

DOI: 10.19650/j.cnki.cjsi.J2514025

数字孪生驱动小波注意力迁移网络的 齿轮箱故障诊断方法*

朱朋, 邓蕾, 汤宝平, 张小龙, 刘永刚

(重庆大学高端装备机械传动全国重点实验室 重庆 400044)

摘要:针对仿真数据驱动的故障诊断方法存在动力学模型响应与实测数据分布差异较大、模型泛化性差的问题,提出了一种数字孪生驱动小波注意力迁移网络的齿轮箱故障诊断方法。首先,基于集中参数法建立了齿轮传动系统虚拟模型,实现物理实体的映射,并采用实测正常数据对虚拟模型中关键敏感参数进行优化修正,进而融合齿轮故障失效机理模型生成丰富的孪生故障数据集。其次,设计了离散小波注意力特征提取网络模型,该模型的设计融合了离散小波变换的多尺度信号分解和通道注意力机制动态聚焦强相关故障特征的特点,能在小波域层面上有效地提取孪生数据与实测数据的域不变故障特征。然后,考虑了孪生与实测数据的边缘分布及条件分布差异,结合最大均值差异和动态局部最大均值差异提出了联合子域自适应准则,匹配孪生与实测数据的联合分布差异,实现齿轮箱孪生模型向真实物理实体的迁移故障诊断。最后,在多级平行齿轮箱实验台上对所提方法进行试验验证,试验结果表明,所提方法在所有迁移任务下均获得了较优的诊断结果,平均分类精度可达98.10%,能够在含标签的高质量故障数据稀缺条件下有效地实现孪生数据向实测数据的迁移故障诊断。

关键词: 数字孪生;小波注意力网络;联合子域自适应;数据稀缺;齿轮箱故障诊断

中图分类号: TH17 **文献标识码:** A **国家标准学科分类代码:** 510.40

Digital twin driven wavelet attention transfer network for gearbox fault diagnosis

Zhu Peng, Deng Lei, Tang Baoping, Zhang Xiaolong, Liu Yonggang

(State Key Laboratory of Mechanical Transmission for Advanced Equipment, Chongqing University, Chongqing 400044, China)

Abstract: Aiming at the problems of large discrepancies between dynamic model response and the distribution of measured data, as well as the poor generalization in fault diagnosis methods driven by simulation data, a gearbox fault diagnosis method based on digital twin driven wavelet attention transfer network is proposed. Firstly, a virtual model of the gear transmission system was built using the lumped parameter method to map the physical system. Key parameters were optimized with measured normal data, and fault mechanism models were integrated to generate a rich twin fault dataset. Secondly, a discrete wavelet attention-based feature extraction network was designed, combining the multi-scale signal decomposition capability of discrete wavelet transform with the channel attention mechanism that dynamically focuses on strongly correlated fault features. This model effectively extracts domain-invariant fault features from both twin and measured data in the wavelet domain. Then, to address differences in marginal and conditional distributions between twin and measured data, a joint subdomain adaptation criterion was proposed by combining maximum mean discrepancy (MMD) and dynamic local MMD. This criterion measures the joint feature distribution discrepancy between the two domains, enabling the transfer of the gear twin model to real-world fault diagnosis. Finally, the proposed method was experimentally validated on a multi-stage parallel gearbox test bench. Results showed that it achieved superior diagnostic performance across all transfer tasks, with an average classification accuracy of 98.10%. The method effectively enables fault diagnosis transfer from twin data to measured data under conditions of limited labeled high-quality fault data.

Keywords: digital twin; wavelet attention network; joint subdomain adaptation; data scarcity; gearbox fault diagnosis

收稿日期:2025-05-11 Received Date: 2025-05-11

* 基金项目:国家自然科学基金创新研究群体(T2421001)、国家自然科学基金(52275087)、山西省创新生态服务支撑专项-创新平台基地建设专项(202404010911010Z)、高端装备机械传动全国重点实验室自主研究课题(SKLM-TZZK-2024Z07)项目资助

0 引 言

齿轮箱作为机械传动系统的重要组成部分,具有传动比大、效率高等优点,被广泛应用于航空航天、风力发电等行业。长期运行在变速、重载等恶劣工况环境下,齿轮箱中齿轮不可避免地会发生磨损、裂纹或缺齿等故障。一旦齿轮发生故障,轻者导致齿轮箱振动加剧,性能下降,严重则会导致重大事故发生。因此,对齿轮箱开展故障诊断研究具有重要意义^[1]。

近年来,基于深度学习的齿轮箱智能故障诊断方法因其具有强大的特征提取能力,受到国内外众多学者的广泛关注^[2-3]。这类方法获得令人满意的诊断精度的前提是能获取目标齿轮箱大量的含标签的高质量数据集^[4]。然而,在实际工业场景中,齿轮箱服役工况多变,不同工况下数据特征分布存在差异。此外,齿轮箱绝大部分时间是在正常状态下运行,含标签的高质量故障样本获取费时费力,难以对建立的深度学习模型进行有效地训练,导致故障诊断方法准确率较低。

得益于物联网、大数据、人工智能等新一代信息通信技术的发展,使得高保真建模技术发展到了一个新阶段,并由此产生了数字孪生技术^[5],该技术的出现为解决上述故障诊断问题提供了一种有效的途径。夏景演等^[6]通过 ADAMS 软件建立了齿轮箱的虚拟模型,并采用生成对抗网络进行孪生数据的特征增强,有效地实现了小样本下齿轮箱的故障诊断。Jiang 等^[7]采用振动响应机理模型来获取不同故障形式的齿轮仿真数据集,然后结合设计的增强无监督领域自适应策略实现仿真向实测数据的迁移故障诊断。针对大多数刚出厂的旋转机械故障数据缺乏,难以进行智能故障诊断,Yu 等^[8]提出了动力学模型嵌入的智能机械故障诊断框架,并在平行齿轮箱上进行了所提方法的有效性验证。Yan 等^[9]通过动力学仿真模型来获取仿真数据集,并采用子域自适应机制和边缘感知正则化技术来缩小仿真与实测数据的差异,实现了数据不平衡下齿轮箱故障诊断。Feng 等^[10]提出了一种数字孪生驱动的智能健康管理方法,有效地实现了不同齿轮磨损程度下的监测与评估。上述研究表明,数字孪生建模与迁移学习技术的融合为解决样本稀缺条件下的齿轮箱故障诊断难题提供了一个行之有效的方案。但是,仍然存在一些瓶颈:基于卷积神经网络及其变体构建的故障特征提取模型受限于深度学习网络的“黑盒”属性,存在可解释性差、孪生与实测数据的迁移故障特征提取不充分的缺点;大多数方法中孪生数据故障特征向实测数据的迁移主要通过适配不同域的边缘分布或条件分布差异来实现,然而,由于孪生与实测数据存在较大的特征分布差异,采用单一的领域适配方法迁移能力有限,不

能较好地促进孪生向实测数据的迁移故障诊断。

为了解决上述问题,提出了一种数字孪生驱动小波注意力迁移网络(wavelet attention transfer network, WATN)的齿轮箱故障诊断方法。首先,一个采用实测正常数据修正后的高保真齿轮箱数字孪生模型被建立,用于生成丰富的孪生故障数据集;其次,搭建了小波注意力网络,在小波域层面对孪生与实测数据的迁移故障特征进行有效地提取;最后,设计了联合子域自适应度量准则来匹配孪生与实测数据的联合分布差异,促进孪生体向物理实体的迁移故障诊断,解决故障数据稀缺条件下齿轮箱诊断难题。

1 齿轮箱数字孪生模型构建

高保真齿轮箱数字孪生模型构建主要由虚拟空间中的实体建模和虚实交互两个部分组成,相关流程如图 1 所示。在虚拟空间中,结合齿轮传动系统动力学模型和失效机理模型,进行物理实体的虚拟模型构建。在虚实交互方面,采用多目标寻优算法求解建立的目标函数,实现虚拟模型的优化更新,提高建模保真度。

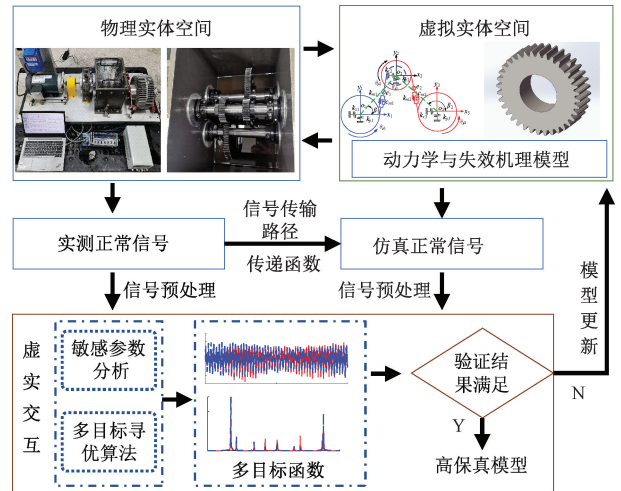


图 1 齿轮箱数字孪生体构建流程

Fig. 1 Gearbox digital twin construction process

1.1 两级平行齿轮箱虚拟实体模型建立

齿轮箱物理实体主要由 2 对直齿轮、3 根转轴及 3 对轴承组成。作为物理实体模型的有效映射,基于动力学建模的虚拟实体模型构建是数字孪生的核心部分。集中参数法具有计算效率较高、求解结果可靠等特点,被广泛应用于齿轮箱的动力学建模中。建立的两级直齿圆柱齿轮传动系统动力学模型如图 2 所示。

建模的假设条件为:每个齿轮为刚性轮;当一对齿轮对啮合时,通过沿共切线方向设置的弹簧阻尼元件连接;转轴为忽略质量的短刚性轴;滚动轴承等效为具有线性

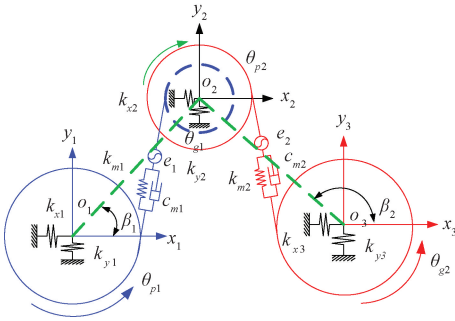


图2 直齿圆柱齿轮传动系统动力学模型

Fig. 2 Dynamic model of spur gear transmission system

弹簧和粘性阻尼值的元件。通过分别在主、从动齿轮应用拉格朗日方程,可以获得两级圆柱齿轮传动系统十二自由度的运动微分方程,如式(1)所示。

$$\begin{cases}
 m_{p1}\ddot{x}_1 = -k_{x1}x_1 - c_{x1}\dot{x}_1 + F_{pg1}\sin(\alpha_0 + \beta_1) \\
 m_{p1}\ddot{y}_1 = -k_{y1}y_1 - c_{y1}\dot{y}_1 - F_{pg1}\cos(\alpha_0 + \beta_1) \\
 (m_{g1} + m_{p2})\ddot{x}_2 = -k_{x2}x_2 - c_{x2}\dot{x}_2 - \\
 \quad F_{pg1}\sin(\alpha_0 + \beta_1) - F_{pg2}\sin(\alpha_0 + \beta_2) \\
 (m_{g1} + m_{p2})\ddot{y}_2 = -k_{y2}y_2 - c_{y2}\dot{y}_2 + \\
 \quad F_{pg1}\cos(\alpha_0 + \beta_1) + F_{pg2}\cos(\alpha_0 + \beta_2) \\
 m_{g2}\ddot{x}_3 = -k_{x3}x_3 - c_{x3}\dot{x}_3 + F_{pg2}\sin(\alpha_0 + \beta_2) \\
 m_{g2}\ddot{y}_3 = -k_{y3}y_3 - c_{y3}\dot{y}_3 - F_{pg2}\cos(\alpha_0 + \beta_2) \\
 I_M\dot{\theta}_M = T_M - k_{\theta 1}(\theta_M - \theta_{p1}) - c_{\theta 1}(\dot{\theta}_M - \dot{\theta}_{p1}) \\
 I_{p1}\ddot{\theta}_{p1} = k_{\theta 1}(\theta_M - \theta_{p1}) + c_{\theta 1}(\dot{\theta}_M - \dot{\theta}_{p1}) - R_{bp1}F_{pg1} \\
 I_{g1}\ddot{\theta}_{g1} = -k_{\theta 2}(\theta_{g1} - \theta_{p2}) - c_{\theta 2}(\dot{\theta}_{g1} - \dot{\theta}_{p2}) + R_{bg1}F_{pg1} \\
 I_{p2}\ddot{\theta}_{p2} = k_{\theta 2}(\theta_{g1} - \theta_{p2}) + c_{\theta 2}(\dot{\theta}_{g1} - \dot{\theta}_{p2}) - R_{bp2}F_{pg2} \\
 I_{g2}\ddot{\theta}_{g2} = -k_{\theta 3}(\theta_{g2} - \theta_L) - c_{\theta 3}(\dot{\theta}_{g2} - \dot{\theta}_L) + R_{bg2}F_{pg2} \\
 I_L\ddot{\theta}_L = k_{\theta 3}(\theta_{g2} - \theta_L) + c_{\theta 3}(\dot{\theta}_{g2} - \dot{\theta}_L) - T_L
 \end{cases} \quad (1)$$

式中: x_i 和 y_i ($i = 1, 2, 3$) 分别为输入轴、中间轴和输出轴沿 x 和 y 方向的位移; k_{xi} 、 c_{xi} 和 c_{yi} ($i = 1, 2, 3$) 分别为对应轴上的支撑轴承在 x 和 y 方向上的刚度和阻尼; m_{p1} 、 m_{g1} 、 m_{p2} 和 m_{g2} 分别表示第 1、2 级啮合的主动齿轮、从动齿轮的质量; 同理, θ_{p1} 、 θ_{g1} 、 θ_{p2} 和 θ_{g2} 分别表示相应齿轮的旋转角位移; θ_M 和 θ_L 分别为电机和负载的质量惯性矩; I_{p1} 、 I_{g1} 、 I_{p2} 、 I_{g2} 、 I_M 和 I_L 分别表示为对应齿轮、电机和负载的转动惯量; $c_{\theta i}$ 和 $k_{\theta i}$ ($i = 1, 2, 3$) 分别表示对应轴的扭转阻尼和刚度; T_M 和 T_L 分别表示电机驱动力矩与负载力矩; α_0 、 β_1 和 β_2 分别表示齿轮副的压力角、第 1 对齿轮轴位角、第 2 对齿轮轴位角; F_{pg1} 和 F_{pg2} 分别表示第 1、2 对齿轮的动态啮合力, 考虑齿侧间隙的动态啮合力表示为:

$$F_{pg1} = k_m f_1(t) + c_{m1} \delta_1(t) \quad (2)$$

$$F_{pg2} = k_m f_2(t) + c_{m2} \delta_2(t) \quad (3)$$

式中: k_{mi} 和 c_{mi} ($i = 1, 2$) 分别表示第 i 对齿轮的时变啮合刚度和时变啮合阻尼, 两者的关系 $c_{mi} = 2\zeta \sqrt{k_{mi} \frac{m_{pi} m_{gi}}{m_{pi} + m_{gi}}}$, ζ 为阻尼比, 初始取值为 0.05, k_{mi} 详细计算过程见文献[11]; $\delta_1(t)$ 和 $\delta_2(t)$ 分别表示第 1、2 对齿轮啮合沿压力线上的弯曲变形位移; $f(t)$ 表示尺侧间隙非线性函数, 如式(4)所示。

$$f(t) = \begin{cases} \delta(t) - b_l, & \delta(t) > b_l \\ 0, & |\delta(t)| \leq b_l \\ \delta(t) + b_l, & \delta(t) < -b_l \end{cases} \quad (4)$$

式中: b_l 表示为单边齿侧间隙, 取值为 0.1 mm。第 1、2 对齿轮相对位移可以表示为如式(5)和(6)所示。

$$\delta_1(t) = (R_{bp1}\theta_{p1} - R_{bg1}\theta_{g1}) + (x_2 - x_1)\sin(\alpha_0 + \beta_1) + (y_1 - y_2)\cos(\alpha_0 + \beta_1) - e_1(t) \quad (5)$$

$$\delta_2(t) = (R_{bp2}\theta_{p2} - R_{bg2}\theta_{g2}) + (x_2 - x_3)\sin(\alpha_0 + \beta_2) + (y_3 - y_2)\cos(\alpha_0 + \beta_2) - e_2(t) \quad (6)$$

式中: R_{bpi} 和 R_{bgi} ($i = 1, 2$) 是第 i 对齿轮的主动和从动轮的基圆半径; $e(t)$ 表示传递误差函数, 主要由偏心和齿廓误差组合波形的表达式组成, 如式(7)所示。

$$\begin{cases}
 e(t) = e_e + e_p \\
 e_e = e_{A1}\sin(2\pi f_p t + \varphi_p) + e_{A2}\sin(2\pi f_g t + \varphi_g) \\
 e_p = \sum_{i=1}^N A_i \sin(2\pi i f_m t + \varphi)
 \end{cases} \quad (7)$$

式中: e_{A1} 和 e_{A2} 分别为主动、从动齿轮的偏心率; f_p 和 f_g 分别表示为主动、从动齿轮的所在转轴的转频; f_m 为啮合频率; φ_p 、 φ_g 和 φ 分别表示为相应的相位差; A 表示为齿廓误差的振幅。

上述建立的动力学模型只能预测正常状态下齿轮箱振动响应, 要想获得其他失效形式下振动数据, 还需结合失效机理模型。磨损、缺齿、断齿等典型齿轮故障失效模型已得到众多学者研究^[12-13], 主要是考虑不同故障形式下齿轮时变啮合刚度、时变啮合阻尼、齿侧间隙等重要影响因素变化来进行不同失效机理规律的特征。将不同故障形式下失效机理模型嵌入到建立的齿轮传动系统动力学模型, 获得具有故障预测功能的齿轮箱虚拟实体模型。

忽略故障齿轮与传感器之间传递路径的影响可能会导致仿真信号与实测信号存在较大差异。为了从实测正常信号的功率谱中提取齿轮传动系统的传递函数, 采用自适应杂波分离算法进行频谱背景估计; 然后, 对估计的频谱背景应用最小相位假设, 可获得估计的传递函数, 相关理论原理见文献[14]。将估计的传递函数 $tf(t)$ 与虚拟实体模型的原始仿真信号 $x(t)$ 进行卷积, 即可获取调整后的仿真信号, 如式(8)所示。

$$x_y[t] = x(t) * tf(t) = \text{IFFT}[x(f)tf(f)] \quad (8)$$

式中: t 和 f 分别表示时、频域; $*$ 表示卷积运算;

IFFT[·] 为逆傅里叶变换。

1.2 物理实体与虚拟实体模型交互

动力学建模过程中初始参数的设置更多依赖于经验,然而,实际装配工艺和服役工况会对齿轮箱性能产生影响,降低虚拟模型保真度^[10]。因此,为了提高孪生与实测数据的一致性,以实测正常信号为基准,建立了基于余弦相似度的物理实体和虚拟实体模型交互机制。余弦相似度被用于评估孪生信号与实测信号的差异,能有效反映孪生模型的保真度高,其数学表达式如式(9)所示。

$$R(X_D, X_R) = \frac{X_D \cdot X_R}{\|X_D\| \times \|X_R\|} \quad (9)$$

式中: X_D 和 X_R 分别表示同一工况正常状态下的孪生与实测信号。

余弦相似度值越大,表明孪生仿真信号与实测信号越相似。基于此,建立的目标函数如式(10)所示。

$$\begin{cases} f_1(\mathbf{P}) = 1 - R(X_D^t, X_R^t) \\ f_2(\mathbf{P}) = 1 - R(X_D^f, X_R^f) \end{cases} \quad (10)$$

$$\begin{cases} \min: [f_1(\mathbf{P}), f_2(\mathbf{P})] \\ \text{s.t.}: \forall p_i \in \mathbf{P}, L(p_i) \leq p_i \leq U(p_i) \end{cases} \quad (11)$$

式中: \mathbf{P} 表示待优化的参数向量; $L(\cdot)$ 和 $U(\cdot)$ 分别表示左、右取值区间;上角标 t 和 f 分别表示时域波形和傅里叶变换后的频谱。

在求解建立的目标函数最优解时,引入了多目标蚱蜢优化算法^[15]。该算法具有求解效率高、收敛速度快等优点,是一种元启发式仿生优化算法。指标 f_1 和 f_2 分别表征孪生与实测信号的时域波形相似度和频域中啮合频率及其谐波的幅值相似度。

2 提出的小波注意力迁移故障诊断网络

2.1 物理实体与虚拟实体模型交互

经典的智能故障诊断方法受限于卷积神经网络及其变体的“黑盒”属性,特征提取过程可解释性较差,制约了模型性能的进一步提升。小波变换作为一种多分辨率信号分析方法,不仅具有优异的非平稳信号分解能力,还具有完备的理论支撑。因此,将离散小波变换(discrete wavelet transform, DWT)嵌入到卷积神经网络结构中进行所提方法的特征提取网络结构的设计,增强模型的局部可解释性,进一步提高故障特征提取能力。DWT 采用高通和低通滤波器组将信号分解成高频系数和低频系数近似值,如式(12)所示。

$$\begin{cases} x_{j,L}[n] = \sum_{k=0}^{K-1} x_{j-1,L}[2n-k]g[k] \\ x_{j,H}[n] = \sum_{k=0}^{K-1} x_{j-1,L}[2n-k]h[k] \end{cases} \quad (12)$$

式中: $g[n]$ 和 $h[n]$ 分别表示低通和高通滤波器; x_L 和 x_H 分别为原始信号的低频和低频系数近似值。

结合 DWT 和通道注意力机制,提出的小波注意力机制(wavelet attention mechanism, WAM)模块结构示意图如图 3 所示。首先,假设输入的特征图 $X_{in} \in R^{C \times L}$ 的通道为 C 、长度为 L ,根据给定的小波基函数 haar^[16],通过式(12)对每个通道的特征信号进行 DWT 分解,可获得高频和低频分量的特征图,如图 3 所示。由于小波分解尺度为一级,因此高、低频分量特征图的长度变为 $L/2$ 。为了使提出的网络模型充分学习信号的低频和高频特征,将高频和低频特征图按照通道方向进行拼接,获得分解后的融合特征图 $X_{mid} \in R^{2C \times L/2}$ 。

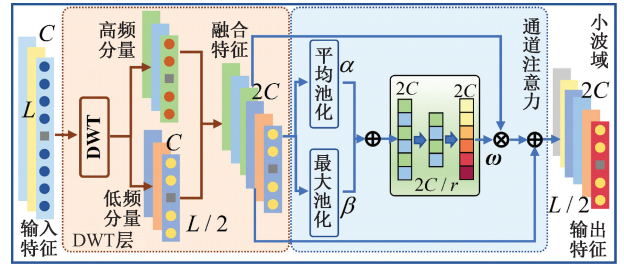


图 3 WAM 模块结构

Fig. 3 Diagram of the WAM module architecture

经 DWT 层分解后获得的融合特征不可避免地存在冗余信息,影响网络模型的性能。为了提高所提网络模型动态捕捉小波域特征有用频段信息、抑制无效特征的能力,通道注意力机制被引入到 DWT 层后进行强相关故障特征学习,如图 3 所示。首先,将融合特征图 X_{mid} 采用平均池化和最大池化的操作来进行局部和全局信息聚合,并引入可学习参数 α 和 β 来平衡局部和全部信息,数学表达式为:

$$\mathbf{d} = \alpha \cdot f_{\text{Avg}}(X_{mid}) + \beta \cdot f_{\text{Max}}(X_{mid}) \quad (13)$$

式中: \mathbf{d} 为通道向量; $f_{\text{Avg}}(\cdot)$ 和 $f_{\text{Max}}(\cdot)$ 分别表示平均池化和最大池化运算。

其次,采用两个全连接(fully connected, FC)层进行非线性变换,以学习到不同通道的重要性。其中,第 1 个全连接层(FC_1)将通道表示向量压缩为维度为 $(2C/r) \times 1$ 的隐藏层向量, r 为通道降维率,取值为 16,并采用线性整流函数(ReLU)作为非线性变换;第 2 个全连接层(FC_2)将隐藏层向量恢复为原始维度为 $2C \times 1$ 的权重向量,激活函数为 Sigmoid 函数。

$$\boldsymbol{\omega} = \text{Sigmoid}(FC_2(\text{ReLU}(FC_1(\mathbf{d})))) \quad (14)$$

式中: $\boldsymbol{\omega}$ 为通道特征的重要性权重。

最后,将融合特征图进行通道加权与残差连接运算,获得输出特征图 X_{out} , 即:

$$X_{out} = X_{mid} + \boldsymbol{\omega} \cdot X_{mid} \quad (15)$$

2.2 联合子域自适应

由于孪生数据与实测数据特征分布存在差异,导致直接采用孪生数据集训练的网络模型泛化能力差,难以直接应用于物理实体模型故障诊断中。因此,需采用基于领域自适应的度量策略来匹配孪生与实测数据的特征分布差异。最大均值差异(maximum mean discrepancy, MMD)作为一种广泛使用的域适应度量准则,数学表达式如式(16)所示。

$$\mathcal{L}_M = MMD(X^S, X^T) = \left\| \frac{1}{N_S} \sum_{i=1}^{N_S} \phi(X_i^T) - \frac{1}{N_T} \sum_{i=1}^{N_T} \phi(X_i^T) \right\|_H^2 \quad (16)$$

式中: X^S 和 X^T 分别表示源域与目标域数据; N_S 和 N_T 定义为相应的样本数据量; H 为再生核希尔伯特空间; $\phi: X^S, X^T \rightarrow H$ 是高斯核函数。

MMD 只考虑了不同域的边缘分布差异,忽略条件分布差异,迁移能力有限。针对该问题,考虑不同域的相同子类故障之间的相关性,Zhu 等^[17]提出了局部最大均值差异(local MMD, LMMD):

$$\mathcal{L}_L = LMMD(X^S, X^T) = \frac{1}{C_m} \sum_{c=1}^{C_m} \left\| \sum_{i=1}^{N_S} \omega_i^{sc} \phi(X_i^S) - \sum_{j=1}^{N_T} \omega_j^{tc} \phi(X_j^T) \right\|_H^2 \quad (17)$$

式中: C_m 表示故障类别数; ω_i^{sc} 和 ω_j^{tc} 分别表示源域与目标域样本属于第 c 类故障的权重;对于源域样本,使用真实标签进行概率计算,对于目标域标签,采用网络预测的伪标签进行计算。然而,直接使用网络预测的概率可能会由于伪标签预测的潜在不准确而导致子域适配能力的下降。为了解决该问题,基于信息熵理论,提出了动态局部领域自适应度量准则(dynamic LMMD, DLMMMD),如式(18)所示。

$$\mathcal{L}_{DL} = DLMMMD(X^S, X^T) = \frac{1}{C_m} \sum_{c=1}^{C_m} \left\| \sum_{i=1}^{N_S} \omega_i^{sc} \phi(X_i^S) - \sum_{j=1}^{N_T} h(X_j^T) \omega_j^{tc} \phi(X_j^T) \right\|_H^2 \quad (18)$$

$$h(X) = - \frac{1}{\sum_{c=1}^{C_m} p(\hat{y}^{tc}) \log p(\hat{y}^{tc})} \quad (19)$$

式中: $p(\cdot)$ 表示网络模型分类器预测的目标域伪标签 \hat{y} 的概率分布。当模型预测中出现不确定性,导致所有类别的概率分布相似时,其信息熵值就越高,信息熵倒数 $h(\cdot)$ 越小,模型被赋予更小的权重。

考虑孪生和实测数据的边缘及条件分布对齐,结合 MMD 和 DLMMMD 领域自适应策略,提出了联合子域自适应准则,数学表达式为:

$$\mathcal{L}_{JD} = \mu \mathcal{L}_M + (1 - \mu) \mathcal{L}_{DL} \quad (20)$$

式中: μ 表示动态平衡因子, $\mu = |\mathcal{L}_M| / (|\mathcal{L}_M| + |\mathcal{L}_{DL}|)$ 。

2.3 所提小波注意力迁移故障诊断网络模型

为了更好地提取孪生与实测数据的可迁移故障特征,将提出的 WAM 模块嵌入到卷积神经网络结构中,设计了齿轮箱故障诊断 WATN 网络。同时,分类层的最后一个全连接层被选择作为自适应层,采用提出的联合子域自适应准则进行孪生与实测数据特征差异度量,所提方法 WATN 的网络结构图如图 4 所示。

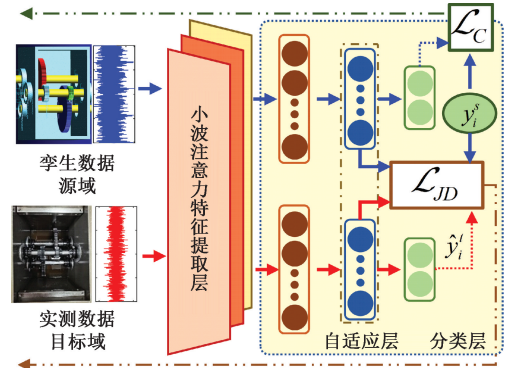


图4 提出的 WATN 结构

Fig. 4 Diagram of the proposed WATN structure

WATN 网络结构主要由提出的 WAM 模块交替嵌入卷积神经网络组成,具体参数说明如表 1 所示,其中每个卷积层后均进行了批归一化(batch normalization, BN)和线性整流函数(rectified linear unit, ReLU)激活操作,WAM 模块后的丢弃操作(Dropout)用于抑制网络训练过拟合。WATN 的分类层由两个全连接层组成,其中 ReLU 激活和丢弃操作应用在第 1 个全连接层后。

对于源域数据集的监督训练,交叉熵损失函数 L_{CE} 被用于计算其分类损失。

$$\mathcal{L}_C = \frac{1}{N_S} \sum_{i=1}^{N_S} L_{CE}(g(X_i^S), y_i^S) \quad (21)$$

式中: $g(X)$ 为训练网络模型的预测输出; y^S 为源域真实标签。

考虑源域分类损失、源域与目标域联合子域自适应损失,所提方法的总的优化目标函数为:

$$\mathcal{L}_T = \mathcal{L}_C + \xi \mathcal{L}_{JD} \quad (22)$$

式中: ξ 表示权重超参数。

3 齿轮箱故障诊断试验验证

3.1 物理齿轮箱试验台描述及工况设置

为了验证所提方法的有效性,采用实验室动力传动系统试验台进行试验验证。试验台主要由电机、速度控制器、一级行星齿轮箱、两级平行轴齿轮箱、磁粉制动器等组成,如图 5 所示。

表 1 所提模型 WATN 网络结构参数说明

Table 1 Description of the WATN network structure parameters of the proposed model

层类型	核尺寸	步长	其他运算	输出维度
输入				(4 096,1)
卷积层 1	3	1	BN+ReLU	(4 094,16)
WAM 层 1			Dropout (0.1)	(2 047,32)
卷积层 2	3	1	BN+ReLU	(2 045,32)
WAM 层 2			Dropout (0.1)	(1 023,64)
卷积层 3	3	1	BN+ReLU	(1 021,64)
WAM 层 3			Dropout (0.1)	(511,128)
卷积层 4	3	1	BN+ReLU	(509,128)
WAM 层 4			Dropout (0.1)	(255,256)
卷积层 5	3	1	BN+ReLU	(253,256)
自适应平均池化层				(1,256)
全连接层 1			ReLU+ Dropout (0.3)	128
全连接层 2				4

表 2 二级平行齿轮箱结构参数

Table 2 Two-stage parallel gearbox structural parameters

齿轮参数	主动轮 p1	从动轮 g1	主动轮 p2	从动轮 g2
模数/mm	1.5	1.5	1.5	1.5
齿数	100	29	36	90
压力角/(°)	20	20	20	20
齿宽/mm	15	15	15	15
基圆半径/mm	70.5	20.4	25.4	63.4
质量/kg	2.071	0.371	0.465	1.676
转动惯量/(kg·m ²)	4.2×10 ⁻³	4.0×10 ⁻⁵	9.6×10 ⁻⁵	2.8×10 ⁻⁵

表 3 不同工况参数设置

Table 3 Parameter settings for different working conditions

工况参数	工况 W1	工况 W2	工况 W3	工况 W4
转速/Hz	25	30	35	35
电压/V	8	4	4	0



图 6 不同故障模式齿轮图片

Fig. 6 Pictures of different gear failure modes

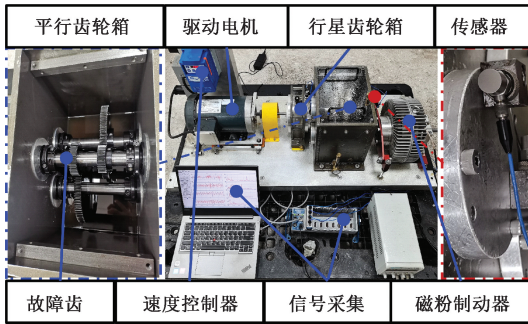


图 5 齿轮故障模拟试验装置

Fig. 5 Gear fault simulation test device

加速度传感器被安装在平行轴齿轮箱中间轴的轴承座的垂直方向上,信号采样频率为 50 kHz。正常及模拟故障的齿轮被安装在平行轴齿轮箱的中间轴上,故障齿轮齿数为 36,支撑传动轴的轴承型号为 ER-16K。此外,行星齿轮箱的减速比为 4.57。平行轴齿轮箱结构参数如表 2 所示。支撑轴承的刚度为 8.5×10^7 N/m,阻尼系数为 2.1×10^4 N·s/m。

通过改变速度控制器的输入转速和磁粉制动器的控制电压进行变工况环境模拟,总共模拟了 4 种变工况环境,如表 3 所示。

此外,除了正常运行状态(normal condition, NO),还进行了 3 种不同故障的模拟,即断齿(chipped fault, CF)、磨损(wear fault, WF)以及缺齿(missing fault, MF)故障,对应的图片如图 6 所示。

3.2 数字孪生模型验证及孪生数据仿真

采用龙格库塔法求解 3.1 节建立的两级平行齿轮传动系统动力学方程组,其中求解函数采用 MATLAB 软件中的 ODE45,时间步长设置为 5.0×10^{-6} 。然后,将求解得到的正常振动响应信号与传递函数进行卷积运算,获得考虑信号传输路径的正常仿真信号。再以实测正常信号作为基准,采用多目标蚱蜢优化算法同时优化传递误差函数的振幅及相位、输入轴转频、阻尼比系数等敏感参数,并求解敏感参数向量的最优值。为了量化评价孪生与实测数据的相似性,给出了 4 种工况下信号的时域、频域余弦相似度值,如表 4 所示。

表 4 不同工况下孪生与实测正常数据的余弦相似度
Table 4 Cosine similarity between twin and measured normal data under different working conditions

余弦相似度	工况 W1	工况 W2	工况 W3	工况 W4
时域	0.739 0	0.683 6	0.662 4	0.888 1
频域	0.884 0	0.768 4	0.838 5	0.739 6

由表4可知,孪生与实测数据的时、频域余弦相似度值分别在0.66和0.73以上,表明虚拟实体模型与物理实体模型存在较强的正相关性。为了更直观的比较孪生与实测数据的相似性,随机给出了W1和W3工况下的孪生模型正常仿真数据与对应滤波后的实测数据的时、频域对比图,如图7所示。

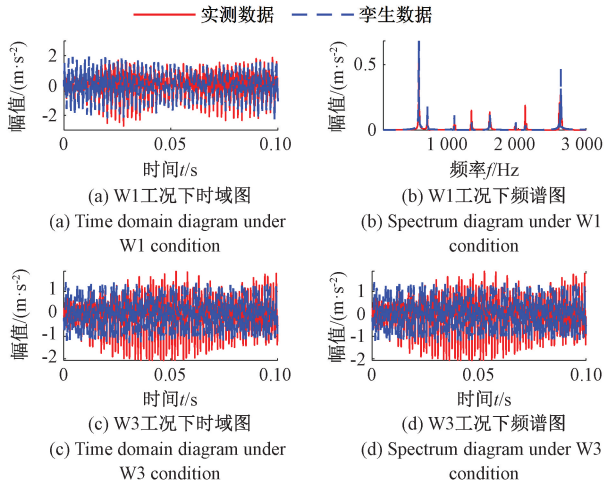


图7 不同工况下的孪生模型正常数据与实测数据比较
Fig. 7 Comparison of normal data and measured data of the twin model under different working conditions

由图7中不同工况的时域波形图可知,孪生与实测数据在振动幅值和周期角度方面较为一致,变化规律基本相同。在频域图中,可以清晰地发现齿对啮合的基频及其倍频一致性较好,有效地反映了齿对的啮合特性。以上分析进一步表明构建的动力学模型具有较好的保真度。

将高保真动力学模型分别与CF、WF、MF的齿轮故障机理模型进行融合,生成不同故障种类下的仿真数据集。随机选择W3工况下WF模式的孪生故障信号时域波形和频谱进行分析,如图8所示。

由图8(a)可以明显发现,故障时域信号存在周期性的冲击成分,间隔时间约为中间轴转频的倒数;图8(b)频谱图中除了啮合频率及其谐波频率,还可发现以故障齿轮旋转频率为间隔的啮合调制边频带。这些现象符合齿轮分布式故障模式的振动响应机理。因此,建立的数字孪生模型能有效地生成可靠的孪生故障数据,可用于智能诊断模型训练。

3.3 迁移故障诊断对比方法及实验设置

为了说明所提模型的有效性与优越性,离散小波神经网络(discrete wavelet neural network, DWNN)、离散小波注意力神经网络(discrete wavelet attention neural network, DWANN)、深度域自适应网络(deep domain confusion, DDC)、领域对抗神经网络(domain adversarial

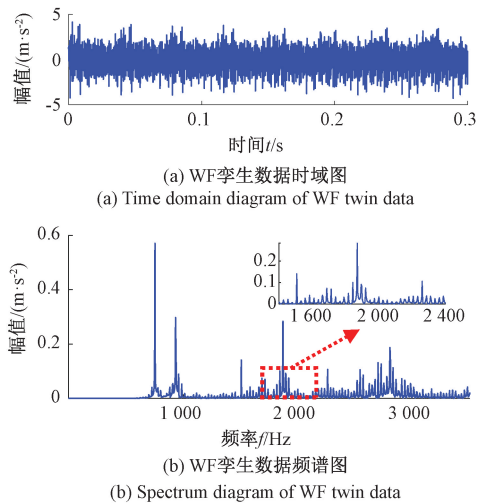


图8 W3工况下WF模式的孪生故障数据时频图
Fig. 8 Time-frequency diagram of twin fault data in WF mode under W3 condition

neural network, DANN)、深度子域自适应网络(deep subdomain adaptation network, DSAN)被选择进行对比实验验证。其中,DWANN网络结构与所提模型WATN主干结构完全一致,只是未进行领域自适应,用于说明孪生向实测数据迁移的必要性;DWNN是在DWANN网络结构的基础上去除通道注意力模块;DDC是在DWANN网络基础上,采用MMD进行孪生与实测差异的度量^[18];DANN采用领域对抗域自适应策略^[19]来缩小不同领域的分布差异;DSAN是将DDC网络中MMD替换为LMMD度量准则^[17],其余保持不变。

在实验数据源域设置方面,齿轮每种健康状态下含标签的孪生样本为200个,每个样本包含4096个点。目标域设置方面,每种健康状态包含300个不含标签的样本,其中200个样本用于训练,剩下的100个样本用于测试。为了降低模型训练的随机性影响,对每种故障诊断任务都重复试验7次,取平均值。所有模型训练迭代次数保持一致,最大迭代次数被设置为2000,批量大小被设置为32。采用Adam梯度优化方法,学习率设置为 1.0×10^{-4} ,正则化系数设置为 1.0×10^{-5} 。为抑制噪声对网络训练的影响,超参 ξ 设置与DANN一样,随着训练次数增加从0逐渐增加到1。

3.4 迁移诊断对比结果分析

各个模型在4种工况下的孪生向实测数据迁移诊断结果如表5所示。

由表5可知,所提模型WATN在4种迁移任务下均获得最高的分类精度,平均分类精度为98.10%,优于其他对比模型。未采用领域自适应的DWNN和DWANN模型分类性能较差,平均分类精度分别为43.36%和

表 5 各模型目标域测试集的分类精度
Table 5 Classification accuracy of each model's target domain test set (%)

方法	DWNN	DWANN	DDC	DANN	DSAN	WATN
W1	38.32	41.64	51.42	61.43	75.00	97.22
W2	25.00	28.39	67.57	62.20	75.00	95.42
W3	63.82	75.00	52.42	74.78	75.00	100.00
W4	46.32	47.28	99.22	95.46	99.75	99.77
平均精度	43.36	48.07	67.65	73.46	81.18	98.10

48.07%。一方面说明孪生故障数据向实测数据迁移的必要性,另一方面说明通道注意力机具有动态聚焦有用频段故障特征的能力,提高模型泛化性。在迁移学习对比模型中,只考虑边缘分布差异的 DDC 平均分类精度最低,为 67.65%。基于对抗策略的 DANN 模型在网络训练过程中存在收敛困难的风险,平均分类精度为 73.46%。DSAN 模型考虑了不同域的故障子类信息,迁移能力有所提高,平均分类精度为 81.18%,但是,直接使用网络预测的伪标签存在潜在的不确定性,导致子域适配能力的不足,迁移任务 W1 ~ W3 上平均分类精度仅达到 75.00%。所提模型 WATN 通过联合子域度量准则优化了孪生与实测数据的联合分布差异,迁移性能最优。分析结果证明了所提模型在数字孪生驱动的齿轮箱故障诊断中有效性和优越性。

为了进一步说明所提模型的优越性,采用 t 分布随机近邻嵌入 (t-distributed stochastic neighbor embedding, t-SNE) 技术对 W1 任务下不同方法学习到的特征进行可视化分析,如图 9 所示,其中“S”为源域特征,“T”为目标域特征。

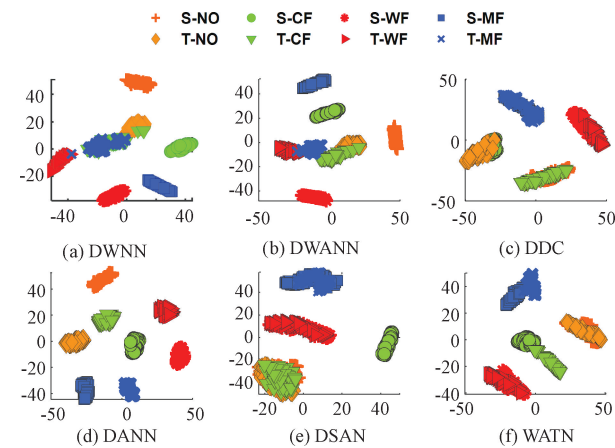


图 9 不同模型在 W1 任务中特征可视化图

Fig. 9 Feature visualization of different models under W1 task

由图 9 可知,所提方法 WATN 在不同域的相同子类故障特征分布较为紧凑,类间故障特征分布较为稀疏,特征聚类效果较好。其余对比模型存在不同程度的特征混叠、误分类现象,聚类效果不理想。

3.5 选择的小波基有效性分析

为了验证所提方法中 WAM 层小波基选择的有效性,随机选择 W1 工况进行实验验证分析。较为常用的小波基函数 db8、db16、dmev、sym8、coif8 被选择作为比较对象,不同小波基函数下所提方法 WATN 的平均分类精度结果如图 10 所示。

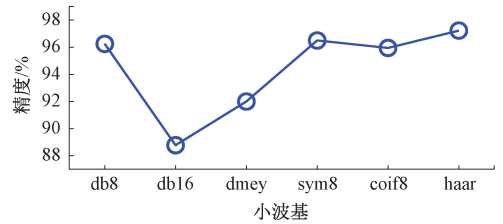


图 10 小波基有效性分析结果展示 (W1 工况)

Fig. 10 Results of wavelet basis validity analysis (W1 condition)

由图 10 可知,采用不同的小波基函数,WATN 获得了不同的诊断效果;db 系列小波函数受消失矩的影响较大,db16 小波基下 WATN 诊断效果较差;在所有对比的小波基函数中,采用具有正交性和对称性的 haar 小波基的 WATN 获取了最好的分类精度。以上分析证明了 WAM 层中小波基 haar 在提取齿轮箱振动信号迁移故障特征方面的有效性。

4 结 论

本研究提出了一种数字孪生驱动小波注意力迁移网络的齿轮箱故障诊断方法,旨在解决现有仿真数据驱动的故障诊断方法存在动力学模型响应与实测数据分布差异较大、模型泛化性差的问题。主要结论为:建立了一个基于实测正常数据修正的二级平行圆柱齿轮传动系统高保真虚拟模型,结合齿轮失效机理模型能有效地生成丰富的孪生故障数据集;设计的小波注意力特征提取网络在小波域中学习域不变特征,并从通道的维度上动态聚焦强相关故障特征,提升了网络泛化性;与经典的迁移学习模型 DDC、DANN 和 DSAN 进行对比分析,验证了所提方法 WATN 具有较好的分类精度和聚类效果。此外,为了应对实际工业场景中可能出现的噪声干扰强、故障模式多样性等问题,未来的工作将进一步致力于具有降噪功能的可解释深度网络构建、通用领域自适应算法开发等研究课题。

参考文献

- [1] 余晓霞, 汤宝平, 魏静, 等. 强背景噪声条件下自适应图卷积神经网络的航空发动机附件机匣故障诊断方法[J]. 仪器仪表学报, 2021, 41(8): 78-86.
YU X X, TANG B P, WEI J, et al. Aero-engine accessory case fault diagnosis method based on adaptive graph convolutional neural network under strong background noise conditions[J]. Chinese Journal of Scientific Instrument, 2021, 41(8): 78-86.
- [2] SU Z Q, ZHANG X L, WANG G Y, et al. The semisupervised weighted centroid prototype network for fault diagnosis of wind turbine gearbox[J]. IEEE/ASME Transactions on Mechatronics, 2024, 29(2): 1567-1578.
- [3] 吴启航, 丁晓喜, 何清波, 等. 齿轮箱故障边缘智能诊断方法及应用研究[J]. 仪器仪表学报, 2024, 45(1): 70-80.
WU Q H, DING X X, HE Q B, et al. Edge intelligent fault diagnosis method in the application of gearbox[J]. Chinese Journal of Scientific Instrument, 2024, 45(1): 70-80.
- [4] 高文超, 陈一帆, 陈诗雨, 等. 应用于齿轮箱故障诊断的小样本图像生成方法[J]. 电子测量与仪器学报, 2025, 39(3): 246-255.
GAO W CH, CHEN Y F, CHEN SH Y, et al. Small sample image generation method for gearbox fault diagnosis[J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(3): 246-255.
- [5] 刘大同, 郭凯, 王本宽, 等. 数字孪生技术综述与展望[J]. 仪器仪表学报, 2018, 39(11): 1-10.
LIU D T, GUO K, WANG B K, et al. Review and prospects of digital twin technology[J]. Chinese Journal of Scientific Instrument, 2018, 39(11): 1-10.
- [6] 夏景演, 黄如意, 陈祝云, 等. 孪生数据与特征增强融合驱动的装备小样本诊断方法[J]. 中国科学: 技术科学, 2023, 53(7): 1202-1213.
XIA J Y, HUANG R Y, CHEN ZH Y, et al. Intelligent fault diagnosis method using small fault samples driven by digital data and feature enhancement[J]. Science China Technological, 2023, 53(7): 1202-1213.
- [7] JIANG F, LIN W Q, WU ZH Q, et al. Fault diagnosis of gearbox driven by vibration response mechanism and enhanced unsupervised domain adaptation[J]. Advanced Engineering Informatics, 2024, 61: 102460.
- [8] YU X L, YANG Y, DU M G, et al. Dynamic model-embedded intelligent machine fault diagnosis without fault data[J]. IEEE Transactions on Industrial Informatics, 2023, 19(12): 11466-11476.
- [9] YAN SH, ZHONG X, SHAO H D, et al. Digital twin-assisted imbalanced fault diagnosis framework using subdomain adaptive mechanism and margin-aware regularization[J]. Reliability Engineering & System Safety, 2023, 239: 109522.
- [10] FENG K, JI J C, ZHANG Y CH, et al. Digital twin-driven intelligent assessment of gear surface degradation[J]. Mechanical Systems and Signal Processing, 2023, 186: 109896.
- [11] LIANG X H, ZUO M J, PANDEY M. Analytically evaluating the influence of crack on the mesh stiffness of a planetary gear set[J]. Mechanism and Machine Theory, 2014, 76: 20-38.
- [12] WAN ZH G, CAO H R, ZI Y Y, et al. An improved time-varying mesh stiffness algorithm and dynamic modeling of gear-rotor system with tooth root crack[J]. Engineering Failure Analysis, 2014, 42: 157-177.
- [13] FENG K, BORGHESEANI P, SMITH W A, et al. Vibration-based updating of wear prediction for spur gears[J]. Wear, 2019, 426: 1410-1415.
- [14] MATANIA O, BACHAR L, KHEMANI V, et al. One-fault-shot learning for fault severity estimation of gears that addresses differences between simulation and experimental signals and transfer function effects[J]. Advanced Engineering Informatics, 2023, 56: 101945.
- [15] KALITA K, RAMESH J V N, CEPOVA L, et al. Multi-objective exponential distribution optimizer (MOEDO): A novel math-inspired multi-objective algorithm for global optimization and real-world engineering design problems[J]. Scientific Reports, 2024, 14(1): 1816.
- [16] SHAO ZH F, JIANG H, ZHANG X F, et al. PLL-WCAN: Pseudo-label progressive learning guided wavelet class-aware adaptive network for gearbox cross-domain fault diagnosis[J]. Mechanical Systems and Signal Processing, 2025, 230: 112624.
- [17] ZHU Y CH, ZHUANG F ZH, WANG J D, et al. Deep subdomain adaptation network for image classification[J]. IEEE Transactions on Neural Networks and

Learning Systems, 2020, 32(4): 1713-1722.

- [18] TZENG E, HOFFMAN J, ZHANG N, et al. Deep domain confusion: Maximizing for domain invariance[J]. ArXiv preprint arXiv: 1412. 3474, 2014.
- [19] GANIN Y, LEMPITSKY V. Unsupervised domain adaptation by backpropagation[C]. 32nd International Conference on Machine Learning, 2016: 451-462.

作者简介



朱朋, 2022 年于重庆交通大学获得硕士学位, 现为重庆大学博士研究生, 主要研究方向为数字孪生驱动的高端装备智能运维。

E-mail: peng188154@163.com

Zhu Peng received his M.Sc. degree from Chongqing

Jiaotong University in 2022. He is currently a Ph. D. candidate at Chongqing University. His main research interests include digital twin-driven prognostics and health management (PHM) of high-end equipment.



邓蕾(通信作者), 2010 年获得重庆大学博士学位, 现为重庆大学副教授, 博士生导师, 主要研究方向为旋转机械智能运维与无线传感器网络。

E-mail: denglei@cqu.edu.cn

Deng Lei (Corresponding author) received her Ph. D. from Chongqing University in 2010. She is currently an associate professor and a doctoral supervisor at Chongqing University. Her main research interests include PHM of rotating machinery and decision-making, and wireless sensor networks.