenxu. MBGA-FuseNet: Bearing Fault Diagnosis Framework Based on Dual-branch Graph Network and Multi-domain Adaptive Fusion[J]. China Mechanical Engineering, 2026, 37(9): 2166-2176.

（上接第2152页）

[14] 刘园. 隐性故障影响下的数控机床元动作单元多状态可靠性控制[D]. 西安：西安科技大学，2022.
LIU Yuan. Multi-state Reliability Control of Meta-action unit of Numerical Control Machine tool under the Influence of Recessive Fault[D]. Xi'an: Xi'an University of Science and Technology, 2022.
[15] YI Lingzhi, LONG Jiao, WANG Yahui, et al. Stator-rotor Fault Diagnosis of Induction Motor Based on Time-frequency Domain Feature Extraction[J]. Metrology and Measurement Systems, 2023: 773-790.
[16] 伍小莉. 不平衡—不对中—支座松动耦合故障下转子系统的动力学特性分析[D]. 成都：西南交通大学，2019.
WU Xiaoli. Dynamic Characteristics Analysis on Rotor System with Unbanlance-misalignment-pedestal Looseness Coupling Faults[D]. Chengdu: Southwest Jiaotong University, 2019.
[17] WANG Chen, CAI Chengming, DU Yanpeng, et al. Fault Diagnosis of Transformer Winding Looseness Based on Vibration Signal and GOA-KELM Model[J]. Journal of Physics: Conference Series, 2022, 2290(1): 012089.
[18] CHENG Li, LI Ning, CHEN Xuefeng, et al. The Influence of Crack Breathing and Imbalance Orientation Angle on the Characteristics of the Critical Speed of a Cracked Rotor[J]. Journal of Sound and Vibration, 2011, 330(9): 2031-2048.

（编辑　王旻玥）

作者简介：葛红玉*（通信作者），女，1982年生，副教授。研究方向为智能制造装备可靠性。E-mail：ghy-xkd@xust.edu.cn。

本文引用格式：
葛红玉，孙佳瑞，李艺，等. 机械元动作单元多故障耦合的动力学建模与特性分析[J]. 中国机械工程，2026，37(9)：2141-2152.
GE Hongyu, SUN Jiarui, LI Yi, et al. Dynamic Modeling and Characteristic Analyses of Multi-Fault Coupling in Mechanical Meta-action Unit[J]. China Mechanical Engineering, 2026, 37(9): 2141-2152.