

多尺度局部特征和 Transformer 全局学习融合的 发动机剩余寿命预测

陈俊英¹ 席月芸¹ 李朝阳¹

摘要 飞机发动机剩余寿命 (Remaining useful life, RUL) 的准确预测对确保其安全性和可靠性至关重要。在基于多传感器检测数据预测时, 需解决局部特征提取问题以全面捕捉设备在不同时间尺度下的退化趋势, 并需解决时间序列中各元素之间长期依赖性的全局学习问题。因此, 提出了结合多尺度局部特征增强单元 (Multi-scale local feature enhancement unit, MSLFU_BLOCK) 和 Transformer 编码器的预测模型, 称之为 MS_Transformer。MSLFU_BLOCK 利用堆叠的因果卷积逐层从时间序列数据中提取多尺度局部信息, 同时避免了传统卷积计算中固有的未来数据泄漏问题。随后, Transformer 编码器通过其自注意机制进一步捕获时间序列数据中的短期和长期依赖关系。通过将多尺度局部特征增强单元与 Transformer 编码器相结合, 提出的 MS_Transformer 全面捕捉了时间序列数据中的局部和全局模式。在广泛使用的 C-MAPSS 基准数据集上进行的消融和预测实验验证了模型的合理性和有效性。与 13 个先进预测模型的比较分析表明, MS_Transformer 模型在操作条件更复杂的 FD002 和 FD004 数据集上的 RMSE 和 Score 指标优于其他模型, 同时在四个数据集上的平均性能最优。该研究为发动机剩余寿命预测提供了更为可靠的解决方案。

关键词 剩余寿命预测, 航空发动机, Transformer, 多尺度特征, 局部特征

引用格式 陈俊英, 席月芸, 李朝阳. 多尺度局部特征和 Transformer 全局学习融合的发动机剩余寿命预测. 自动化学报, 2024, 50(9): 1818–1830

DOI 10.16383/j.aas.c230634

Prediction of Aeroengine Remaining Life by Combining Multi-scale Local Features and Transformer Global Learning

CHEN Jun-Ying¹ XI Yue-Yun¹ LI Zhao-Yang¹

Abstract Accurate prediction of the remaining useful life (RUL) of aeroengine is crucial for ensuring their safety and reliability. In the process of predicting RUL based on multi-sensor detection data, it is necessary to address the issue of local feature extraction to comprehensively capture the degradation trends of equipment at different time scales, as well as the global learning problem of long-term dependencies among elements in the time series. Therefore, we propose a predictive model named MS_Transformer, which combines the multi-scale local feature enhancement unit (MSLFU_BLOCK) and a Transformer encoder. The MSLFU_BLOCK leverages stacked causal convolutional layers to progressively extract multi-scale local information from the time series data, simultaneously mitigating concerns related to future data leakage inherent in conventional convolutional computations. Subsequently, the Transformer encoder, with its self-attention mechanism, further captures short-term and long-term dependencies within the time series data. By integrating the MSLFU_BLOCK with the Transformer encoder, the proposed MS_Transformer comprehensively captures both local and global patterns within time series data. Extensive ablation and prediction experiments were performed on the widely utilized C-MAPSS benchmark dataset to validate the rationality and effectiveness of the proposed model. Comparative analyses with thirteen advanced prediction models demonstrate that the MS_Transformer model outperforms others, particularly on the more complex FD002 and FD004 datasets, based on RMSE and Score metrics. The average performance across all four datasets indicates the superiority of the proposed approach. The research provides a more reliable solution for predicting the RUL of engines.

Key words Remaining life prediction, aeroengine, Transformer, multi-scale features, local features

Citation Chen Jun-Ying, Xi Yue-Yun, Li Zhao-Yang. Prediction of aeroengine remaining life by combining multi-scale local features and Transformer global learning. *Acta Automatica Sinica*, 2024, 50(9): 1818–1830

收稿日期 2023-10-12 录用日期 2024-04-12

Manuscript received October 12, 2023; accepted April 12, 2024

国家自然科学基金 (62103316), 陕西省自然科学基金 (2023-JC-YB-562) 资助

Supported by National Natural Science Foundation of China (62103316) and Natural Science Foundation of Shaanxi

Province (2023-JC-YB-562)

本文责任编辑 诸兵

Recommended by Associate Editor ZHU Bing

1. 西安建筑科技大学信息与控制工程学院 西安 710055

1. College of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an 710055

预测与健康管理系统 (Prognostics and health management, PHM) 对于保障航空发动机的安全性和可靠性有重要意义, 在许多工业领域中受到广泛关注^[1]. PHM 包括异常检测、故障诊断、剩余使用寿命预测等. 其中, 剩余使用寿命 (Remaining useful life, RUL) 预测作为 PHM 的主要任务之一, 旨在利用设备的运行状态信息揭示设备性能退化规律, 从而预测其有效剩余寿命. 在航空发动机领域, 剩余寿命预测尤为重要. 航空发动机的运行环境极其苛刻, 长期运行过程中, 由于疲劳、磨损、腐蚀等因素的影响, 发动机性能会逐渐退化, 最终可能导致故障和事故. 因此, 精确预测航空发动机的剩余寿命对于确保航空安全至关重要.

目前发动机剩余使用寿命预测的主流方法主要有基于失效机理分析的方法、数据驱动的方法和两者融合的方法^[2]. 基于失效机理分析的方法依赖于数学模型和经验知识^[3], 这种方法易受到系统复杂性的影响. 随着传感器技术的进步, 数据驱动的方法逐渐成为主流, 这些方法利用收集到的大量实际运行数据, 通过机器学习算法来建立预测模型. 传统机器学习方法包括极限学习机^[4]、支持向量机 (Support vector machine, SVM)^[5]、随机森林和梯度提升决策树^[6]等. 尽管这些方法的训练较为容易, 但在特征提取方面常依赖于领域专家的先验知识来选择和提取与剩余寿命相关的特征. 同时特征工程和建立回归模型的过程相互独立, 限制了传统机器学习模型的性能.

随着数据量的持续增长和算力的持续提升, 当代数据驱动方法逐渐摒弃了传统的人工特征提取与机器学习相结合的方式, 转而采用深度学习技术实现自动特征提取和分类、回归等任务. 其中, 递归神经网络 (Recurrent neural network, RNN) 和长短期记忆 (Long short-term memory, LSTM) 深度学习模型专门用于捕获和建模时间序列数据中的时间依赖性. 它们擅长保留过去的信息并将其传播到后续步骤, 增强对序列中模式的理解和预测. Zheng 等^[7]介绍了一种基于 LSTM 的 RUL 估计方法, 利用传感器序列信息揭示传感器数据中隐藏的模式. Huang 等^[8]采用了双向 LSTM 网络来预测系统的剩余寿命. YU 等^[9]提出了一种基于双向 RNN 和自编码器的相似曲线匹配方法来估计机械系统的 RUL, 目的是提高 RUL 估计的鲁棒性和准确性. Liu 等^[10]提出了一种结合聚类分析和 LSTM 的航空发动机 RUL 预测模型. 尽管 RNN 和 LSTM 在处理时间序列数据方面表现出色, 但它们面临着并行化的挑战和处理长序列困难等局限性.

许多研究尝试利用深度卷积神经网络 (Deep convolutional neural networks, DCNN) 的高度并行计算和局部特征提取能力来处理时间序列数据. Li 等^[11]引入了一种使用 DCNN 进行预测的新型数据驱动方法. 原始收集的数据经过标准化处理后, 直接输入到 DCNN, 不需要事先具备预测或信号处理方面的专业知识. Yang 等^[12]提出了一种基于两个卷积神经网络的 RUL 预测方法, 一个作为监测初始性能退化的分类网络, 另一个作为预测剩余寿命的网络. Li 等^[13]提出了一种基于多传感器数据的集成深度多尺度特征融合网络 (Integrated deep multi-scale feature fusion network, IDMFFN). 使用不同大小的卷积滤波器来学习不同尺度的特征, 然后将这些多尺度特征进行连接, 并利用基于门控循环单元 (Gated recurrent unit, GRU) 的高级特征融合块进行 RUL 预测. 尽管 CNNs 在某些方面具有优势, 但在时序任务中仍存在一定的局限性, 因为它们倾向于专注局部特征提取, 这对需要考虑全局依赖关系的时间预测任务而言是不够的.

为了结合 LSTM 的时间建模能力和 CNN 的空间特征提取能力, Al-Dulaimi 等^[14]提出了一种用于 RUL 估计的混合深度神经网络模型 (Hybrid deep neural network, HDNN). 该深度学习模型将 LSTM 和 CNN 并行集成, 利用 LSTM 路径提取时间特征, 同时利用 CNN 提取空间特征. 在 C-MAPSS (Commercial modular aero-propulsion system simulation) 数据集^[15]上验证了该模型的优越性. Ayodeji 等^[16]提出利用扩展卷积来扩大感受野, 增强时间序列的全局特征提取. 然后采用增强的 LSTM 网络捕获长期依赖关系, 并进一步提取代表不同操作条件的多尺度特征. Li 等^[17]提出了一种基于卷积块注意模块 (Convolutional block attention module, CBAM) 的改进 CNN-LSTM 模型. 利用多层 CNN 提取飞机发动机运行数据的特征, 然后利用 CBAM 处理通道和空间维度上的注意机制, 识别与 RUL 相关的关键变量. 最后, LSTM 学习特征和服务时间之间的隐藏关系, 进而实现预测. 这些 CNN-LSTM 融合模型在捕捉发动机数据中的复杂模式方面, 相较于单独的 CNN 或 LSTM 模型, 表现出了明显的优势.

注意力机制的引入使得深度学习模型能够有效地专注于 RUL 预测任务的重要数据特征, 从而提高航空发动机预测的准确性. Liu 等^[18]直接将特征注意机制应用于输入数据, 在训练过程中为更重要的特征动态分配更大的注意权重. 随后, 采用双向门控循环单元 (Bidirectional gated recurrent

units, BGRU) 从加权输入数据中提取长期依赖关系, CNN 从 BGRU 的输出序列中捕获局部特征. 最后, 全连接网络用于特征抽象表示和预测剩余寿命. Xu 等^[19] 提出了一种并行的一维 CNN 和池化层来从多个信号中提取和融合特征. Zhang 等^[20] 引入了一种具有时间自注意机制的双向 GRU (Bidirectional gated recurrent unit-temporal self attention mechanism, BiGRU-TSAM) 用于 RUL 预测, 每个时间步根据其重要性被分配一个自学习权值. Wang 等^[21] 提出了一种完全基于注意机制的模型, 利用多头自注意力提取时间序列之间的依赖关系. Xu 等^[22] 将全局注意与自注意机制和时间卷积网络 (Temporal convolutional network, TCN) 相结合, 提出了一种端到端的深度学习 RUL 预测方法. Zhao 等^[23] 提出了一种多尺度集成深度自注意网络 (Multi scale integrated deep self network, MSIDSN), 该网络使用多尺度块结合自注意策略在不同尺度上选择性地提取特征. RNN 模块用于提取退化特征, 然后融合这些特征以准确预测飞机发动机的剩余寿命. Zhu 等^[24] 提出了一种旋转机械剩余寿命预测方法, 该方法使用具有自注意机制的残差混合网络构建健康指标. 为了解决长序列预测任务中提取局部和全局特征的挑战, Li 等^[25] 提出了一种多任务时空增强网络 (Multi task spatio-temporal augmented net, MTSTAN). 该算法利用通道关注机制增强不同传感器数据的局部特征, 采用带跳跃连接的因果增强卷积网络实现时间序列上的全局特征提取. 实验结果表明, 增强局部和全局时间序列特征能有效提高预测精度. Zhang 等^[26] 提出了一种基于注意力机制的时间卷积网络 (Attention based temporal convolutional network, ATCN) 用于剩余寿命预测. 采用了改进的自注意机制对不同时间步的输入特征进行加权, 同时使用时间卷积网络捕捉长期依赖关系并基于加权特征提取特征表示.

随着研究的深入, 学者们开始探索更为复杂和高效的时间序列预测方法. Transformer 架构作为一种自注意力机制模型, 最初是为自然语言处理任务设计的^[27], 已经有效扩展到时间序列数据^[28-29] 的处理. Zerveas 等^[30] 利用 Transformer 的多头注意机制同时考虑来自多个时间步长的上下文信息, 使模型能够学习不同的注意模式, 以改进对多变量时间序列数据的处理. Li 等^[31] 观察到 Transformer 架构的点积自注意机制存在局限, 特别是其对局部上下文的敏感性不足. 为了解决这个问题, 提出了卷积自注意机制, 引入查询和键来更好地将局部上下文融入注意机制. 进一步开发了 LogSparse Trans-

former, 以提高对具有细粒度和强烈长期依赖性的时间序列的预测精度. 认识到 Transformer 在局部特征提取方面存在局限性, Mo 等^[32] 将 Transformer 编码器作为预测模型的核心, 以捕捉时间序列中的短期和长期依赖关系. 通过引入门控卷积单元增强模型在每个时间步中融入局部上下文的能力. 这些改进旨在增强模型捕捉全局依赖性和复杂局部特征的能力, 使其更适用于时间序列数据建模. Guo 等^[33] 提出了一种在飞机发动机中进行 RUL 预测的新方法. 使用一维卷积网络设计了一个具有沙漏形状结构的多尺度特征提取器. 随后, 使用一个增加了金字塔注意机制的 Transformer 编码器和解码器, 进一步从融合的多尺度特征中提取特征进行 RUL 预测. 这种方法旨在充分发挥卷积和 Transformer 架构的优势, 更有效地进行 RUL 预测中的特征提取.

为了解决 Transformer 在应用于时间序列数据时局部特征提取的局限性, 提出了一种新颖的预测模型, 命名为 MS_Transformer. 该模型将多尺度局部特征增强单元与 Transformer 编码器结合, 以有效捕捉数据中的局部和全局依赖关系. 多尺度局部特征增强单元通过堆叠多个因果卷积层组合不同时间尺度的特征. 通过考虑各个尺度的时间上下文, 增强了更具信息性的特征表示的整合. 因果卷积用于从时间序列中提取局部特征, 确保卷积核的感知域仅覆盖过去的时间步, 防止未来数据泄露. 与此同时, Transformer 编码器利用注意机制同时处理整个数据序列. 这使得模型能够捕捉时间序列中的长期依赖关系, 提供对时间模式的全面理解. 在四个广泛使用的基准数据集上进行了大量实验, 验证了 MS_Transformer 方法的有效性和鲁棒性.

本文的主要工作有以下几点:

1) 设计了一个多尺度局部特征增强单元, 用于在多个尺度上提取特征. 该单元增强了模型描述局部特征的能力, 使其能够捕捉不同粒度级别的信息. 在局部特征提取过程中采用了因果卷积操作, 确保模型仅依赖于历史数据进行预测, 解决了潜在的未来自数据泄漏问题.

2) 提出了一种用于预测飞机发动机剩余寿命的预测模型. 该模型巧妙地融合了多尺度局部特征与 Transformer 全局学习的优势, 从而能更全面地捕捉数据中细粒度局部细节与全局模式.

3) 在 C-MAPSS 基准数据集上进行了大量的消融和预测实验, 验证了 MS_Transformer 模型的合理性和有效性, 展示了其准确预测 RUL 的能力.

1 基于 MS_Transformer 的剩余寿命预测方法

1.1 MS_Transformer 预测模型

本文提出的 MS_Transformer 剩余寿命预测模型的结构如图 1 所示, 主要包括多尺度局部特征增强单元 (MSLFU_BLOCK)、Transformer 编码器和全连接输出层. 表征发动机状态的传感器监测数据通过数据归一化等预处理, 采用滑动窗口方法提取时间窗数据, 输入到 MSLFU_BLOCK 提取时间序列数据中的局部特征模式. MSLFU_BLOCK 采用因果卷积逐层提取多尺度局部信息, 确保模型仅使用历史数据而不会利用未来数据, 克服以往单一尺度固定窗口的局限性并获得多尺度信息的建模能力. 将 Transformer 编码器作为骨干网络进一步捕捉时间序列中的短期和长期依赖关系, 以提高模型的预测性能. 全连接输出层将 Transformer 编码器学习到的特征映射成输出预测值. 接下来将详细描述每个模块的结构和作用.

1.2 因果卷积

时间序列数据具有严格的时间顺序性, 其中每

个数据点的前后顺序对于了解其因果关系至关重要. 传统卷积并未针对时间顺序进行优化, 其卷积核可以访问时间序列中未来的数据点, 导致未来数据泄漏问题. 为了避免模型从未来的数据点中获取信息进而给出不合理的预测, 本文采用因果卷积来处理时序数据, 即限制卷积核只能访问过去的点, 从而确保模型在建模时间序列时不会依赖未来数据. 两种卷积运算过程示意图如图 2 所示. 图中 Input 为输入时间序列层, Hidden layer 对输入层进行一次卷积运算, Output 层对 Hidden layer 的输出又进行一次卷积, 可以看到 Hidden layer 和 Output 层的输出对应于不同感受野、不同时间尺度的特征. 右图示意的因果卷积只使用历史数据, 而左图示意的传统卷积用到了未来数据. 假设输入序列为 \mathbf{X} , 卷积核为 \mathbf{h} , 使用一次因果卷积后的输出可表示为式 (1).

$$\mathbf{y}_t = (\mathbf{X} * \mathbf{h})_t = \sum_{i=0}^{k-1} \mathbf{h}_i \times \mathbf{X}_{t-i} \quad (1)$$

其中, $*$ 表示卷积运算. \mathbf{y}_t 表示输出序列的第 t 个元素. \mathbf{X}_{t-i} 表示输入序列的第 $t-i$ 个元素, i 取值从 0 到 $k-1$, k 表示卷积核的大小, \mathbf{h}_i 表示卷积核的第 i 个权重. 如不考虑批大小, \mathbf{X} 的维度为 $[\text{win-}$

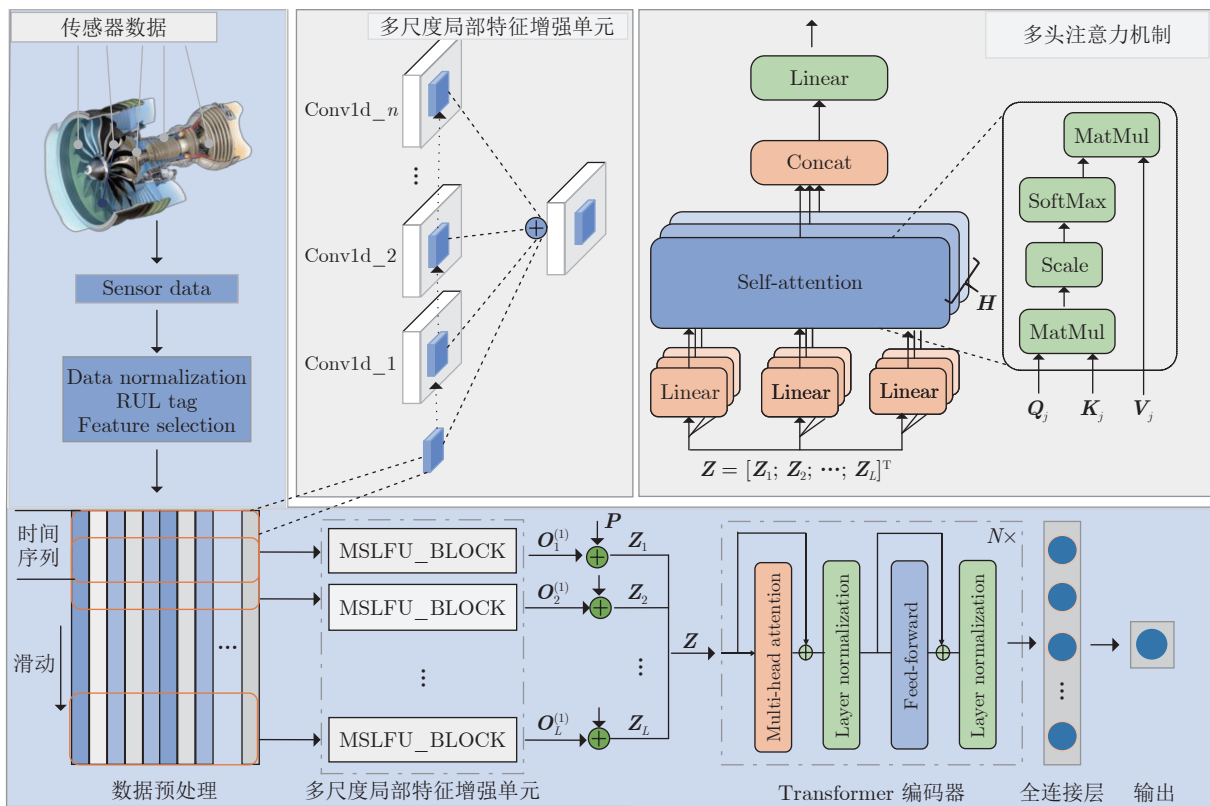


图 1 MS_Transformer 剩余寿命预测模型结构图

Fig. 1 Architecture of the MS_Transformer remaining life prediction model

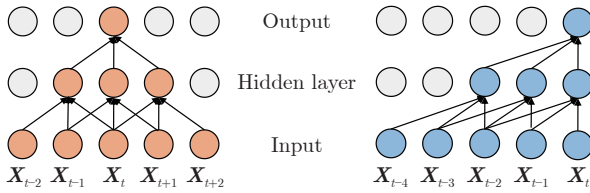


图 2 传统卷积与因果卷积示意图

Fig.2 Diagrams of traditional convolution and causal convolution

dow_length, feature_size], 卷积核 h 的维度为 $[d, \text{feature_size}, k]$, 输出 y 的维度为 $[\text{window_length}, d]$, 其中 window_length 为时间序列窗口长度, feature_size 为传感器的特征数, d 是 Transformer encoder 的编码长度. 在边缘位置进行适当填充, 保证输出和输入维度一致. 随着神经网络的深度增加, 因果卷积神经网络中的每一层会逐渐扩大其感受野, 以更全面地捕捉来自前一层神经元的历史信息. 在图 2 所示的因果卷积神经网络中, 通过堆叠 2 个卷积核大小为 3 的因果卷积层来构建. 经过 2 次卷积操作后, 输出将包含一个感受野跨度为 5 的历史信息范围.

1.3 多尺度局部特征增强单元

传统的卷积操作擅长提取图像等类型数据的局部信息, 对于时间序列数据处理的适应性较差. 为了从时间序列数据中提取局部特征, 本文引入了一维因果卷积以捕捉相邻时间点之间的相关局部模式, 通过多层因果卷积获得不同时间尺度的特征.

进一步将这些特征组合起来得到多尺度局部特征描述, 这一处理模块被称为多尺度局部特征增强单元 (MSLFU_BLOCK). MSLFU_BLOCK 的原理示意图如图 3 所示, 展开的卷积示意图如图 3(a) 所示, 上一层的输出用作下一层的输入, 不同因果卷积层提取了不同的时间尺度特征. 将每层提取的特征相加作为 MSLFU_BLOCK 的输出. 具体的运算过程如图 3(b) 所示, 其中 Conv1 表示因果卷积运算, Chomp 模块通过零填充的方法实现输出序列与输入序列的长度相同.

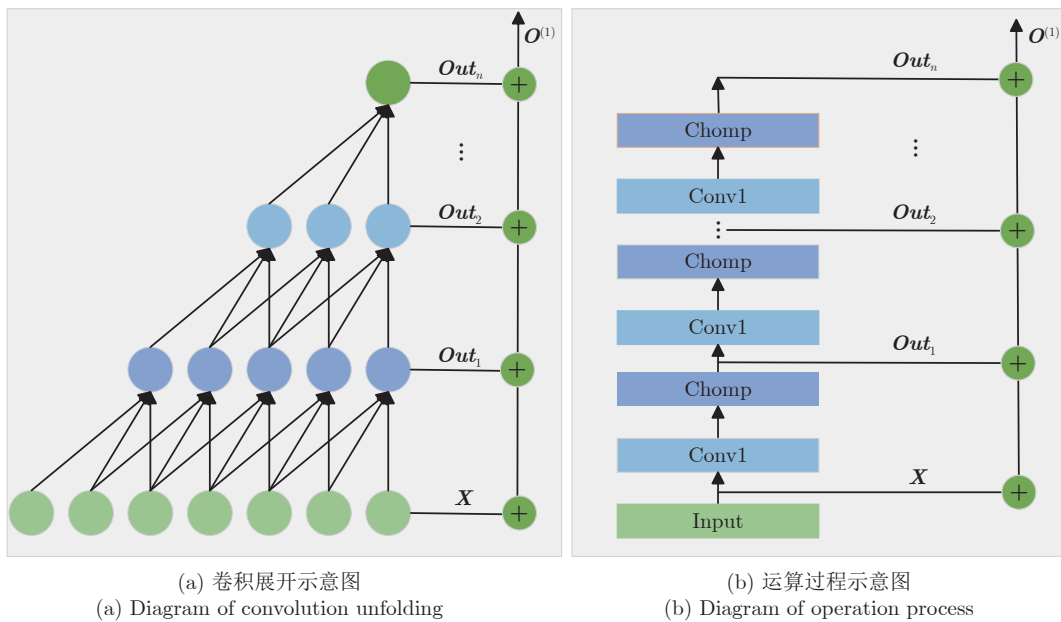
最终, MSLFU_BLOCK 的输出 $O^{(1)}$ 是对应卷积结果 Out_1 到 Out_n 的累加, 并加上输入 X , 计算过程表示为式 (2).

$$O^{(1)} = X + \sum_{i=1}^n Out_i \quad (2)$$

综上所述, 多尺度局部特征增强单元结合了一维因果卷积和层次结构的设计, 用于处理时间序列建模任务, 能够捕捉时间模式的局部相关性. 它具有三个显著特点: 1) 输出包含了时间序列的多尺度局部特征; 2) 确保不会有未来数据泄露问题; 3) 可接受任意长度的时间序列作为输入, 在映射至下一层时保持输出序列与输入序列长度相同.

1.4 Transformer 编码器

在本研究中, 使用 Transformer 编码器对多尺度局部特征增强单元提取的特征进行进一步处理. 编码器由 N 个相同的 Block 组成, 如图 1 所示.



(a) 卷积展开示意图

(a) Diagram of convolution unfolding

(b) 运算过程示意图

(b) Diagram of operation process

图 3 多尺度局部特征增强单元示意图

Fig.3 Diagrams of multi-scale local feature enhancement unit

1.4.1 位置编码

Transformer 编码器模型引入了位置编码 \mathbf{P} , 将序列中的时间先后信息融入到模型中. 位置编码是一个与输入序列维度相同的矩阵. 通过将位置编码与 MSLFU_BLOCK 的输出相加, 模型能够学习到序列中不同位置的相对关系. \mathbf{P} 中的元素使用式 (3) 和 (4) 计算.

$$\mathbf{P}_{(i,2s)} = \sin\left(\frac{i}{10000^{2s/d}}\right) \quad (3)$$

$$\mathbf{P}_{(i,2s+1)} = \cos\left(\frac{i}{10000^{2s/d}}\right) \quad (4)$$

其中, i 表示序列中的位置, 从 1 开始递增. s 表示维度的索引, 从 0 开始递增. d 是 MSLFU_BLOCK 输出的特征维度.

1.4.2 多头注意力

在多头注意力中, MSLFU_BLOCK 的输出和位置编码叠加后的序列通过多个独立的注意力头将特征映射到不同的子空间, 每个头都学习不同的关注权重, 以捕捉时间序列中不同的关系和特征. 每个注意力头的核心是自注意力机制 (Self-attention), 此时输入 \mathbf{Z} 为 $\mathbf{O}^{(1)} + \mathbf{P}$, 输入分别乘以第 j 个头查询、键和值的权重矩阵 $\mathbf{W}_j^q, \mathbf{W}_j^k, \mathbf{W}_j^v \in \mathbf{R}^d$ 得到相应的 $\mathbf{Q}_j, \mathbf{K}_j, \mathbf{V}_j$, 计算如式 (5) ~ (7) 所示.

$$\mathbf{K}_j = \mathbf{Z}\mathbf{W}_j^k \quad (5)$$

$$\mathbf{V}_j = \mathbf{Z}\mathbf{W}_j^v \quad (6)$$

$$\mathbf{Q}_j = \mathbf{Z}\mathbf{W}_j^q \quad (7)$$

然后引入缩放的点积注意力机制, 计算如式 (8) 所示.

$$\text{Attention}(\mathbf{Q}_j, \mathbf{K}_j, \mathbf{V}_j) = \text{softmax}\left(\frac{\mathbf{Q}_j\mathbf{K}_j^T}{\sqrt{d_k}}\right)\mathbf{V}_j \quad (8)$$

其中, d_k 是查询和键向量的维度.

多个注意力头将输入映射到不同的子空间进行自注意力机制运算, 输出被合并得到最终的多头注意力输出. 最终多头注意力的输出可表示为式 (9).

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}\left(\{\text{head}_j\}_{j=1}^H\right)\mathbf{W}^\circ \quad (9)$$

其中, $\text{head}_j = \text{Attention}(\mathbf{Q}_j, \mathbf{K}_j, \mathbf{V}_j)$, $\mathbf{W}^\circ \in \mathbf{R}^{Hd_k \times d}$, H 为注意力头个数, $d_k = d/H$.

1.4.3 前馈网络部分

前馈网络是 Transformer 模型中的另一个重要

组件, 由两个线性变换单元和一个非线性激活函数 (通常是 ReLU) 组成. 其输入 \mathbf{R} 为 $\text{LayerNorm}(\mathbf{Z} + \text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}))$, LayerNorm 表示层归一化操作. 前馈网络的计算过程可通过式 (10) 来表示.

$$\text{FFN}(\mathbf{R}) = \mathbf{W}_2 \cdot \text{ReLU}(\mathbf{W}_1\mathbf{R} + b_1) + b_2 \quad (10)$$

其中, \mathbf{W}_1 和 \mathbf{W}_2 为权值矩阵, b_1 和 b_2 为偏置向量. 再经过残差连接和层归一化后的输出 $\mathbf{O}^{(2)}$ 为式 (11).

$$\mathbf{O}^{(2)} = \text{LayerNorm}(\mathbf{R} + \text{FFN}(\mathbf{R})) \quad (11)$$

1.5 RUL 估计

最后通过全连接层将 Transformer 编码后的特征向量映射成 RUL 估计值. 此时, 经过 N 个 BLOCK 编码器的输出表示为 $\mathbf{O}^{(2N)}$, $\mathbf{O}^{(2N)} \in \mathbf{R}^d$. 则全连接层的输出可通过式 (12) 表示.

$$y = \text{Sigmoid}\left(\mathbf{W}_0\mathbf{O}^{(2N)} + b_0\right) \quad (12)$$

其中, y 是输入序列的 RUL 估计值, 应用 Sigmoid 函数将模型的输出限定在 0 和 1 之间, \mathbf{W}_0 和 b_0 是需要优化的参数. 此时得到的 RUL 值乘以 125 实现反归一化, 即变换到原空间评估其预测性能.

2 实验分析

本节对提出的 MS_Transformer 模型在 C-MAPSS 基准数据集上的性能进行全面评估. 随后, 将详细介绍数据预处理步骤、使用的评估指标、参数设置、消融实验的设计, 以及与先进方法的比较结果. 所有实验均在一台配备了 Intel(R) Core(TM) i5-1035G1 CPU (4 核处理器) 和 NVIDIA GeForce MX350 显卡的计算机上进行. 使用 PyTorch 框架进行模型的训练和测试. 损失函数采用了预测 RUL 值与实际值之间的均方误差.

2.1 数据预处理

实验数据来源于 NASA 的 C-MAPSS^[15] 航空发动机仿真模型. 该数据集由美国国家航空航天局提供, 且被广泛用于验证工业剩余寿命预测模型性能的研究. 该数据集包含 FD001、FD002、FD003 和 FD004 四个不同的子集, 各个子集的参数设置和操作条件有所不同, 模拟了发动机在不同情况下的性能和健康状态, 以便评估 RUL 预测模型在各种挑战条件下的性能, 其相关参数如表 1 所示.

每个子集由训练集和测试集组成, 其中训练集记录了航空发动机的多个状态参数从正常到故障的完整周期内每个时间序列的采样值; 测试集则包含了故障前某一时间点之前的状态参数值以及对应的剩余寿命. 属性包含发动机编号、循环次数、操作条

表 1 C-MAPSS 数据集的属性
Table 1 Attributes of the C-MAPSS dataset

参数	FD001	FD002	FD003	FD004
训练集中发动机个数	100	260	100	249
测试集中发动机个数	100	259	100	248
操作条件	1	6	1	6
错误模式	1	1	2	2
训练集大小	20632	53760	24721	61250
测试集大小	13097	33992	16597	41215

件和 21 个传感器测量数据, 共计 26 个数据字段. 以 FD001 子集为例, 传感器 1、5、6、10、16、18 和 19 的测量值在整个使用寿命期间都保持恒定, 因此它们不提供有关 RUL 预测的有用的退化信息. 最终选择了与发动机剩余寿命相关的 14 个传感器特征作为预测输入, 包括低压压气机出口总温、高压压气机出口总温、低压涡轮出口总温、高压压气机出口总压、风扇物理转速、核心机物理转速、高压压气机出口静压、燃油流量与高压压气机出口总压比值、

风扇换算转速、涵道比、引气焓值、高压涡轮冷却引气流量、低压涡轮冷却引气流量和使用周期. 其他三个数据集采用相同的特征选择方法.

用于预测建模的 14 个传感器数据均采用最大最小归一化方法. 这种处理有助于后续的数据操作和网络训练过程收敛. 在实验中, 分别在四个数据集 (FD001、FD002、FD003 和 FD004) 中随机选择一个发动机, 并将其监测参数值的变化曲线展示在图 4 中. 从图中可以看出, 不同数据集中的发动机状态参数的变化趋势各异, 因此在这些参数上建立统一的预测模型是一项具有挑战性的任务.

考虑到发动机初期性能较为稳定, 退化趋势不明显, 运行一段时间后发生退化. 据此, 使用分段线性回归模型, 如 Zheng 等^[7]将设备的剩余寿命划分为常数阶段和线性递减阶段. 根据目前已有的研究成果, 其中常数阶段的剩余寿命值设置为 125.

2.2 评估指标

为了评价预测性能, 使用发动机剩余寿命预测

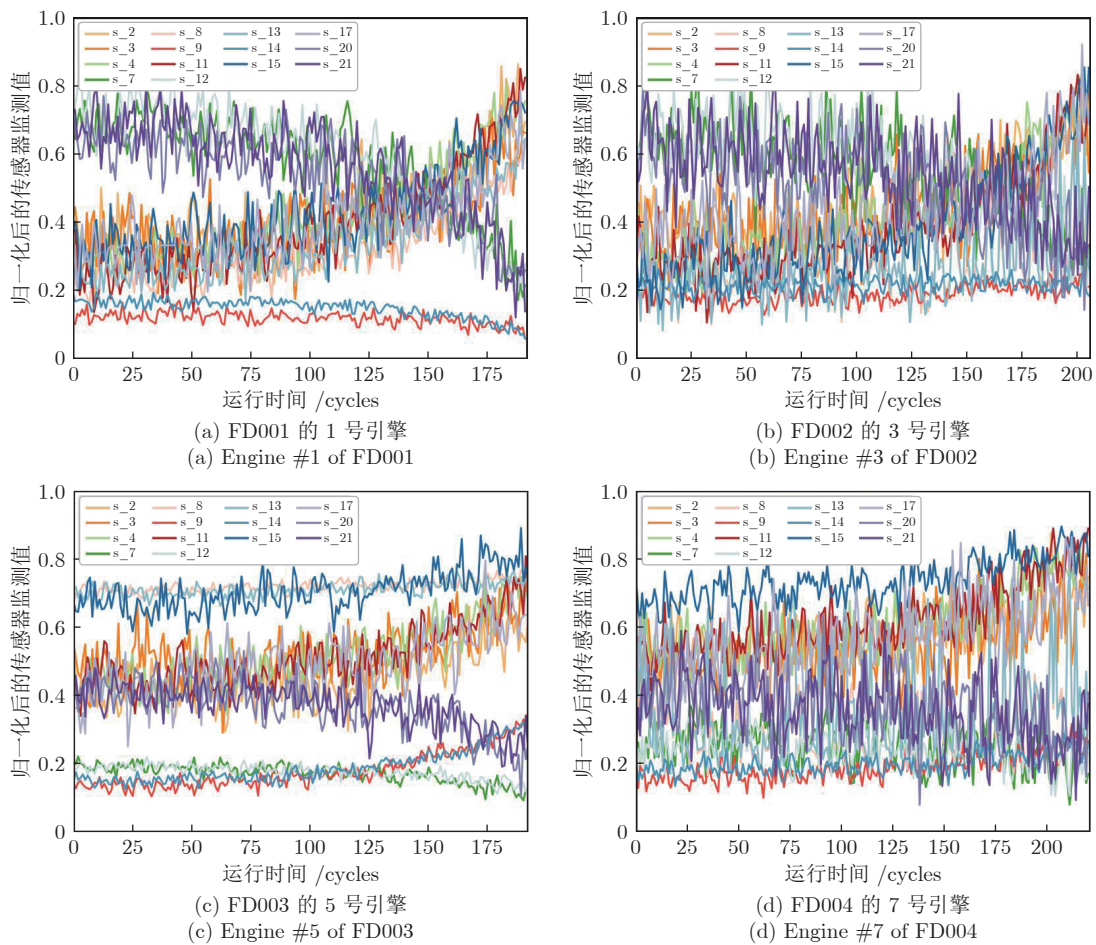


图 4 状态参数变化曲线示意图

Fig. 4 Illustration of state parameter variation curve

和回归算法评价中常用的惩罚得分和均方根误差 (Root mean square error, RMSE) 指标。

RMSE 表示预测值和真实值之间残差的样本标准差, 用于衡量预测结果的离散程度, 其计算如式 (13) 所示。

$$RMSE = \sqrt{\frac{1}{|D_{\text{test}}|} \sum_{x_i \in D_{\text{test}}} (y_i - \hat{y}_i)^2} \quad (13)$$

其中, y_i 为第 i 个测试发动机 x_i 对应的真实寿命, \hat{y}_i 为对应的预测寿命, D_{test} 表示测试集, $|D_{\text{test}}|$ 为测试样本个数。将发动机剩余寿命预测误差定义为 $\Delta y_i = \hat{y}_i - y_i$ 。根据预测误差的大小, 将预测结果分为提前预测、滞后预测和准确预测 3 种, 当 $\Delta y_i < -13$ 时对应于提前预测, 当 $\Delta y_i > 10$ 时对应于滞后预测, 当 $-13 \leq \Delta y_i \leq 10$ 时对应于准确预测。考虑到滞后预测比提前预测带来的后果更为严重, 惩罚得分 Score 指标对滞后预测进行了惩罚, 其计算方法如式 (14) 所示。

$$Score = \begin{cases} \sum_{i=1}^{|D_{\text{test}}|} e^{-\left(\frac{\Delta y_i}{13}\right)} - 1, & \Delta y_i < 0 \\ \sum_{i=1}^{|D_{\text{test}}|} e^{\left(\frac{\Delta y_i}{10}\right)} - 1, & \Delta y_i \geq 0 \end{cases} \quad (14)$$

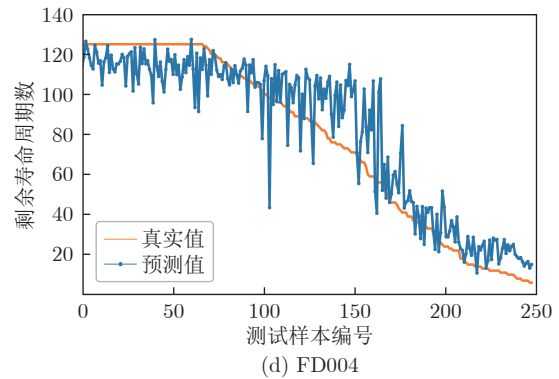
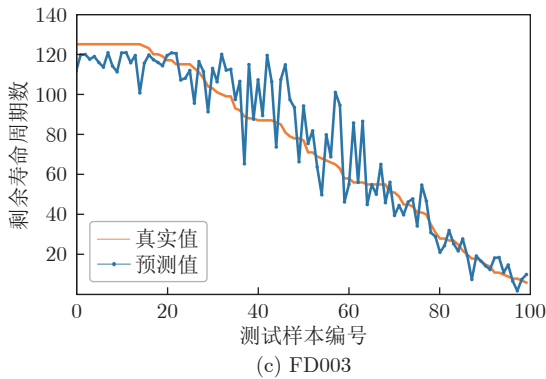
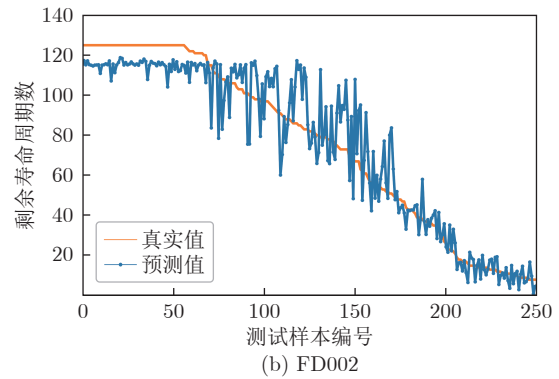
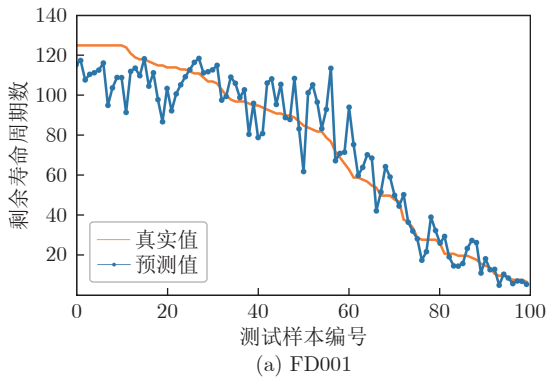


图 5 测试集上预测值与真实值对比图

Fig. 5 Comparison of predicted values and ground truth on the test set

2.3 超参数设置

在训练过程中, 为了展示所提模型的适应性, 在四个数据集上均采用统一的参数设置: 批量大小 (Batch size) 为 8, 学习率 (Learning rate) 为 0.002, 训练周期 (Training epoch) 为 30, 丢失率 (Dropout rate) 为 0.01, 时间窗口大小 (Time window size) 为 60, 编码器层数为 1, 因果卷积层数为 2。对于长度小于 60 的发动机记录, 采用线性插值方法对时间序列进行扩展, 使其满足所提模型最小输入长度为 60 的需求。

2.4 结果验证

以下从两个方面测试所提预测模型在 C-MAPSS 数据集上的性能。

1) 评估整个测试集上所有发动机的 RUL 预测性能

首先, 使用训练好的 MS_Transformer 模型在 C-MAPSS 四个测试集上进行了 RUL 预测。为了更好地观察预测结果, 将发动机编号按照剩余寿命递减的顺序重新排列, 图 5 展示了模型的预测结果与实际 RUL 之间的对比。横轴表示重新排序后的发动机测试样本编号, 纵轴表示 RUL 值。

从图 5 可以看出, 模型的 RUL 预测值与实际

RUL 在大多数情况下具有一致性. 这表明模型对于 RUL 的趋势有着良好的捕捉能力.

基于上述实验所用的计算资源, 本文所提方法在 FD001 ~ FD004 四个数据集上的训练时间分别为 0.51 h、2.97 h、0.67 h 和 3.69 h. 同时, 在四个数据集上的平均预测时间 (包括数据预处理时间) 为 32 ms. 对于发动机寿命预测这类实时性要求不高的应用场景, 预测时间完全满足需求.

2) 在单个发动机的全测试谱上评估模型

除了用上述评估方法验证模型性能外, 还对模型在不同数据库不同操作条件下的发动机全测试谱上的性能进行评估. 图 6(a) ~ 6(d) 展示了从测试数据集中随机选择的发动机的模型预测性能, 其中“Predicted RUL”表示模型的预测结果, 而“Actual RUL”则表示 C-MAPSS 提供的真实 RUL 值. 目的是验证模型在整个发动机使用寿命期间的预测表现, 更全面地考察模型在不同数据情境下的可靠性和泛化能力.

从图 6(a) ~ 6(d) 中可以观察到, 在分段线性近似方法的早期寿命值设置为恒定值时, 模型预测值也比较平稳; 当发动机的剩余使用寿命出现下降趋势时, 模型的预测结果也趋于下降; 直观上验证了预测模型的有效性. 在发动机寿命由恒定值转变为

线性下降时, 预测值存在一定的偏差. 然而, 随着发动机寿命的递减, 可以观察到模型的 RUL 估计逐渐改善, 这一趋势在曲线尾部更为显著. 这是因为随着发动机接近寿命终点, 退化特征更为显著, 模型能够更准确地捕捉数据的模式, 从而实现更准确的预测结果, 这一结果符合当发动机剩余使用寿命越短时期望预测结果越准确的需求.

2.5 与先进方法的比较

为了对所提方法进行定量评估, 选择了 13 种先进的深度学习方法进行对比实验, RMSE 和 Score 指标值如表 2 所示. 通过对比实验结果可以看出, 在较复杂的 FD002 和 FD004 数据集上, MS_Transformer 方法取得了最佳性能, 这表明 MS_Transformer 针对复杂数据的建模能力较强. 而且, 在四个数据集上的 RMSE 和 Score 的平均值也表现出最佳. 与之前最先进方法 MHT 相比, 该方法将卷积和 Transformer 架构的优势结合起来, 以增强在 RUL 预测中的特征提取能力, MS_Transformer 将 RMSE 指标降低了 7.04%, 在 Score 指标上降低了 20.63%. 所提出的 MS_Transformer 模型与 GCU_Transformer 模型均以 Transformer 的编码器作为主干网络, 在 FD002 数据集上, MS_

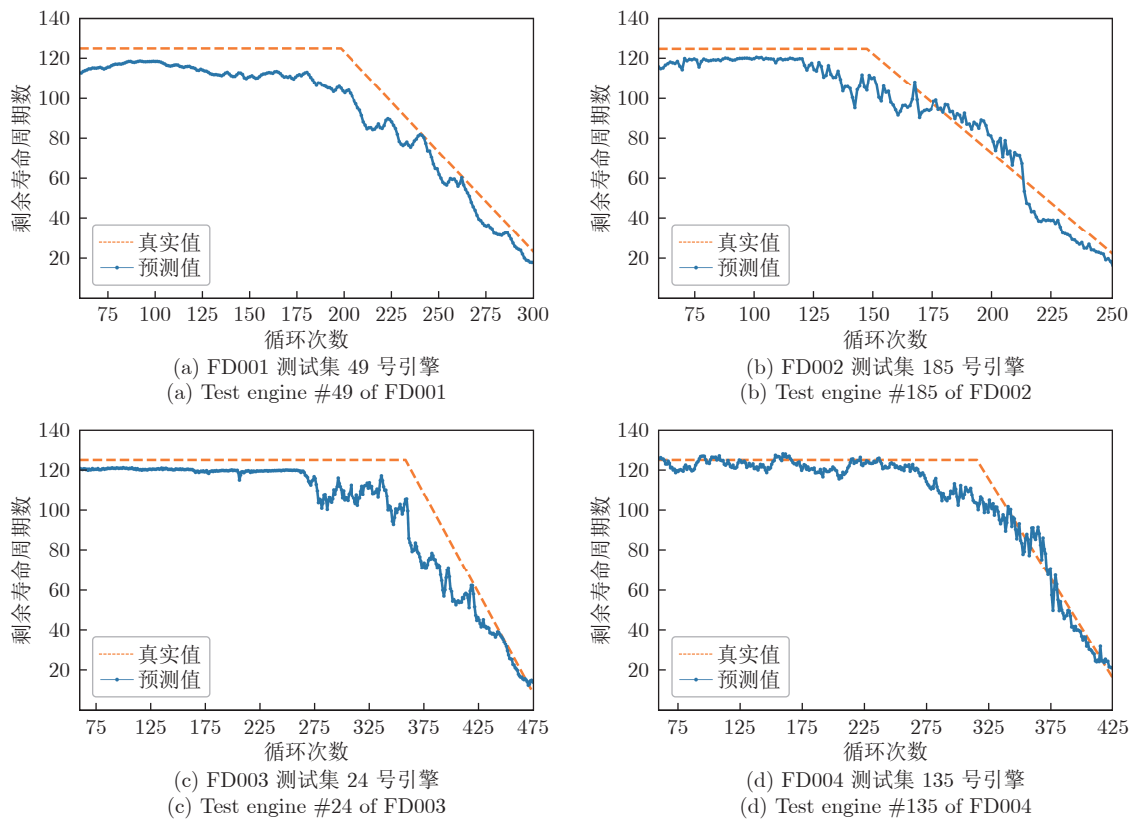


图 6 发动机剩余使用寿命全周期预测评估

Fig.6 Full lifecycle predictive evaluation of engine remaining useful life

表 2 与先进方法相比较
Table 2 Comparison with state-of-the-art methods

方法	FD001		FD002		FD003		FD004		Average	
	RMSE	Score	RMSE	Score	RMSE	Score	RMSE	Score	RMSE	Score
LSTM (2017) ^[7]	16.14	338.00	24.49	1718.00	16.18	852.00	28.17	2238.00	21.25	1286.50
DCNN (2018) ^[11]	12.61	274.00	22.36	4020.00	12.64	284.00	23.31	5027.00	17.73	2401.25
HDNN (2019) ^[14]	13.02	245.00	15.24	1282.42	12.22	287.72	18.16	1527.42	14.66	835.64
AGCNN (2021) ^[18]	12.42	225.51	19.43	1492.00	13.39	227.09	21.50	3392.00	16.68	1334.15
GCU_Transformer (2021) ^[32]	11.27	—	22.81	—	11.42	—	24.86	—	17.59	—
BiGRU-TSAM (2022) ^[20]	12.56	213.35	18.94	2264.13	12.45	232.86	20.47	3610.34	16.10	1580.17
IDMFFN (2022) ^[13]	12.18	204.69	19.17	1819.42	11.89	205.54	21.72	3338.84	16.24	1392.12
MTSTAN (2023) ^[24]	10.97	175.36	16.81	1154.36	10.90	188.22	18.85	1446.29	14.38	741.06
Encoder-Attention (2023) ^[21]	10.35	183.75	15.82	1008.08	11.34	219.63	17.35	1751.23	13.72	790.67
MSIDSN (2023) ^[23]	11.74	205.55	18.26	2046.65	12.04	196.42	22.48	2910.73	16.13	1339.83
ATCN (2024) ^[26]	11.48	194.25	15.82	1210.57	11.34	249.19	17.80	1934.86	14.11	897.22
MHT (2024) ^[33]	11.92	215.20	13.70	746.70	10.63	150.50	17.73	1572.00	13.50	671.10
MachNet (2024) ^[34]	11.04	176.82	24.52	3326.00	10.59	161.26	28.86	5916.00	18.75	2395.02
Ours	11.79	224.36	11.98	608.88	11.95	225.05	14.47	1072.38	12.55	532.67

Transformer 对应的 RMSE 降低了 47.48%; 在 FD004 数据集上, RMSE 降低了 41.79%. GCU_Transformer 模型在 FD001 和 FD003 数据集上采用了和 FD002 和 FD004 数据集不同的超参数设置, 而 MS_Transformer 模型在四个数据集上采用了相同的超参数设置. MS_Transformer 模型相比 GCU_Transformer 模型在 FD001 和 FD003 数据集上的 RMSE 分别增加了 4.61% 和 4.64%, 但在四个数据集上的平均 RMSE 降低了 28.65%, 结果说明了 MS_Transformer 模型采用的局部特征增强单元更有效.

实验结果表明, 该方法在更为复杂的 FD002 和 FD004 数据集上取得了最优的性能. 然而, 在操作条件简单的 FD001 和 FD003 数据集上, 其性能没有达到最佳水平. 这表明, 所提出的方法更适合复杂操作条件的发动机剩余寿命预测任务.

2.6 消融实验

消融实验的结果汇总于表 3 中. 本预测模型 MS_Transformer 主要由三个关键组件构成: 多尺度局部特征增强单元 (MSLFU_BLOCK)、Trans-

former 编码器和具有 Sigmoid 函数映射的输出层. 在本节中, 为深入探究多尺度局部特征增强单元在模型中所起的关键作用, 从 MS_Transformer 架构中剔除了 MSLFU_BLOCK 部分, 相应的模型表示为 MS_Transformer (w/o MS). 在此基础上, 进一步移除了输出层中的 Sigmoid 函数, 并将改动后的模型记为 MS_Transformer (w/o s & MS). 另外, 为了评估因果卷积的实际效能, 以常规卷积代替因果卷积, 并将对应的模型表示为 MS (CNN)_Transformer.

从表 3 中可以看出, 在所有数据集上, 相比去除多尺度局部特征提取模块的模型, MS_Transformer 模型具有更低的 RMSE 值和 Score 值, 这表明多尺度局部特征提取模块能够有效地提高模型对数据的拟合度, 减少了预测误差, 提高了预测的准确性和稳定性. 值得注意的是, 当在 MS_Transformer (w/o MS) 模型中加入多尺度局部特征提取模块后, 模型在四个数据集上的性能提升幅度并不相同. 具体来说, RMSE 指标在四个数据集上分别下降了 10.68%、24.08%、17.30% 和 21.70%. 这暗

表 3 消融实验结果
Table 3 Results of ablation experiment

方法	FD001		FD002		FD003		FD004	
	RMSE	Score	RMSE	Score	RMSE	Score	RMSE	Score
MS_Transformer	11.79	224.36	11.98	608.88	11.95	225.05	14.47	1072.38
MS (CNN)_Transformer	12.82	254.36	13.72	1098.09	13.80	325.05	15.93	1372.87
MS_Transformer (w/o MS)	13.20	275.59	15.78	1430.90	14.45	445.51	18.48	1754.22
MS_Transformer (w/o s & MS)	13.91	298.18	15.91	1497.41	16.10	552.61	19.03	1992.69

示了不同数据集的局部特征对于模型性能提升的影响不同. 数据集的不同特性导致多尺度局部特征提取模块的作用程度有所不同, 但每个数据集上均能较大幅度地提高模型性能. 这些实验结果证实了多尺度局部特征提取模块在提升模型性能方面的关键作用, 以及其在不同数据集上的较好适用性.

通过将多尺度局部特征提取模块中的因果卷积替换为常规卷积, 即采用常规卷积代替因果卷积的方法 MS (CNN)_Transformer, 并未带来预测性能的提升. 相反, 在所有数据集上, 相较于 MS_Transformer 模型, MS (CNN)_Transformer 模型的 RMSE 和 Score 值均有所上升, 表明预测精度下降. 这一结果说明了因果卷积在多尺度局部特征提取模块中的有效性. 这可能是因为因果卷积在处理时间序列数据时, 更能有效地保持时间顺序信息的完整性, 从而帮助模型更精确地捕捉数据中的局部时间上下文特征. 通常, 常规卷积在处理时间序列数据时可能引发未来信息的泄露, 进而对模型的性能产生不利影响.

MS_Transformer (w/o s & MS) 模型在四个数据集上的 RMSE 和 Score 指标值均出现了不同程度的增加, 这也证实了 Sigmoid 缩放在 RUL 估计中的有效性.

2.7 参数敏感性分析

2.7.1 滑动窗口长度

表 4 给出了不同滑动窗口长度对应的性能评价指标值. 对于时间序列数据集, 窗口长度是深度学习模型所必需的重要参数, 它直接影响着模型的预测精度. 然而, 目前尚缺乏确定最佳窗口长度的明

确标准. 因此, 本研究根据最近相关文献的实验结果尝试了一系列窗口长度值 ($L = 30, 40, 50, 60, 70$). 如表 4 所示, 当窗口长度从 30 增加到 60, 四个数据集上的 RMSE 值整体上都降低了. 同时, 除了 FD001 数据集窗口长度从 30 到 50 和 FD003 数据集窗口长度为 40 的情况外, Score 指标也主要呈现下降趋势. 然而, 当进一步将窗口长度增加到 70 时, 可以观察到 FD002 和 FD004 数据集的性能略有改善, 而 FD001 和 FD003 数据集上的预测性能明显下降. 考虑到模型的性能、计算资源的利用率以及对数据记录长度的需求, 选择窗口长度为 60.

2.7.2 MSLFU_BLOCK 中因果卷积层数

本实验探究了在相同的 Transformer 编码器下, MSLFU_BLOCK 中因果卷积的层数对预测性能的影响, 表 5 给出了不同层数对应的预测指标值.

从表 5 可见, 在所有数据集上, 因果卷积层数的递增并不总是带来模型性能的线性提升. 具体来说, 当因果卷积层数从一层扩展至两层时, RMSE 和 Score 值均出现了明显的下降, 这表明了模型性能在此阶段得到了显著的提升. 然而, 当因果卷积层数进一步增加至三层时, RMSE 和 Score 值却出现了上升的趋势. 这一现象可能与输入的时间序列数据特性有关. 由于该数据所包含的特征信息有限, 过多的因果卷积层数可能导致模型在训练过程中过拟合. 这同时也暗示了多尺度局部特征提取模块的局部感受野应与数据特性相契合. 此外, 随着因果卷积层数的递增, 卷积层所需训练的参数数量和计算成本也会相应增加. 因此, 综合考虑模型性能、过拟合风险以及计算成本等因素后, 最终将因果卷积层数设定为 2 层.

表 4 不同窗口长度对应的预测指标值

Table 4 Predictive metric values corresponding to different window lengths

滑动窗口长度	FD001		FD002		FD003		FD004	
	RMSE	Score	RMSE	Score	RMSE	Score	RMSE	Score
$L = 30$	12.89	264.78	14.38	1011.04	13.73	279.99	17.20	1858.13
$L = 40$	12.67	268.07	13.42	854.88	12.21	213.13	16.74	1676.82
$L = 50$	11.93	212.96	12.94	724.12	12.31	255.20	15.56	1375.81
$L = 60$	11.79	224.36	11.98	608.88	11.95	225.05	14.47	1072.38
$L = 70$	12.23	242.86	11.75	587.67	12.59	266.96	14.26	1093.49

表 5 不同因果卷积层数对应的预测指标值

Table 5 Predictive metric values corresponding to different numbers of causal convolution layers

因果卷积层数	FD001		FD002		FD003		FD004	
	RMSE	Score	RMSE	Score	RMSE	Score	RMSE	Score
1	12.28	270.33	12.64	749.42	12.60	278.99	16.31	1887.34
2	11.79	224.36	11.98	608.88	11.95	225.05	14.47	1072.38
3	13.02	270.33	14.98	1225.98	14.19	367.16	17.30	2185.46

表 6 不同数量的 encoder layer 对应的预测指标值
Table 6 Predictive metric values corresponding to different numbers of encoder layers

编码器层数	FD001		FD002		FD003		FD004	
	RMSE	Score	RMSE	Score	RMSE	Score	RMSE	Score
1	11.79	224.36	11.98	608.88	11.95	225.05	14.47	1072.38
2	11.35	210.25	12.78	785.32	11.56	230.32	16.72	1785.03
3	11.95	223.25	12.58	735.32	11.86	235.46	15.72	1685.03

2.7.3 Encoder layer 的个数

本实验探究了在相同的 MSLFU_BLOCK 下, Transformer 编码器的层数对预测性能的影响, 表 6 给出了不同 Encoder layer 个数对应的预测指标值。

每个 Encoder layer 都可以提取不同层次的特征, 通过堆叠多个 Encoder layer 可以逐渐提取更高级别的特征, 从而增加模型的特征描述能力。但同时, 增加 Encoder layer 的个数会增加模型的复杂度、训练时间以及过拟合风险。因此, 在选择 Encoder layer 的个数时需要综合考虑模型性能、训练时间、计算资源和过拟合问题等因素。实验过程中将 Encoder layer 的个数依次从 1 增加到 3, 对应的预测性能指标值 RMSE 和 Score 显示在表 6 中。实验结果表明, 随着 Encoder layer 数量的增加, 四个数据集上的 RMSE 和 Score 指标都没有明显提升, 说明 MSFLU_BLOCK 结合 1 层的 Encoder layer 已有足够的特征描述能力。

3 结束语

1) 针对时间序列预测问题需解决局部和全局特征有效提取的问题, MS_Transformer 模型采用了 Transformer 编码器作为骨干网络, 通过基于点积自关注的全局上下文感知机制获取时间序列的全局特征。为了克服 Transformer 编码器在捕获相邻时间序列的合并局部上下文方面的限制, 引入了多尺度局部特征增强单元 (MSLFU_BLOCK)。这两个关键组件的整合保证了 MS_Transformer 模型能够全面考虑局部和全局特征, 提高了发动机剩余寿命的预测准确性。

2) 为了解决时间序列中局部特征提取问题, 引入 MSLFU_BLOCK 提取局部上下文的多时间尺度特征, 多尺度局部特征使得模型能够更好地捕捉时间序列中的局部细节, 使用的因果卷积避免了未来数据泄露, 由多层因果卷积实现的 MSLFU_BLOCK 有助于提高预测性能。

3) 通过与 13 种先进方法的对比实验, MS_Transformer 模型在操作条件更复杂的 FD002 和 FD004 数据集上获得了最优的 RMSE 和 Score 值。在所有四个数据集上, 该模型的平均 RMSE 和 Score

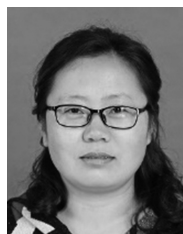
最优, 验证了本研究提出的预测模型在发动机剩余使用寿命预测方面的有效性和优越性。并表明 MS_Transformer 在不同数据集上均能够稳健地实现较好的性能。

本研究在不考虑预测性能对发动机操作条件复杂性的潜在敏感性的情况下, 基于相关运行数据建立了飞机发动机剩余使用寿命预测模型。下一步的研究工作将包括通过特征可视化等方法, 深入分析不同操作场景下影响飞机发动机寿命的关键因素。此外, 将提出的方法扩展到具有类似预测性维护需求的其他领域, 探索其对不同类型设备的适应性, 并评估其在各种应用场景中的稳健性和有效性。

References

- 1 Qiao W, Zhang P, Chow M Y. Condition monitoring, diagnosis, prognosis, and health management for wind energy conversion systems. *IEEE Transactions on Industrial Electronics*, 2015, **62**(10): 6533–6535
- 2 Moghaddass R, Zuo M J. An integrated framework for online diagnostic and prognostic health monitoring using a multistate deterioration process. *Reliability Engineering & System Safety*, 2014, **124**: 92–104
- 3 Pei Hong, Hu Chang-Hua, Si Xiao-Sheng, Zhang Jian-Xun, Pang Zhe-Nan, Zhang Peng. Review of machine learning based remaining useful life prediction methods for equipment. *Journal of Mechanical Engineering*, 2019, **55**(8): 1–13 (裴洪, 胡昌华, 司小胜, 张建勋, 庞哲楠, 张鹏. 基于机器学习的剩余寿命预测方法综述. *机械工程学报*, 2019, **55**(8): 1–13)
- 4 Javed K, Gouriveau R, Zerhouni N. A new multivariate approach for prognostics based on extreme learning machine and fuzzy clustering. *IEEE Transactions on Cybernetics*, 2015, **45**(12): 2626–2639
- 5 Khelif R, Chebel-Morello B, Malinowski S, Laajili E, Fnaiech F, Zerhouni N. Direct remaining useful Life estimation based on support vector regression. *IEEE Transactions on Industrial Electronics*, 2017, **64**(3): 2276–2285
- 6 Wu D, Jennings C, Terpenney J, Gao R X, Kumara S A. Comparative study on machine learning algorithms for smart manufacturing: Tool wear prediction using random forests. *Journal of Manufacturing Science and Engineering*, 2017, **139**(7): Article No. 071018
- 7 Zheng S, Ristovski K, Farahat A, Gupta C. Long short-term memory network for remaining useful life estimation. In: Proceedings of IEEE International Conference on Prognostics and Health Management. Dallas, TX, USA: IEEE, 2017. 88–95
- 8 Huang C G, Huang H Z, Li Y F. A bidirectional LSTM prognostics method under multiple operational conditions. *IEEE Transactions on Industrial Electronics*, 2019, **66**(11): 8792–8802
- 9 Yu W, Kim I Y, Mechefske C. An improved similarity-based prognostic algorithm for RUL estimation using an RNN autoencoder scheme. *Reliability Engineering & System Safety*, 2020, **199**: Article No. 106926
- 10 Liu J, Lei F, Pan C, Hu D, Zuo H. Prediction of remaining use-

- ful life of multi-stage aero-engine based on clustering and LSTM fusion. *Reliability Engineering & System Safety*, 2021, **214**: Article No. 107807
- 11 Li X, Ding Q, Sun J Q. Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliability Engineering & System Safety*, 2018, **172**: 1–11
 - 12 Yang B, Liu R, Zio E. Remaining useful life prediction based on a double-convolutional neural network architecture. *IEEE Transactions on Industrial Electronics*, 2019, **66**(12): 9521–9530
 - 13 Li X, Jiang H, Liu Y, Wang T, Li Z. An integrated deep multiscale feature fusion network for aeroengine remaining useful life prediction with multi-sensor data. *Knowledge-Based Systems*, 2022, **235**: Article No. 107652
 - 14 Al-Dulaimi A, Zabih S, Asif A, Mohammadi A. A multi-modal and hybrid deep neural network model for remaining useful life estimation. *Computers in Industry*, 2019, **108**: 186–196
 - 15 Saxena A, Goebel K, Simon D, Eklund N. Damage propagation modeling for aircraft engine run to failure simulation. In: Proceedings of 2008 International Conference on Prognostics and Health Management. Denver, CO, USA: IEEE, 2008. 1–9
 - 16 Ayodeji A, Wang Z, Wang W, Qin W, Yang C, Xu S, et al. Causal augmented ConvNet: A temporal memory dilated convolution model for long-sequence time series prediction. *ISA Transactions*, 2021, **123**: 200–217
 - 17 Li H, Wang Z, Li Z. An enhanced CNN-LSTM remaining useful life prediction model for aircraft engine with attention mechanism. *PeerJ Computer Science*, 2022, **8**: Article No. 1084
 - 18 Liu H, Liu Z, Jia W, Lin X. Remaining useful life prediction using a novel feature-attention-based end-to-end approach. *IEEE Transactions on Industrial Informatics*, 2021, **17**(2): 1197–1207
 - 19 Xu X, Li X, Ming W, Chen M. A novel multi-scale CNN and attention mechanism method with multi-sensor signal for remaining useful life prediction. *Computers & Industrial Engineering*, 2022, **169**: Article No. 108204
 - 20 Zhang J, Jiang Y, Wu S, Li X, Luo H, Yin S. Prediction of remaining useful life based on bidirectional gated recurrent unit with temporal self-attention mechanism. *Reliability Engineering & System Safety*, 2022, **221**: Article No. 108297
 - 21 Wang X, Li Y, Xu Y, Liu X, Zheng T, Zheng B. Remaining useful life prediction for aero-engines using a time-enhanced multi-head self-attention model. *Aerospace*, 2023, **10**(1): Article No. 80
 - 22 Xu Z, Zhang Y, Miao J, Miao Q. Global attention mechanism based deep learning for remaining useful life prediction of aero-engine. *Measurement*, 2023, **217**: Article No. 113098
 - 23 Zhao K, Jia Z, Jia F, Shao H. Multi-scale integrated deep self-attention network for predicting remaining useful life of aero-engine. *Engineering Applications of Artificial Intelligence*, 2023, **120**: Article No. 105860
 - 24 Zhu J, Jiang Q, Shen Y, Xu F, Zhu Q. Res-HSA: Residual hybrid network with self-attention mechanism for RUL prediction of rotating machinery. *Engineering Applications of Artificial Intelligence*, 2023, **124**: Article No. 106491
 - 25 Li H, Cao P, Wang X, Yi B, Huang M, Sun Q, et al. Multi-task spatio-temporal augmented net for industry equipment remaining useful life prediction. *Advanced Engineering Informatics*, 2023, **55**: Article No. 101898
 - 26 Zhang Q, Liu Q, Ye Q. An attention-based temporal convolutional network method for predicting remaining useful life of aero-engine. *Engineering Applications of Artificial Intelligence*, 2024, **127**: Article No. 107241
 - 27 Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, et al. Attention is all you need. *Neural Information Processing Systems*, 2017, **30**: 5998–6008
 - 28 Xu J, Wu H, Wang J, Long M. Anomaly transformer: Time series anomaly detection with association discrepancy. In: Proceedings of 10th International Conference on Learning Representations. Virtual Event: ICLR, 2022.
 - 29 Tuli S, Casale G, Jennings N R. TranAD: Deep transformer networks for anomaly detection in multivariate time series data. arXiv: 2201.07284, 2022.
 - 30 Zerveas G, Jayaraman S, Patel D, Bhamidipaty A, Eickhoff C. A transformer-based framework for multivariate time series representation learning. In: Proceedings of 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Virtual Event: 2021. 2114–2124
 - 31 Li S, Jin X, Xuan Y, Zhou X, Chen W, Wang Y, et al. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. arXiv: 2331.8422, 2019.
 - 32 Mo Y, Wu Q, Li X, Huang B. Remaining useful life estimation via transformer encoder enhanced by a gated convolutional unit. *Journal of Intelligent Manufacturing*, 2021, **32**(7): 1997–2006
 - 33 Guo J, Lei S, Du B. MHT: A multiscale hourglass-transformer for remaining useful life prediction of aircraft engine. *Engineering Applications of Artificial Intelligence*, 2024, **128**: Article No. 107519
 - 34 Jaenal A, Ruiz-Sarmiento J R, Javier G J. MachNet, a general deep learning architecture for predictive maintenance within the industry 4.0 paradigm. *Engineering Applications of Artificial Intelligence*, 2024, **127**: Article No. 107365



陈俊英 西安建筑科技大学信息与控制工程学院副教授。2010 年获得西安交通大学计算机科学与技术专业博士学位。主要研究方向为机器学习, 智能检测和视觉感知信息处理。本文通信作者。

E-mail: chenjy@xauat.edu.cn

(**CHEN Jun-Ying** Associate professor at the College of Information and Control Engineering, Xi'an University of Architecture and Technology. She received her Ph.D. degree in Computer Science and Technology from Xi'an Jiaotong University in 2010. Her research interest covers machine learning, intelligent detection, and visual perception information processing. Corresponding author of this paper.)



席月芸 西安建筑科技大学信息与控制工程学院硕士研究生。主要研究方向为预测与健康管理系统。

E-mail: yueyun@xauat.edu.cn

(**XI Yue-Yun** Master student at the College of Information and Control Engineering, Xi'an University of Architecture and Technology. Her research interest covers prediction and health management systems.)



李朝阳 西安建筑科技大学信息与控制工程学院硕士研究生。主要研究方向为工业缺陷目标检测。

E-mail: nicholas@xauat.edu.cn

(**LI Zhao-Yang** Master student at the College of Information and Control Engineering, Xi'an University of Architecture and Technology. His main research interest is industrial defect detection.)