Time Series Anomaly Detection using Skip-Step Contrastive Predictive Coding

Kexin Zhang*
Institute of Cyber-Systems and Control
Zhejiang University
Hangzhou, P. R. China
zhangkexin@zju.edu.cn

Qingsong Wen†
DAMO Academy, Alibaba Group
Bellevue, WA, USA
qingsong.wen@alibaba-inc.com

Chaoli Zhang
DAMO Academy, Alibaba Group
Hangzhou, P. R. China
chaoli.zcl@alibaba-inc.com

Liang Sun
DAMO Academy, Alibaba Group
Bellevue, WA, USA
liang.sun@alibaba-inc.com

Yong Liu
Institute of Cyber-Systems and Control
Zhejiang University
Hangzhou, P. R. China
yongliu@iipc.zju.edu.cn

Abstract

Self-supervised learning (SSL) shows impressive performance in many tasks lacking sufficient labels. In this paper, we study SSL in time series anomaly detection (TSAD) by incorporating the characteristics of time series data. Specifically, we build an anomaly detection algorithm consisting of global pattern learning and local association learning. The global pattern learning module builds encoder and decoder to reconstruct the raw time series data to detect global anomalies. To complement the limitation of the global pattern learning that ignores local associations between anomaly points and their adjacent windows, we design a local association learning module, which leverages contrastive predictive coding (CPC) to transform the identification of anomaly points into positive pairs identification in contrastive learning. Motivated by the observation that adjusting the distance between the history window and the time point to be detected directly impacts the detection performance in the CPC framework, we further propose a skip-step CPC scheme in the local association learning module which adjusts the distance for better construction of the positive pairs and detection results. The experimental results show that the proposed algorithm achieves superior performance on multiple benchmark datasets in comparison with 11 state-of-the-art algorithms.

1 Introduction

Time series anomaly detection (TSAD) is one of the challenging tasks in many real-world applications like server monitoring, process control, influenza detection, and so on [1; 2; 3; 4; 5]. Traditional methods usually use feature engineering to generate features and then design feature-based detection algorithms to achieve anomaly detection. As real systems become more and more complex,

*This work was done when Kexin Zhang was a Research Intern at Alibaba Group
†Corresponding author

NeurIPS 2022 Workshop on Self-Supervised Learning - Theory and Practice.

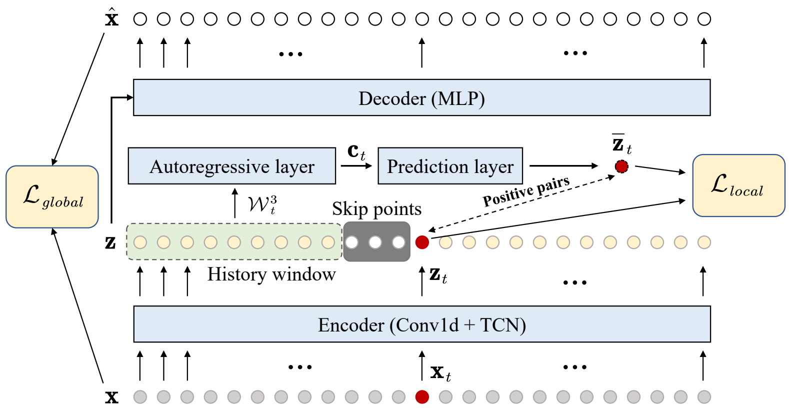

Figure 1: The model structure of the proposed time series anomaly detection algorithm, which consists of two major modules: an autoencoder global pattern learning module and a skip-step CPC-based local association learning module.

feature engineering becomes more difficult and traditional methods fail to achieve good detection performance. Deep learning (DL) is a powerful technique for feature learning, which directly extracts features from the data and reduces the efforts in the design of hand-crafted features. Generally, building a successful deep learning model requires a large amount of labeled data. Unfortunately, labeled anomaly data is very limited for time series data, which motivates us to seek a method for anomaly detection using unlabeled data more effectively.

Recently, self-supervised learning (SSL), which allows for learning representations without ground-truth labels, has exerted great power in the fields of computer vision and natural language processing. Contrastive learning (CL) is an important branch of SSL, and it has been applied widely in image classification tasks [6]. The key idea behind CL is to minimize the distance between similar samples and maximize the distance between dissimilar samples. Similar samples are usually obtained through data augmentation or context sampling, and such samples are defined as positive samples in CL. On the contrary, dissimilar samples are negative samples. Recently, CL has also been adopted in time series analysis. Due to the space limitation, we summarize the related work on time series anomaly detection and the application of contrastive learning in time series in Appendix A.1.

In this paper, we investigate contrastive learning for time series anomaly detection. We observe that due to the rarity of anomalies in time series, it is difficult to use anomaly points to construct positive pairs in the contrastive learning framework straightforwardly. In other words, the contrastive loss will be large when the constructed positive sample pairs contain anomaly points. Based on this observation, we introduce contrastive predictive coding (CPC) [7] to the TSAD task. The CPC aims to learn representations by predicting the future in latent space through a given history window. The distance between history window and future time points is an important factor affecting the detection results because different distances represent different positive pairs. Therefore, we study how positive samples are constructed and propose a skip-step CPC that is more suitable for the TSAD task.

Based on the previous analysis, we propose a new TSAD algorithm consisting of two major modules: an autoencoder (AE) global pattern learning module and a skip-step CPC-based local association learning module. The global pattern learning module tries to reconstruct the raw data through an encoder and a decoder, in which the anomaly points have significantly larger reconstruction errors. A limitation of the global module is that it ignores local associations between anomaly points and their adjacent windows. To address this issue, a local association learning module is designed to leverage contrastive predictive coding (CPC) to transform the identification of anomaly points into positive pairs identification tasks in contrastive learning. The final anomaly score is then calculated by fusing the global anomaly score and local anomaly score, where each time point is determined as normal or abnormal with a threshold.

2 Methodology

Suppose the time series samples has C measurements with length N, and we denote by $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) \in \mathbb{R}^{N \times C}$, where $\mathbf{x}_t \in \mathbb{R}^C$ is the observation at time t. In time series anomaly detection, our goal is to determine whether $\mathbf{x}_t$ is an anomaly point. Figure 1 shows the structure of

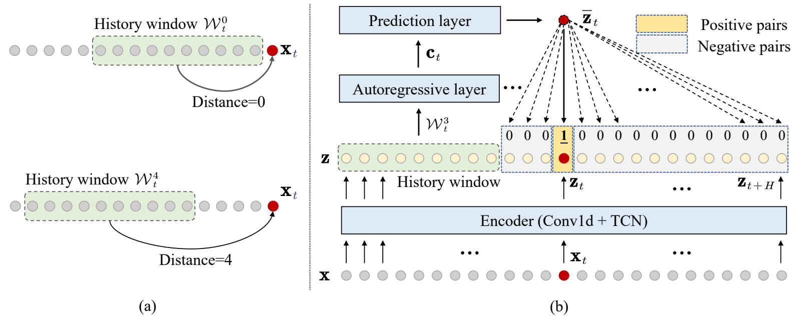

Figure 2: Illustration of the proposed skip-step CPC scheme: (a) constructing positive pairs with different distances; (b) the strategy for calculating local contrastive score.

the proposed time series anomaly detection algorithm. First, a single 1-D convolutional layer maps the input time point $\mathbf{x}_t$ to embedding space $\mathbf{h}_t = \mathrm{Conv1d}(\mathbf{x}_t)$, where $\mathbf{h}_t \in \mathbb{R}^d$. Next, a temporal convolutional network (TCN) [8] is used to further map $\mathbf{h}_t$ to more informative representation $\mathbf{z}_t = \mathrm{TCN}(\mathbf{h}_t)$. Then, the model is divided into two branches. The first branch uses MLP as a decoder to reconstruct the raw input $\mathbf{x}_t$, i.e., $\hat{\mathbf{x}}_t = \mathrm{MLP}(\mathbf{z}_t)$. This branch learns global information in the data, which makes anomaly points exhibit large reconstruction errors. However, this module ignores the local associations between anomaly points and their adjacent windows, which is very useful for determining whether a time point is normal or abnormal [9]. Therefore, we introduce another new branch, called Skip-Step CPC, to capture the local association between a time point and its adjacent time window. The second branch uses an autoregressive model to summarize the historical window $\mathcal{W}_t$ in the new latent space and produces a context representation $\mathbf{c}_t$. Unlike the original CPC, we found that keeping a certain distance between $\mathcal{W}_t$ and $\mathbf{x}_t$ can better capture local associations.

A simple illustration is shown in Figure 2(a). We define $\mathcal{W}_t^d$ to be a historical window with the distance d from $\mathbf{x}_t$, where d represents the number of time points between the last time point of $\mathcal{W}_t$ and $\mathbf{x}_t$. We expect that when $\mathbf{x}_t$ is not an anomaly point, this branch maximizes the mutual information between the $\mathbf{x}_t$ and the context representation of $\mathcal{W}_t^d$. On the contrary, if $\mathbf{x}_t$ is an anomaly point, it is difficult to extract the underlying patterns in common.

Both the AE-based and Skip-Step CPC-based branches are trained to jointly optimize the final loss as

$$\mathcal{L} = \mathcal{L}_{AE} + \mathcal{L}_{CPC}. \tag{1}$$

The first term in Eq. (1) is the reconstruction error loss, defined as

$$\mathcal{L}_{AE} = \frac{1}{NC}\sum_{i=1}^{N}\sum_{j=1}^{C}(x_i^j - \hat{x}_i^j)^2, \tag{2}$$

where x_i^j is the ground truth and $\hat{x}_i^j$ is the output of the AE-based branch. The second term in Eq. (1) is the InfoNCE loss $\mathcal{L}_{CPC}$, which can maximize a lower bound of mutual information [7], defined as

$$\mathcal{L}_{CPC} = \mathbb{E}_{\mathcal{X}}[-\log \frac{\exp(\mathbf{z}_t^T \mathbf{W}^d \mathbf{c}_t^d)}{\sum_{\mathbf{z}_j \in \mathcal{X}} \exp(\mathbf{z}_j^T \mathbf{W}^d \mathbf{c}_t^d)}], \tag{3}$$

where $\mathcal{X}$ is a set of N random training samples. Each sample contains a history window and a time point. The loss in Eq. (3) is the categorical cross-entropy of classifying the positive sample correctly.

Finally, the output anomaly score for $\mathbf{x}_t$ is calculated by fusing global and local anomaly scores as

$$S(t) = S_{local}(t) * S_{global}(t). \tag{4}$$

The global anomaly score for $\mathbf{x}_t$ is calculated by reconstruction error, defined as

$$S_{global}(t) = (\mathbf{x}_t - \hat{\mathbf{x}}_t)^2. \tag{5}$$

The local anomaly score for $\mathbf{x}_t$ is computed by the cross entropy loss, as shown in Figure 2(b). Specifically, we first make a prediction $\bar{\mathbf{z}}_t = \mathbf{W}^d \mathbf{c}_t^d$, then the distribution for the time point is computed based on the similarity between the prediction $\bar{\mathbf{x}}_t$ and the representations of all time points in the predefined windows $\mathcal{W}_p$. Therefore, the local contrastive score is calculated as

$$S_{local}(t) = -\log \frac{\exp(\mathbf{z}_t^T \bar{\mathbf{z}}_t)}{\sum_{\mathbf{z}_j \in \mathcal{W}_p} \exp(\mathbf{z}_j^T \bar{\mathbf{z}}_t)}. \tag{6}$$

Table 1: Results of time series anomaly detection on three popular benchmark datasets. The best results are highlighted.

Dataset	SMD			PSM			MSL		
Metric	P	R	F1	P	R	F1	P	R	F1
Deep-SVDD	78.54	79.67	79.10	95.41	86.49	90.73	91.92	76.63	83.58
LSTM	78.55	85.28	81.78	76.93	89.64	82.80	85.45	82.50	83.95
CL-MPPCA	82.36	76.07	79.09	56.02	99.93	71.80	73.71	88.54	80.44
ITAD	86.22	73.71	79.48	72.80	64.02	68.13	69.44	84.09	76.07
LSTM-VAE	75.76	90.08	82.30	73.62	89.92	80.96	85.49	79.94	82.62
BeatGAN	72.90	84.09	78.10	90.30	93.84	92.04	89.75	85.42	87.53
OmniAnomaly	83.68	86.82	85.22	88.39	74.46	80.83	89.02	86.37	87.67
InterFusion	87.02	85.43	86.22	83.61	83.45	83.52	81.28	92.70	86.62
THOC	79.76	90.95	84.99	88.14	90.99	89.54	88.45	90.97	89.69
TS-CP^2	87.42	66.25	75.38	82.67	78.16	80.35	86.45	68.48	76.42
AnomalyTrans	89.40	95.45	92.33	96.91	98.90	97.89	92.09	95.15	**93.59**
Ours	91.75	97.34	**94.46**	98.36	98.74	**98.55**	90.84	94.73	92.75

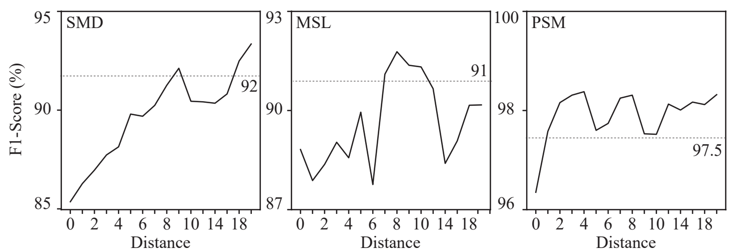

Figure 3: Anomaly detection results at different distances in the proposed skip-step CPC scheme.

3 Experiments and Discussion

In this section, we evaluate our detection algorithm on three popular benchmark datasets with 11 DL-based baseline models. The detailed experimental settings are described in Appendix A.2 and the results are summarized in Table 1. It can be observed that detection methods considering context information of time series are more likely to achieve good performance, such as AnomalyTrans which considers local associations. In summary, our proposed method achieves better performance than all 11 baseline models on datasets SMD and PSM, and achieves results close to the state-of-the-art on the MSL dataset. The superiority of our model comes from incorporating both reconstruction-based and prediction-based strategies from the global pattern and local context information perspectives to make better decisions.

As aforementioned, adjusting the distance between the history window and the time point to be detected in the proposed skip-step CPC scheme has a direct impact on the detection performance in the proposed method. The underlying reason behind this is that the adjustment of the distance essentially changes the construction of the positive pairs in the CPC framework. To illustrate it, we conduct experiments with different distances on the three benchmark datasets and depict the results in Figure 3. It can be observed that as the distance increases, the F1-score of SMD also increases, and better detection performance is achieved when the distance is greater than 8. The best detection performance for the MSL is obtained when the distance is 8. The F1-score of PSM increases first and then remains stable, and it is least sensitive to distance. The results in Figure 3 show that we need to set different distances to construct reasonable positive pairs under different datasets for better anomaly detection results. Moreover, two different cases are provided in Appendix A.3 to further illustrate the effect of different positive sample distances on anomaly detection results.

4 Conclusion and Future Work

This paper investigates the application of SSL for TSAD and we propose a new detection method that combines the AE-based global pattern learning module and the skip-step CPC-based local association learning module. The proposed method achieves superior performance to the state-of-the-art results. In future, we plan to extend the current research, including the adaptive construction of positive and negative samples and the theoretical basis of the combination of SSL and time series analysis.

References

[1] Q. Wen, L. Yang, T. Zhou, and L. Sun, “Robust time series analysis and applications: An industrial perspective,” in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2022, pp. 4836–4837.

[2] A. Blázquez-García, A. Conde, U. Mori, and J. A. Lozano, “A review on outlier/anomaly detection in time series data,” *ACM Computing Surveys (CSUR)*, vol. 54, no. 3, pp. 1–33, 2021.

[3] J. Gao, X. Song, Q. Wen, P. Wang, L. Sun, and H. Xu, “RobustTAD: Robust time series anomaly detection via decomposition and convolutional neural networks,” *KDD Workshop MileTS*, 2020.

[4] A. A. Cook, G. Mısırlı, and Z. Fan, “Anomaly detection for iot time-series data: A survey,” *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6481–6494, 2019.

[5] L. Li, J. Yan, Q. Wen, Y. Jin, and X. Yang, “Learning robust deep state space for unsupervised anomaly detection in contaminated time-series,” *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 2022.

[6] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A simple framework for contrastive learning of visual representations,” in *Proceedings of the 37th International Conference on Machine Learning (ICML)*, 2020.

[7] A. van den Oord, Y. Li, and O. Vinyals, “Representation learning with contrastive predictive coding,” *CoRR*, vol. abs/1807.03748, 2018.

[8] S. Bai, J. Z. Kolter, and V. Koltun, “An empirical evaluation of generic convolutional and recurrent networks for sequence modeling,” *CoRR*, vol. abs/1803.01271, 2018.

[9] J. Xu, H. Wu, J. Wang, and M. Long, “Anomaly transformer: Time series anomaly detection with association discrepancy,” in *International Conference on Learning Representations (ICLR)*, 2022.

[10] L. Shen, Z. Li, and J. Kwok, “Timeseries anomaly detection using temporal hierarchical one-class network,” in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, 2020, pp. 13 016–13 026.

[11] Y. Su, Y. Zhao, C. Niu, R. Liu, W. Sun, and D. Pei, “Robust anomaly detection for multivariate time series through stochastic recurrent neural network,” in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery* & *Data Mining*, 2019, p. 2828–2837.

[12] L. Shen, Z. Yu, Q. Ma, and J. T. Kwok, “Time series anomaly detection with multiresolution ensemble decoding,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 11, 2021, pp. 9567–9575.

[13] J. Audibert, P. Michiardi, F. Guyard, S. Marti, and M. A. Zuluaga, “Usad: Unsupervised anomaly detection on multivariate time series,” in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery* & *Data Mining*, 2020, p. 3395–3404.

[14] K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Soderstrom, “Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding,” in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery* & *Data Mining*, 2018, p. 387–395.

[15] T. Schneider, C. Qiu, M. Kloft, D. Aspandi-Latif, S. Staab, S. Mandt, and M. Rudolph, “Detecting anomalies within time series using local neural transformations,” *CoRR*, vol. abs/2202.03944, 2022.

[16] C. U. Carmona, F.-X. Aubet, V. Flunkert, and J. Gasthaus, “Neural contextual anomaly detection for time series,” in *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, 2022, pp. 2843–2851.

[17] C. Zhang, T. Zhou, Q. Wen, and L. Sun, "TFAD: A decomposition time series anomaly detection architecture with time-frequency analysis," in *31st ACM International Conference on Information and Knowledge Management (CIKM)*, 2022.

[18] J.-Y. Franceschi, A. Dieuleveut, and M. Jaggi, "Unsupervised scalable representation learning for multivariate time series," in *Proceedings of the 33rd International Conference on Neural Information Processing Systems (NeurIPS)*, 2019.

[19] S. Tonekaboni, D. Eytan, and A. Goldenberg, "Unsupervised representation learning for time series with temporal neighborhood coding," in *International Conference on Learning Representations (ICLR)*, 2021.

[20] X. Yang, Z. Zhang, and R. Cui, "TimeCLR: A self-supervised contrastive learning framework for univariate time series representation," *Knowledge-Based Systems*, vol. 245, p. 108606, 2022.

[21] Y. Jiao, K. Yang, D. Song, and D. Tao, "Timeautoad: Autonomous anomaly detection with self-supervised contrastive loss for multivariate time series," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 3, pp. 1604–1619, 2022.

[22] Q. Wen, L. Sun, F. Yang, X. Song, J. Gao, X. Wang, and H. Xu, "Time series data augmentation for deep learning: A survey," in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, 2021, pp. 4653–4660.

[23] Z. Yue, Y. Wang, J. Duan, T. Yang, C. Huang, Y. Tong, and B. Xu, "TS2Vec: Towards universal representation of time series," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 8, 2022, pp. 8980–8987.

[24] 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 *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, 2021, pp. 2352–2359.

[25] L. Yang and S. Hong, "Unsupervised time-series representation learning with iterative bilinear temporal-spectral fusion," in *Proceedings of the 39th International Conference on Machine Learning (ICML)*, vol. 162, 17–23 Jul 2022, pp. 25 038–25 054.

[26] G. Woo, C. Liu, D. Sahoo, A. Kumar, and S. Hoi, "CoST: Contrastive learning of disentangled seasonal-trend representations for time series forecasting," in *International Conference on Learning Representations (ICLR)*, 2022.

[27] A. Abdulaal, Z. Liu, and T. Lancewicki, "Practical approach to asynchronous multivariate time series anomaly detection and localization," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery* & *Data Mining*, 2021, p. 2485–2494.

A Appendix

A.1 Related Works

A.1.1 time series Anomaly Detection

Time series anomaly detection is a very vital and challenging task in practice. Most methods use the unsupervised learning paradigm because obtaining enough labeled data is time-consuming and laborious. The methods can be roughly categorized into clustering-based, reconstruction-based, prediction-based, and association-based.

The clustering-based methods detect anomalies by measuring the distance between time points and normal pattern cluster centers. THOC [10] uses the fuses features from all intermediate layers of dilated recurrent neural network by a differentiable hierarchical clustering mechanism and detects the anomalies by a novel score that measures the distance in the multiple hyperspheres. The reconstruct-based methods typically use an AE architecture to reconstruct the raw data and then detect anomalies based on the reconstruction error [11; 12; 13]. For example, OmniAnomaly [11] captures the normal patterns by learning their robust representations with stochastic variable connection and planar normalizing flow and uses the reconstruction probabilities to determine anomalies. USAD [13] proposes an encoder-decoder architecture within an adversarial training framework that allows it to isolate anomalies while providing fast training. The prediction-based methods employ an autoregressive model to predict the future time points and calculate the prediction error [14; 15]. For example, LNT [15] applies predictive contrastive learning (CPC) to produce good semantic time series representations and makes predictions of the context at different time horizons. The association-based method is a new anomaly detection approach that learns associations between a time point and its adjacent time points. AnomalyTrans [9] proposes the transformer-based model with a new designed anomaly-attention mechanism, which can model the prior-association and series-association simultaneously to embody the association discrepancy, and the learned association discrepancy is a criterion for detecting anomalies. NCAD [16] defines three kinds of windows, the context window, the full window, and the suspect window. The model encodes the context window and the full window using the same network and computes a distance score between them in the embedding space. A high score is given if the instances with an anomaly are in the suspect window. While the aforementioned algorithms mostly work in the time domain for anomaly detection, a recent TFAD model [17] exploits both time and frequency domains, together with decomposition and data augmentation mechanisms for performance improvement.

A.1.2 Supervised Contrastive learning in time series

As one of the SSL methods, CL has been widely adopted in time series analysis. Categorizing by the generation of positive and negative samples, the paradigms roughly include the temporal-context-based, augmentation-based, mixture-based, and time-frequency-based methods.

The idea behind the temporal-context-based methods [18; 19] is an assumption that given an anchor (a window or a time point), the windows (or time points) adjacent to it are more likely to be the positive samples. The windows (or time points) far away from it should be negative samples. Augmentation-based methods employ data augmentation techniques to generate the required positive and negative samples rather than sampling directly from the data. These methods basically follow the learning paradigm of SimCLR [6], but design new augmentation approaches for time series data, such as DTW-based augmentation [20] and randomly sample segments and exchange locations [21]. More data augmentation approaches for time series can be found in [22]. The mixture-based methods combine context-based and augmentation-based methods, making the model learn informative representations. The recent works include TS2vec [23] and TSTCC [24]. The last method is the time-frequency based method that applies the CL framework not only in the time domain but also in the frequency domain. BTSF [25] incorporates the spectral information in feature representation and devises a novel iterative bilinear temporal-spectral fusion to explicitly encode the affinities of abundant time-frequency pairs. CoST [26] is a new representation learning framework that comprises both time domain and frequency domain contrastive losses to learn discriminative trends and seasonal representations.

A.2 Experimental Settings

We evaluate our method on three datasets. (1) SMD (Server Machine Dataset) [11] is a new 5-week-long dataset and is collected from a large Internet company with 38 dimensions. (2) MSL (Curiosity) [14] is a spacecraft anomaly detection dataset and it has 55 dimensions. (3) PSM (Pooled Server Metrics) [27] is collected internally from multiple application server nodes at eBay with 26 dimensions. We use random sampling to obtain training samples, and each sample contains a historical window and a future window. The historical window is with a fixed size of 50 and the future window is set to 20 for all datasets. The time points are labeled as anomalies if their anomaly scores are larger than the predefined threshold θ. The widely-used adjustment strategy [11; 9] that an anomaly segment is considered correctly detected as long as any point in this segment is detected is adopted in this paper. We use precision P, recall R, and F1-score $F1$ as evaluation metrics. F1-score is the harmonic mean of precision and recall. A high value indicates better performance. We compare our model with 11 DL-based methods, including Deep-SVDD , LSTM , CL-MPPCA , ITAD , LSTM-VAE , BeatGAN , OmniAnomaly , InterFusion , THOC, TS-CP^2, and AnomalyTrans. The results of the baselines are collected from [9].

A.3 Case analysis

We take the most distance-sensitive dataset as an example, i.e., SMD. The distances are set to 0 and 19, respectively. The first is the false negative and false positive detection case as shown in Figure 4. We observe that local contrastive scores exhibit different patterns at anomaly locations when different distances of positive samples are given. Although the reconstruction error score is correct when the distance is 0, the wrong local score leads to false negative detection. on the contrary, the local score is correct when the distance is 19, which means the model learns the correct pattern. If we only use the global reconstruction score, it would result in a false positive detection (a spike with right circles in Figure 4). Therefore, the correct local contrastive score and correct global reconstruction score amplify the final anomaly score, making the anomalous time points easier to detect. The second is a case that contains both true detection and false positive detection, as shown in Figure 5. In this case, we observed that although the local detection module provides the correct score at distance 0, it also leads to excessively large scores (spikes with right circles) elsewhere, detecting some normal points as anomaly points. However, the local anomaly score only shows a larger value at the location of the anomaly when the distance is 19.

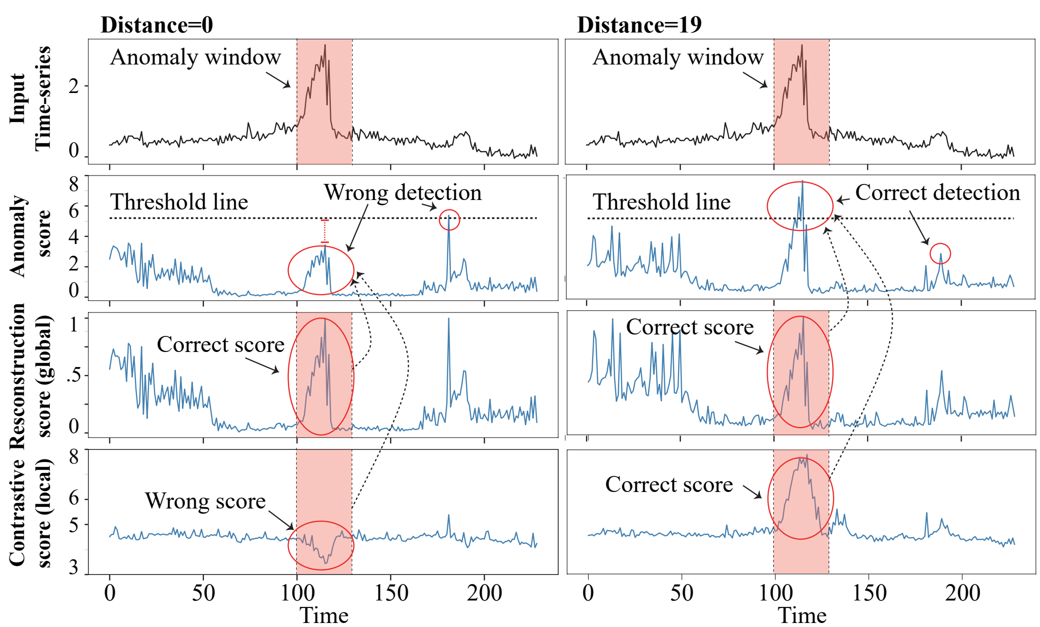

Figure 4: Case 1 of the distance studies in the proposed skip-step CPC scheme. Left figure: false negative around time index 110 and false positive around time index 180 under distance=0. Right figure: the correct results are obtained under distance=19.

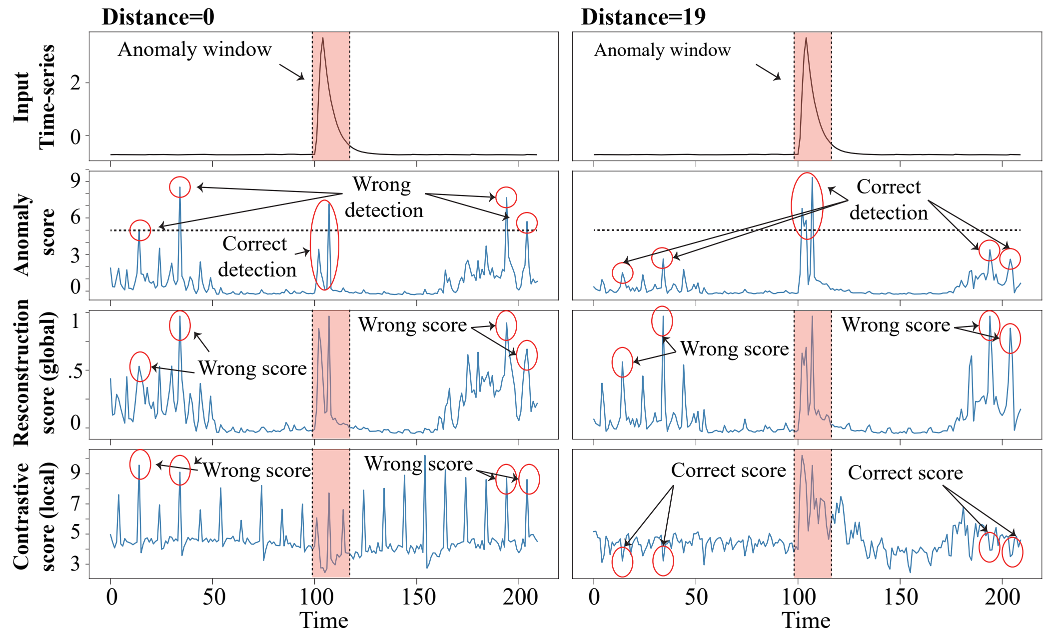

Figure 5: Case 2 of the distance studies in the proposed skip-step CPC scheme. Left figure: false positive around time indices 15, 35, 190, and 210 under distance=0. Right figure: the correct results are obtained under distance=19.