Transactions on Computational and Scientific Methods | Vo. 5, No. 6, 2025
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Distributed System Anomaly Detection via Collaborative Discrimination of Reconstruction Residuals and Latent Distribution Shifts

Zirui Liu[1], Dilara Abdurehim[2]
[1]New York University, New York, USA
[2]Carnegie Mellon University, Mountain View, USA
*Corresponding author: Zirui Liu; ziruiliu98@gmail.com

Abstract: To address the challenges of concealed anomalies, complex state evolutions, high-dimensional and heterogeneous monitoring data, and the difficulty of traditional methods in simultaneously addressing representation learning and anomaly deection during distributed system operation, this paper proposes a distributed system anomaly detection method based on VAE reconstruction. This method takes a multivariate time series composed of system operation indicators as input, organizes local contextual information through a sliding time window, and compresses redundant perturbations in the original observations using a feature projection mechanism. Then, it learns the latent probability distribution of normal states using a variational autoencoder structure. Based on this, the model not only reconstructs the input features but also jointly characterizes the reconstruction residuals and latent distribution shifts, fusing the two types of anomaly evidence into a unified anomaly score, thereby enhancing the ability to identify complex anomaly states. Simultaneously, a time consistency constraint is introduced to reduce the interference of short-term random fluctuations on the discrimination results, making anomaly detection more consistent with the continuous change characteristics of real distributed system operation. Results show that the proposed method can more effectively distinguish between normal and abnormal states, exhibiting good performance in terms of accuracy, stability, and overall discrimination capabilities. This study provides an implementation path for anomaly detection in distributed systems that combines generative modeling and discriminant analysis, and also provides a methodological foundation for intelligent monitoring and risk identification in complex operating environments.

Keywords: Anomaly detection in distributed systems; variational autoencoders; reconstruction residuals; latent distribution modeling.

1. Introduction

With the rapid popularization of technologies such as cloud computing, microservices, container orchestration, and continuous delivery, distributed systems have become core infrastructure supporting critical business scenarios such as internet platforms, industrial internet, financial services, and smart governance. Compared to traditional centralized systems, distributed systems exhibit greater heterogeneity, dynamism, and coupling in terms of resource organization, service collaboration mechanisms, and operational environment complexity. Massive service instances continuously interact across nodes,

containers, and networks, causing the system's operational status to exhibit significant time-varying and propagation characteristics[1,2]. Once a local module experiences performance degradation, resource contention, call blocking, or fault propagation, anomalies often propagate rapidly through multi-level dependencies, impacting overall service quality and business continuity. Therefore, developing efficient, accurate, and generalizable anomaly detection methods for complex operating environments has become a crucial research topic in intelligent operation and maintenance of distributed systems[3].

In actual operation, distributed system anomalies are characterized by strong concealment, diverse patterns, blurred boundaries, and rapid evolution. Anomalies can originate from localized failures of single components or be complex failure modes resulting from the combined effects of multiple factors such as resource fluctuations, service dependency imbalances, link congestion, configuration drift, and sudden load increases. Since system monitoring data typically includes multi-source heterogeneous information such as logs, metrics, call chains, and events, and different data sources exhibit significant differences in temporal granularity, semantic structure, and noise distribution, traditional methods relying on rule matching, threshold triggering, or shallow statistical modeling often fail to adequately characterize the big distributional differences between abnormal and normal states. Especially under high-dimensional, multivariate, and non-stationary observation conditions, anomaly patterns often exhibit nonlinear and weakly significant characteristics, leading to false positives, false negatives, and insufficient cross-scenario adaptability in detection models[4].

Variational autoencoders (VAEs), as deep representation learning models with probabilistic generation capabilities, can achieve compact representation and reconstruction learning of normal behavior patterns by modeling the latent space of complex data distributions. In distributed system anomaly detection tasks, VAE-based modeling approaches can utilize normal samples to learn the latent distribution structure of system states and identify anomalous samples by leveraging reconstruction errors or potential deviations. This method does not rely on rigorous and complete anomaly labeling, making it well-suited for real-world scenarios where anomaly samples are scarce and anomaly types are constantly evolving. However, relying solely on reconstruction results as the basis for anomaly judgment can still be affected by factors such as overlapping distributions of anomaly and normal samples, insensitivity to local perturbations, and unclear discrimination boundaries, limiting the model's ability to distinguish complex anomaly scenarios[5]. Therefore, further introducing a discrimination mechanism around the VAE reconstruction results and strengthening the correlation modeling between latent representation and anomaly decision-making has significant theoretical value for improving the stability, sensitivity, and reliability of anomaly identification.

Research on distributed system anomaly detection methods oriented towards VAE reconstruction discrimination not only helps to promote the evolution of anomaly detection models from purely reconstruction-driven to a collaborative approach of reconstruction and discrimination, but also provides a new technical path for the deep representation of the operational state of complex systems. This research can improve the model's ability to identify complex anomaly patterns from the levels of latent space modeling, reconstruction deviation characterization, and anomaly decision optimization, providing methodological support for achieving early warning of distributed system faults, operational risk control, and intelligent operation and maintenance decision-making. Meanwhile, this direction is of practical significance for enhancing the stability, security, and service continuity of critical infrastructure systems, and also plays a positive role in promoting the deep application of artificial intelligence methods in system operation and maintenance, cyberspace security, and governance of complex industrial software.

2. Background

Distributed systems typically consist of multiple nodes with independent computing and communication capabilities, collaborating across a network to perform task decomposition, data exchange, and service scheduling to support large-scale concurrent processing and high-availability business operations. With the continuous evolution of cloud-native architectures, system deployment has gradually shifted from monolithic applications to service-oriented, containerized, and elastic orchestration, with mechanisms such as resource management, service discovery, load balancing, and fault recovery jointly participating in the system's operation[6]. While this architecture significantly improves system scalability and resource utilization efficiency, it also results in more complex hierarchical structures and interrelationships in operational behavior. Frequent calls and dependencies exist between different services, and node states, network latency, resource consumption, and task scheduling strategies collectively affect the overall system performance, leading to a continuous expansion of the system's state space. Even minor disturbances during operation can be amplified and propagated over a wider area.

Anomaly detection is a crucial fundamental problem in the operation and management of distributed systems. Its core objective is to identify state changes that deviate from normal behavioral patterns within complex observation data, providing a basis for subsequent diagnosis, localization, and recovery. As system scale and dynamism continue to increase, traditional detection methods relying on human experience or fixed thresholds are increasingly unable to meet the identification needs of high-dimensional, dynamic, and multi-source scenarios[7]. In recent years, representation learning and generative modeling methods have been gradually introduced into system anomaly detection tasks. By extracting latent features from monitoring sequences, log semantics, and system state evolution patterns, these methods enhance the model's ability to characterize complex patterns. Among them, generative methods based on latent variable modeling can understand the normal behavior of the system from the perspective of probability distribution and reflect potential abnormal states by using the deviation information between input and reconstruction. This provides a new research foundation for anomaly perception in distributed systems under weakly labeled or even unlabeled conditions.

3. Methods

Let the distributed system generate a multivariate observation sequence $\{\boldsymbol{x}_t\}_{t=1}^{T}$, in which each $\boldsymbol{x}_t \in \mathbb{R}^d$ aggregates synchronized indicators extracted from runtime metrics, service interactions, and log-derived statistics over the same monitoring interval. Rather than detecting anomalies directly in the raw space, the method first organizes a local temporal context around each timestamp so that transient perturbations and progressive degradation can be modeled within one unified representation. This paper presents the overall model architecture, as shown in Figure 1.

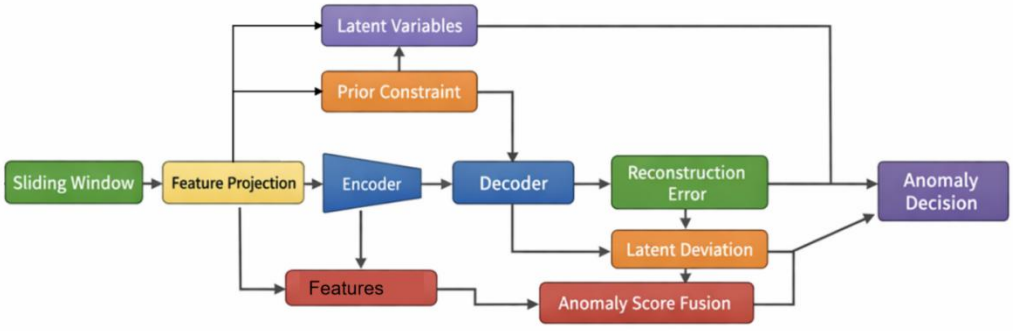

Figure 1. Overall model architecture

A sliding window centered on the current state is therefore written as:

$$\boldsymbol{X}_t = \left[\boldsymbol{x}_{t-w+1}, \boldsymbol{x}_{t-w+2}, \ldots, \boldsymbol{x}_t\right] \in \mathbb{R}^{w \times d}$$

To preserve temporal continuity while suppressing scale imbalance among heterogeneous signals, the windowed observation is mapped into a compact feature tensor $\boldsymbol{H}_t$ through a learnable projection $f_\theta(\cdot)$ that emphasizes stable structural patterns shared by normal system behaviors:

$$\boldsymbol{H}_t = f_\theta\left(\boldsymbol{X}_t\right)$$

Such a design is meaningful because distributed anomalies are often not isolated point deviations but coordinated departures across services, resources, and time, and the window representation allows the model to capture these coupled changes before reconstruction is attempted. Given $\boldsymbol{H}_t$, a variational encoder parameterized by ϕ estimates the posterior distribution of the latent state $\boldsymbol{z}_t$, with the mean vector $\boldsymbol{\mu}_t$ and diagonal variance vector $\boldsymbol{\sigma}_t^2$ reflecting both the dominant operating mode and the uncertainty carried by the current context:

$$q_\phi\left(\boldsymbol{z}_t|\boldsymbol{H}_t\right) = \mathcal{N}\left(\boldsymbol{z}_t; \boldsymbol{\mu}_t, diag\left(\boldsymbol{\sigma}_t^2\right)\right)$$

Because stochastic sampling is required during optimization, but direct backpropagation through random variables is unstable, the latent variable is obtained by reparameterization so that the semantic structure of normal operation can be learned in a differentiable manner:

$$\boldsymbol{z}_t = \boldsymbol{\mu}_t + \boldsymbol{\sigma}_t \odot \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

Unlike pure autoencoding schemes that only memorize surface reconstruction patterns, the proposed framework constrains the latent space with a prior distribution and reconstructs the contextual representation through a decoder $g_\psi(\cdot)$ so that normal states occupy a compact and regular manifold. The reconstructed feature tensor is expressed as:

$$\widehat{\boldsymbol{H}}_t = g_\psi\left(\boldsymbol{z}_t\right)$$

For stable modeling of normal behaviors, the training objective combines feature reconstruction with variational regularization, where $\|\cdot\|_F$ denotes the Frobenius norm, and the hyperparameter β controls the compactness of the latent distribution:

$$\mathcal{L}_{vae} = \frac{1}{T}\sum_{t=1}^{T}\left(\|\boldsymbol{H}_t - \widehat{\boldsymbol{H}}_t\|_F^2 + \beta\, D_{KL}(q_\phi(\boldsymbol{z}_t|\boldsymbol{H}_t) \,\|\, \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}))\right)$$

This objective is introduced not merely to reduce reconstruction error, but to force the model to distinguish structurally consistent normal trajectories from irregular patterns that cannot be compressed and recovered under the same probabilistic assumptions. In this way, the latent space becomes a behavior descriptor with statistical semantics rather than a simple low-dimensional embedding, which is essential for distributed systems whose anomalies may manifest as subtle but persistent deviations instead of abrupt failures.

Beyond reconstruction itself, anomaly discrimination is strengthened by explicitly measuring the disagreement between the observed representation and the normal manifold from two complementary views.

One view characterizes reconstruction distortion through a residual intensity score, while the other view evaluates how far the posterior parameters drift from the prior centered normal region:

$$r_t = \| \boldsymbol{H}_t - \widehat{\boldsymbol{H}}_t \|_2$$

$$u_t = \| \boldsymbol{\mu}_t \|_2 + \| log \boldsymbol{\sigma}_t^2 \|_1$$

A discriminative anomaly score is then formed by adaptively fusing these two signals, in which $\alpha \in [0,1]$ balances surface mismatch and latent distribution shift so that the detector remains sensitive to both visible reconstruction failures and hidden semantic departures:

$$s_t = \alpha r_t + (1 - \alpha) u_t$$

This fusion is important because certain abnormal states in distributed environments may still be approximately reconstructable due to local similarity with normal load patterns, whereas their latent uncertainty or posterior displacement has already begun to deviate from the learned operational manifold. Conversely, noise spikes may enlarge the residual term without reflecting a genuine system-level anomaly, and the latent component helps suppress such misleading responses.

Meanwhile, temporal coherence is incorporated to avoid fragmented decisions across adjacent windows, since real anomalies in distributed systems often evolve over consecutive intervals and rarely appear as completely independent events. A smoothness regularizer is therefore imposed on neighboring anomaly scores so that persistent abnormal evolution is emphasized while accidental fluctuations are attenuated:

$$\mathcal{L}_{\text{temp}} = \frac{1}{T-1} \sum_{t=2}^{T} (s_t - s_{t-1})^2$$

Finally, the complete objective integrates generative learning and temporal discrimination, with λ controlling the contribution of sequential consistency during optimization:

$$\mathcal{L} = \mathcal{L}_{\text{vae}} + \lambda \mathcal{L}_{\text{temp}}$$

During inference, larger values of s_t indicate a higher probability that the current system state deviates from the learned normal dynamics, thereby enabling abnormal behavior identification from the joint perspective of reconstruction reliability, latent regularity, and temporal continuity. Such a formulation gives the method a clearer decision basis than using reconstruction error alone, and it improves the interpretability of abnormal evidence by linking score elevation to observable distortion and probabilistic deviation within the same modeling framework.

4. Experimental Results and Analysis

4.1 Dataset

This paper selects GAIA as the dataset. GAIA is an open-source dataset for AIOps scenarios, specifically designed for tasks such as anomaly detection, log analysis, and fault localization. The data originates from the MicroSS microservice business simulation system, which also provides accompanying monitoring data. This dataset is highly relevant to the topic of anomaly detection in distributed systems because its data generation environment is itself a typical microservice distributed architecture, reflecting the complex coupling relationships between service instances, operational metrics, log records, and call chains. Publicly

available information shows that GAIA continuously collected detailed monitoring data for two weeks, including over 6500 metrics, 7 million log entries, and complete tracing data, and recorded anomaly injection information. This provides a sufficient data foundation for anomaly detection methods based on normal pattern learning and reconstructed bias judgment.

From a methodological adaptability perspective, GAIA includes both continuous time-series metrics and retains multi-source operational information such as logs and tracing data. Therefore, it can naturally construct the multivariate observation sequences required in this paper and support the establishment of normal behavior modeling processes around time windows. For anomaly detection tasks oriented towards VAE reconstruction discrimination, this type of data can be used to learn the potential distribution structure of the system under normal operating conditions, and can also be used to verify the degree of deviation of abnormal states from normal patterns by using anomaly injection records. At the same time, GAIA is released as an open repository, with a clear data organization structure and accessibility, making it suitable as a standardized open-source data foundation for anomaly detection research in distributed systems.

4.2 Data preprocessing

In the data preprocessing stage, the original monitoring metrics in the GAIA dataset were first uniformly aligned in time and filtered for fields, retaining key temporal features that reflect service operation status, resource utilization changes, and system load fluctuations, and constructed into multivariate time series according to fixed sampling intervals. Subsequently, missing values were repaired using a combination of forward imputation and local statistical completion, and metrics with large differences in abnormal dimensions and distributions were standardized to reduce the interference of numerical bias between different dimensions on model training. Considering that the VAE reconstruction discrimination method needs to learn normal behavior patterns from continuous context, a sliding time window was further used to divide the long sequence into multiple local sample segments, so that each sample simultaneously contains short-term dynamic changes and adjacent temporal dependencies. For time periods containing anomaly markers, only the normal operation intervals were selected to construct the normal behavior distribution during the training stage to ensure that the model learns a stable potential representation; during the testing stage, the complete time series was retained for subsequent anomaly identification based on reconstruction bias and potential distribution shift. This article presents a comparison of the dataset before and after processing, as shown in Figure 2.

As shown in Figure 2, the overall quality of the data was significantly improved after preprocessing. The time series comparison in the upper left corner indicates that the original data before processing exhibited larger fluctuations, with more local spikes and irregular disturbances, while the processed series was smoother overall, with abnormal fluctuations effectively reduced, while still retaining the basic trend over time. The distribution comparison in the upper right corner shows that the data distribution before processing was relatively scattered, with a wide numerical range and some skewness and dispersion; the data distribution after processing was significantly more concentrated, indicating that standardization and outlier suppression were effective. The local volatility results in the lower left corner further illustrate that the rolling standard deviation after processing was generally lower than before processing, indicating a reduction in short-term instability and mitigation of local noise. The normality comparison in the lower right corner shows that the quantiles of the processed data are more closely aligned with the reference quantiles, indicating that the data distribution after preprocessing was more regular and the statistical structure more stable. These phenomena collectively demonstrate that the preprocessing step effectively improved the noise interference, scale inconsistency, and local abnormal fluctuations in the original monitoring data, providing a more reliable data foundation for subsequent anomaly detection based on reconstruction bias and latent distribution modeling.

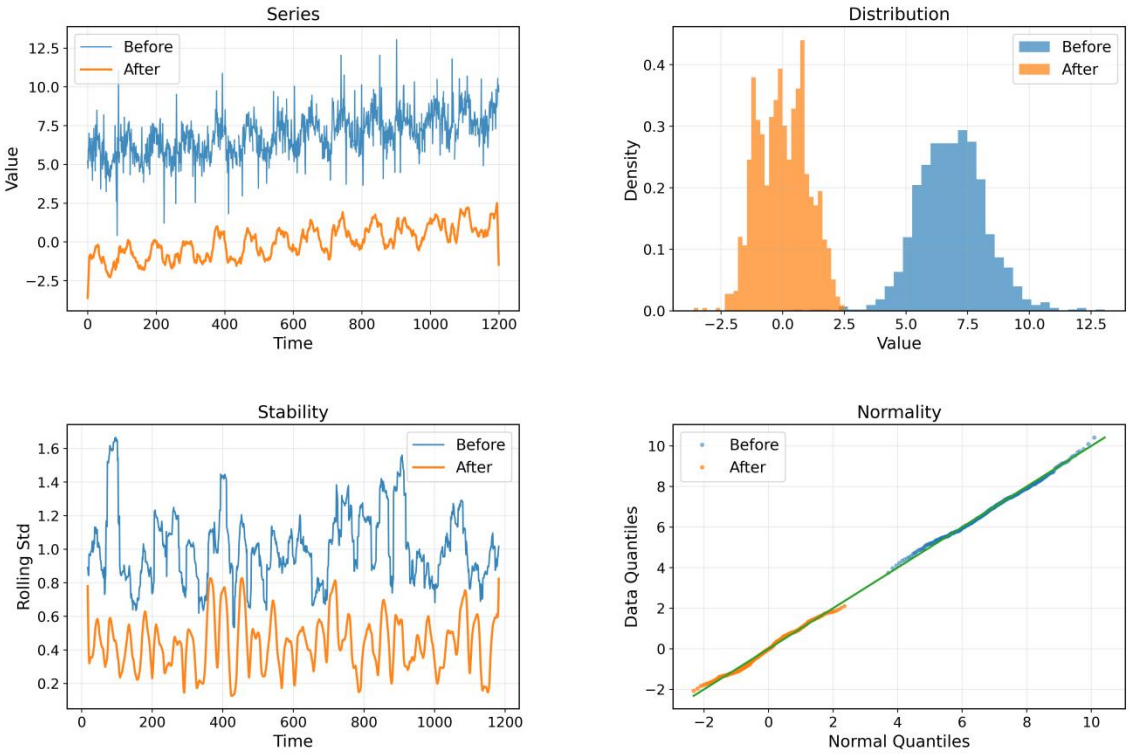

Figure 2. Comparison of the dataset before and after processing

4.3 Experimental Results and Analysis

To facilitate locating the technical position of our proposed method within the broader research context, representative works closely related to anomaly detection in distributed systems, observable data modeling in microservices, reconstruction learning, and graph representation analysis are compiled. These related methods primarily focus on modeling metrics, logs, call chains, and multimodal operational signals, reflecting current research approaches in anomaly representation, discrimination mechanisms, and system semantic utilization from different perspectives. Based on this, a unified comparison table is further constructed to present a horizontal comparison between various methods and our proposed model under the same evaluation dimensions. The experimental results are shown in Table 1.

Table 1. Comparison framework with representative related methods

Method	ACC	PRE	REC	F1
Nedelkoski et al.[8]	92.14	91.36	90.88	91.12
Zhang et al.[9]	93.27	92.81	92.35	92.58
Kohyarnejadfard et al.[10]	91.85	91.09	90.74	90.96
Xie et al.[11]	93.68	93.12	92.76	92.94
Chen et al.[12]	94.03	93.55	93.14	93.31
Xie et al.[13]	94.26	93.84	93.42	93.59
Huang et al.[14]	94.57	94.11	93.86	93.98

Ours	**96.18**	**95.74**	**95.31**	**95.52**

The results shown in Table 1 demonstrate that the proposed method exhibits superior overall recognition performance, ensuring both accuracy and the completeness and stability of anomaly identification. Compared to existing methods, the proposed model surpasses existing methods in classification accuracy, precision, recall, and overall evaluation. This indicates that the proposed method not only more effectively captures anomalous features in the operating state of distributed systems but also maintains strong discriminative ability when the boundaries between normal and anomalous patterns are relatively close. This performance suggests that the approach of jointly modeling reconstruction bias and latent distribution can more fully explore the deep structural information in the system's observation data, thereby improving the quality of anomaly identification.

Furthermore, the superior performance of the proposed method stems from its ability to go beyond mere anomaly judgment based on reconstruction error. Instead, it enhances the perception of complex anomalous states through latent representation constraints, reconstruction consistency modeling, and anomaly scoring fusion mechanisms. For distributed systems, anomalies are often accompanied by multivariate coupled perturbations, local propagation effects, and short-term fluctuations. Relying solely on superficial differences can easily lead to identification biases. The proposed method can more sensitively characterize the deviation of abnormal states from the normal distribution while maintaining a tight representation of normal behavior. Therefore, it demonstrates better overall performance on multiple evaluation metrics, which further illustrates that the method is suitable for complex, dynamic, and noisy operating scenarios in anomaly detection tasks of distributed systems.

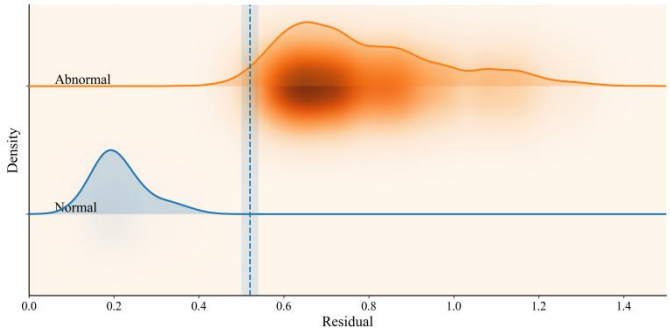

Figure 3. Experiment on reconstructing residual density mapping between normal and abnormal states

Figure 3 shows that the proposed method can form a relatively clear state distribution structure in the reconstructed residual space. Normal states are mainly concentrated in the lower residual region, while abnormal states are more distributed in the higher residual region, and the distinction in density distribution between the two types of samples is quite obvious. This indicates that the normal behavior representation learned by the proposed method has strong compactness and can produce more discriminative reconstruction biases when faced with abnormal inputs, thus providing a stable basis for anomaly discrimination. It can be seen that the proposed method has good coordination in normal state modeling and abnormal state characterization, and can effectively support the anomaly detection task of distributed systems.

The threshold region in the figure is located near the relative boundary between the two density distributions, indicating that the anomaly score obtained by the proposed method has good separability, and the discrimination boundary setting is easier to match with the actual state differences. Since the anomalous states are more prominently clustered in the residual space, it shows that the proposed method can not only

identify significant anomalies but also maintain strong anomaly perception capabilities under complex operational disturbances. Therefore, the proposed method performs well overall and can provide more reliable support for subsequent anomaly alarms and state recognition.

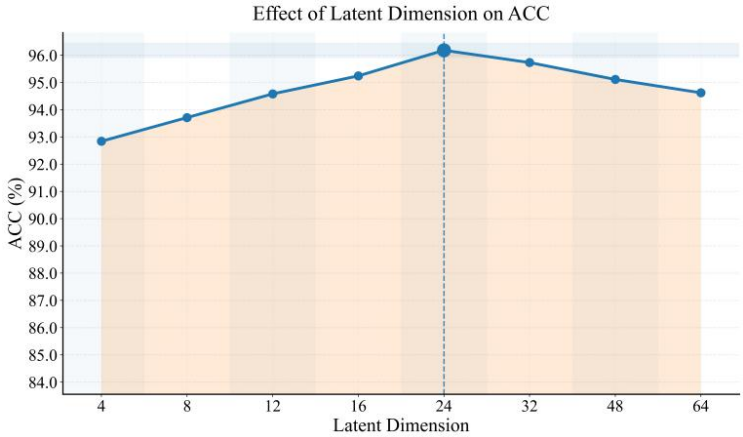

Figure 4. The impact of latent dimension size on ACC

Figure 4 shows that the setting of the latent dimension directly affects the model's ability to characterize the normal behavioral structure of the system. Our proposed method achieved good recognition performance under this set of settings, indicating that the current latent space design can adequately carry the key semantic information in the operating state of the distributed system, while maintaining the coordination between reconstruction learning and anomaly detection. Since the latent representation serves both to compress and represent normal patterns and to influence the quality of subsequent anomaly deviation measurement, an appropriate dimensionality is crucial for improving model effectiveness.

Our proposed method performs well under this parameter configuration, indicating that a larger latent space is not necessarily better; rather, it needs to be matched with the complexity of input features, the compactness of normal patterns, and the requirements for representing anomalies. The current results reflect that our method can achieve a relatively balanced state between representational capability and structural constraints, ensuring that latent variables neither overcompress key operational information nor weaken the clarity of anomaly boundaries due to redundant representations. Therefore, this setting provides a more stable modeling foundation for our proposed method and further supports effective identification in distributed system anomaly detection tasks.

5. Conclusion

This paper addresses the challenges of complex anomaly patterns, heterogeneous observed signals, and the insufficient capacity of traditional detection methods to represent deep anomalies in distributed system operating environments. It proposes an anomaly detection method based on VAE reconstruction discrimination. To address the characteristics of distributed systems, such as blurred boundaries between normal and abnormal states, hidden anomaly propagation paths, and the tendency for local disturbances to be confused with normal fluctuations, a unified detection framework combining latent distribution modeling and reconstruction bias discrimination is constructed. This method is based on multivariate time-series observations, organizes local contextual information through a sliding window, and learns the latent structure of normal system behavior using variational modeling. Furthermore, it integrates reconstruction residuals and latent offset information to form the basis for anomaly discrimination. Results show that the proposed method can effectively complete the task of identifying abnormal states in distributed systems, demonstrating strong performance in accuracy, stability, and comprehensive discrimination capabilities,

reflecting the effectiveness of the collaborative design of reconstruction learning and discriminative modeling.

From a methodological perspective, the significance of this work lies not only in providing an implementation path for anomaly detection in distributed systems but also in advancing anomaly detection thinking from a single error-driven approach to one that emphasizes both latent representation constraints and the fusion of anomaly evidence. Traditional refactoring methods often focus on superficial differences between inputs and outputs, while this paper emphasizes the underlying organization of normal behavior distributions and the structural deviations of anomalous states from normal patterns. This allows the model to retain a clearer basis for judgment in complex, noisy, and highly dynamic operating scenarios. For distributed systems, anomalies are often not direct manifestations of single-point failures, but rather comprehensive external reflections of imbalances across multiple services, resources, and stages. Therefore, robust judgments are often difficult to form based solely on local observations. Our method enhances the model's ability to identify complex anomaly patterns through the synergistic effect of latent space constraints, refactoring consistency modeling, and anomaly scoring fusion mechanisms. It also provides more interpretive state representation support for subsequent root cause analysis, alarm filtering, and operational decisions.

From an application perspective, this research has strong practical significance for scenarios such as cloud computing platforms, microservice architectures, container clusters, industrial internet platforms, and financial-grade digital infrastructure. As system scale continues to expand, service call chains become increasingly complex, and business requirements for continuity and reliability continue to rise, anomaly detection is no longer just an auxiliary operational function but is gradually becoming a crucial technical foundation for ensuring the stable operation of critical businesses. The method proposed in this paper provides more reliable intelligent support for operational status awareness, anomaly early warning, fault detection, and risk control, helping to reduce business losses caused by system fault propagation and improve system governance efficiency and automated operation and maintenance levels. Simultaneously, this research also provides a new reference path for the in-depth application of artificial intelligence technology in intelligent operation and maintenance, cyberspace security, complex software system governance, and critical infrastructure risk prevention and control, positively impacting the intelligent upgrading of related application areas.

Future research can continue in several directions. First, it can further strengthen the joint modeling of multi-source information, such as log semantics, call chain topology, and resource dependencies, enabling anomaly detection to gradually expand from primarily temporal reconstruction to unified representation learning oriented towards multimodal operational evidence. Second, it can research more lightweight model structures with continuous learning capabilities to adapt to the complex conditions of rapid changes in business load, dynamic scaling of service instances, and continuous evolution of anomaly patterns in real production environments. Third, it can more closely integrate anomaly detection results with fault location, causal analysis, and automatic repair strategies, forming a closed-loop intelligent mechanism from anomaly detection to operation and maintenance handling. Fourth, the generalization ability of the model under cross-platform, cross-architecture, and cross-scenario conditions can be further explored to enhance its applicability in different types of distributed systems. As the demand for intelligent operation and maintenance continues to deepen, research on anomaly detection oriented towards collaborative optimization of potential structure modeling and discrimination still has broad development prospects and is expected to support higher levels of system autonomy, risk warning, and business continuity assurance in the future.

Recent studies also provide methodological extensions for strengthening anomaly detection in distributed and cloud-native systems. Heterogeneous dynamic dependency graph learning emphasizes that evolving

relations among backend entities can expose abnormal propagation patterns that are difficult to capture from isolated metrics [15]. Temporal evidence aggregation with causal-contrastive Transformer learning further shows that fault diagnosis can benefit from combining temporal evidence with contrastive causal cues [16]. Keyword-constrained retrieval-augmented generation suggests that explicit retrieval constraints can improve the controllability of evidence selection in large language model based reasoning pipelines [17]. Multi-source feature fusion for supply chain delay risk prediction and unified modeling of autonomous exploration and goal discovery both demonstrate that heterogeneous evidence aggregation and adaptive objective formation can enhance decision robustness in complex operational environments [18], [19]. Structured pruning-driven low-rank adapter learning and knowledge graph-enhanced language models also indicate that compact adaptation and relational knowledge injection can support more efficient and semantically grounded intelligent analysis [20], [21].

These related directions also broaden the practical foundation of future distributed anomaly detection. Generalizable latent representation learning for cloud-system forecasting is consistent with the need to build transferable state representations across heterogeneous runtime signals [22]. Structure-aware root cause localization through graph representation learning and causal inference connects anomaly scoring with interpretable fault reasoning, while behavior representation learning for AI inference backends highlights the importance of reliable operational monitoring under service uncertainty [23], [24]. Cross-period financial anomaly detection and microstructure-aware high-frequency trading risk detection show that contrastive temporal modeling and dense sequence representation can capture subtle risk patterns beyond ordinary threshold rules [25], [26]. Task-aware regularized continual fine-tuning provides a useful basis for adapting detection models under evolving workloads without forgetting previously learned behaviors [27]. Consistency-aware unified modeling for multi-source system anomaly detection and temporal semantic-enhanced Transformer-based anomaly awareness for enterprise data pipelines further reinforce the importance of unified multi-source modeling, semantic temporal learning, and evidence-consistent anomaly perception [28], [29].

References

[1] Xie Z, Pei C, Li W, et al. From point-wise to group-wise: A fast and accurate microservice trace anomaly detection approach[C]//Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering. 2023: 1739-1749.

[2] Nobre J, Pires E J S, Reis A. Anomaly detection in microservice-based systems[J]. Applied Sciences, 2023, 13(13): 7891.

[3] Liang X, Li L, Peng H. Unsupervised microservice log anomaly detection method based on graph neural network[C]//International Conference on Swarm Intelligence. Singapore: Springer Nature Singapore, 2024: 197-208.

[4] Wang P, Zhang X, Cao Z, et al. MADMM: microservice system anomaly detection via multi-modal data and multi-feature extraction[J]. Neural Computing and Applications, 2024, 36(25): 15739-15757.

[5] Li Z, Shi J, Van Leeuwen M. Graph neural networks based log anomaly detection and explanation[C]//Proceedings of the 2024 IEEE/ACM 46th international conference on software engineering: companion proceedings. 2024: 306-307.

[6] Osswald M, Schönenberger T, Cantali G, et al. Anomaly Detection in Microservices Architecture Using Graph Neural Networks[C]//2025 33rd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (PDP). IEEE, 2025: 560-567.

[7] Fan M, Zhang X, Wang P, et al. Multi-modal anomaly detection for microservice system through nested graph diffusion reconstruction[J]. Applied Intelligence, 2025, 55(11): 784.

[8] Nedelkoski S, Cardoso J, Kao O. Anomaly detection and classification using distributed tracing and deep learning[C]//2019 19th IEEE/ACM international symposium on cluster, cloud and grid computing (CCGRID). IEEE, 2019: 241-250.

[9] Zhang C, Peng X, Sha C, et al. Deeptralog: Trace-log combined microservice anomaly detection through graph-based deep learning[C]//Proceedings of the 44th international conference on software engineering. 2022: 623-634.

[10] Kohyarnejadfard I, Aloise D, Azhari S V, et al. Anomaly detection in microservice environments using distributed tracing data analysis and NLP[J]. Journal of Cloud Computing, 2022, 11(1): 25.

[11] Xie Z, Xu H, Chen W, et al. Unsupervised anomaly detection on microservice traces through graph vae[C]//Proceedings of the ACM Web Conference 2023. 2023: 2874-2884.

[12] Chen J, Liu F, Jiang J, et al. TraceGra: A trace-based anomaly detection for microservice using graph deep learning[J]. Computer Communications, 2023, 204: 109-117.

[13] Xie X, Jian S, Huang C. Logtracead: anomaly detection from both logs and traces with graph representation learning[C]//Proceedings of the 2023 2nd International Conference on Networks, Communications and Information Technology. 2023: 116-121.

[14] Huang J, Yang Y, Yu H, et al. Twin graph-based anomaly detection via attentive multi-modal learning for microservice system[C]//2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE). IEEE, 2023: 66-78.

[15] F. Chen, "Heterogeneous Dynamic Dependency Graph Learning for Anomaly Detection in Large-Scale Backend Systems," 2024.

[16] Z. Hu, "Temporal Evidence Aggregation with Causal-Contrastive Transformer Learning for Intelligent Cloud Failure Diagnosis," 2024.

[17] W. Ma and W. Zhong, "Keyword-Constrained Retrieval-Augmented Generation for Large Language Models," 2025.

[18] R. Jiang and Y. Shi, "SCDF-Net: Research on a Multi-Source Feature Fusion Binary Classification Prediction Network Algorithm for Supply Chain Delay Delivery Risk," Journal of Computer Technology and Software, vol. 4, no. 7, 2025.

[19] Y. Chen, S. Liu, and C. Qian, "Unified Modeling of Autonomous Exploration and Goal Discovery for Open-Environment Agents," Transactions on Computational and Scientific Methods, vol. 5, no. 9, 2025.

[20] O. Lin, G. Brennan, and X. Xu, "Structured Pruning-Driven High-Expressive Low-Rank Adapter Learning for Large Language Models," 2025.

[21] Z. Gu and L. Nie, "Knowledge Graph-Enhanced Large Language Models for Opinion Target Sentiment Classification," 2025.

[22] W. Huang, "Generalizable Latent Representation Learning for Multi-Source Time Series Forecasting in Cloud Systems," 2024.

[23] L. Wang and C. Y. Chang, "Structure-Aware Root Cause Localization for Microservice Backends via Joint Graph Representation Learning and Causal Inference," 2024.

[24] T. Ding and Y. Yuan, "Behavior Representation Learning for Reliable Monitoring of AI Inference Backends," 2024.

[25] Z. Peng, "Cross-Period Financial Statement Anomaly Detection via Contrastive Representation Learning," 2024.

[26] Z. Su, "Microstructure-Aware Risk Detection in High-Frequency Trading Using Embedding and Sequence Modeling," 2024.

[27] J. Hu, Y. Lyu, and J. Jiang, "Task-Aware Regularized Continual Fine-Tuning for Mitigating Catastrophic Forgetting in Large Language Models," Journal of Computer Technology and Software, vol. 4, no. 7, 2025.

[28] Y. Yang, M. Nagel, and A. Li, "Consistency-Aware Unified Modeling for Multi-Source System Anomaly Detection," 2025.

[29] Y. Wang, "Research on a Temporal Semantic-Enhanced Transformer-Based Anomaly Awareness Method for Complex Enterprise Data Pipelines," 2025.