

双流自适应时空特征融合网络的机械设备 剩余寿命预测方法*

朱江艳¹ 马 军^{1,2} 吴建德³ 熊 新^{1,2}

(1. 昆明理工大学信息工程与自动化学院 昆明 650500; 2. 昆明理工大学云南省先进装备智能控制
及国际联合实验室 昆明 650500; 3. 云南大学现代工学院 昆明 650500)

摘要: 复杂系统的状态监测数据是多个传感器采集的多源时空信息, 为了充分利用时间退化特征和测量变量间的空间关联信息, 提出了基于双流结构的自适应时空特征融合神经网络 (temporal-spatial feature fusion neural network, TSTFNN), 通过并行的时间流和空间流结构分别提取时间依赖特征和空间关联特征。设计了卷积自注意机制, 以弥补传统点积自注意力忽略时间序列连续性的缺点, 使模型更好地捕捉时间序列的连续性和细节变化。同时, 应用多尺度卷积神经网络挖掘不同测量变量之间的空间关联特征, 提升模型的全局感知能力。在特征融合阶段, 引入自适应加权机制, 实现时间和空间特征的动态融合。为了优化模型的预测效果, 构建了由约束均方误差损失和特征平衡损失组成的联合损失函数, 来增强时空特征的协同学习能力。最后, 基于 NASA 的 C-MAPSS 基准数据集的实验结果表明, 所提方法在多源数据 RUL 预测精度方面优于多种 SOTA 模型。

关键词: 剩余寿命预测; 时空特征融合; 卷积自注意; 多尺度卷积神经网络; 自适应加权

中图分类号: V229⁺.2; TH133.33; TP183 文献标识码: A 国家标准学科分类代码: 460

Dual-stream adaptive temporal-spatial feature fusion network for remaining useful life prediction of mechanical equipment

Zhu Jiangyan¹ Ma Jun^{1,2} Wu Jiande³ Xiong Xin^{1,2}

(1. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China;
2. Yunnan Key Laboratory of Intelligent Control and Application, Kunming University of Science and Technology,
Kunming 650500, China; 3. School of Engineering, Yunnan University, Kunming 650500, China)

Abstract: In complex systems, condition monitoring data consist of multi-source spatiotemporal information collected from multiple sensors. To effectively capture temporal degradation patterns and spatial correlations among measured variables, we propose an adaptive temporal-spatial feature fusion neural network (TSTFNN) based on a dual-stream structure. This framework incorporates parallel temporal and spatial streams to extract temporal dependencies and spatial correlations separately. To overcome the limitations of traditional dot-product self-attention, which often neglects time-series continuity, a convolutional self-attention mechanism is implemented, enhancing the capacity of model to capture sequential continuity and subtle temporal variations. A multi-scale convolutional neural network further extracts spatial correlation features across variables, improving global perception capabilities. During feature fusion, an adaptive weighting mechanism enables dynamic integration of temporal and spatial features. To optimize predictive performance, a joint loss function, combining constrained mean squared error (MSE) and feature balance loss, is introduced, facilitating the collaborative learning of temporal and spatial features. Finally, experimental results based on NASA's C-MAPSS benchmark dataset demonstrate that the proposed method outperforms various state-of-the-art (SOTA) models in terms of multi-source data RUL prediction accuracy.

Keywords: residual life prediction; spatial-temporal feature fusion; convolutional self-attention; multi-scale CNN; adaptive weighting

0 引言

作为智能工业的重要组成部分,预测与健康管理 (prognostics and health management, PHM) 对于确保设备运行的安全性和可靠性具有重要意义^[1-3]。作为 PHM 的一项关键任务,准确预测设备的剩余使用寿命 (remaining useful life, RUL) 对于实际应用过程中的维护决策具有重要意义^[4-7]。

在机械运行和维护策略不断演变的过程中,随着传感器和数据获取手段的发展,状态监测数据的类型和数量大幅增加。除振动信号外,声音、电压、电流、温度、力和功率等信号也越来越多地用于机械状态的诊断分析。这些多源数据为设备的剩余使用寿命提供了丰富的信息来源。然而,复杂机械系统内部各子系统之间普遍存在显著的空间相关性,且该相关性会随工况、载荷及外部环境的变化而动态演化^[8-11]。同时,传感器测量的时间序列具有强时间依赖性,且随时间和环境变化。因此,在多源监测数据背景下,RUL 预测不仅是一个传统的时间序列建模问题,更需要深入挖掘不同测量变量之间的相互依赖关系,从而将其建模任务拓展为一个融合时间依赖与变量间关联的时空特征预测问题^[12]。

针对上述需求,已有大量研究致力于基于时空特征的 RUL 预测方法探索。早期研究多采用串行结构对时空特征进行建模,即先提取空间或变量相关特征,再对时间依赖性进行建模。Tian 等^[13]通过构建空间相关性注意力衡量不同变量在不同时间数据中的重要性,再应用时间注意力增强长短期记忆 (long short term memory, LSTM) 在时间维度上的信息提取能力。Wang 等^[14]应用两个卷积神经网络来学习每个序列的时间和空间特征。Kong^[12]等采用空间图神经网络 (graph neural networks, GNN) 融合设备的空间特征和结构信息,提出改进的注意力机制融合时间特征。Peng 等^[15]采用 LSTM 和一维卷积神经网络 (convolutional neural network, CNN) 与全连接层的组合,同时提取时间和空间特征。Zhao 等^[16]提出名为 CBLSTM 的卷积双向长短记忆网络来提取时空特征,实现刀具磨损预测。尽管上述方法在一定程度上提升了预测性能,但串行建模方式本身仍存在局限:一方面,特征在分阶段处理过程中可能发生信息衰减,难以充分挖掘时间与空间特征之间的互补关系;另一方面,该类方法对特征提取顺序高度敏感,在面对多传感器数据复杂、动态依赖关系时灵活性不足。因此,如何构建更有效的时空信息融合模型,增强时间和空间特征的互补性,提高多传感器数据的综合利用效率,仍然是 RUL 预测研究中亟待解决的问题。

相比串行结构,并行双流架构在时空特征的提取

和融合上具备显著优势,避免了串行结构中可能出现的信息丢失和特征次序依赖问题。双流结构最早由 Simonyan 等^[17]于 2014 年提出,并应用于人体动作识别,将模型分为空间和时间流来分别处理特征信息。此后,并行架构在许多领域中得到了广泛应用,包括人体动作识别^[18-19]、图像分割^[20]、图像去雾^[21]、语音识别^[22]等。在 PHM 领域,双流结构也展现出了极大的潜力,Wang 等^[14]设计了并行卷积神经网络,同步挖掘多变量传感器信号中的空间相关性和时间依赖性来实现 RUL 预测。He 等^[23]通过多核 CNN 和长短期记忆神经网络的并行模块构建了时空融合模型,并用于风力发电机故障诊断。Xu 等^[24]和 Gao 等^[25]将双流结构与自注意力机制结合,平衡数据中时间和空间特征的贡献度,实现了更高效的特征提取和融合。Zhu 等^[26]基于双流结构应用 CNN-Transformer 模型来提取监测序列中的时间-空间特征,提升了模型对复杂数据的表征能力。尽管上述方法在缓解串行结构不足方面取得了一定进展,但仍面临两方面挑战:1) 标准 Transformer 的自注意力机制通过点积计算查询和键之间的相似度,未能充分考虑时间序列中时间点之间的连续性,难以有效提取细粒度的局部特征模式^[27-28];2) 在 RUL 预测任务中,多传感器监测数据的时空相关性是随设备运行状态和外部环境的变化而不断变化,固定的特征融合策略忽略了不同测量变量在时间和空间维度上相互作用的动态变化。

针对上述问题,提出一种结合双流结构与改进自注意力机制的剩余寿命预测方法。分别提取时间和空间特征,并设计自适应加权机制动态调整融合权重以适应数据的动态变化,实现时空信息的深度耦合。本文提出了一种双流自适应时空信息融合架构。采用双流结构分别提取时间和空间特征,并通过自适应加权机制实现时间和空间特征动态融合,增强了模型对复杂时空依赖关系的捕捉能力。融合卷积和注意力机制,设计卷积注意力,增强模型对时间序列细粒度特征的提取。构建了基于门控机制的自适应时空特征融合策略。针对设备的运行状态和传感器数据的时空变化特征,动态调整两类特征的权重,使得模型能够根据任务需求智能地分配不同特征的贡献,并引入了特征平衡损失函数,提高特征融合的有效性。

1 多源传感历史数据描述

针对多源传感监测的随机退化设备,假设设备上安装了 S 个传感器,每个传感器记录一个一维时间序列 (如振动、温度等特征)。设各时间序列均包含 T 个时间步,则第 s 个传感器在时刻 t 的观测值表示为 x_t^s ,将所有传感器在时刻 t 获取的特征组成特征向量 $\mathbf{x}_t = \{x_t^1, \dots, x_t^S\} \in \mathbf{R}^S$,进

而可将传感器在 T 个时间步内的历史监测数据表示为矩阵 $\mathbf{X}_T^s = (x^1, x^2, \dots, x^s)^T = (x_1, x_2, \dots, x_T) \in \mathbf{R}^{s \times T}$ 。矩阵 \mathbf{X}_T^s 同时包含时间维度与传感器空间维度的变化关系,如图 1 所示,用于表征设备在整个运行过程中的多源退化特征。RUL 预测任务即基于传感器在 T 个时间步内的监测数据 X_T^s , 估计设备在当前时刻的剩余寿命记为 $y(t)$:

$$y(t) = F(\mathbf{X}_T^s) \quad (1)$$

2 基于时空特征融合神经网络 (TSTFNN) 的预测模型

为充分挖掘监测数据中的空间和时间相关特征,构建了基于双流结构的自适应 TSTFNN,结构如图 2 所示。

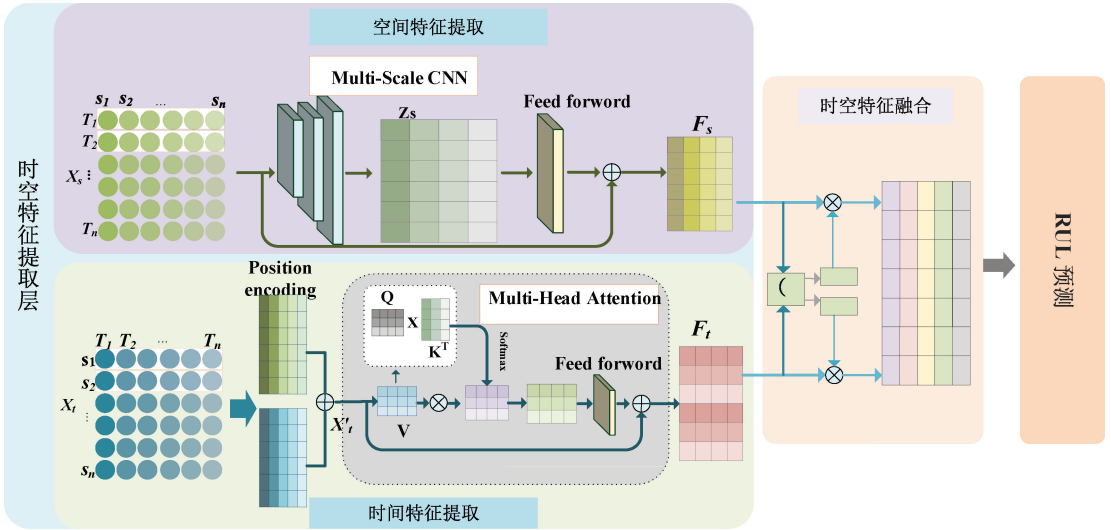


图 2 基于 TSTFNN 模型的 RUL 预测框架

Fig. 2 RUL prediction framework based on TSTFNN model

2.1 时空特征提取模块

为了充分利用监测数据中的时间和空间信息特征,设计了一个双通道特征提取层,应用并行的空间特征提取层和时间特征提取层来捕获不同传感器和时间步长的权重特征。其中时间特征提取层主要由卷积自注意力 Transformer 构成,空间特征提取层主要由多尺度 CNN 构成。

1) 基于卷积自注意力 Transformer 的时间特征学习

在标准 Transformer 的自注意力层中^[29],多头自注意力子层通过线性投影将时间维度数据 X_T 同时转换为不同的查询矩阵 \mathbf{Q}_h , 键矩阵 \mathbf{K}_h 和值矩阵 \mathbf{V}_h 。其中, $h = 1, \dots, H$ 为注意力头数。然后,应用缩放点积注意力计算输出向量:

$$\mathbf{Q}_h = \text{attention}(\mathbf{Q}_h, \mathbf{K}_h, \mathbf{V}_h) = \text{Softmax}\left(\frac{\mathbf{Q}_h \mathbf{K}_h^T}{\sqrt{d_k}}\right) \mathbf{V}_h \quad (2)$$

该网络框架主要由时空特征提取模块、时空特征融合模块和 RUL 预测模块构成。

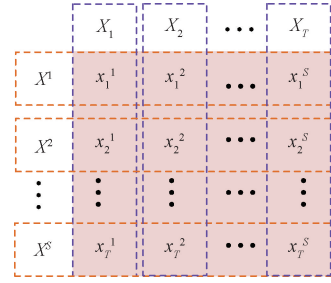


图 1 时间维度数据 X_T 与空间维度数据 X^s 示意图

Fig. 1 Diagram of time dimension X_T and space dimension data X^s

然后,将 O_1, O_2, \dots, O_H 连接并再进行线性投影,得到多头注意力的输出。

$$\text{Multihead} = \text{Concat}(O_1, O_2, O_3) W_H \quad (3)$$

然而,自注意力层中查询和键之间的相似性是根据它们的点值计算的,没有充分利用局部上下文,如图 3(a) 所示。按照信号中的采样点(红点)来计算相似度则会造成信号点周围的发展趋势信息遗漏,忽略了各样本点之间的关系。

参考文献[30],本文提出构建卷积自注意机制,以解决传统点积自注意力机制在处理时间序列数据时忽视连续性的问题,从而更有效地捕捉时间序列中的细粒度特征和局部变化。

$$\hat{\mathbf{Q}}_h = \text{Conv}(X) \quad (4)$$

$$\hat{\mathbf{K}}_h = \text{Conv}(X) \quad (5)$$

通过使用因果卷积,建立起时间序列中各样本点之

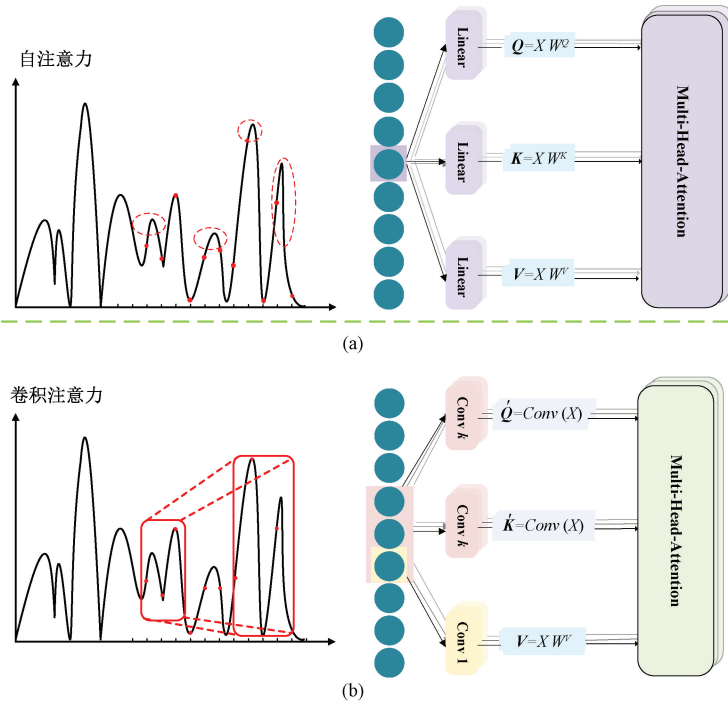


图 3 标准自注意力和卷积自注意力的对比

Fig. 3 Comparison of standard and convolutional self-attention

间的关系,在 self-attention 中局部信息,从而提高预测准确性。其中,当 $K=1$ 时,卷积自注意力将退化为规范自注意力,因此可以将其视为一种泛化模型。本文方法实现位置编码如下:

$$P_{pos}^{2i} = \sin\left(\frac{pos}{10\,000^{\frac{2i}{d_v}}}\right) \quad (6)$$

$$P_{pos}^{2i+1} = \cos\left(\frac{pos}{10\,000^{\frac{2i}{d_v}}}\right) \quad (7)$$

式中: pos 表示位置信息; i 表示嵌入特征向量的维度。

2) 基于多尺度 CNN 的空间特征学习

CNN 具有很强的局部特征捕获能力,可以根据内核的大小提取不同的特征,较小的卷积核可以有效地保留细粒度信息,较大的卷积核能捕获到整体特征。本文为为了挖掘多传感器监测数据中更深层次的退化特征并提取变量之间的耦合关系,设计了多尺度卷积层来捕获不同传感器变量之间的空间相关性。其结构如图 4 所示。应用 3 个具有不同卷积核大小的分支来提取不同尺度的特征,其中卷积核为 1×1 的分支旨在保留数据中最原始的信息,并应用残差连接来加速网络的训练。

空间多尺度特征提取主要基于卷积和池化操作来实现。卷积层使用窗口长度为 k 的卷积内核(也称为过滤器)对空间维度数据 X^s 执行特征提取。并在整个输入空间上移动卷积核来对局部区域执行卷积操作,直到遍历

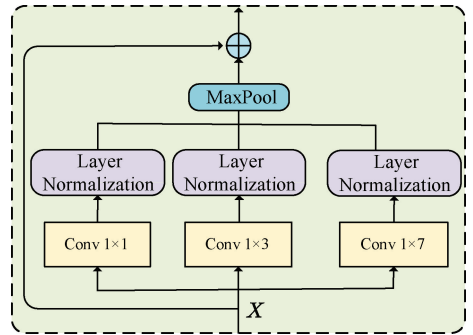


图 4 多尺度 CNN 的架构

Fig. 4 Multi-scale CNN architecture

所有输入空间,并获得相应的特征图,其过程如图 5 所示。输出特征 z_i 可以表示为:

$$z_i = \sigma(W^T y_{i:i+k-1} + b) \quad (8)$$

式中: $W \in R^k$ 表示滤波器向量; b 表示偏置项; $y_{i:i+k-1}$ 是从第 i 个变量开始的输入信号 X^s 的 k 长度子信号; $\sigma(\cdot)$ 是非线性激活函数,本文选择线性整流单元 (ReLU) $\sigma(x) = \max(0, x)$ 作为非线性激活函数,防止梯度消失,加速模型收敛。

输出标量 z_i 可以看作是过滤器对相应子信号的激活。将过滤器从开始变量滑动到结束变量,其中第 j 个滤波器的特征可以表示为:

$$Z_j = [z_1, z_2, \dots, z_{L-k+1}] \quad (9)$$

之后,卷积层生成的特征图进一步经过池化层,从而提取最重要的位置不变特征。提取到的空间特征为:

$$F_s = \maxpool(Z) \quad (10)$$

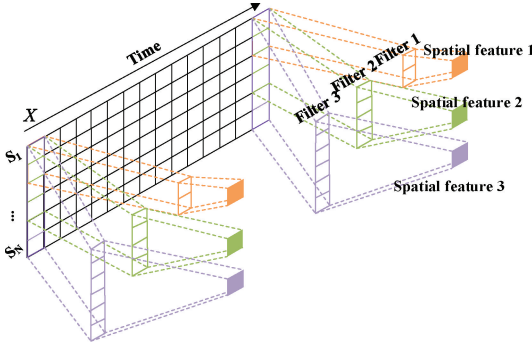


图5 基于多尺度 CNN 的多尺度空间特征提取

Fig. 5 Multi-scale spatial feature extraction based on multi-scale CNN

2.2 时空特征融合

本文采用自适应加权机制,根据输入特征动态调整时间和空间特征的融合比例,使得两种特征能够在不同情况下有效地贡献特征。空间特征 $F_s \in \mathbf{R}^{K \times D}$ 和时间特征 $F_T \in \mathbf{R}^{T \times D}$ 通过式(11)的计算来实现特征融合。

$$F = \alpha F_s + (1 - \alpha) F_T \quad (11)$$

其中动态权重 α 由自适应权重函数 $g(\cdot)$ 根据时间和空间特征生成:

$$\alpha = g(F_s, F_T) \quad (12)$$

自适应权重函数 $g(\cdot)$ 采用前馈神经网络实现:

$$g(F_s, F_T) = \sigma(\mathbf{W} \cdot [F_s, F_T] + b) \quad (13)$$

式中: $\sigma(\cdot)$ 是 sigmoid 激活函数; \mathbf{W} 是可学习的权重矩阵; b 是偏置; $[\cdot]$ 表示特征的拼接操作。通过训练,自适应加权机制能够学习到不同输入场景下的 α , 实现更灵活的时空特征融合。

2.3 RUL 预测层

考虑到退化特征与 RUL 之间复杂的非线性映射关系,本文引入前馈层来有效地整合融合后的特征 F 以嵌入这些非线性退化特征。为了进一步提高模型的收敛性能,采用高斯误差线性单元(Gaussian error linear unit, GeLU)作为前馈层的激活函数。最终,RUL 预测值可以表示为:

$$F_{ff} = GeLU(FFN(W_{ff} \cdot F + b_{ff})) \quad (14)$$

$$\hat{y} = sigmoid((W \cdot F_{ff}) + b) \quad (15)$$

式中: W, W_{ff} 表示前馈网络的权重参数; b, b_{ff} 表示前馈网络的偏置。应用 sigmoid 函数将模型的输出限定在 0~1 之间,由于在数据预处理阶段,RUL 标签采用统一的截断策略,将健康阶段的最大剩余寿命固定为 125,并利用该最大值对 RUL 进行归一化处理,使其缩放至 $[0, 1]$ 范围。因此,在反归一化时需将模型输出乘以 125,以将预测结

果映射回真实的 RUL 数值范围。

2.4 损失函数设计

根据剩余寿命预测任务的特点和时空特征融合需求开发了一种联合损失函数来训练所提出的 TSTFNN 网络。首先,结合工业实践,RUL 预测中应避免预测值大于实际值的情况,为解决这个问题,文献[26]提出了约束 MSE (constrain the MSE, CMSE) 损失函数,加大对预测值大于真实值情况的惩罚,来引导模型在关键误差方向上进行优化。CMSE 损失函数的表达式为:

$$Loss(y, \hat{y})_{CMSE} = \frac{1}{2N} \sum_{i=1}^N (y - \hat{y})^2 + \frac{\lambda}{2N} \sum_{i=1}^N (\max(\hat{y} - y)) \quad (16)$$

其次,为了防止自适应权重机制过度依赖某一类特征,引入了特征平衡损失函数约束时间特征和空间特征的权重 α , 确保 α 不过于偏向 0 或 1,促进时间特征和空间特征的均衡利用。

$$Loss(y, \hat{y})_{reg} = (\alpha \cdot (1 - \alpha)) \quad (17)$$

最终的损失函数为 CMSE 损失函数与特征平衡损失函数结合得到的联合损失函数。

$$L = Loss_{CMSE} + Loss_{reg} \quad (18)$$

3 案例研究与分析

3.1 数据说明

本文使用的 C-MAPSS 数据集是 NASA 通过商业模块化航空航天推进系统模拟生成的涡扇发动机退化数据^[31],以模拟发动机在不同运行条件和故障模式下的退化轨迹。如图 6 所示,建模对象主要为发动机气路部分,由 4 个具有不同故障模式和操作条件的子数据集组成,如表 1 所示。FD002 和 FD004 数据集具有更复杂的操作条件和故障模式,RUL 更难预测。每个数据集都包含一个训练集和一个测试集,训练集包含整个运行到失效过程中每个时间步的数据。且每个数据集都包含 21 维传感器数据。

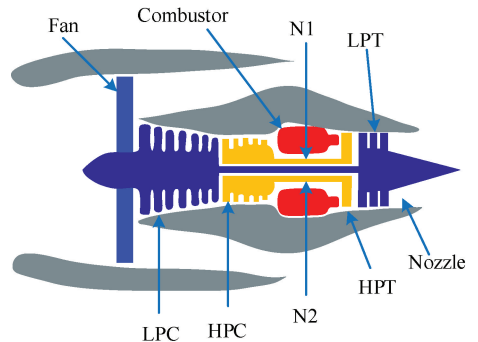


图6 涡扇发动机简图

Fig. 6 Simplified diagram of a turbofan engine

表 1 C-MAPSS 数据集
Table 1 The C-MAPSS dataset

子数据集	训练集	测试集	操作条件	故障模式
FD001	100	100	1	1
FD002	260	259	6	1
FD003	100	100	1	2
FD004	249	248	6	2

3.2 数据预处理

所用数据集中包含 21 维传感器数据,但其中某些传感器数据样本不随时间变化,或者传感器数据与其他传感器数据不相关。因此,需要对传感器进行初步筛选。本文通过计算传感器数据之间的相关性来进行传感器筛选,结果如图 7 所示。由图 7 可以看出,传感器 1、5、6、10、16、18 和 19 无法表征发动机的退化过程,予以剔除,仅使用剩余 14 个传感器的数据进行后续 RUL 预测。为减小数据单位和尺度差异对 RUL 预测性能的影响,采用最小-最大缩放方法对不同传感器数据进行归一化处理。为了捕获不同时间数据点之间的依赖关系,应用滑动窗口处理将顺序数据转换为时间窗口数据。实验表明,FD001 和 FD003 数据集上的时间窗口长度为 30 时,FD002 和 FD004 数据集上时间窗口长度为 40 时模型的性能最好。

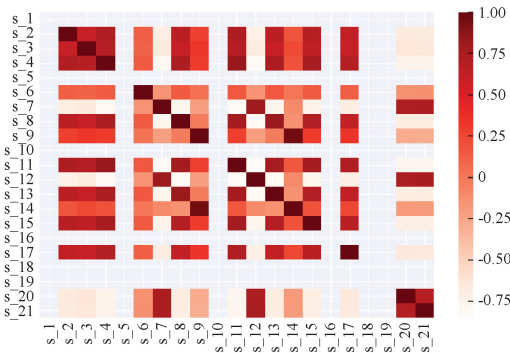


图 7 传感器序列自相关图

Fig. 7 Sensor sequence autocorrelation map

3.3 评估指标

为了评估模型的有效性,本文采用均方根误差 (RMSE) 和评分函数 (Score) 两个常用指标客观评价所提 TSTFNN 模型的预测性能。

1) RMSE 是回归任务中常用的评估指标之一,表示为:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y - \hat{y})^2} \quad (19)$$

2) Score 是 2008 年预测与管理 (PHM08) 数据挑战赛中提出作为 RUL 预测领域评估指标,并已被研究人员

广泛使用^[32-34]。两种评价指标之间的差异如图 8 所示。与 RMSE 相比,Score 对延迟预测的惩罚比提前预测更严重。Score 表示为:

$$Score = \begin{cases} \sum_{i=1}^N \exp\left(-\frac{\hat{y} - y}{13} - 1\right), \hat{y} < y \\ \sum_{i=1}^N \exp\left(\frac{\hat{y} - y}{10} - 1\right), \hat{y} > y \end{cases} \quad (20)$$

式中: N 表示测试样本的总量; y 表示真实 RUL 值; \hat{y} 表示预测值。

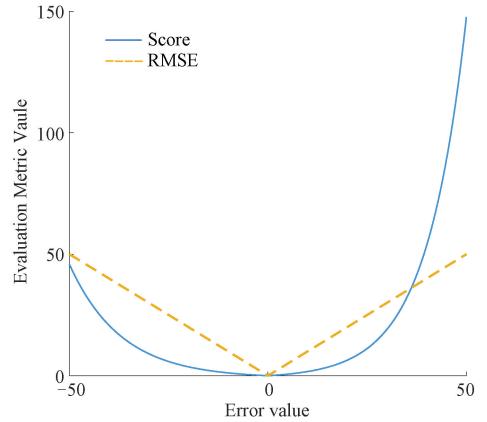


图 8 Score 函数和 RMSE 的图形表示

Fig. 8 Graphical representation of the Score function and RMSE

3.4 模型参数设置

引入网格搜索方法获取模型的超参数,以增强 RUL 预测性能。在训练过程中应用了 Adam 优化器,在编码层应用了 dropout,值设置为 0.2。学习率设置为 0.001,批大小设置为 256。训练过程设置为最多 100 个 epoch。表 2 为所提方法的超参数。

表 2 模型超参数

Table 2 Hyperparameters of the proposed method

层	结构	参数
空间特征提取层	Multi-CNN	Kernel_size = 1, 3, 7 Stride = 1
时间特征提取层	ConvTransformer	Self-attention heads = 6 Encoder blocks N = 2 Kernel_size = 7

3.5 主要结果与分析

1) CMSE 损失函数惩罚系数 λ 对模型性能的影响

为进一步探究 CMSE 中惩罚系数对模型性能的影响,本文设计了惩罚系数灵敏度分析实验。在固定其他超参数的前提下,调整 λ 的取值,通过在多个数据集上比较模型在预测精度与约束合规性两方面的表现,分析 λ

对损失函数优化过程及模型鲁棒性的影响。当 $\lambda = 0$ 时, 约束损失函数的惩罚项为 0, 损失函数由 MSE 函数与特征平衡损失函数 $Loss_{reg}$ 的组成。本文在 4 个 C-MAPSS 子数据集 (FD001、FD002、FD003、FD004) 上进行对比实验。具体地, 在保持其他超参数不变的情况下, 选取多组不同的 λ 值 ($\lambda \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$) 分别训练模型, 并评估其在 RUL 预测任务中的表现。图 9 所示为不

同惩罚系数 λ 在 4 个数据集上的预测结果。从图 9 可以看出, 在这 4 个数据集中, 惩罚系数对 RUL 预测性能的影响有所不同, 合适的惩罚系数能够有效提升模型的预测性能, 使得 RMSE 和 Score 的值最小。因此, 为了使模型到达最佳预测效果, 本文在 FD001 数据中设置 $\lambda = 0.2$, FD002、FD003 数据集中设置 $\lambda = 0.6$, FD004 数据集中设置 $\lambda = 0.8$ 。

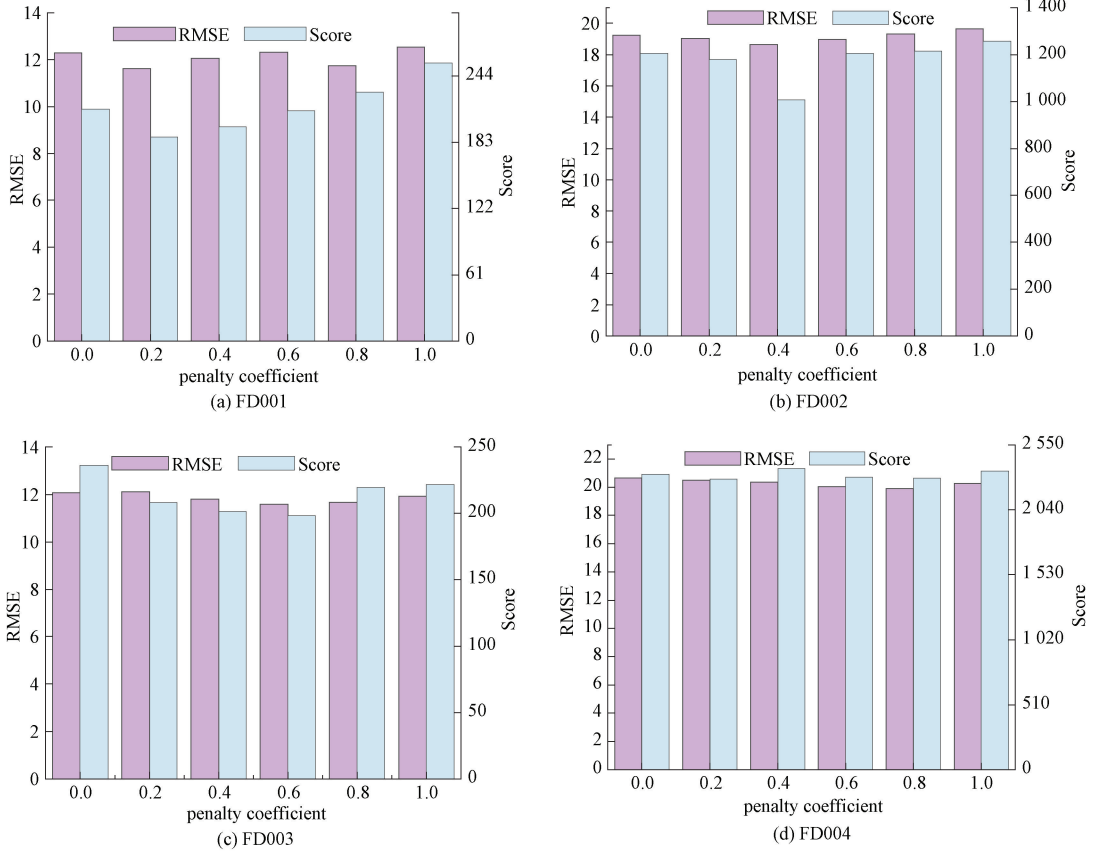


图 9 CMSE 损失函数惩罚系数 λ 对模型性能的影响

Fig. 9 Impact of CMSE loss function penalty coefficient λ on model performance

2) 本文所提方法的实验结果与分析

图 10 所示为所提模型在 4 个数据集上的预测 RUL 值与实际 RUL 值及误差。通过对比图 10 可以发现, TSTFNN 模型表现出较高的预测精度和优异的测试性能, 模型预测结果与发动机的真实 RUL 基本一致。从图 10 可以看出 FD002 和 FD003 数据集的中值预测误差基本为 0, FD001 和 FD004 数据集的中值预测误差接近于 0。FD001 和 FD003 具有更集中的预测误差分布, 因此其 RMSE 低于其他两个子数据集。虽然 FD002 和 FD004 数据集的误差分布上下界存在显著差异, 但总体分布中没有异常值, 表明 TSTFNN 模型具有优异的预测稳定性。总的来说, 通过比较 TSTFNN 模型预测结果与真实 RUL 的误差, 证明了 TSTFNN 具有优异的预测性能。

3) 与最新方法的比较

为验证所提出的 TSTFNN 的有效性和优越性, 本文将其实验结果与几种先进方法进行了比较: 1) 未同时考虑时间和空间特征的方法, ELSTMNN^[35]、DCNN^[36]、DSAN^[37]、BiGRU-TSAM^[38]; 2) 采用串联方式同时考虑时间和空间特征的方法, STFA^[12]、SCTA-LSTM^[13]、STRUL^[14]。表 3 和 4 所示为基准数据集 C-MAPSS 的实验对比结果。结果表明, 所提出的 TSTFNN 在整体性能上优于其他方法。在工况较为简单的 FD001 数据集上, 尽管 STFA 模型的 RMSE 最低, 但 TSTFNN 在 Score 值上降低了 3.28%。在复杂工况的 FD002 数据集上, TSTFNN 的 RMSE 至少降低了 2.76%, 而在 Score 值上对比同时考虑时间和空间特征的方法至少降低了 19.27%。此外, 同

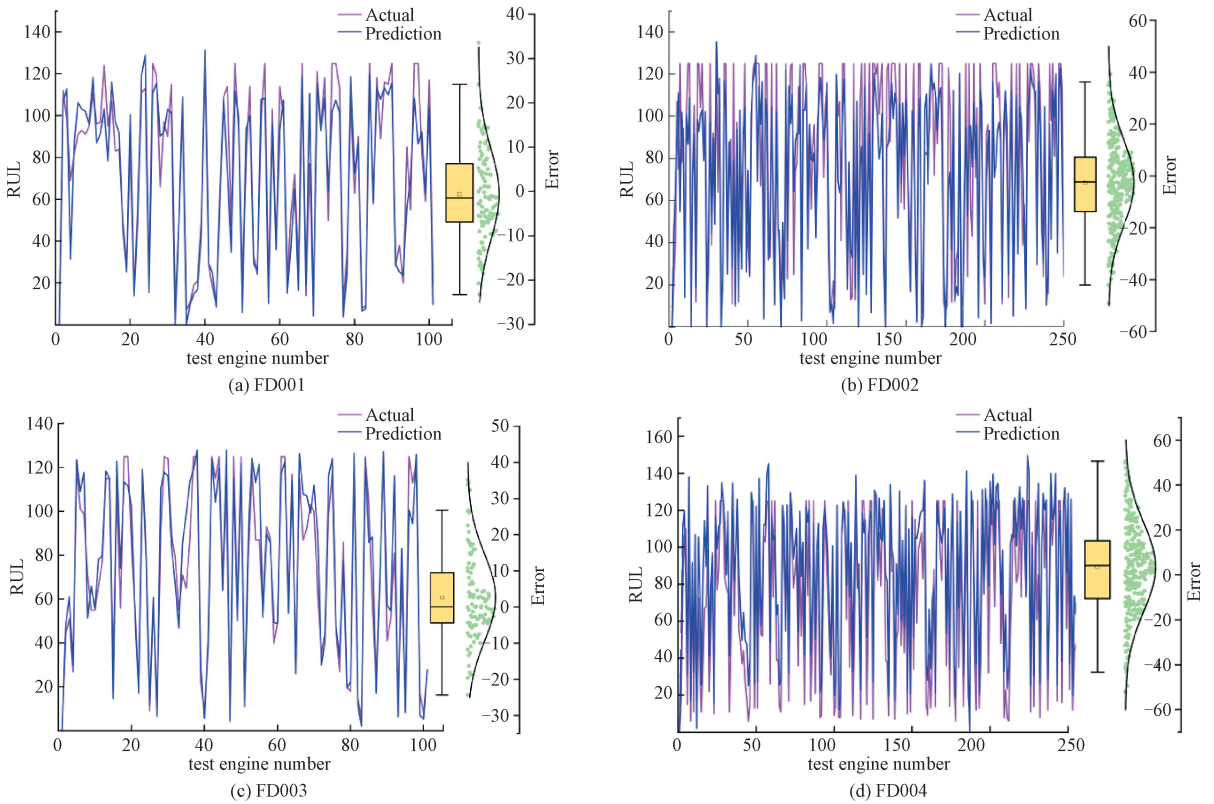


图 10 测试集中预测 RUL 与实际 RUL 的比较

Fig. 10 Comparison of predicted RUL and actual RUL in the test set

时考虑时间和空间特征的方法在整体性能上优于未同时考虑这两类特征的方法。综上所述,TSTFNN 模型在 4 个数据集中都展示出了最佳的预测精度,证明了所提框架的优越性。

表 3 所提方法与最新方法的 RMSE 比较

Table 3 RMSE comparison between the proposed method and the state of the art method

模型	FD001	FD002	FD003	FD004
ELSTMNN ^[35] 2020	18.22	—	23.21	—
DCNN ^[36] 2018	12.61	22.36	12.64	23.31
DSAN ^[37] 2022	13.40	22.06	15.12	21.03
BiGRU-TSAM ^[38] 2022	12.56	18.94	12.45	20.47
STFA ^[12]	11.35	19.17	11.64	21.41
SCTA-LSTM ^[13] 2023	12.10	19.43	13.39	21.50
STRUL ^[14] 2023	12.85	19.24	13.74	22.344
TSTFNN	11.62	18.64	11.57	19.89

3.6 消融实验

为了验证卷积自注意和双流结构在多元信号特征提取方面的优势,分别对卷积自注意和双流结构设计了消融实验。

表 4 所提方法与最新方法的 Score 比较

Table 4 Score comparison between the proposed method and the state of the art method

模型	FD001	FD002	FD003	FD004
ELSTMNN ^[35] 2020	571	—	839	—
DCNN ^[36] 2018	273.7	10 412	284.2	12 466
DSAN ^[37] 2022	242	2 869	497	2677
BiGRU-TSAM ^[38] 2022	213.35	2 264.13	232.86	3 610.34
STFA ^[12]	194.44	2 493.09	224.53	2 760.13
SCTA-LSTM ^[13] 2023	226	1 492	227	3 392
STRUL ^[14] 2023	224	1 950	252	3 080
TSTFNN	188.07	1 204.46	198.32	2 287.78

1) 卷积自注意的消融研究

(1) Convolution self-attention 卷积核大小分析

对提出的卷积自注意力机制进行消融研究。在保持模型其他参数不变的情况下研究了卷积自注意力在不同核大小 $k \in \{1, 3, 5, 7\}$ 下模型的性能,其中, $k=1$ 时,卷积自注意退化为标准自注意,即 Transformer 模型。结果如表 5 所示,核大小为 $k \in \{3, 5, 7\}$ 的卷积 Transformer 模型在 RMSE 和 Score 方面略好于标准自注意。卷积核 k 越大,模型的性能越好,尤其在多故障数据集 FD002 和

FD004 中的体现更明显。在 FD004 数据集上, $k=1$ 的模型 (Transformer 模型) 相比 $k=7$ 的模型 Score 降低了

13.72%。这得益于卷积的加入使得更多的局部上下文信息用于更精确的实现查询-键的匹配。

表 5 不同 k 值得模型性能

Table 5 Model performance for different values of k

k	FD001		FD002		FD003		FD004	
	RMSE	Score	RMSE	Score	RMSE	Score	RMSE	Score
1	13.13	210.53	20.53	1091.16	13.22	245.33	23.05	2 651.87
3	12.84	205.94	19.87	1025.42	12.55	222.08	21.54	2 531.18
5	12.22	197.78	19.21	1003.34	12.03	218.43	20.25	2 381.51
7	11.62	188.07	18.64	1204.46	11.57	211.54	19.89	2 287.78

(2) self-attention 与卷积自注意力的可视化分析

为了更直观的分析卷积自注意与标准自注意在捕获信息方面的差异, 本文在 FD004 数据集中分别对 self-attention 和卷积自注意力的注意力矩阵进行可视化, 其结果如图 11 所示。self-attention 倾向于将注意力集中在

部分序列上, 导致另一部分序列的关注度较低。相较之下, 卷积的加入引入了更多的局部信息, 使卷积自注意力能够覆盖整个序列并合理分配权重。该注意力权重的可视化进一步验证了卷积自注意力在信息捕获上的优越性。

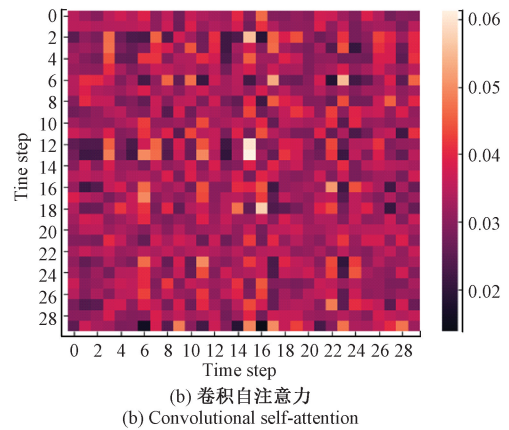
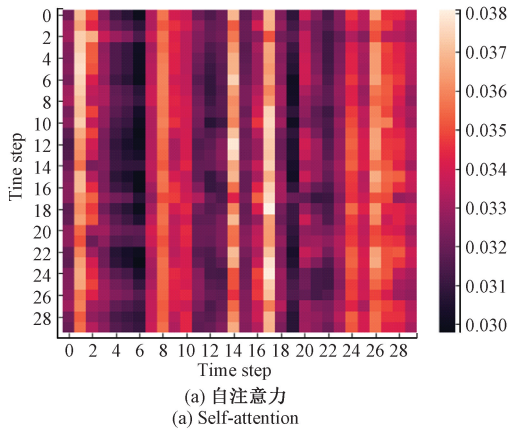


图 11 两种注意力 attention-score 的对比

Fig. 11 The attention-score of two attention

2) 时空特征表达的消融研究

为验证双流结构中引入的时-空特征在提升模型预测精度方面的有效性, 本文设计了 4 种对比网络结构, 并与 TSTFNN 进行比较: (1) 仅提取时间特征的分支 (TFNN); (2) 仅提取空间特征的分支 (SFNN); (3) 串联时间-空间结构 (CSTFNN); (4) 并行方式提取特征, 以拼接方式直接融合时间-空间特征的模型 (TSTFNN_1)。通过对这 5 种网络结构的实验结果进行对比, 分析它们在不同数据集上的表现, 尤其是在复杂工况下的表现, 从而验证时空特征融合对提升剩余寿命预测精度的有效性。

实验结果如图 12 所示。相较于同时利用时间和空间特征的双流模型, 单独提取时间或空间特征的模型表现出明显的性能下降。这表明时间和空间特征在剩余寿

命预测中的贡献是互补的, 结合两者可以提供更全面的信息, 从而提升预测准确性。具体而言, 在复杂工况下 (如 FD002 和 FD004 数据集), 串联时间-空间特征的模型性能下降更为明显。特别是在 FD002 数据集中, 串联结构的 Score 值增加了 23.09%, 表明特征串联会导致时空信息融合上存在不当, 从而降低了模型的预测精度。与自适应加权融合方式相比, 直接融合时间-空间特征的模型 (TSTFNN_1) Score 值增加了 12.21%, RMSE 增加了 4.93%。因此, 通过双流结构并行提取时间与空间特征, 能够更有效地捕捉不同维度的信息, 且自适应动态分配时间特征和空间特征的融合权重能够优化特征的协同作用, 使模型在不同情境下更灵活地调整其关注点, 从而提升整体预测性能。

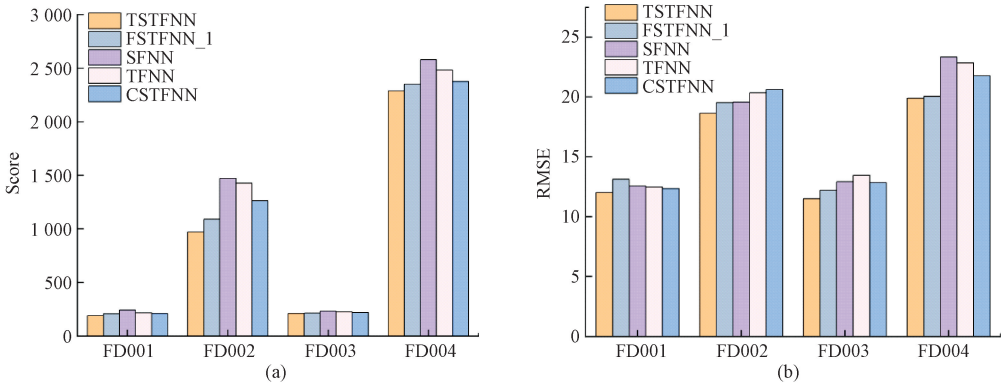


图 12 4 种不同模型的实验结果

Fig. 12 Experimental results for four different models

4 结 论

本文提出了一种基于双流结构的时空特征融合神经网络 TSTFNN, 用于充分融合多传感器监测数据中的时间依赖性和空间关联性, 以实现更为精确的 RUL 预测。具体而言, 设计的卷积自注意机制通过引入卷积操作利用更多的局部上下文信息, 提升了查询-键匹配的精确度, 增强了时间特征的提取能力; 多尺度卷积网络则有效捕捉了不同测量变量之间的空间关联特征, 增强了模型对复杂数据的全局理解与处理能力。此外, 通过动态分配时间特征和空间特征的融合权重, 模型能够优化特征协同作用, 使其在不同情境下灵活调整关注点, 从而提升整体预测性能。实验结果显示, TSTFNN 在 NASA 的 C-MAPSS 基准数据集上表现优异, 相较于仅依赖时间或空间特征的模型, 双流结构融合时空特征的模型显著提高了预测精度。值得指出的是, 本文所研究的 RUL 预测任务属于监督学习范畴, 依赖训练数据与测试数据分布一致的前提。然而, 在实际工业应用中, 由于设备工况、使用环境和监测条件的多样性, 这一前提往往难以满足, 限制了模型的直接迁移能力。因此, 未来的研究将致力于结合迁移学习、自监督学习或领域自适应方法, 以缓解训练-测试分布偏移对预测性能的影响, 实现跨设备、跨工况的鲁棒 RUL 预测。

参考文献

- [1] LIU Z, LIU H, JIA W, et al. A multi-head neural network with unsymmetrical constraints for remaining useful life prediction [J]. *Advanced Engineering Informatics*, 2021, 50: 101396.
- [2] ZHANG J, JIANG Y, WU S, et al. Prediction of remaining useful life based on bidirectional gated recurrent unit with temporal self-attention mechanism [J]. *Reliability*

Engineering & System Safety, 2022, 221: 108297.

- [3] 刘丽, 裴行智, 雷雪梅. 基于时间卷积注意力网络的剩余寿命预测方法 [J]. *计算机集成制造系统*, 2022, 28(8): 2375-2386.
LIU L, PEI X ZH, LEI X M. Remaining useful life prediction method based on temporal convolutional attention network [J]. *Computer Integrated Manufacturing System*, 2022, 28(8): 2375-2386.
- [4] YANG B, LIU R, ZIO E. Remaining useful life prediction based on a double-convolutional neural network architecture [J]. *IEEE Transactions on Industrial Electronics*, 2019, 66(12): 9521-9530.
- [5] CAO Y, DING Y, JIA M, et al. A novel temporal convolutional network with residual self-attention mechanism for remaining useful life prediction of rolling bearings [J]. *Reliability Engineering & System Safety*, 2021, 215: 107813.
- [6] 刘业峰, 王帅, 刘晶晶, 等. 基于 TCN-SA 和 Bi-GRU 的轴承剩余寿命预测 [J]. *计算机集成制造系统*, 2024: 1-11.
LIU Y F, WANG SH, LIU J J, et al. Bearing remaining life prediction based on TCN-SA and Bi-GRU [J]. *Computer Integrated Manufacturing System*, 2024: 1-11.
- [7] 陈仁祥, 张雁峰, 徐向阳, 张鹏博, 杨宝军. 基于子空间域对抗判别网络的不同型号滚动轴承剩余寿命预测 [J]. *仪器仪表学报*, 2024, 45(3): 119-127.
CHEN R X, ZHANG Y F, XU X Y, et al. Remaining life prediction of different types of rolling bearings based on subspace domain adversarial discriminant network [J]. *Chinese Journal of Scientific Instrument*, 2024, 45(3): 119-127.
- [8] PANG Y, HE Q, JIANG G, et al. Spatio-temporal

- fusion neural network for multi-class fault diagnosis of wind turbines based on SCADA data [J]. *Renewable Energy*, 2020,161:510-524.
- [9] 韩莹,陈熙. 一种基于融合特征聚类和随机配置网络的轴承剩余寿命预测方法[J]. *电子测量与仪器学报*, 2024,38(4):128-139.
- HAN Y, CHEN X. Remaining useful life prediction of system relative density high dimensional kernel estimation fusing multi-sensor data [J]. *Journal of Electronic Measurement and Instrumentation*, 2024, 38 (4) : 128-139.
- [10] 张加劲. 基于注意力机制和 CNN-BiLSTM 模型的航空发动机剩余寿命预测[J]. *电子测量与仪器学报*, 2022,36(8):231-237.
- ZHANG J J. Remaining life prediction of aeroengine based on attention mechanism and CNN-BiLSTM model [J]. *Journal of Electronic Measurement and Instrumentation*, 2022,36(8):231-237.
- [11] 徐震震,薛林,马凯,等. 基于时空特征的航空发动机剩余使用寿命预测[J]. *电子测量技术*, 2023, 23(46):63-67.
- XU ZH ZH, XUE L, MA K, et al. Remaining useful life prediction of aeroengine based on spatio-temporal features[J]. *Electronic measurement and Instrumentation*, 2023, 23(46):63-67.
- [12] KONG Z, JIN X, XU Z, et al. Spatio-temporal fusion attention: A novel approach for remaining useful life prediction based on graph neural network [J]. *IEEE Transactions on Instrumentation and Measurement*, 2022, 71:1-12.
- [13] TIAN H, YANG L, JU B. Spatial correlation and temporal attention-based LSTM for remaining useful life prediction of turbofan engine [J]. *Measurement*, 2023, 214:112816.
- [14] WANG T, LI X, WANG W, et al. A spatiotemporal feature learning-based RUL estimation method for predictive maintenance [J]. *Measurement*, 2023, 214:112824.
- [15] PENG C, CHEN Y, CHEN Q, et al. A remaining useful life prognosis of turbofan engine using temporal and spatial feature fusion[J]. *Sensors*, 2021,21:418.
- [16] ZHAO R, YAN R, WANG J, et al. Learning to monitor machine health with convolutional bi-directional LSTM networks[J]. *Sensors*, 2017,17(2):273.
- [17] SIMONYAN K, ZISSERMAN A. Two-stream convolutional networks for action recognition in videos [C]. 27th International Conference on Neural Information Processing Systems, 2014:568-576.
- [18] HUSSAIN A, KHAN S U, KHAN N, et al. Low-light aware framework for human activity recognition via optimized dual stream parallel network [J]. *Alexandria Engineering Journal*, 2023,74:569-583.
- [19] LIU Z, CHENG J, LIU L, et al. Dual-stream cross-modality fusion transformer for RGB-D action recognition [J]. *Knowledge-Based Systems*, 2022,255:109741.
- [20] TANG M, ZHU L, XU Y, et al. Dual-stream reinforcement network for few-shot image segmentation[J]. *Digital Signal Processing*, 2023,134:103911.
- [21] CHEN J, YANG G, DING X, et al. Robust detection of dehazed images via dual-stream CNNs with adaptive feature fusion [J]. *Computer Vision and Image Understanding*, 2022,217:103357.
- [22] CHEN H, ZHANG P. A dual-stream deep attractor network with multi-domain learning for speech dereverberation and separation [J]. *Neural Networks*, 2021,141:238-248.
- [23] HE Q, PANG Y, JIANG G, et al. A spatio-temporal multiscale neural network approach for wind turbine fault diagnosis with imbalanced SCADA data [J]. *IEEE Transactions on Industrial Informatics*, 2021, 17(10): 6875-6884.
- [24] XU D, XIAO X, LIU J, et al. Spatio-temporal degradation modeling and remaining useful life prediction under multiple operating conditions based on attention mechanism and deep learning [J]. *Reliability Engineering & System Safety*, 2023,229:108886.
- [25] GAO H, LI Y, ZHAO Y, et al. Dual channel feature attention-based approach for RUL prediction considering the spatiotemporal difference of multisensor data [J]. *IEEE Sensors Journal*, 2023,23(8):8514-8525.
- [26] ZHU J, MA J, WU J. A regularized constrained two-stream convolution augmented Transformer for aircraft engine remaining useful life prediction[J]. *Engineering Applications of Artificial Intelligence*, 2024, 133:108161.
- [27] GUO J, HAN K, WU H, et al. CMT: Convolutional Neural Networks Meet Vision Transformers [C], 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022:12165-12175.
- [28] PENG Z, GUO Z, HUANG W, et al. Conformer: Local features coupling global representations for recognition and detection[J]. *IEEE Transactions on Pattern Analysis*

- and Machine Intelligence, 2023,45(8):9454-9468.
- [29] ASHISH V, NOAM S, NIKI P, et al. Attention is all you need[J]. ArXiv preprint arXiv:1706.03762, 2017.
- [30] YAN S L X J X Z C Y W X. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting [C]. Proceedings of the 33rd International Conference on Neural Information Processing Systems, 2019:5243-5253.
- [31] SAXENA A, GOEBEL K, SIMON D, et al. Damage propagation modeling for aircraft engine run-to-failure simulation [C]. 2008 International Conference on Prognostics and Health Management, 2008:1-9.
- [32] WANG B, LEI Y, LI N, et al. Multiscale convolutional attention network for predicting remaining useful life of machinery [J]. IEEE Transactions on Industrial Electronics, 2021,68(8):7496-7504.
- [33] DU X, JIA L, HAQ I. Fault diagnosis based on SPBO-SDAE and transformer neural network for rotating machinery[J]. Measurement, 2022,188:110545.
- [34] LI X, DING Q, SUN J. Remaining useful life estimation in prognostics using deep convolution neural networks [J]. Reliability Engineering & System Safety, 2018, 172: 1-11.
- [35] CHENG Y, WU J, ZHU H, et al. Remaining useful life prognosis based on ensemble long short-term memory neural network [J]. IEEE Transactions on Instrumentation and Measurement, 2021,70:1-12.
- [36] LIX, DING Q, SUN J. Remaining useful life estimation in prognostics using deep convolution neural networks [J]. Reliability Engineering & System Safety, 2018, 172: 1-11.
- [37] XIA J, FENG Y, TENG D, et al. Distance self-attention network method for remaining useful life estimation of aeroengine with parallel computing [J]. Reliability Engineering & System Safety, 2022,225:108636.
- [38] ZHANG J, JIANG Y, WU S, et al. Prediction of remaining

useful life based on bidirectional gated recurrent unit with temporal self-attention mechanism [J]. Reliability Engineering & System Safety, 2022,221:108297.

作者简介



朱江艳,2025 年于昆明理工大学获得博士学位,主要研究方向为时间序列分析与应用、故障预测与健康管理等。

E-mail:zhujiangyan1@stu.kust.edu.cn

Zhu Jiangyan received her Ph. D. degree from Kunming University of Science and

Technology in 2025. Her main research directions in time series analysis and applications, fault prediction, and health management.



马军(通信作者),2016 年于昆明理工大学获得博士学位,现为昆明理工大学教授,主要研究方向为工业预测与故障诊断。

E-mail:mjun@kust.edu.cn

Ma Jun (Corresponding author) received his Ph. D. degree from Kunming University of Science and Technology in 2016. Now he is a professor at Kunming University of Science and Technology. His main research directions are industrial prediction and fault diagnosis.



吴建德,现为云南大学教授,主要研究方向为机械设备剩余使用寿命预测、人工智能等。

E-mail:wjiande@kust.edu.cn

Wu Jiande now is a professor at Yunnan University. His main research interests include the prediction of the remaining useful life of mechanical equipment and artificial intelligence.



熊新,现为昆明理工大学高级工程师,主要研究方向为机械智能故障诊断。

E-mail:305428501@qq.com

Xiong Xin now is a senior engineer at Kunming University of Science and Technology. His main research interest includes mechanical intelligent fault diagnosis.