CROSS: Contrastive Representation Learning for Optimal Source-to-Source Transfer Learning

Seunghwan Song*
Department of Industrial & Management Engineering
Korea National University of Transportation
Chungju, South Korea
seunghwan@ut.ac.kr

Joo-Hyuk Oh
Samsung Electronics Company Ltd.
Hwaseong, South Korea
jh1.oh@samsung.com

Jun-Geol Baek*
Department of Industrial and Management Engineering
Korea University
Seoul, South Korea
jungeol@korea.ac.kr

***Abstract*— Deep learning-based prognostics and health management (PHM) has advanced remaining useful life (RUL) prediction; however, its performance remains limited by domain-specific dependence and noisy degradation data. The present paper introduces Contrastive Representation learning for Optimal Source-to-Source transfer (CROSS), a two-stage self-supervised framework that enhances domain generalization in degradation modeling. In the initial phase, a dual-tiered alarm system integrates the augmented Dickey–Fuller (ADF) test and acceleration analysis to identify the primary prediction time (FPT). This approach eliminates transient fluctuations and precisely delineates degradation initiation. In Stage 2, a negative-free contrastive learning module encodes temporal continuity between historical and future subsequences, thereby producing invariant representations for cross-domain transfer. Experiments on benchmark datasets demonstrate that CROSS outperforms baseline models, achieving the lowest mean absolute error (MAE) and the most precise FPT alignment with ground-truth failure points. The findings substantiate the hypothesis that the amalgamation of statistical FPT detection with contrastive representation learning furnishes a resilient, domain-invariance approach for intelligent industrial PHM applications.**

Keywords—Prognostics and Health Management, Remaining Useful Life Prediction, Contrastive Representation Learning, Time-Series Representation, Source-to-Source Adaptation

I. INTRODUCTION

Industrial systems are increasingly relying on predictive intelligence to ensure the reliability of their equipment and minimize downtime. The aim of prognostics and health management (PHM) is to estimate the remaining useful life (RUL) of components, enabling maintenance to be scheduled before failures occur [1], [2]. In contrast, traditional reactive maintenance (RM) only acts after breakdowns occur, whereas preventive maintenance (PM) follows fixed intervals regardless of actual degradation [3]. Neither approach is efficient in dynamic manufacturing environments.

Recent advances in deep learning have made it possible to use data to predict health outcomes by mapping sensor signals directly to health indicators [4]. However, three persistent challenges remain: (1) Domain dependence: models trained under one operating condition rarely generalize to another; (2) Data imbalance: healthy data dominate, while data on degradation are scarce; (3) Noise variability: environmental fluctuations obscure degradation patterns. Conventional transfer learning methods, such as conditional domain adversarial networks (CDAN) [5] and domain-adversarial networks (DAN) [6], align distributions between domains, but they ignore the temporal evolution of degradation, which is essential for accurate RUL prediction.

*Co-corresponding author

To address these issues, we propose Contrastive Representation learning for Optimal Source-to-Source transfer (CROSS). Unlike single-domain adaptation, CROSS first creates invariant degradation representations across multiple source domains, which are then transferred to the target domain.

The framework has two stages. The Stage 1 is first prediction time (FPT) identification. This involves a two-level alarm mechanism that integrates the augmented Dickey-Fuller (ADF) test for non-stationarity detection [7] and acceleration-based trend analysis to suppress false alarms caused by transient fluctuations. The intersection of both alarms defines the FPT, marking the transition from healthy to degraded states.

The next Stage 2 is RUL prediction. For each degradation segment, historical and future sub-sequences are encoded through a shared convolutional network. The framework minimizes the cosine distance between the predicted and actual future embeddings using negative-free contrastive loss to promote temporal continuity. Learned representations are transferred across sources using partial weight freezing to preserve degradation-invariant features while adapting to new conditions.

The main contributions are as follows:

- A robust, two-level alarm method that unifies FPT identification across machines and suppresses random noise.
- A negative-free contrastive objective that captures intrinsic temporal continuity in degradation.
- A source-to-source transfer strategy that aligns degradation structures prior to fine-tuning using limited target data.

The remainder of this paper is organized as follows. Section II reviews related works. The proposed CORSS framework is described in Section III. In Section IV, the performance of the proposed method is evaluated using a benchmark datasets. Finally, Section V concludes the paper and discusses future research directions.

II. RELATED WORKS

A. RUL Predcition and Recent PHM

Estimating the RUL of critical components is a core objective of PHM. Traditional approaches have relied on physical degradation models or statistical filtering techniques, such as particle filters and Kalman filter variants [1], [2]. Recent research applies deep learning architectures, such as convolutional, recurrent, and attention-based, to learn degradation dynamics directly from multi-sensor signals [4], [8]. While these methods are highly accurate on individual machines, their reliance on domain-specific data limits their deployment in heterogeneous environments.

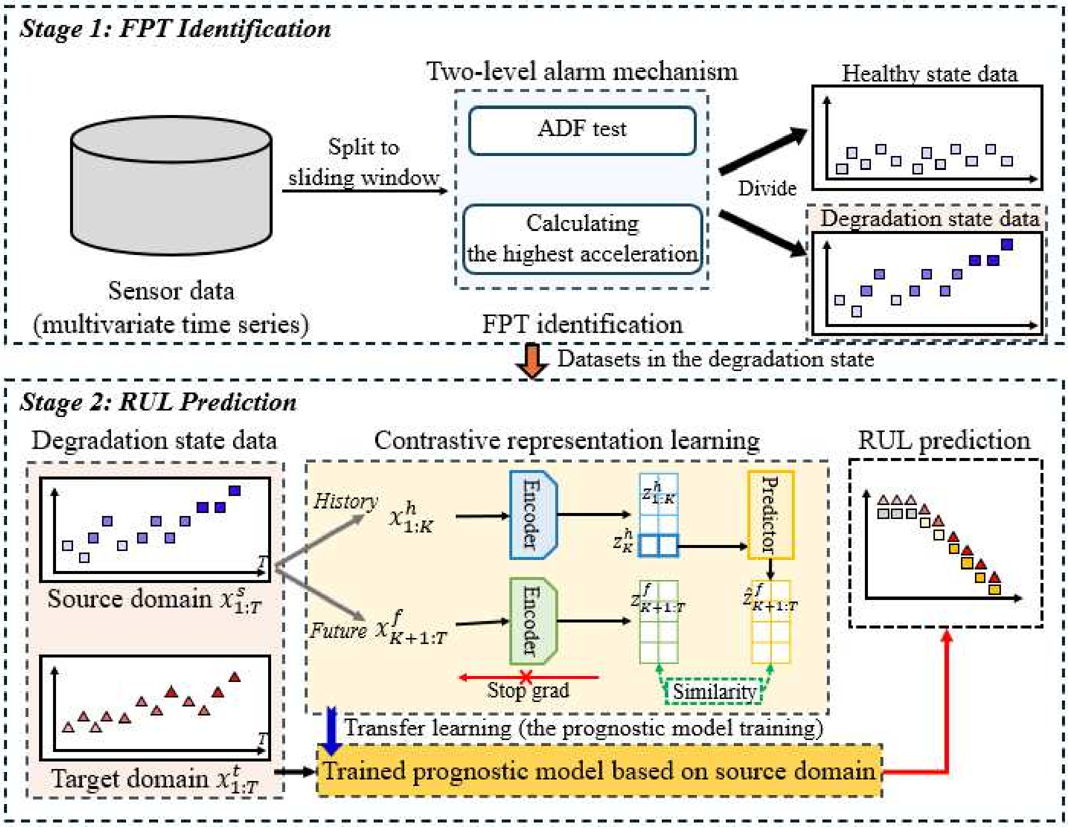

Fig. 1. The overall workflow for contrastive representation learning for optimal source-to-source transfer.

Transfer learning (TL) has been widely used to address domain mismatch by aligning feature spaces between different working conditions. Methods such as CDAN [5] and DAN [6] encourage invariance by minimizing the discrepancy between source and target distributions through adversarial training. However, these methods typically operate at the distribution level, disregarding the temporal degradation trajectory underlying RUL estimation. Recent studies have explored kernel-based and moment-matching strategies, but these still assume stationarity and are susceptible to overfitting in the presence of sensor noise or partial degradation records [3].

B. Time-series Representation Learning

The goal of representation learning is to obtain task-agnostic embeddings from sequential data. Self-supervised frameworks, such as temporal neighborhood coding (TNC) [9], temporal and contextual contrasting (TS-TCC) [10], and TS2Vec [11], demonstrate that context prediction can yield transferable features. However, industrial degradation differs from generic sequences in that it is irreversible, non-stationary, and often domain-dependent. Therefore, models that assume stationary or cyclic signals cannot reliably capture gradual fault propagation.

C. Contrastive Learning for PHM

Conventional contrastive frameworks, such as the simple framework for contrastive learning of visual representations (SimCLR) [12] and momentum contrast (MoCo v3) [13], rely on explicit negative pairs to separate dissimilar instances in the embedding space. However, in degradation modeling, two similar temporal trajectories may be incorrectly treated as negatives, which can lead to representation collapse and unstable optimization.

Negative-free paradigms, such as bootstrap your own latent (BYOL) [14] and masked hierarchical cluster-wise contrastive learning (MHCCL) [15], address this problem by aligning predictions between twin encoders without creating negatives. While these approaches generally improve the stability of self-supervised learning, they rarely consider domain shifts or the nonstationary progression of industrial degradation. This limits their applicability to PHM tasks.

Building on these observations, the proposed CROSS framework aims to overcome two limitations of prior transfer-learning and self-supervised studies: the lack of explicit FPT detection and the absence of source-to-source alignment. CROSS couples a statistical, ADF-based, two-level alarm for robust FPT identification with a negative-free, temporal, contrastive encoder [7]. This enables reliable detection of degradation onset and learning of domain-invariant, temporal representations. This design improves cross-domain RUL prediction under noisy and limited data conditions, bridging the gap between contrastive representation learning and practical industrial prognostics.

III. Proposed Method

The proposed CROSS framework consists of two sequential stages: First, a two-level alarm mechanism is used to identify the FPT. Second, RUL prediction is performed through contrastive representation learning and partial weight transfer. Fig. 1 shows an overview of the workflow.

CROSS consists of two main modules: (1) a two-level alarm mechanism to locate the First Prediction Time (FPT), and (2) a contrastive representation encoder for degradation modeling and transfer learning.

A. Stage 1: FPT Identification

In industrial signals, most of the data correspond to a healthy state, which provides little information about degradation. Therefore, detecting the point at which a component begins to deteriorate—the FPT—is critical for identifying meaningful degradation segments. However, transient fluctuations, sensor drift, and random noise often trigger false alarms when using a single indicator or threshold.

To address this issue, the CROSS system uses a two-level alarm mechanism integrating the ADF test for statistical non-stationarity and acceleration-based trend analysis for sustained changes in signal magnitude. For each sliding window $W_i = \{x_{t_i}, \ldots, x_{t_i+L}\}$, the ADF test evaluates the null hypothesis of a unit root as (1), following the classical formulation introduced by Dickey and Fuller [16].

$$\Delta x_t = \alpha + \beta t + \gamma x_{t-1} + \sum_{i=1}^{p} \phi_i \Delta x_{t-i} + \varepsilon_t \qquad (1)$$

where $\gamma < 0$ indicates stationarity.

An alarm is triggered when the ADF $p-value < 0.1$, suggesting that the local distribution has shifted from the previous neighborhood. Next, the local acceleration is computed as (2).

$$a_t = \frac{(x_{t+1} - 2x_t + x_{t-1})}{\Delta t^2} \qquad (2)$$

If the average a_t over n consecutive windows exceeds the threshold τ_a, a trend alarm is raised.

The intersection of statistical and trend alarms defines the FPT, which marks the transition from a healthy state to a degraded state. This dual criterion filters out short-term noise and stabilizes FPT detection across heterogeneous operating conditions.

B. Stage 2: RUL Prediction

After the degradation segments have been isolated, CROSS uses contrastive representation learning to capture temporal continuity and enable cross-domain transfer. Each segment is divided into historical (X_h) and future (X_f) sub-sequences of equal length. These sub-sequences are encoded by a shared, multi-scale, one-dimensional (1D) convolutional network $F_\theta(\cdot)$:

$$z_h = F_\theta(X_h),\ z_f = F_\theta(X_f) \qquad (3)$$

A predictor $G_\phi(\cdot)$maps the last historical embedding to a predicted future representation $\hat{z}_f = G_\phi(z_h)$. The negative-free contrastive loss function promotes temporal consistency by minimizing the angular distance between the predicted and actual embeddings.

$$\mathcal{L}_{\text{CROSS}} = 1 - \frac{\langle \hat{z}_f, z_f \rangle}{\|\hat{z}_f\|_2 \|z_f\|_2} \qquad (4)$$

Gradients are stopped through z_f, as in BYOL [14], to prevent representation collapse. After training on the source domain, freeze the lower convolutional layers and fine-tune the upper layers on another source domain to achieve source-to-source transfer before adapting to the final target. A lightweight regression head $R_\psi(\cdot)$ predicts the RUL as (5):

$$\hat{y}_{RUL} = R_\psi(z_f) \qquad (5)$$

The overall loss combines contrastive and smoothness terms as (6).

$$\mathcal{L}_{total} = \mathcal{L}_{\text{CROSS}} + \lambda \parallel \nabla_\theta \mathcal{L}_{\text{CROSS}} \parallel_2^2 \qquad (6)$$

where λ controls the trade-off between alignment and stability. This integrated design enables Stage 1 to accurately detect degradation onset and allows Stage 2 to learn domain-invariant temporal embeddings that can be transferred across operating conditions. These mechanisms enable CROSS to robustly predict RUL across domains, even under noisy or partially observed degradation trajectories.

IV. Experiments

A. Experimental Setup

1) Datasets: CROSS was validated on two public benchmarks dataset. XJTU-SY [17]: Vibration data from LDK UER204 bearings under three load levels (25.6kHz sampling, 32,768 points/min). PHM 2012 [18]: IEEE PHM Challenge bearing run-to-failure signals (25.6kHz, 2,560 points/10s). Fig. 2 illustrates the experimental setups of both public datasets The summary of the dataset is shown in Table I..

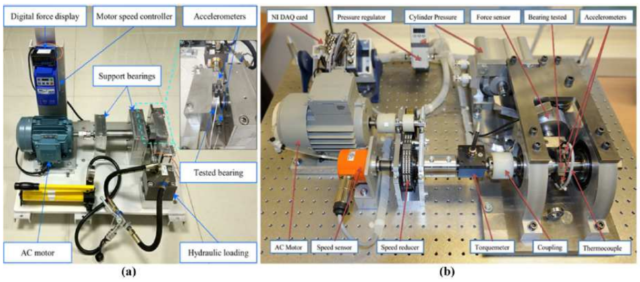

Fig. 2 Public bearing test datasets: (a) XJTU-SY; (b) PHM2012..

TABLE I. Summary of Datasets

Dataset	Name	Bearing Entities
XJTU-SY	XB1	XB11, XB12, XB13, XB15
	XB2	XB21, XB22, XB23, XB24, XB25
PHM2012	PB1	PB11, PB12, PB13, PB14, PB15, PB16, PB17
	PB2	PB21, PB22, PB23, PB24, PB26, PB27

To evaluate source-to-source transfer, each dataset was divided into multiple cross-domain tasks (e.g., XB1 → XB2, PB1 → PB2). Training used a sliding window size 20, a batch size 128, an epoch size 300, and a learning rate 1×10^{-4}. Fifty percent of the encoder layers were frozen during source transfer.

2) Baseline and metrics: For FPT identification, three classic detectors were compared: (1) the 2σ principle method, (2) the gradient method with the linear rectification technique (LRT), and (3) the gradient method with the moving average filter (MAF). Performance was evaluated by the difference between the detected FPT and the ground truth failure time for each sequence.

Three prognostic methods with different transfer strategies were compared for RUL prediction. Baseline (without transfer learning), CDAN [5], and DAN [6]. The evaluation metric was mean absolute error (MAE).

B. Results of FPT Identification

The proposed two-level alarm robustly detects the FPT, even under noisy conditions. Fig. 3 shows that each colored vertical marker corresponds to an FPT estimated by a different method according to the colors in the legend: proposed, 2σ, gradient (LRT), and gradient (MAF).

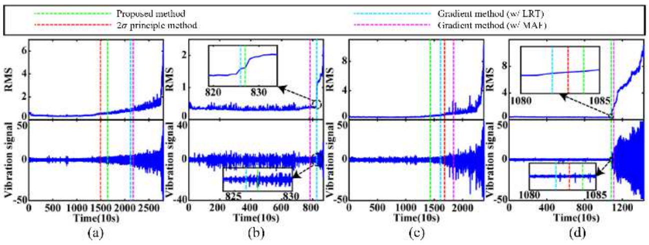

Fig. 3. FPT identification results comparison in XJTU-SY dataset: (a) XB11, (b) XB12, (c) XB13, (d) XB14.

Across the XJTU-SY dataset, the markers corresponding to the proposed method align most closely with the ground-truth failure references. In contrast, the 2σ method tends to trigger prematurely, and the gradient-based methods remain sensitive to local fluctuations despite smoothing.

These results confirm that fusing statistical (ADF) and kinetic (acceleration) indicators provide stable and accurate detection of degradation onset. The experimental evidence shows that the proposed method can effectively capture the degradation trend.

TABLE I. CROSS-DOMAIN RUL PREDICTION RESULTS

Transfer Scenario	Test Bearing	Proposed Method	Baseline	CDAN	DAN
XB1-2	XB21	**0.1031**	0.1920	0.2091	0.1770
	XB22	**0.1147**	0.2219	0.1871	0.2022
	XB23	**0.1435**	0.2538	0.2418	0.2311
	XB24	0.2077	0.3070	0.2335	**0.1996**
	XB25	**0.1684**	0.2111	0.1812	0.1852
XB2-1	XB11	**0.1823**	0.2227	0.2171	0.2096
	XB12	**0.1491**	0.2207	0.1787	0.2153
	XB13	**0.1250**	0.2332	0.1733	0.1814
	XB15	**0.0965**	0.1685	0.1385	0.1226
PB1-2	PB21	**0.1516**	0.1641	0.2522	0.2051
	PB22	0.2376	0.4753	**0.2147**	0.2865
	PB23	**0.2082**	0.3412	0.2751	0.3662
	PB24	**0.2292**	0.2651	0.3288	0.2721
	PB26	**0.2151**	0.2652	0.2854	0.2367
	PB27	**0.1956**	0.2312	0.2078	0.3122
PB2-1	PB11	**0.2292**	0.3821	0.2631	0.2869
	PB12	**0.1535**	0.2145	0.2620	0.2641
	PB13	**0.1517**	0.2861	0.2393	0.2582
	PB14	0.3665	0.4310	0.3887	**0.3592**
	PB15	**0.1521**	0.1942	0.2554	0.2417
	PB16	0.1821	0.2151	0.2668	0.2615
	PB17	0.1255	0.2571	0.2063	0.2082

C. Cross-Domain RUL Prediction

After FPT segmentation, CROSS performs contrastive representation learning and source-to-source transfer. Fig. 4 shows representative RUL trajectories across domains, and Table II presents the quantitative results. The proposed model yields smoother and more accurate RUL curves than CDAN and DAN, the latter of which often exhibits delayed or stepwise degradation prediction.

CROSS achieves the lowest MAE across all scenarios, indicating superior domain-invariant representation learning. The largest relative improvement is seen in transfer tasks with a significant operating-condition mismatch (e.g., PB1 → PB2), where conventional adversarial alignment (DA) demonstrates unstable adaptation.

Additionally, the overall trend of CROSS predictions closely follows the true degradation trajectory. Predicted RUL values decrease smoothly as operation time increases, and degradation states correspond well to FPT boundaries identified in Stage 1. This consistency shows that the learned temporal representations capture continuous damage propagation. This allows for reliable RUL forecasting under domain shift.

V. CONCLUSION

This study presented CROSS, a two-stage framework for robust degradation modeling and RUL prediction. Stage 1 involves a two-level alarm that combines the ADF test and acceleration analysis to effectively identify the FPT. This process filters out transient noise and produces reliable degradation segments. Stage 2 involves a negative-free contrastive representation that learns smooth, domain-invariant temporal embeddings. This enables efficient source-to-source transfer for RUL estimation. Experiments on benchmark demonstrated that, compared with CDAN and DAN, CROSS consistently reduced MAE by 15-30% while achieving the most accurate FPT alignment with ground-truth degradation onset. These results confirm the advantage of combining statistical detection and contrastive learning for domain generalization in prognostics.

Future work will extend CROSS toward multi-sensor fusion and transformer-based encoders to capture cross-modal degradation dynamics. Additionally, integrating online adaptation and continual learning will enable the real-time deployment of CROSS within industrial PHM systems.

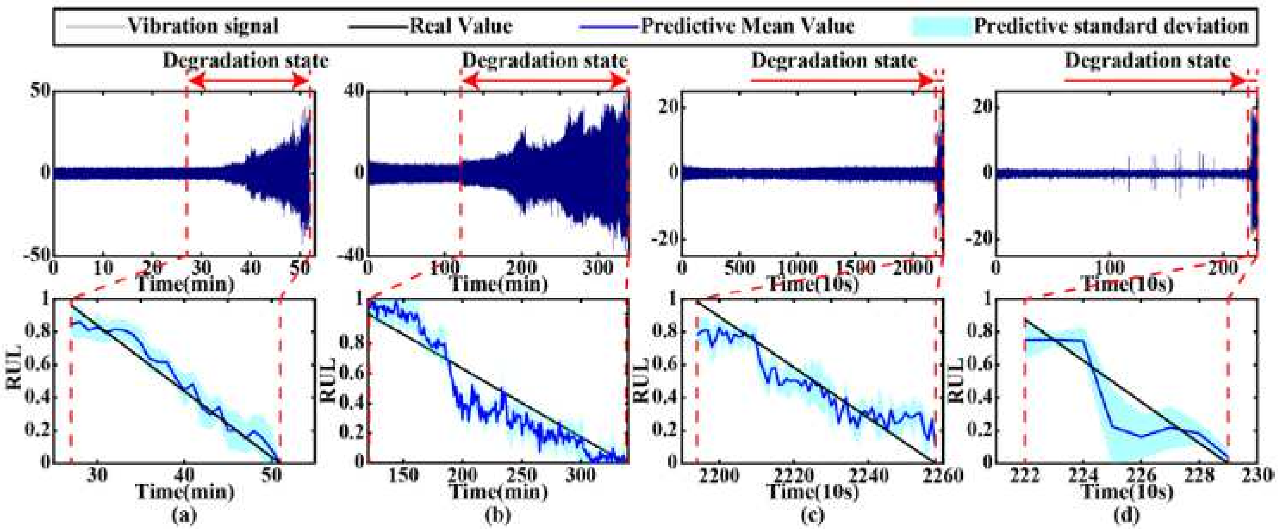

Fig. 4. RUL prediction trends on target domains: (a) XB15 in XB2→XB1, (b) XB25 in XB1→XB2, (c) PB17 in PB2→PB1, (d) PB17 in PB2→PB1.

ACKNOWLEDGMENT

This was supported by Korea National University of Transportation in 2025.

REFERENCES

[1] A. Heng, S. Zhang, A. C. C. Tan, and J. Mathew, "Rotating machinery prognostics: State of the art, challenges, and opportunities," *Mech. Syst. Signal Process.*, vol. 23, no. 3, pp. 724–739, 2009.

[2] M. Pecht, *Prognostics and Health Management of Electronics*, Hoboken, NJ, USA: Wiley, 2008.

[3] M. Yazdi, "Maintenance strategies and optimization techniques," in Advances in Computational Mathematics for Industrial System Reliability and Maintainability, Cham, Switzerland: Springer Nature Switzerland, 2024, pp. 43–58.

[4] F. Wu, J. Xu, and Y. Li, "Remaining useful life prediction based on deep learning," *Sensors*, vol. 24, no. 11, p. 3454, 2024.

[5] M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Conditional adversarial domain adaptation," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, pp. 1647–1657, 2018.

[6] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 1–35, 2016.

[7] Y. Liu and R. Wang, "A comprehensive overview of remaining useful life prediction," *Eng. Rep.*, vol. 7, 2025.

[8] A. Siffer, P.-A. Fouque, A. Termier, and C. Largouët, "Anomaly detection in streams with extreme value theory," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, Aug. 2017, pp. 1067–1075..

[9] S. Tonekaboni, D. Eytan, and A. Goldenberg, "Unsupervised representation learning for time series with temporal neighborhood coding," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2021.

[10] E. Eldele, M. Ragab, Z. Chen, M. Wu, C.-K. Kwoh, X. Li, and C. Guan, "Time-series representation learning via temporal and contextual contrasting," in *Proc. Int. Joint Conf. Artificial Intelligence (IJCAI)*, 2021, pp. 2862–2870..

[11] Z. Yue, Y. Zhang, and T. Tan, "TS2Vec: Towards universal representation of time series," in *Proc. AAAI Conf. Artificial Intelligence (AAAI)*, vol. 36, no. 8, 2022, pp. 9352–9360.

[12] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn. (ICML), vol. 119, Proc. Mach. Learn. Res. (PMLR)*, 2020, pp. 1597–1607.

[13] X. Chen, S. Xie, and K. He, "An empirical study of training self-supervised vision transformers," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2021, pp. 9640–9649.

[14] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, and M. Valko, "Bootstrap your own latent: A new approach to self-supervised learning," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 33, 2020, pp. 21271–21284.

[15] Q. Meng, H. Qian, Y. Liu, L. Cui, Y. Xu, and Z. Shen, "MHCCL: Masked hierarchical cluster-wise contrastive learning for multivariate time series," in *Proc. AAAI Conf. Artificial Intelligence (AAAI)*, vol. 37, no. 8, 2023, pp. 9153–9161.

[16] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *J. Amer. Stat. Assoc.*, vol. 74, no. 366, pp. 427–431, 1979.

[17] XJTU-SY Bearing Dataset, Xi'an Jiaotong Univ., Xi'an, China, 2018. [Online]. Available: https://biaowang.tech/xjtu-sy-bearing-dataset

[18] IEEE PHM 2012 Data Challenge, *Proc. IEEE Int. Conf. on Prognostics and Health Management (PHM)*, Denver, CO, USA, 2012.