Journal of Computer Technology and Software

ISSN: 2998-2383

Vol. 3, No. 5, 2024

Temporal Contrastive Representation Learning for Unsupervised Anomaly Detection in High-Dimensional Cloud Environments

Sibo Wang
Rice University, Houston, USA
coldbrew737@gmail.com

Abstract: This paper proposes an unsupervised anomaly detection method based on contrastive learning to address challenges in cloud computing environments, such as high data dimensionality, complex structure, and lack of labels. The method segments raw time series monitoring data into subsequences using a sliding window mechanism and applies various data augmentation strategies to construct positive and negative sample pairs, guiding the model to learn discriminative embeddings without supervision. A temporal attention mechanism is integrated to capture key dynamic features in the sequence, enhancing the model's ability to represent long-term dependencies and local fluctuations. Anomaly scores are calculated by measuring similarities in the embedding space, enabling an efficient detection process without the need for labels. The method is evaluated on a cloud monitoring dataset across different augmentation strategies, parameter settings, and temporal modeling configurations. Experimental results show that it outperforms several recently published unsupervised models in F1 Score, AUC, and KS Score, demonstrating its effectiveness and engineering adaptability in handling high-dimensional dynamic data within cloud platform scenarios.

Keywords: unsupervised learning; time series modeling; embedding representation; system monitoring

1. Introduction

In today's data-driven technological ecosystem, cloud computing has become the foundation for supporting complex applications and large-scale data processing tasks. As enterprises increasingly pursue digital and intelligent transformation, the workloads on cloud platforms continue to grow. System structures are becoming more complex, and the runtime environment is becoming more dynamic. This high degree of virtualization and distribution brings not only higher demands for scalability and flexibility but also exposes the platform to more potential security and stability risks. Abnormal events such as performance bottlenecks, network attacks, and resource misallocations can significantly impact service quality and even lead to business interruptions. Therefore, efficiently identifying and locating anomalies in large-scale and dynamic environments has become a core issue in ensuring the stable operation of cloud platforms.

However, anomaly detection in cloud computing presents significant challenges[1]. First, the monitored data is high-dimensional and originates from multiple sources, including logs, metrics, and tracing data, resulting in strong heterogeneity and high dimensionality. Second, due to frequent system evolution and rapid component updates, the boundary between normal and abnormal states is often unclear, and abnormal patterns are difficult to define in advance. Additionally, a large portion of the data lacks labeled annotations, making traditional supervised detection methods less applicable. Against this background, it is crucial to find a method that can extract essential features and identify hidden anomaly patterns without relying on labels[2].

In recent years, unsupervised learning methods have attracted considerable attention in the field of anomaly detection. Their main advantage lies in the ability to discover structural features and behavioral patterns from data without the need for manual labeling. Among them, contrastive learning, as a representative unsupervised representation learning strategy, guides the model to learn discriminative feature representations by constructing positive and negative sample pairs. It demonstrates strong feature extraction capabilities and broad transferability. Contrastive learning enhances the model's sensitivity to structural or semantic differences, thereby improving its ability to distinguish abnormal states. Applying this approach to unsupervised anomaly detection in cloud environments shows great potential in addressing challenges such as high dimensionality and label scarcity.

Moreover, data in cloud computing scenarios exhibit clear temporal properties and dynamic evolution. System performance metrics, network traffic, and user behavior traces often contain complex temporal dependencies and local fluctuation patterns. Effectively modeling these temporal structures has a direct impact on anomaly detection performance. Integrating contrastive learning with temporal modeling strategies may enhance the ability to capture anomaly trends, fluctuation signals, and periodic changes. This can improve the model's robustness and generalization in dynamic environments[3]. Therefore, contrastive learning strategies that incorporate both temporal awareness and structural discrimination are a promising direction for anomaly detection in cloud computing.

In summary, there is an urgent need for an efficient algorithmic framework that can perform unsupervised modeling while extracting key features from high-dimensional, multimodal, and temporal data. Unsupervised anomaly detection methods based on contrastive learning have emerged in response to this demand. These methods align with the current data-driven paradigm and provide new technical support for improving intelligent monitoring and operational security of cloud platforms. Their development is of great significance for ensuring the reliability, stability, and service quality of cloud computing systems[4].

2. Related work

With the continuous advancement of cloud computing technologies, more enterprises and organizations are migrating their core services to the cloud. This shift helps them respond to rapidly changing market demands and increasing data processing pressure. Cloud platforms offer elastic scalability, automated resource scheduling, and multi-tenant shared architectures, which significantly improve computing efficiency and operational flexibility[5]. However, these advantages also introduce greater system complexity and runtime uncertainty. Frequent data interactions and dynamic changes among heterogeneous components make the platform more vulnerable to various abnormal events. In high-concurrency and high-traffic production environments, sudden performance degradation, service failures, and attack behaviors pose serious threats to system stability. Therefore, building efficient and accurate anomaly detection mechanisms in cloud environments is crucial for ensuring business continuity and data security. It also serves as a fundamental support for enhancing resource utilization and service quality.

Monitoring systems in cloud platforms typically collect large volumes of multi-dimensional time series data, such as CPU usage, memory consumption, network traffic, and service call chains[6]. This data is generated continuously densely and dynamically. In such environments, anomalies often appear as rare and subtle shifts or sudden state changes, lacking clear boundaries from normal patterns and easily overwhelmed by noise. Additionally, the types of anomalies are highly diverse and may result from misconfigurations, external attacks, workload surges, or hardware failures. These factors introduce high uncertainty. Traditional rule-based methods or supervised learning models face significant limitations in practice. In particular, when data labeling is expensive and abnormal samples are scarce, building classifiers or regressors that rely on labeled data becomes impractical. Under conditions without labels and weak structure, automatically discovering potential anomaly signals from raw data becomes a core challenge in current research.

In recent years, unsupervised learning has emerged as a key approach to address these issues. Among them, contrastive learning, which learns discriminative features by modeling relationships between samples, has shown strong adaptability and representational power in anomaly detection tasks. By designing appropriate mechanisms to construct positive and negative sample pairs, contrastive learning can guide models to capture semantic similarity and difference without requiring label supervision. This leads to more effective embedding representations. In cloud environments, where data often contains periodic fluctuations, local noise, and structural redundancy, contrastive learning helps identify atypical patterns hidden in dynamic complexity by emphasizing structural consistency and distribution boundaries. More importantly, this method offers good scalability and transferability[7]. It can adapt to different platform architectures and data characteristics, improving both stability and reliability in real-world applications while maintaining algorithm generalizability[8].

3. Architectural Approach

The network architecture illustrates an unsupervised anomaly detection method based on contrastive learning. The process includes sliding window segmentation, data augmentation, and embedding learning through an encoder. A dual-branch structure constructs positive and negative sample pairs to guide the model in learning discriminative representations without labels. A temporal attention mechanism is introduced to capture key dynamic information within the sequences. Finally, anomaly scores are estimated through similarity computation, aligning with the method proposed in the paper. The model architecture is shown in Figure 1.

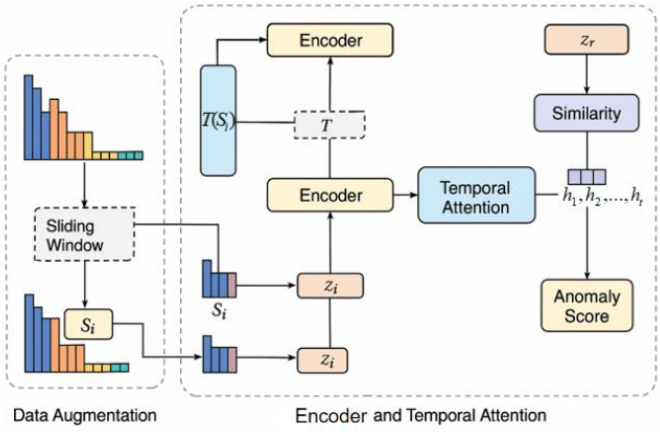

Figure 1. Contrastive Learning for Unsupervised Detection in Cloud Anomalies

This paper proposes an unsupervised anomaly detection method based on contrastive learning, which aims to automatically learn discriminative representations from high-dimensional, multi-source, and time-series cloud computing monitoring data to achieve effective identification of abnormal behaviors. First, let the input raw monitoring data be $X = \{x_1, x_2, ..., x_T\}$, where each $x_t \in R^d$ represents a d --dimensional observation vector at the time step t. In order to capture the local pattern and global structure of the data, a sliding window mechanism is used to divide the sequence into subsequence fragments $S = \{x_i, x_{i+1}, ..., x_{i+l-1}\}$ of length l, which are used as model input for feature modeling. The process can be formalized as:

$$S_i = x[i : i + l], i = 1,2,...,T - l + 1$$

Each subsequence is then mapped to a low-dimensional embedding space through the encoder $f_\theta(\cdot)$ to obtain the latent representation $z_i = f_\theta(S_i) \in R^m$. In order to perform unsupervised contrastive learning, a positive and negative sample pair generation strategy needs to be designed. For each original subsequence S_i, a positive sample view $S_i^+ = T(S_i)$ is constructed through the data enhancement function $T(\cdot)$, and negative samples $S_j^-(j \neq i)$ are randomly selected from other subsequences. The enhanced subsequences are input into the encoder respectively to obtain the positive sample representation $z_i^+ = f_\theta(S_i^+)$ and the negative sample representation $z_i^- = f_\theta(S_i^-)$. Subsequently, a contrast loss function based on temperature scaling is used to optimize the semantic distance relationship in the embedding space. The loss function is as follows:

$$L_i = -\log \frac{\exp(sim(z_i, z_i^+)/r)}{\sum_{j=1}^{N} \exp(sim(z_i, z_j^-)/r)}$$

Among them, $sim(\cdot,\cdot)$ represents the cosine similarity function, and r is the temperature parameter, which controls the sensitivity of the similarity distribution.

In the encoder structure design, a time-sensitive neural network module is used to model the dynamic evolution characteristics within the subsequence. Considering the periodicity and local disturbance characteristics of cloud computing monitoring data, the model introduces a time-sequence attention mechanism in the embedding space learning to emphasize the key time points. Suppose the input subsequence is embedded as $H = [h_1, h_2, ..., h_t]$, and each $h_t \in R_m$, then the context is calculated through the attention mechanism as:

$$a_t = \frac{\exp(q^T h_t)}{\sum_{k=1}^{l} \exp(q^T h_k)}, z = \sum_{t=1}^{l} a_t h_t$$

Where q is a learnable query vector, a_t represents the importance weight of the t th time point, and finally z aggregates the key dynamic information in the temporal structure.

After completing the contrastive learning training, the model obtains the potential representation ability of the input sequence distribution structure. In order to perform anomaly scoring, this paper defines anomaly measurement functions based on reconstruction error and embedding similarity. Given a subsequence S_i to be detected, which is represented by z_i, and looking for the closest reference representation z_r in the training set, its anomaly score is defined as:

$$Score(z_i) = 1 - sim(z_i, z_r)$$

This score measures the degree of deviation of the current sample from the normal mode in the embedding space, thereby achieving label-free anomaly detection. The entire method framework forms a closed-loop structure between data enhancement, feature learning, and anomaly measurement, with good adaptability and robustness, and is suitable for complex, high-dimensional, and dynamic anomaly detection needs in cloud platform environments.

4. Dataset & Experimental Analysis

4.1 Dataset

This study uses the AWS CloudWatch Dataset, which is widely adopted in real-world cloud computing environments, as the experimental data source. The dataset consists of multiple monitoring metrics, including CPU utilization, network traffic, and disk read/write rates. It covers the operational states of virtual machines, containers, and service layers. The data exhibits high dynamism and multi-dimensional characteristics, effectively reflecting complex behavioral patterns in cloud platforms.

The dataset is sampled at the minute level and spans system operation data over various periods. It shows clear temporal structures and periodic fluctuations. Several synthetic anomalies are injected into the data, such as service congestion, resource exhaustion, and network interruption. These are suitable for validating the model's ability to distinguish between normal and abnormal patterns using contrastive learning.

Considering the real application scenarios of cloud platform data, this dataset does not rely heavily on labels, which aligns well with the objectives of unsupervised anomaly detection tasks. Its high dimensionality, heterogeneity, multi-scale properties, and dynamic evolution are closely matched with the design of the proposed model in terms of input structure, augmentation strategy, and temporal modeling. This

provides a strong foundation for evaluating the method's applicability in real systems.

4.2 Experimental Results

This paper first conducts a comparative experiment, and the experimental results are shown in Table 1.

Table1: Comparative experimental results

Model	F1 Score	Precision	Recall	AUC	KS Score
TimeConAD (Ours)	0.872	0.891	0.854	0.938	0.682
Anomaly Transformer[9]	0.841	0.859	0.826	0.915	0.641
DONUT+[10]	0.773	0.799	0.748	0.875	0.578
GDN[11]	0.802	0.813	0.791	0.891	0.605
MO-GAAL[12]	0.741	0.782	0.708	0.862	0.561

The experimental results show that the proposed TimeConAD model outperforms other baseline methods across multiple evaluation metrics. In particular, it achieves significantly higher scores in F1 Score, Precision, and Recall, which are core indicators for measuring detection accuracy and coverage. This demonstrates that the method offers stronger discrimination and robustness when dealing with the challenges of dynamic changes, missing labels, and complex anomaly types in cloud computing scenarios. By introducing a contrastive learning mechanism, TimeConAD can effectively learn the underlying structural differences between normal and abnormal patterns under unsupervised conditions, leading to superior anomaly detection performance.

Compared with mainstream methods such as Anomaly Transformer and GDN, TimeConAD achieves more notable improvements in AUC and KS Score. AUC reflects the model's global detection ability, while KS Score measures the distributional difference between normal and abnormal samples. Both are critical in anomaly detection tasks. TimeConAD constructs subsequences using a sliding window, generates multi-view inputs through data augmentation, and applies temporal attention to extract key sequential features. This enables the model to capture subtle disturbances and local shifts in anomaly patterns, significantly enhancing overall detection effectiveness.

For structure-aware models such as GDN and MO-GAAL, although they show advantages in structural learning, their generalization ability is limited when applied to high-dimensional time series data without labels. TimeConAD addresses this issue by integrating representation alignment and anomaly distribution modeling through unsupervised contrastive learning. It learns high-quality sequence embeddings without relying on labeled data. This mechanism is particularly suitable for monitoring data in cloud platforms that are frequently updated and lack accurate annotations, showing strong adaptability to real-world scenarios.

This paper also experiments on the sensitivity of data enhancement strength to contrastive learning effects. The experimental results are shown in Figure 2.

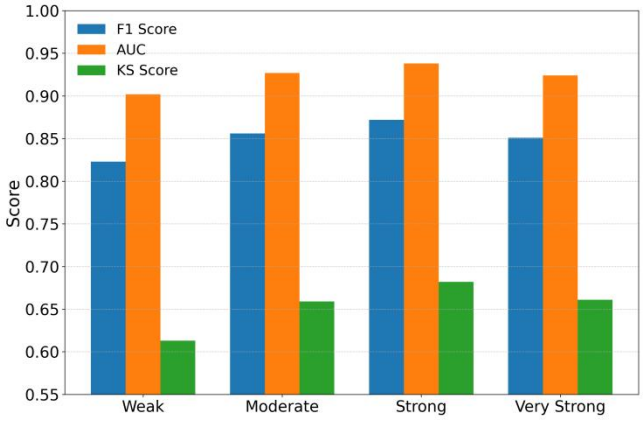

Figure 2. Sensitivity analysis of data augmentation intensity on contrastive learning effects

The experimental results show that data augmentation strategies with varying strengths have a significant impact on the performance of contrastive learning in anomaly detection tasks. As the augmentation strength increases from weak to strong, the model's performance in the F1 Score, AUC, and KS Score initially improves and then slightly declines. This suggests that moderate perturbations help the model learn more discriminative features between normal and abnormal states. In particular, the model performs best when the augmentation strength is set to "Strong," with all evaluation metrics reaching their highest levels. This indicates that the positive sample views generated at this stage effectively guide the embedding space to learn more distinguishable representations.

The improvements in AUC and KS Score indicate that augmentation not only increases the distance between different classes in the representation space but also enhances the model's sensitivity to anomaly boundaries. This is especially important in cloud platform monitoring data, where normal fluctuations occur frequently. Proper augmentation helps the model distinguish between noisy normal samples and mildly abnormal ones, thereby improving detection accuracy and robustness. In contrast, overly strong augmentation may distort key features of the original sequence, resulting in representation shifts and a slight drop in detection performance.

The changes in the KS Score reveal that augmentation strength directly affects the separability of positive and negative sample distributions. Stronger augmentation reinforces the consistency among positive samples and increases the distribution gap from negative samples. This improves the model's discriminative power during distribution learning. However, when the augmentation exceeds a reasonable threshold, it may disrupt the original temporal structure and weaken the ability of contrastive loss to enforce semantic boundaries. This can reduce the precision of anomaly detection.

This paper also conducts comparative experiments on the impact of sample view generation strategies on robustness. The experimental results are shown in Figure 3.

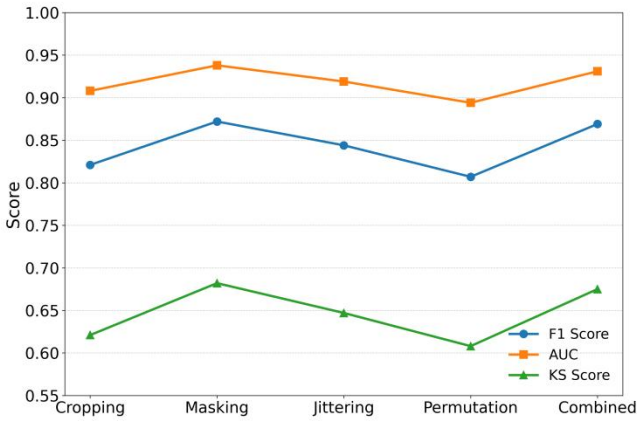

Figure 3. Analysis of the impact of sample view generation strategy on robustness

The experimental results indicate that different sample view generation strategies have a significant impact on model robustness in unsupervised contrastive learning. Among these strategies, Masking and Combined augmentation achieve the best performance, especially in the F1 Score and KS Score, which reflect anomaly detection accuracy and sample distribution separability. This suggests that masking local temporal segments or applying multiple perturbations helps the model better capture subtle differences between normal and abnormal states, enhancing generalization in high-dimensional time series data.

The improvement in AUC shows that the Masking strategy guides the model to focus on key structures and long-term dependencies. It increases the diversity of positive samples while preserving their semantic integrity. This is especially important for monitoring data in cloud platforms, which often contain periodic fluctuations and local anomalies. In contrast, strategies such as Cropping and Permutation introduce heavy disturbances to temporal structures. This may cause the loss of feature information and weaken the model's ability to identify boundary cases, reducing overall robustness.

Jittering achieves moderate performance in the F1 Score and AUC, but its KS Score is relatively low. This indicates that although it enhances input diversity, it has a limited ability to increase the distribution gap between positive and negative samples. As a result, it struggles to form clear anomaly boundaries. Permutation, which disrupts the original time order, introduces strong perturbations but breaks causal relationships in the sequence. This leads to weaker representation quality in the embedding space and degrades performance in time-dependent anomaly detection tasks.

This paper also analyzes the impact of the number of temporal attention layers on sequence modeling capabilities. The experimental results are shown in Figure 4.

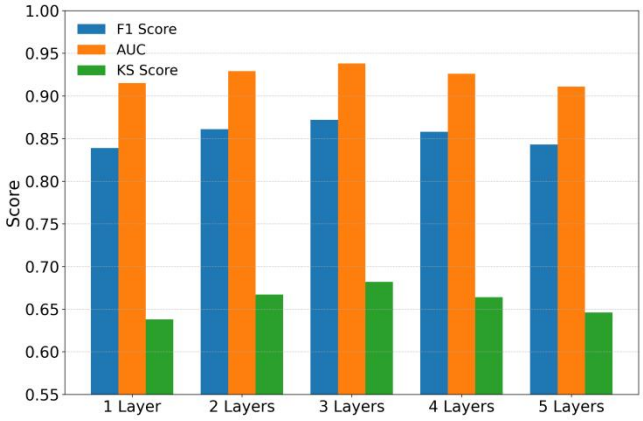

Figure 4. Attention Layer Depth Evaluation

The experimental results show that the number of temporal attention layers has a clear impact on the model's ability to capture sequential patterns. When the attention mechanism is set to three layers, the model achieves the best performance in F1 Score, AUC, and KS Score. This indicates that a moderate increase in attention depth enhances the model's perception of complex temporal structures. It allows the model to better capture long-term dependencies and subtle fluctuations in monitoring data, improving its ability to distinguish abnormal behaviors. In cloud platform environments, system states often show nonlinear dynamics. Three layers of attention strike a good balance between representation capacity and the risk of overfitting.

When the number of attention layers is small, such as only one layer, the model's ability to capture temporal dependencies is limited. It cannot fully model multi-scale time patterns, resulting in overall lower performance. In particular, the KS Score reveals the model's difficulty in establishing clear boundaries between normal and abnormal sample distributions. This can lead to uncertain anomaly judgments in high-frequency cloud scenarios, reducing the system's responsiveness to potential risks.

As the number of layers increases to four and five, model performance begins to decline. Although deeper attention structures can theoretically learn more complex representations, the lack of supervision in unsupervised tasks may cause training instability. This increases the risk of overfitting. In anomaly detection without labels, deeper attention stacking may cause the model to focus more on local noise rather than key patterns. This harms the quality of representations and reduces generalization.

The overall trend shows that temporal attention mechanisms play an important role in modeling dynamic sequences. However, the number of layers should be carefully balanced based on the task characteristics. For high-dimensional, dynamic, and weakly structured monitoring data in cloud computing, using three attention layers provides a good trade-off between expressive power and training stability. This supports more robust semantic learning within the contrastive learning framework.

Finally, this study investigates the effect of learning rate settings on the stability of the training process, as shown in Figure 5.

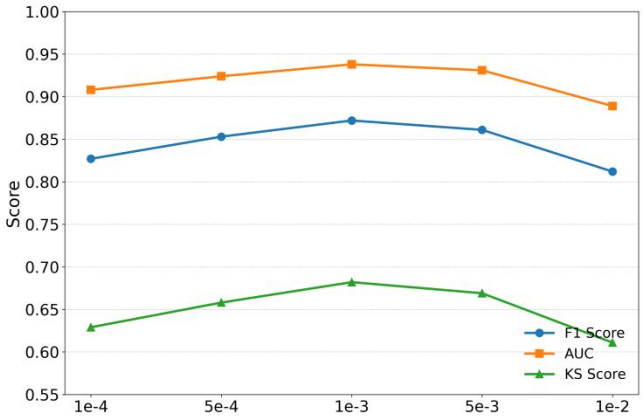

Figure 5. The impact of learning rate setting on the stability of the training process

The experimental results show that the learning rate has a significant impact on both training stability and final model performance. When the learning rate is set to 0.001, the model achieves peak values in F1 Score, AUC, and KS Score. This indicates that this setting allows effective gradient propagation while avoiding oscillation or failed convergence caused by large step sizes. These results confirm that a moderate learning rate helps the embedding space form gradually in an unsupervised contrastive learning framework, which enhances the model's ability to distinguish anomalies.

When the learning rate is too low, such as 0.0001, the model remains stable but converges slowly. This results in weak discriminative power of the learned embeddings. In high-dimensional and dynamic monitoring data, a low learning rate may fail to capture subtle anomaly patterns in time, leading to poor boundary construction. This effect is especially visible in the KS Score, which suggests the model cannot effectively separate the distribution of positive and negative samples.

In contrast, a high learning rate, such as 0.01, speeds up early training but often causes instability in the loss landscape. This makes the optimization path unstable. In contrastive learning, such instability can blur the representations of positive and negative samples, disrupting the structure of the embedding space. As a result, the model's ability to detect abnormal sequences declines. Both the F1 Score and AUC show a downward trend in the experiments, confirming that an excessively high learning rate damages the model's robustness and weakens its applicability to real-world cloud data.

Performance analysis under different learning rate settings reveals that the learning rate directly influences how the embedding space is shaped during training. For high-frequency, sparse-distribution monitoring data in cloud environments, choosing a learning rate that ensures both training stability and effective contrastive feature extraction is key to achieving efficient convergence and accurate unsupervised anomaly detection.

5. Conclusion

This study addresses key challenges in anomaly detection within cloud computing environments by proposing an unsupervised method based on contrastive learning. The approach integrates sliding window segmentation, diversified data augmentation, and a temporal attention encoding module. It effectively handles the modeling of high-dimensional, heterogeneous, and unlabeled monitoring data. By constructing positive and negative sample pairs to guide feature learning, the model can automatically acquire discriminative representations under unsupervised conditions, enabling accurate detection of complex anomalous behaviors. Experimental results demonstrate strong performance across multiple key metrics, confirming the practicality and robustness of the proposed framework in real-world cloud scenarios.

Starting from the goal of ensuring system stability, the study designs an anomaly detection framework with structural generalization and temporal modeling capabilities. It performs well in environments with frequent changes and limited labeled data. The model does not rely on predefined rules or manual feature extraction, making it applicable to various types of monitoring data, including service metrics, resource usage, and network conditions. This provides a practical technical path for intelligent alerting and anomaly diagnosis in real operations. The introduction of contrastive learning further enhances the model's transferability, supporting cross-platform and cross-scenario deployment. This adds flexibility and efficiency to the development and scaling of anomaly detection algorithms.

The proposed method has broad application potential in the cloud computing domain. It can support resource scheduling at the infrastructure level and extend to behavior modeling, security analysis, and performance bottleneck identification in microservice architectures. In emerging architectures such as edge computing, container orchestration, and serverless computing, system behaviors are more complex, and anomaly patterns are more diverse. The contrastive learning model proposed here shows strong structural compatibility and data adaptability. It provides theoretical support and engineering foundations for building future autonomous operations and intelligent monitoring systems.

Future research may further explore the method's scalability in federated environments, multi-source platforms, and online learning scenarios. For example, integrating incremental representation updates, graph-structured information, or multi-task learning modules could improve adaptability and multi-objective recognition. Incorporating semantic modeling techniques such as large language models into the anomaly detection pipeline may enhance the model's ability to understand complex system behaviors. This could promote the evolution from static anomaly detection to causal analysis and intelligent decision-making, supporting the continued development of cloud computing and intelligent operations.

References

[1] Xu, H., et al. (2022). "CUTS: Contrastive Unsupervised Time Series Representation Learning via Time-Based Negative Sampling." Proceedings of NeurIPS 2022.

[2] Zhou, H., et al. (2021). "Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting." Proceedings of AAAI 2021.

[3] Eldele, E., et al. (2021). "Time-Series Representation Learning via Temporal and Contextual Contrasting." IJCAI 2021.

[4] Laptev, N., et al. (2015). "Generic and Scalable Framework for Automated Time-series Anomaly Detection." Proceedings of KDD 2015.

[5] Liu, J., et al. (2023).Multi-scale Association Discrepancy for Time-Series Anomaly Detection.Proceedings of the 29th ACM SIGKDD Conference (KDD 2023).

[6] Yamaguchi, Y., et al. (2021).Improving Variational Autoencoders for Unsupervised Anomaly Detection in Time Series.Pattern Recognition Letters, Volume 147.

[7] Zhang, H., et al. (2022).Robust Graph-based Anomaly Detection in Multivariate Time Series.IEEE Transactions on Neural Networks and Learning Systems (TNNLS).

[8] Chen, C., et al. (2021).Improved Generative Adversarial Networks for Unsupervised Anomaly Detection in Time Series.Knowledge-Based Systems, Volume 229.

[9] Xu, Y., et al. (2022).Anomaly Transformer: Time Series Anomaly Detection with Association Discrepancy.International Conference on Learning Representations (ICLR 2022).

[10] Ren, H., et al. (2019).Time-Series Anomaly Detection Service at Microsoft.Proceedings of the 25th ACM SIGKDD Conference (KDD 2019).

[11] Deng, A., et al. (2021).Graph Neural Network-Based Anomaly Detection in Multivariate Time Series.Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2021).

[12] Liu, Y., et al. (2020).Generative Adversarial Active Learning for Unsupervised Outlier Detection.IEEE Transactions on Neural Networks and Learning Systems (TNNLS).