Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Multisensor contrast neural network for prediction of remaining useful life of rolling bearing under scarce labeled data*

Binkun LIU[1,2,3,4], Zhenyi XU[†‡2,3,4], Yu KANG[1,3], Yang CAO[1,3], Yunbo ZHAO[†‡1,3]

[1]*Department of Automation, University of Science and Technology of China, Hefei 230027, China*

[2]*Key Laboratory of System Control and Information Processing, Ministry of Education, Shanghai 200240, China*

[3]*Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei 230088, China*

[4]*Jianghuai Advance Technology Center, Hefei 230000, China*

[†]E-mail: xuzhenyi@mail.ustc.edu.cn; ybzhao@ustc.edu.cn

Received Aug. 29, 2024; Revision accepted Dec. 30, 2024; Crosschecked

Unedited

Abstract: Predicting remaining useful life (RUL) of bearing under scarce labeled data is significant for intelligent manufacturing. Current approaches typically encounter the challenge that different degradation stages have similar behaviors in multisensor scenarios. Given that cross-sensor similarity improves the discrimination of degradation features, we propose a multisensor contrast method for RUL prediction under scarce RUL-labeled data, in which we exploit cross-sensor similarity to mine multisensor similar representations that indicate machine health condition from rich unlabeled sensor data in co-occurrence space. Specifically, we use Resnet18 to span the features of different sensors into a co-occurrence space. We then obtain multisensor similar representations of abundant unlabeled data through alternate contrast based on cross-sensor similarity in co-occurrence space, which indicate the machine degradation stage. Finally, we fine-tune these similar representations with attention to achieve RUL prediction with limited labeled sensor data. The proposed method is evaluated on a publicly available bearing dataset, and the results show that the mean absolute percentage error is reduced by at least 0.058, and the score is improved by at least 0.122 compared to the state-of-the-art methods.

Key words: Self-supervised; Remaining useful life prediction; Contrast Learning
https://doi.org/10.1631/FITEE.2400753 **CLC number:**

1 Introduction

Prediction of remaining useful life (RUL) for bearing aims to forecast the duration of time from current operation until failure of the bearings (Wen et al., 2018). As a critical component of intelligent machine health management (Tao et al., 2018; Souza et al., 2021; Wang et al., 2022), RUL prediction for bearings can assist companies in reducing maintenance costs and preventing significant losses due to accidents, thereby improving their competitiveness.

Traditional approaches for RUL prediction for bearings can be broadly classified as model-based and statistics-based methods. Model-based methods (Morales-Espejel and Gabelli, 2020) require extensive domain knowledge to build physical models that accurately reflect machine degradation. However, obtaining domain knowledge is challenging, and building accurate physical models is difficult due to the complex system structure and operating environment. On the other hand, statistics-based methods (Li et al., 2022a) focus on building a stochastic

‡ Corresponding authors
* Project supported by the Dreams Foundation of Jianghuai Advance Technology Center (NO. 2023-ZM01Z002) and Open Project Program of Key Laboratory of Ministry of Education of System Control and Information Processing (No. SCIP20230109)
© Zhejiang University Press 2025

model that describes the degradation process to predict RUL based on monitored machine degradation variables. Nevertheless, these methods have limited processing ability on low-quality data.

In recent years, the use of deep learning for RUL prediction for bearings has become a research trend due to its effectiveness in improving prediction accuracy. Data-driven approaches (Wang et al., 2020, 2021) utilize deep learning models to establish potential relationships between machine monitoring data and RUL labels or degradation labels. These approaches have powerful degradation feature extraction capabilities, which enable them to effectively handle massive and complex structured data and reduce the need for domain knowledge.

Compared to model-based and statistical-based approaches for RUL prediction for bearings, data-driven approaches heavily rely on RUL-labeled data. As manufacturing levels continue to improve and machine reliability gradually increases, it is often challenging to obtain a sufficient amount of failure data with RUL labels over a short-term period or too expensive to obtain data for expensive machines. Therefore, the data-driven approaches still face challenges in practical applications due to the scarcity of degradation data with RUL labels.

Benefiting from recent significant progress in addressing the lack of labeled data in areas such as computer vision (Tian et al., 2020; Korbar et al., 2018; Wang et al., 2023; Zhu and Pu, 2021), self-supervised techniques provide solutions for RUL prediction for bearings under scarce labeled data, but current efforts suffer from poor discrimination of degradation features. Typically, current methods (Ding et al., 2022; Akrim et al., 2023; Krokotsch et al., 2022; Kong et al., 2023) treat each sensor signal as a channel and mine the temporal autocorrelation of multisensors from massive unlabeled sensor data during pretraining. The temporal autocorrelation is then used as a representation and fine-tuned with the limited labeled data to achieve RUL prediction. However, the methods using stacked channels have the weakness that there may be similar sensor signals at different degradation states, which is not favorable for RUL prediction. We stack the vertical and horizontal acceleration time-frequency matrices of the bearings and calculate the cosine similarity of the stacked matrices at any two moments, shown in Fig. 1a and 1c. The figures illustrate that there are plenty of red high-similarity regions, which indicates similar sensor signals at different degradation states. To improve the discrimination of degradation features, we employ multisensors similarity. As shown in Fig. 1b and 1d, we first take the dot product of the vertical and horizontal acceleration time-frequency matrices, and then calculate the cosine similarity of the dot product matrices at any two moments. This approach significantly reduces the similarity of sensors in different degradation states compared to Fig. 1a and 1c.

Hence, we propose a multisensor contrast neural network, in which cross-sensor similar features indicating machine health conditions are captured from a large amount of unlabeled sensor data. Specially, the sensor data are mapped to the time–frequency domain through wavelet transform. Then, we devise an alternate contrast process to extract similar features between sensors from a large amount of unlabeled data, which indicate machine health conditions. In the alternate contrast stage, the multi-branch ResNet18 is used as a feature extractor to span the features into a co-occurrence space between sensors. In the co-occurrence space, any sensor is selected as the main sensor, and the remaining ones are regarded as auxiliary sensors. The auxiliary sensor feature extractor uses momentum update to ensure the consistency of features. Then, we calculate the similarity between the main sensor feature and auxiliary sensor features, and the noise contrastive estimation (NCE) loss is optimized so that the main sensor feature and the auxiliary sensor feature at the same moment are the most similar, to obtain the similar feature between the main sensor and the auxiliary sensor. We repeat the above process until each sensor has served as the main sensor, leading to highly discriminative degradation features. Next, the model is fine-tuned by using the scarce data with RUL labels. In the fine-tuning stage, attention is devised to adjust feature weights. Ultimately, RUL prediction under scarce labeled sensor data is achieved with the assistance of rich unlabeled sensor data.

Overall, the main contributions of the proposed method are summarized as follows:

1) A multisensor contrast neural network is proposed, which can utilize a large amount of unlabeled multisensor degradation data to model the degradation process and achieve RUL prediction when RUL

Unedited

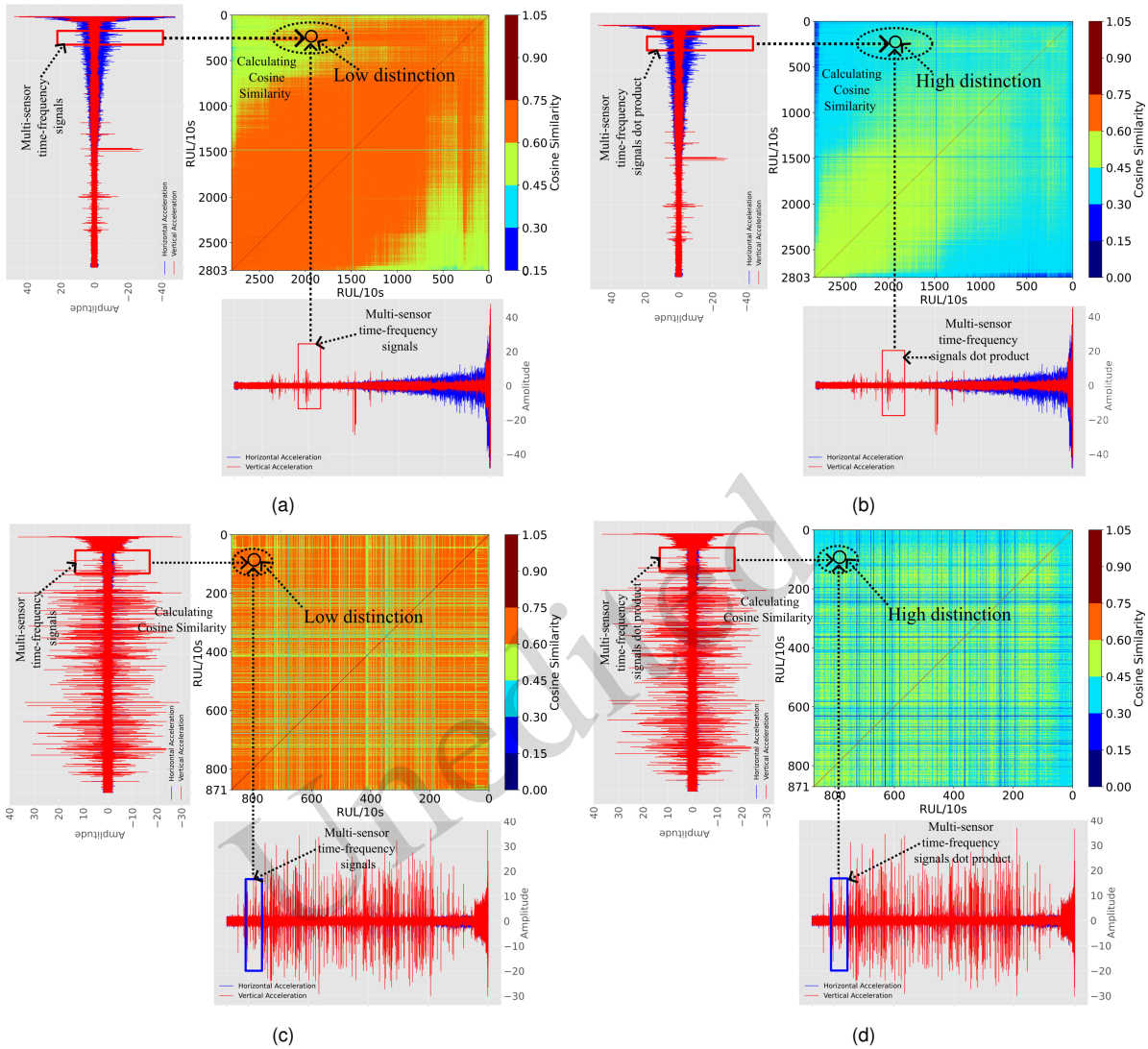


Fig. 1: Stacking multiple sensors causes different degradation states to behave similarly, and cross-sensor correlation improves the discrimination of degradation features. (a) The cosine similarity matrix between stacked signals at any two moments of Bearing 1_1; (b) The correlation between the vertical and horizontal acceleration time-frequency signals is calculated using the dot product, and the cosine similarity matrix between correlations at any two moments of Bearing 1_1; (c) The cosine similarity matrix between stacked signals at any two moments of Bearing 1_2; (d) The correlation between the vertical and horizontal acceleration time-frequency signals is calculated using the dot product, and the cosine similarity matrix between correlations at any two moments of Bearing 1_2.

Unedited

labeled data are scarce. Compared to existing work, this method can extract more discriminative degradation features through cross-sensor contrast.

2) A cross-sensor alternate contrast process is devised to effectively mine highly discriminative degradation features from a large amount of unlabeled sensor data by maximizing the similarity of multisensor features at the same moment. Compared with existing works, the cross-sensor alternate contrast process can greatly improve the distinction of degradation features regarding RUL by mining potential shared degradation features among multiple sensors through alternate contrast.

3) The proposed method is evaluated on the Franche-Comté Electronics Mechanics Thermal Science and Optics-Sciences and Technologies

(FEMTO-ST) Bearing dataset and the results show that the mean absolute percentage error (MAPE) is reduced by at least 0.058, and the score is improved by at least 0.122 compared to the state-of-the-art methods.

2 Related works

2.1 Work based on the FEMTO-ST bearing dataset

The FEMTO-ST bearing dataset is a dataset containing 17 full-cycle bearing degradation data for three operating conditions and is mainly used to study RUL prediction for bearings. Recent research on the FEMTO-ST bearing dataset focuses on improving the accuracy of RUL prediction and the predicted RUL across working conditions.

Recent RUL prediction efforts have been excellent with sufficient data (Yang et al., 2023; Zou et al., 2023). Li et al. (2022b) proposes a two-dimensional long short-term memory (2DLSTM) based fusion network for RUL prediction. The 2DLSTM framework is used to extract deep time features of sensor data one by one and fuse multisensor features using an information fusion unit (IFU) to predict the RUL of bearings. Zuo et al. (2023) proposes a hybrid attention-based multiwavelet coefficient fusion method to evaluate the RUL of bearings. A hybrid attention-based convolutional LSTM (HA-ConvLSTM) network is used to self-adaptively extract features from the original signal after wavelet packet transformation to evaluate RUL. These efforts have demonstrated excellent performance.

Another research area of interest is to achieve accurate RUL prediction across operating conditions (Dong et al., 2023; Behera and Misra, 2023), and these approaches show excellent performance when relying on source domains with sufficient labeled RUL data. Deng et al. (2023) proposes a calibration-based hybrid transfer learning framework to improve data fidelity and model generality while demonstrating superiority in prediction accuracy and uncertainty quantification. To predict the RUL of bearings under invisible operating conditions, Ding et al. (2023) proposes an adversarial out-of-domain augmentation framework to generate pseudo-domains, thus increasing the diversity of available samples to improve the generalization of inaccessible target domains.

2.2 Self-supervised learning

Deep neural network represented by convolutional neural network and the transformer has made remarkable strides in the fields of computer vision and natural language processing, but it usually relies on sufficient labeled data. In some special fields, such as intelligent manufacturing and medicine, it is often very difficult to collect enough labeled data. Self-supervised learning can transform unsupervised learning based on unlabeled data into supervised learning based on labeled data by leveraging certain properties of unlabeled data to set pseudo-supervised tasks and learn features that are beneficial to the real task.

Self-supervised methods are mainly classified into generation-based methods and contrast-based methods. The generative methods are mainly concerned with the pseudo-supervisory task of data generation. The contrast-based approachs verify that multiple different input data channels correspond to each other. Tian et al. (2020) constructs a contrast pseudo-supervision task by maximizing the mutual information between different views of the same scene. Korbar et al. (2018) matches the visual and auditory elements of the video to achieve self-supervised learning

At present, there are relatively few research works (Melendez et al., 2019; Saeed et al., 2020; Zhang et al., 2021b,a) on the self-supervised RUL prediction. Ding et al. (2022) proposes a method to learn the multisensor self-sequence temporal correlation by contrasting the similarity between different sensor data augmentations. Krokotsch et al. (2022) considers the higher similarity of adjacent sensor time series, and use the network to estimate the time difference between any two segments of a time series. These methods all perform well but retain the difference between multisensors which may adversely affect RUL prediction.

3 Overview

3.1 Preliminary

Definition 1: (Remaining useful life [RUL]): The RUL corresponds to the duration of machine operation from the start moment t to the failure moment

Unedited

T. This is formally described as follows:

$$T - t|T > t. \tag{1}$$

Definition 2: (RUL prediction): Estimation of the RUL of the machine at the beginning of prediction based on effective information, such as machine health condition. This is formally described as follows:

$$T - t|T > t, Z(t), \tag{2}$$

where T represents the failure time of the machine, t represents the current time, and $Z(t)$ represents the valid information such as the machine health condition. Machine health condition is usually constructed from sensor data to reflect the degree of machine degradation.

3.2 Problem formulation

Given a sensor dataset without RUL labels $X^u \in \mathbb{R}^{N^u \times C \times M}$, the amount of data is N^u, and a sensor dataset with RUL labels $X^l \in \mathbb{R}^{N^l \times C \times M}$, the corresponding RUL is RUL^l, and the amount of data is N^l, $N^u \gg X^l$. C is the number of sensors, M is the length of the time series. Model E is constructed based on the unlabeled sensor dataset $X^u \in \mathbb{R}^{N^u \times C \times M}$ and the labeled sensor dataset $X^l \in \mathbb{R}^{N^l \times C \times M}$ to build machine health condition Z. And then a prediction model f is constructed to establish the association between the machine health condition Z and RUL. The ultimate goal is to predict the corresponding RUL based on the sensor data $x(t) \in \mathbb{R}^{C \times M}$ at the starting moment t.

$$Z(t) = E(x(t)|X^u, X^l) \tag{3}$$

$$\mathrm{RUL} = f(Z(t)) \tag{4}$$

4 Method

4.1 Framework

Fig. 2 shows the framework of multisensor contrast neural network with attention. It mainly includes three parts: data preprocessing, alternate contrast, and fine-tuning. The data preprocessing step transforms the original signal into a time–frequency domain by wavelet transform to extract the features of the time and frequency domains at the same time. Alternative contrast is composed of feature extractor and feature contrast, which alternately captures similar features between multisensors. The feature extractor spans the nonlinear features extracted from the time–frequency domain without RUL labels into a co-occurrence space to learn similar features between sensors. In feature contrast, one sensor is regarded as the main sensor, and another is regarded as the auxiliary sensor. The auxiliary sensor feature extractors use momentum updates to ensure the consistency of features. Then, the similarity between the main sensor feature and the auxiliary sensor feature is calculated in the co-occurrence space, and the features at the same moment for different sensors having the maximum similarity are obtained by optimizing the NCE Loss. We repeat the above process until each sensor serves as the main sensor to obtain the similarity between different sensors. Fine-tuning reuses the parameters of the feature extractor and initializes the predictor, thereby reducing the need for labeled data in the fine-tuning stage. Feature fusion makes full use of features from different sensors, and the attention mechanism adjusts feature weights to achieve RUL prediction.

4.2 Data preprocessing

Considering that in the machine degradation process, not only the amplitude but also the frequency will gradually change, the frequency and time features are revealed through data preprocessing. The methods of transforming the original signal from the time domain to the time–frequency domain mainly include short-time Fourier transform (STFT) and wavelet transform. Since the sensor signal is usually nonstationary and the window of the STFT is fixed, the high-frequency signal is suitable for STFT with a small window and the low-frequency signal is suitable for the STFT with a large window; thus, STFT cannot meet the needs of nonstationary signal changes. The wavelet transform replaces the infinitely long trigonometric function basis in the Fourier transform with a finitely long-decaying wavelet basis, so it can show the corresponding times when the different frequency components appear. The wavelet transform formula is as follows:

Unedited

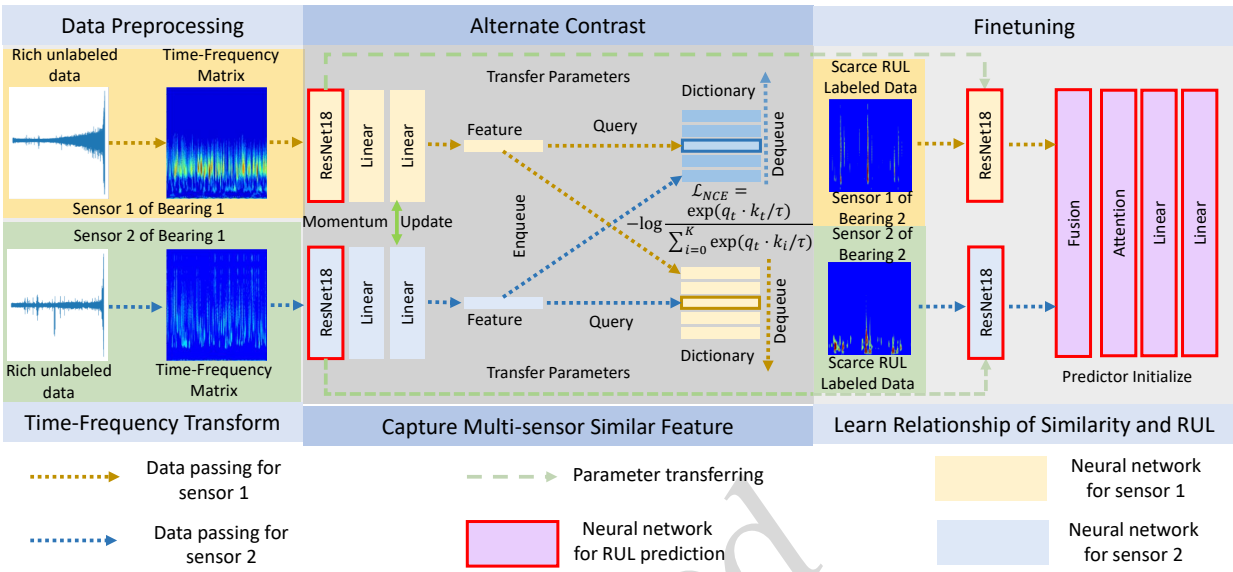


Fig. 2: The structure of the proposed framework. Dictionary: Select any sensor alternately as the main sensor, and the others as auxiliary sensors; a dictionary is formed from the corresponding auxiliary sensor features. Enqueue: Add the auxiliary sensor features of each batch to the corresponding dictionary. Dequeue: The earliest enqueued feature in the dictionary is deleted. Query: Calculate the similarity between the main sensor feature and dictionary and thus obtain the most similar auxiliary sensor feature. Momentum update: Ignore the gradient of the auxiliary sensor network and use momentum to update slowly.

Unedited

$$I_{i,t} = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x_i(t) * \psi\left(\frac{t-\beta}{a}\right) dt, \tag{5}$$

where $x_i(t)$ represents the original signal of the i-th sensor with t as the starting time and $t+N$ as the ending time. a is scale parameter, β is translation parameter, and $\psi\left(\frac{t-\beta}{a}\right)$ represents the mother wavelet function. Here we choose the Gaussian derivative wavelet as the mother wavelet function. Through wavelet transform, the sensor raw signal $x_i(t)$ of the sensor is transformed into a time-frequency matrix $I_{i,t}$.

Bilinear interpolation $\varphi(.)$ reduces the dimension of the time-frequency matrix $I_{i,t}$ to speed up feature extraction.

$$\bar{I}_{i,t} = \varphi(I_{i,t}), \tag{6}$$

where $\bar{I}_{i,t}$ is the time-frequency matrix after wavelet transformation and dimension reduction of the original signal of the i-th sensor with t as the start time and $t+N$ as the end time.

4.3 Alternate contrast

Since only a large amount of unlabeled sensor data is available during the alternate contrast phase, it is necessary to use the characteristics of the sensor data itself as training labels. For multisensor systems, a good machine health condition is invariant across sensors. In our work, it is desired to obtain degenerate features that are invariant across sensors, which indicate machine health conditions. Therefore, the alternate contrast is performed using cross-sensor similarity to maximize the similarity of multisensor features at the same moment while suppressing the similarities of multisensor features at different moments to obtain highly discriminative cross-sensor invariant degradation features.

4.3.1 Select main sensor

A multibranch network is constructed to extract the features of each sensor and span a co-occurrence space. In this paper, the ResNet18 backbone network E as feature extractor and the fully connected layer C as predictor are constructed as a multibranch

network. We select the i-th sensor V_i as the main sensor V_m, the corresponding backbone network f_i is denoted as f_m, and the time-frequency matrix $\bar{I}_{i,t}$ is denoted as $\bar{I}_{m,t}$. The other sensors $V_j, i \neq j$ are regarded as auxiliary sensors V_{a_j}, the corresponding backbone network f_j is recorded as f_{a_j}, and the time-frequency matrix $\bar{I}_{j,t}$ is denoted as $\bar{I}_{a_j,t}$.

4.3.2 Feature dictionary

The feature q_t of the main sensor is the machine health condition recorded by the main sensor at time t. The feature dictionary D_j consists of the features $\{k_{j,0}, k_{j,1}, \cdots, k_{j,K}\}$ of the j-th auxiliary sensor. The meaning of D_j is the machine health condition recorded at $K+1$ different moments by the j-th auxiliary sensor. It is formulated as follows:

$$q_t = f_m\left(\bar{I}_{m,t}\right) = C_m(E_m\left(\bar{I}_{m,t}\right)), \tag{7}$$

$$k_{j,k} = f_{a_j}\left(\bar{I}_{a_j,k}\right) = C_{s_j}(E_{a_j}\left(\bar{I}_{a_j,k}\right)), \tag{8}$$

where f_m, C_m and E_m refer to the backbone network, predictor, and feature extractor corresponding to the main sensor, respectively, and f_{a_j}, C_{s_j} and E_{a_j} refer to the backbone network, predictor and feature extractor corresponding to the j-th auxiliary sensor, respectively.

4.3.3 Feature similarity calculation

Firstly, the degradation feature q_t of the main sensor at moment t and the feature dictionary $D_j = \{k_{j,0}, k_{j,1}, \cdots, k_{j,K}\}$ corresponding to the j-th auxiliary sensor are given.

Then, the similarity $Sim^j_{t,k}$ between the main sensor feature q_t and the degradation feature $k_{j,k}$ at moment k in the dictionary D_j is shown in Eq. 9.

$$Sim^j_{t,k} = q_t \cdot k_{j,k} = \sum_{n=1}^{N} q_t^n k_{j,k}^n, \tag{9}$$

where $Sim^j_{t,k}$ denotes the similarity between the degradation feature of the main sensor at moment t and the degradation feature of the j-th auxiliary sensor at moment k, and $\cdot$ denotes the dot product. N denotes the number of elements in the degenerate features q_t and $k_{j,k}$, and n denotes the n-th element. $q_t^n k_{j,k}^n$ denotes the multiplication of the n-th element in the degenerate feature q_t with the n-th element in the degenerate feature $k_{j,k}$.

Finally, following the above steps, the similarity $\left\{Sim^j_{t,0}, Sim^j_{t,1}, \cdots, Sim^j_{t,K}\right\}$ between the main sensor degradation feature q_t and all the degradation features $k_{j,0}, k_{j,1}, \cdots, k_{j,K}$ in the feature dictionary D_j is calculated.

4.3.4 Loss function

The feature similarities at different moments are normalized using softmax. Let the multisensor feature similarity $Sim^j_{t,t}$ matching label of the same moment t be 1 and the multisensor feature matching labels of different moments be 0. Then the normalized feature similarities are fed into the cross-entropy. The cross-entropy maximizes the similarity of multisensor features at the same moment while suppressing the similarity of sensor features at different moments. In summary, to maximize the similarity of sensor features at the same moment to obtain the degradation features that are invariant across sensors, the NCE Loss function $\mathcal{L}_{NCE}$ is constructed by measuring the sensor similarity by dot product as follows:

$$\mathcal{L}_{NCE} = -\sum_j \log \frac{\exp\left(\frac{Sim^j_{t,t}}{\tau}\right)}{\exp\left(\frac{Sim^j_{t,t}}{\tau}\right) + \sum\limits_{k=0,k\neq t}^{K} \exp\left(\frac{Sim^j_{t,k}}{\tau}\right)} + \mu_c \left\|\theta_m\right\|_2, \tag{10}$$

where τ is the temperature hyperparameter, μ_c is the regularization weight, and $\|\theta_m\|_2$ represents the L2 regularization of the main sensor network f_m.

The feature q_t of the main sensor looks for the most similar sample among the $K+1$ samples of auxiliary sensors, and this process is similar to $K+1$ classification for the feature q_t.

4.3.5 Feature dictionary update by enqueuing and dequeuing

Since the feature q_t is constantly changing during the training process, the feature dictionary needs to be dynamically maintained. Longer dictionaries will make use of auxiliary sensor data. Since the length of the feature dictionary is much larger than the batch size, it is impractical to recalculate all the features in the feature dictionary each time. Therefore, updating is accomplished by queuing in a first-in-first-out mode. In each batch, auxiliary sensor

Unedited

features are enqueued to the corresponding feature dictionary, and the oldest auxiliary sensor feature is dequeued.

4.3.6 Updating auxiliary sensor network parameters

Because the auxiliary sensor backbone network is also being updated, the parameters corresponding to these oldest features may be significantly different from the current parameters. Through the above operations, the dynamic update of the feature dictionary is gradually achieved, while ensuring that the dictionary length can be much larger than the batch size. It will reduce consistency among features in the dictionary by the rapidly changing parameters of the auxiliary sensor network. Therefore, the gradients of the auxiliary sensor networks are ignored and the momentum update is used for parameters update (He et al., 2020).

$$\theta_j \leftarrow m\theta_j + (1-m)\theta_m, \tag{11}$$

where m is the momentum update factor, θ_m represents the main sensor network parameters and θ_j is the j-th auxiliary sensor network parameters.

4.3.7 Alternate mechanism

The above processes are alternated every F_{ex} epochs until each sensor has served as the main sensor to ensure that the network of each sensor can learn good representation.

4.4 Fine-tuning

In the alternate contrast stage, the model has extracted cross-sensor invariant degradation features. To establish an association between cross-sensor invariant degradation features and RUL, the model is finetuned using scarce sensor data with RUL labels. The parameters of the feature extractor are reused and the predictor C_F is initialized.

4.4.1 Feature fusion

Features from different sensors are fused to make full use of them. The fusion method selects concatenation, and the formula is as follows:

$$F_t = Cat\left(\left[E_1\left(\bar{I}_{1,t}\right), E_2\left(\bar{I}_{2,t}\right), \cdots, E_i\left(\bar{I}_{i,t}\right)\right]\right). \tag{12}$$

where Cat is the feature concatenation operator. Specifically, it means that the degenerate features $E_1\left(\bar{I}_{1,t}\right)$, $E_2\left(\bar{I}_{2,t}\right)$, $\cdots$, $E_i\left(\bar{I}_{i,t}\right)$ are concatenated according to the last dimension. E_i refers to the feature extractor corresponding to the i-th sensor.

4.4.2 Attention mechanism

Spatial attention Att_s is devised to adjust feature weights.

$$\alpha_s = Tanh\left(W_s F_t + b_s\right), \tag{13}$$

$$Att_s = Softmax\left(W_i \alpha_s + b_i\right), \tag{14}$$

where W_s and W_i are trainable weight parameters, and b_s and b_i are trainable bias parameters. $Tanh$ and $Softmax$ are both nonlinear activation functions.

$$\bar{F}_t = Att_s \cdot F_t, \tag{15}$$

where $\cdot$ is element-wise multiplication, and $\bar{F}_t$ is the feature after attention reconstruction.

4.4.3 RUL prediction

The predictor C_F performs regression prediction on the reconstructed feature $\bar{F}_t$.

$$\mathrm{RUL}_p = C_F\left(\bar{F}_t\right), \tag{16}$$

where $\mathrm{RUL}_p \in \mathbb{R}^+$ is the RUL predicted by the model, and the loss function $\mathcal{L}_F$ in the finetuning stage is formulated as follows:

$$\mathcal{L}_F = \left(\mathrm{RUL}_p - \mathrm{y}_t\right)^2 + \mu_f \|\theta\|_2, \tag{17}$$

where y_t is the corresponding RUL label, $\|\theta\|_2$ represents L2 regularization, and μ_f is the regularization coefficient.

5 Experiment

5.1 Data and setup

5.1.1 Description of datasets

The FEMTO-ST Bearing dataset (Nectoux et al., 2012) contains accelerated degradation data of bearings. The bearing operating state was recorded every 10 seconds using vertical and horizontal acceleration sensors with a sampling frequency of 25.6 kHz. Each sampling lasted 0.1s. This dataset contains a total of 17 bearing degradation datasets under three operating conditions.

Unedited

5.1.2 Datasets setup

In real-world scenarios, factories usually collect only limited RUL-labeled data and a huge amount of unlabeled degradation data due to expensive collection costs and other reasons. Due to the limited research on RUL prediction in this scenario, there is currently a lack of publicly available datasets that are suitable for this scenario. For the research and comparison with the baseline methods, we make reasonable modifications to the original FEMTO-ST bearing dataset that contains a large amount of RUL data. The modifications made to the dataset are aligned with real-world scenarios and can serve as a reference for subsequent similar research.

The training data for the alternate contrast phase and the finetuning phase of the experiment are shown in Tab. 1. The unlabeled dataset is constructed with a large amount of unlabeled data in the alternate contrast phase, and the RUL labeled dataset is constructed using few RUL data in the finetuning phase. In the alternate contrast phase, the model is pre-trained using only the vibration acceleration data of the bearings. During the alternate contrast phase, any RUL labels are not involved in the training of the model. Then, in the finetuning phase, the model is finetuned using a small number of vibration accelerations with RUL labels. The labels are related to the training phase and not to the sensors themselves.

Taking Condition 1 as an example, the dataset used during the alternate contrast phase includes degraded data from Bearing 1_2, Bearing 2_1, Bearing 2_2, Bearing 3_1, Bearing 3_2. During the entire alternate contrast process, there are no RUL labels involved in the training. During the finetuning phase, the model is finetuned using only the RUL-labeled vibration acceleration data for the last 50% of the Bearing 1_1. Bearing 1_3, Bearing 1_4, Bearing 1_5, Bearing 1_6, and Bearing 1_7 are test sets.

5.1.3 Hyperparameters

In the alternate contrast stage, the regularization factor μ_c is 0.0001, the learning rate is 0.0001, the epoch is 800, and the main sensors alternate every 100 epochs, stochastic gradient descent (SGD) is selected as the optimization algorithm, and the corresponding momentum is 0.9. The batch size is 128, the dictionary size $K+1$ is 3201, the momentum update factor m is 0.999, the temperature coefficient is τ is 0.07, and the feature dimension is 128.

In the finetuning stage, the regularization factor μ_f takes a value of 0.01, the learning rate is 0.0001, the epoch is 200, SGD is selected as the optimization algorithm, the corresponding momentum is 0.9, and the batch size is 128.

5.1.4 Metrics

The evaluation metrics are the mean absolute percentage error (MAPE), and the score.

$$MAPE = \frac{1}{11}\sum_{i=1}^{11}\frac{|ActRUL_i - \widehat{RUL_i}|}{ActRUL_i}, \tag{18}$$

$$\%Er_i = 100 \times \frac{ActRUL_i - \widehat{RUL_i}}{ActRUL_i}, \tag{19}$$

$$A_i = \begin{cases} \exp^{-\ln(0.5)\cdot(Er_i/5)} & \text{if } Er_i \le 0 \\ \exp^{+\ln(0.5)\cdot(Er_i/20)} & \text{if } Er_i > 0 \end{cases}, \tag{20}$$

$$\text{score } = \frac{1}{11}\sum_{i=1}^{11}(A_i), \tag{21}$$

where $ActRUL_i$ and $\widehat{RUL_i}$ are the true and the predicted RUL values respectively, of the i-th bearing. The score reflects the average performance of the model in predicting the final RUL of the 11 bearings. The higher the score, the higher is the model performance. Further, the lower the MAPE, the higher is the model performance.

5.1.5 Baselines

Due to our goal of RUL prediction under scarce labeled data, we choose four self-supervised methods to better show the method's performance.

Self-supervised pretraining via contrast learning (SSPCL (Ding et al., 2022)): SSPCL is based on data augmentation. The most similar variant to the current variant is searched for among all moments of data variants augmented by other data in the pretraining phase.

Self-supervised learning (SSL (Krokotsch et al., 2022)): An SSL model for RUL prediction is proposed. The time interval between any two time series segments is estimated in the pretraining phase, thus learning the temporal correlation of the signals.

Unedited

Table 1: Experimental dataset setup

Working condition	Alternate contrast phase (Training without RUL labels)	Finetuning phase (Training with RUL labels)	Testing data
Condition 1	Bearing 1_2 Bearing 2_1 Bearing 2_2 Bearing 3_1 Bearing 3_2	50% of Bearing 1_1 data	Bearing 1_3 Bearing 1_4 Bearing 1_5 Bearing 1_6 Bearing 1_7
Condition 2	Bearing 1_1 Bearing 1_2 Bearing 2_2 Bearing 3_1 Bearing 3_2	50% of Bearing 2_1 data	Bearing 2_3 Bearing 2_4 Bearing 2_5 Bearing 2_6 Bearing 2_7
Condition 3	Bearing 1_1 Bearing 1_2 Bearing 2_1 Bearing 2_2 Bearing 3_2	50% of Bearing 3_1 data	Bearing 3_3

Deep self-supervised learning (DeepSSL (Akrim et al., 2023)): A DeepSSL model to overcome the lack of labeled data for RUL prediction is proposed. The encoding and decoding architectures are designed using gated recurrent unit (GRU) to perform temporal prediction of the sensor signals in the pretraining phase.

Unlabeled sample learning (USL (Kong et al., 2023)): A contrastive learning framework for RUL prediction is proposed. First, an unlabeled sample augmentation is developed to extend the sample set. Then, an unlabeled sample learning (USL) architecture is proposed to learn degradation information from unlabeled samples to improve the performance of general deep learning models in RUL prediction.

5.1.6 Experimental fairness

Our problem setting involves massive unlabeled sensor data and scarce sensor data with RUL labels. The self-supervised method has training data that are consistent with our method in two phases. Thus, all comparisons of our method with the self-supervised baselines are fair.

5.2 Results

5.2.1 Comparison with self-supervision baselines

Our proposed method and all self-supervised baseline methods are validated on 11 bearings using a large amount of unlabeled bearing data and 50% of the labeled degradation data. Tab. 2 presents the true values of RUL, the predicted values of our method with other baseline methods, the MAPE, and the score (the mean value of Ai). Our method achieves an optimal score of 0.738 and MAPE of 0.108, while the suboptimal score is 0.616 and the suboptimal MAPE is 0.166. Our proposed multi-sensor contrast neural network outperforms SSPCL, SSL, DeepSSL, and USL in terms of MAPE, and score, with at least a 0.058 decrease in MAPE, and a 0.122 increase in score. These results demonstrate the effectiveness of our alternate contrast, which maximizes the similarity of features between multiple sensors at the same moment, emphasizing the discriminative power of similar features across sensors. Self-supervised RUL prediction algorithms, represented by SSL, SSPCL, DeepSSL and USL, typically construct pretraining tasks by stacking sensor signals and exploiting the temporal correlation of sequences. However, these algorithms may be inefficient when the multisensor signals at different degradation stages are relatively similar. This is where our proposed method is needed to improve the discrimination of features and achieve better performance by continuously alternate contrast to maximize the similarity of features between multiple sensors at the same moment.

We notice that compared to other bearings, the RUL prediction effect for Bearing 2_3 is unsatisfactory. However, all the comparison baselines also struggle to achieve good performance, possibly due to the uniqueness of Bearing 2_3 itself, making it difficult to predict effectively given the scarcity of labeled RUL data.

Fig. 3 shows the RUL prediction for Bearing 1_3. As can be observed from Fig. 3, in the early stage of bearing degradation, the prediction of RUL is close to horizontal and our method does not perform very well. Whereas at the end of bearing degra-

Unedited

Table 2: Experimental results for comparison with self-supervised methods with 50% labeled data and sufficient unlabeled data

Metrics	Method	B 1_3[1]	B 1_4	B 1_5	B 1_6	B 1_7	B 2_3	B 2_4	B 2_5	B 2_6	B 2_7	B 3_3	Mean
Actual RUL(s)	-	5730	339	1610	1460	7570	7530	1390	3090	1290	580	820	-
Predicted RUL(s)	SSL	5402	282	1426	1576	5044	3139	1446	2803	1072	524	749	-
	SSPCL	3484	287	1348	1334	3907	**3709**	1167	2634	1156	523	729	-
	DeepSSL	5367	297	1479	1487	6502	2778	1202	2518	1107	443	756	-
	USL	5445	279	1300	1344	6953	2344	1269	2573	1103	492	762	-
	Our method	**5607**	**314**	**1505**	**1483**	**7000**	2670	**1418**	**2807**	**1204**	**543**	**783**	-
A_i[2]	SSL	0.820	0.558	0.673	0.332	0.315	0.133	0.572	0.725	0.557	0.716	0.741	0.558
	SSPCL	0.257	0.588	0.569	0.741	0.187	**0.172**	0.573	0.600	0.698	0.711	0.681	0.525
	DeepSSL	0.803	0.651	0.754	0.774	0.613	0.112	0.626	0.526	0.612	0.441	0.763	0.607
	USL	0.842	0.542	0.513	0.759	0.754	0.092	0.740	0.560	0.605	0.591	0.783	0.616
	Our method	**0.928**	**0.774**	**0.798**	**0.804**	**0.770**	0.107	**0.756**	**0.728**	**0.794**	**0.802**	**0.855**	**0.738**
MAPE	SSL	0.057	0.168	0.114	0.079	0.334	0.583	0.040	0.093	0.169	0.097	0.087	0.166
	SSPCL	0.392	0.153	0.163	0.086	0.484	**0.507**	0.160	0.148	0.104	0.098	0.111	0.219
	DeepSSL	0.063	0.124	0.081	0.018	0.141	0.631	0.135	0.185	0.142	0.236	0.078	0.167
	USL	0.050	0.177	0.193	0.079	0.082	0.689	0.087	0.167	0.145	0.152	0.071	0.172
	Our method	**0.021**	**0.074**	**0.065**	**0.016**	**0.075**	0.645	**0.020**	**0.092**	**0.067**	**0.064**	**0.045**	**0.108**

[1] B denotes Bearing, e.g. B 1_3 denote Bearing 1_3.

[2] The mean value of A_i is the official metric score of the FEMTO-ST dataset. The higher the score, the better is the model performance.

Unedited

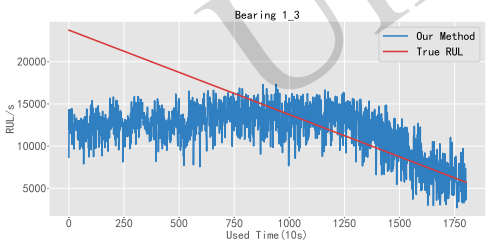


Fig. 3: RUL prediction for bearings. The horizontal axis is the used life and the vertical axis is the predicted RUL.

dation, our method can fit RUL change better. This is because during model training, we only use labeled data from the second half of the bearing degradation, so our method in will perform better in the second half of the stage.

Fig. 4 shows the visualization of the Bearing 1_3 fusion feature after TSNE. Blue and green colors represent the early stages of degradation, yellow and orange indicate the middle stages of degradation, and red shows the end stages of degradation. The degradation trajectory is visible, and the transition from the current degradation stage to the next degradation stage can be very smooth. This proves

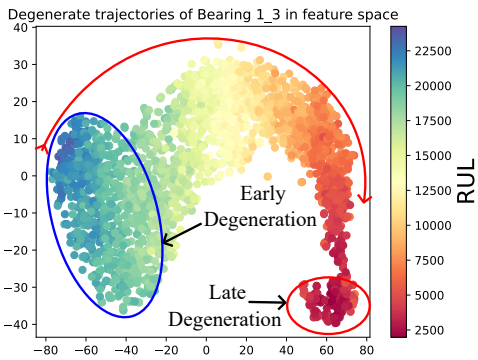


Fig. 4: Bearing 1_3 fusion feature visualization.

that our method captures the degradation pattern of the bearing

5.2.2 Ablation study

To better explain the superiority of our method, ablation experiments are designed to construct the following two variants.

Variant-NP: To verify the effectiveness of the alternate contrast phase in the framework, the alternate contrast phase is skipped and the model is trained directly using a few data with RUL labels, and the remainder of the framework is retained unchanged.

Variant-Loss: To verify the effectiveness of the NCE loss function in the framework, the NCE loss function is replaced with the MSE loss function for calculating the feature differences between different sensors at the same moment, and the remainder of the framework remains unchanged.

Tab. 3 presents a comparative analysis of predicted RULs, the MAPE, and the score among our proposed method and its variants. We conduct these experiments to verify the effectiveness of the alternate contrast, and NCE loss in our proposed method. To verify the effectiveness of alternate contrast, our proposed method is compared with Variant-NP, which does not have the alternate contrast mechanism. The MAPE of our proposed method is reduced by 0.184, compared to Variant-NP, while the score is improved by 0.302. Alternate contrast based on cross-sensor similarity can effectively reduce the need for labeled RUL data by capturing multi-sensor similar features. It also demonstrates that our model does not rely on finetuning to achieve good results.

To verify the effectiveness of NCE loss, our proposed method was compared with Variant-Loss, which uses MSE loss instead of NCE loss. The MAPE of our proposed method is reduced by 0.148, compared to Variant-loss, while the score is improved by 0.282. The MSE loss captures multisensor similarity by closing the distance between multisensor features at the same moment. NCE loss ensures that cross-sensor invariant highly discriminative degradation features are extracted by drawing on the idea of classification to extract multisensor similarity at the same moment while suppressing multisensor feature similarities at different moments.

5.2.3 Hyperparametric sensitivity analysis

Fig. 5 shows the effects of the dictionary size on the model performance. As the dictionary size increases, the prediction error decreases. This is because the larger the dictionary size in the alternate contrast phase, the more difficult it is to match the main sensor features in auxiliary sensor features to the auxiliary sensor features at the same moment, which will result in enhancing the model's ability to extract similar features across sensors.

6 Conclusions

In this paper, for RUL prediction, we propose a multisensor contrast method that utilizes abundant unlabeled sensor data to assist a small amount of sensor data with RUL labels. Typically, current methods stack multisensor signals and mine the multisensor temporal autocorrelation from a large amount of unlabeled sensor data during pretraining, but these suffer from poor discrimination of degradation features. Our approach uses an alternate contrast process to capture similar features among multisensors, which indicate the machine health conditions, and can effectively improve the discrimination of degradation features. The attention mechanism is used for fine-tuning to establish an association between degradation features and RUL. Lastly, we fully evaluate our approach using the open FEMTO-ST bearing dataset, where the testing dataset contains 11 bearing degradation data under three different operating conditions. The proposed model outperforms other state-of-the-art baselines on test data, showing that, for RUL prediction, our proposed model can utilize rich unlabeled sensor data to assist few sensor data with RUL labels.

Our method also has some limitations, for example, our model only considers the bearing functioning under a single operating condition and ignores the effect of variable operating conditions on the RUL. From the application perspective, our proposed multisensor contrast framework can greatly improve the prediction accuracy using only a small amount of labeled degradation data and thus can be widely used for the prediction of a bearing's RUL, thereby reducing the failure maintenance cost of bearings. In the future, we are committed to analyzing and modeling the effects of dynamically changing operating condi-

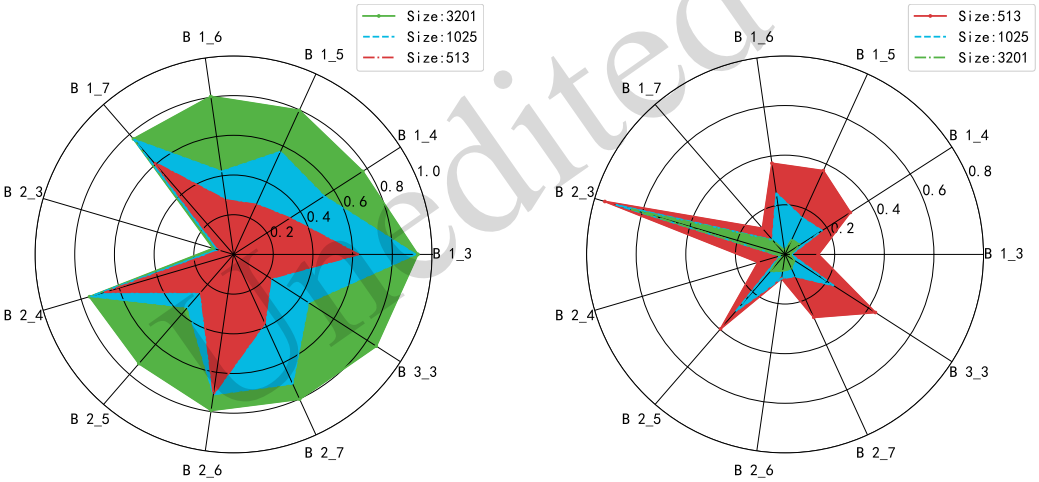
Unedited

Table 3: Ablation results

Metrics	Method	B 1_3[1]	B 1_4	B 1_5	B 1_6	B 1_7	B 2_3	B 2_4	B 2_5	B 2_6	B 2_7	B 3_3	Mean
Actual RUL(s)	-	5730	339	1610	1460	7570	7530	1390	3090	1290	580	820	-
Predicted RUL(s)	Variant-NP	4889	192	1110	770	6924	2559	1172	1445	1110	499	706	-
	Variant-loss	4896	211	1201	1186	6200	2556	1205	1982	1114	454	681	-
	Our method	**5607**	**314**	**1505**	**1483**	**7000**	**2670**	**1418**	**2807**	**1204**	**543**	**783**	-
A_i	Variant-NP	0.601	0.222	0.341	0.194	0.744	0.101	0.581	0.158	0.617	0.616	0.618	0.436
	Variant-loss	0.604	0.270	0.415	0.522	0.534	0.101	0.630	0.289	0.623	0.471	0.556	0.456
	Our method	**0.928**	**0.774**	**0.798**	**0.804**	**0.770**	**0.107**	**0.756**	**0.728**	**0.794**	**0.802**	**0.855**	**0.738**
MAPE	Variant-NP	0.147	0.434	0.311	0.473	0.085	0.660	0.157	0.532	0.140	0.140	0.139	0.292
	Variant-loss	0.146	0.378	0.254	0.188	0.181	0.661	0.133	0.359	0.136	0.217	0.170	0.256
	Our method	**0.021**	**0.074**	**0.065**	**0.016**	**0.075**	**0.645**	**0.020**	**0.092**	**0.067**	**0.064**	**0.045**	**0.108**

[1] B denotes Bearing, e.g. B 1_3 denote Bearing 1_3.

Unedited



(a) Score of our method under different dictionary sizes. (b) MAPE of our method under different dictionary sizes.

Fig. 5: Metrics of our method under different dictionary sizes.

tions on the bearing's RUL, so that it can be adapted to a variety of complex industrial production environments.

Contributors

Binkun LIU designed the research. Zhenyi XU processed the data. Binkun LIU drafted the paper. Zhenyi XU organized the paper. Zhenyi XU and Yunbo ZHAO helped the data control and project management. Yunbo ZHAO, Yang CAO and Yu KANG revised and finalized the paper. Zhenyi XU and Yunbo ZHAO provided the funding acquisition.

Conflict of interest

All the authors declare that they have no conflict of interest.

References

Akrim A, Gogu C, Vingerhoeds R, et al., 2023. Self-supervised learning for data scarcity in a fatigue damage prognostic problem. *Engineering Applications of Artificial Intelligence*, 120:105837.

Behera S, Misra R, 2023. A multi-model data-fusion based deep transfer learning for improved remaining useful life estimation for iiot based systems. *Engineering Applications of Artificial Intelligence*, 119:105712.

Deng Y, Du S, Wang D, et al., 2023. A calibration-based hybrid transfer learning framework for rul prediction of rolling bearing across different machines. *IEEE Transactions on Instrumentation and Measurement*, 72:1-15.

Ding Y, Zhuang J, Ding P, et al., 2022. Self-supervised pretraining via contrast learning for intelligent incipient

fault detection of bearings. *Reliability Engineering & System Safety*, 218:108126.

Ding Y, Jia M, Cao Y, et al., 2023. Domain generalization via adversarial out-domain augmentation for remaining useful life prediction of bearings under unseen conditions. *Knowledge-Based Systems*, 261:110199.

Dong S, Xiao J, Hu X, et al., 2023. Deep transfer learning based on bi-lstm and attention for remaining useful life prediction of rolling bearing. *Reliability Engineering & System Safety*, 230:108914.

He K, Fan H, Wu Y, et al., 2020. Momentum contrast for unsupervised visual representation learning. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, p.9729-9738.

Kong Z, Jin X, Xu Z, et al., 2023. A contrastive learning framework enhanced by unlabeled samples for remaining useful life prediction. *Reliability Engineering & System Safety*, 234:109163.

Korbar B, Tran D, Torresani L, 2018. Cooperative learning of audio and video models from self-supervised synchronization. *Advances in Neural Information Processing Systems*, 31.

Krokotsch T, Knaak M, et al., 2022. Improving semi-supervised learning for remaining useful lifetime estimation through self-supervision. *International Journal of Prognostics and Health Management*, 13(1).

Li Q, Yan C, Chen G, et al., 2022a. Remaining useful life prediction of rolling bearings based on risk assessment and degradation state coefficient. *ISA Transactions*, 129:413-428.

Li Y, Wang H, Li J, et al., 2022b. A 2-d long short-term memory fusion networks for bearing remaining useful life prediction. *IEEE Sensors Journal*, 22(22):21806-21815.

Melendez I, Doelling R, Bringmann O, 2019. Self-supervised multi-stage estimation of remaining useful life for electric drive units. 2019 IEEE International Conference on Big Data (Big Data), p.4402-4411.

Morales-Espejel GE, Gabelli A, 2020. A model for rolling bearing life with surface and subsurface survival: Surface thermal effects. *Wear*, 460-461:203446.

Nectoux P, Gouriveau R, Medjaher K, et al., 2012. Pronostia: An experimental platform for bearings accelerated degradation tests. IEEE International Conference on Prognostics and Health Management, PHM'12, p.1-8.

Saeed A, Salim FD, Ozcelebi T, et al., 2020. Federated self-supervised learning of multisensor representations for embedded intelligence. *IEEE Internet of Things Journal*, 8(2):1030-1040.

Souza JS, Bezerril MC, Silva MA, et al., 2021. Motor speed estimation and failure detection of a small uav using density of maxima. *Frontiers of Information Technology & Electronic Engineering*, 22(7):1002-1009.

Tao F, Qi Q, Liu A, et al., 2018. Data-driven smart manufacturing. *Journal of Manufacturing Systems*, 48:157-169.

Tian Y, Krishnan D, Isola P, 2020. Contrastive multiview coding. Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16, p.776-794.

Wang B, Lei Y, Li N, et al., 2020. Multiscale convolutional attention network for predicting remaining useful life of machinery. *IEEE Transactions on Industrial Electronics*, 68(8):7496-7504.

Wang W, Wang Y, Wang J, et al., 2022. Ensemble enhanced active learning mixture discriminant analysis model and its application for semi-supervised fault classification. *Frontiers of Information Technology & Electronic Engineering*, 23(12):1814-1827.

Wang X, Wang T, Ming A, et al., 2021. Cross-operating condition degradation knowledge learning for remaining useful life estimation of bearings. *IEEE Transactions on Instrumentation and Measurement*, 70:1-11.

Wang Y, Cai F, Pan Z, et al., 2023. Self-supervised graph learning with target-adaptive masking for session-based recommendation. *Frontiers of Information Technology & Electronic Engineering*, 24(1):73-87.

Wen Y, Wu J, Zhou Q, et al., 2018. Multiple-change-point modeling and exact bayesian inference of degradation signal for prognostic improvement. *IEEE Transactions on Automation Science and Engineering*, 16(2):613-628.

Yang L, Liao Y, Duan R, et al., 2023. A bidirectional recursive gated dual attention unit based rul prediction approach. *Engineering Applications of Artificial Intelligence*, 120:105885.

Zhang B, Mao Y, Chen X, et al., 2021a. Self-supervised learning advance fault diagnosis of rotating machinery. Neural Computing for Advanced Applications, p.319-332.

Zhang W, Chen D, Kong Y, 2021b. Self-supervised joint learning fault diagnosis method based on three-channel vibration images. *Sensors*, 21(14):4774.

Zhu S, Pu J, 2021. A self-supervised method for treatment recommendation in sepsis. *Frontiers of Information Technology & Electronic Engineering*, 22(7):926-939.

Zou W, Lu Z, Hu Z, et al., 2023. Remaining useful life estimation of bearing using deep multiscale window-based transformer. *IEEE Transactions on Instrumentation and Measurement*, 72:1-11.

Zuo T, Zhang K, Zheng Q, et al., 2023. A hybrid attention-based multi-wavelet coefficient fusion method in rul prognosis of rolling bearings. *Reliability Engineering & System Safety*, 237:109337.

List of electronic supplementary materials

1. Analysis of dataset
2. Extended experimental analysis

Unedited