2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

A Gradient Noise-based Dynamic Conditional Diffusion Model for Time-Series Anomaly Detection

Xianghe Du, Xueru Song, Shikang Pang, Shuaitao Yang, Yao Tong and Jiahui Lu(✉)

Henan Key Laboratory of Big Data Analysis and Processing, Henan University, Kaifeng 475004, China
xianghedu@henu.edu.cn

Abstract. Time series anomaly detection is crucial in various real-world scenarios, including fault diagnosis, financial fraud detection, and early warning systems. While diffusion models have recently emerged as powerful generative tools for anomaly detection, two key challenges persist: (1) conventional Gaussian noise used during the forward process fails to suppress anomaly-specific frequencies due to spectral mismatches; and (2) most existing methods adopt a unified model to detect all types of anomalies, overlooking the distinct characteristics of trend, seasonal, and mixture anomalies. To address these issues, we propose GNDC-DM, a gradient noise-based dynamic conditional diffusion model for time series anomaly detection. GNDC-DM employs three dedicated channels to detect different types of anomalies individually. In the trend and seasonal channels, we introduce a novel gradient noise that fuses gradient-aligned noise with stochastic Gaussian components, effectively preserving normal patterns while corrupting anomaly distortion. In the mixture channel, we dynamically incorporate trend and seasonal components as conditions to guide the denoising process, making mixed anomalies more distinguishable. Extensive experiments on four benchmark datasets demonstrate the superior performance of our approach, highlighting its ability to improve detection accuracy across various anomaly categories.

Keywords: Time series, anomaly detection, diffusion models.

1 Introduction

Time series anomaly detection refers to the identification of data points that significantly deviate from established normal patterns within temporal sequences. This technique holds critical importance in real-world applications, such as fault detection in industrial systems, financial fraud detection, automotive fault diagnosis, and early anomaly warning for machinery [2, 8, 35]. Accurately detecting anomalies in time series data is essential for ensuring operational safety, reducing risks, and preventing significant economic losses [19, 50]. Due to the high cost of obtaining ground-truth anomaly labels, one of the primary challenges in time series anomaly detection is identifying anomalies in an unsupervised manner. To address this, a variety of unsupervised techniques have been proposed, including Autoencoders (AE) [3], Normalizing Flows [9], Graph Neural Networks (GNN) [10], and Transformer-based models [18, 46]. These methods typically

rely on learning compact representations or reconstruction errors to distinguish normal from anomalous patterns.

In contrast, generative models aim to learn the underlying distribution of normal time series data, offering the potential to detect anomalies based on generation quality and likelihood estimates. By capturing normal patterns, these models can accurately reconstruct normal instances, while failing to do that for anomalous ones. As a result, anomalies can be identified by measuring reconstruction errors. Among the generative approaches, Generative Adversarial Networks (GANs) [24] and Variational Autoencoders (VAEs) [30] have been explored for their ability to generate realistic time series data. However, these models often suffer from limited generation fidelity and training instability, resulting in noisy or unrealistic outputs that can impair detection performance [51]. Therefore, diffusion model-based methods [7, 52] have recently been introduced for time series anomaly detection. These models provide stronger generative capabilities, more stable training, and higher-quality sample reconstruction, making them a promising alternative for robust anomaly identification.

Despite recent advancements, two critical limitations hinder the performance of current methods: (1) Existing diffusion-based approaches predominantly employ Gaussian noise during the forward process to corrupt input data [11, 14, 16, 32]. However, because Gaussian noise does not differentiate between normal and anomalous data, it can result in the anomalous data being reconstructed as effectively as the normal data during the generation process. While, reconstruction-based methods require to smooth anomalous points and obtain larger reconstruction errors for anomalies [7, 52]. The main reason for this is that this noise paradigm inadequately suppresses anomalous patterns due a spectral mismatch: Gaussian noise exhibits uniform energy distribution across all frequency bands, whereas anomalies in time series often localize to specific frequencies—e.g., high-frequency spikes or mid-band deviations [51]. Consequently, its indiscriminate spectral perturbation fails to selectively attenuate anomaly dominated frequencies to enlarge reconstruction errors of anomalies. (2) Most methods adopt a unified detection model to identify different types of anomalies [43], leading to suboptimal detection results. Time series anomalies exhibit intricate seasonal anomalies, trend anomalies, and mixture anomalies [23, 49]. A single-model architecture learns entangled representations that inadequately capture the distinct causal mechanisms underlying these anomaly categories . For instance, seasonal anomalies require frequency-aware detectors, while trend anomalies demand robust baseline estimators. This "one-size-fits-all" approach underperforms compared to specialized detectors tailored for specific anomaly types [32].

To address these issues, we propose a gradient noise-based dynamic conditional diffusion model for time-series anomaly detection (called GNDC-DM). In this framework, we design three specialized channels to independently detect trend anomalies, seasonal anomalies, and mixed anomalies. In the trend and seasonal channels, we introduce a gradient noise to replace the standard Gaussian noise used in conventional diffusion models. This custom noise is designed to effectively attenuate the anomalies, thereby enlarging the reconstruction error and improving the detection performance. It is constructed by combining gradient-oriented noise, which reinforces normal trends (or seasonal patterns), with stochastic Gaussian components, which preserve diversity. The gradient aligned term leverages the fact that anomalies typically exhibit gradient directions that deviate from their neighborhood trends (or seasonal patterns). By aligning with dominant normal gradients, this term naturally suppresses anomalous fluctuations. Meanwhile, the residual Gaussian component introduces controlled randomness, helping to avoid

deterministic bias and maintain sample diversity. By integrating these two components, the proposed noise formulation ensures that normal patterns are minimally distorted during the reconstruction process, while anomalies are systematically shifted toward the normal manifold. This achieves both effective anomaly mitigation and robust preservation of normal characteristics.

In the mixture channel, we combine both trend and seasonal noise at each time step as dynamic conditions to guide the diffusion model in reconstructing time series data. By aligning the generated data with normal trend and seasonal patterns, this design effectively suppresses abnormal components in these dimensions, thereby making mixture anomalies more distinguishable and easier to detect. Unlike static conditions (e.g., global trends or seasonal averages), which may fail to align with the local temporal context due to the evolving nature of the diffusion process, our approach dynamically synchronizes the generation process with the step-wise characteristics of the input data. Specifically, we iteratively extract step-specific trend and seasonal information and incorporate them as conditions for the mixture channel. This ensures that the generated outputs remain locally coherent and closely aligned with the distribution of normal data. As a result, data points containing mixture anomalies tend to exhibit larger reconstruction errors, enabling their effective identification based on reconstruction-based scoring. Our main contributions are summarized as follows:

- We propose a novel time series anomaly detection framework, GNDC-DM, which employs three specialized channels to independently detect trend anomalies, seasonal anomalies, and mixed anomalies.
- We introduce a tailored gradient noise, which combines gradient-oriented constraints with stochastic components to explicitly suppress anomalous patterns while preserving normal structures. This design amplifies reconstruction errors for anomalies and enhances the detection of trend and seasonal anomalies.
- We develop a dynamic conditional denoising network that adaptively fuses trend and seasonal components via learnable cross-attention mechanisms. These combined information serves as dynamic conditions for the reverse diffusion process to corrupt anomalies, thereby improving the detection of mixed anomalies.
- Extensive experiments on four benchmark datasets demonstrate the superior performance of GNDC-DM compared to state-of-the-art time series anomaly detection methods.

2 Related Work

2.1 Time-series anomaly detection

Initial research in time series anomaly detection primarily relied on conventional methods such as density estimation [6] and clustering techniques [40, 42]. The advent of deep learning has since spurred the development of numerous models that have significantly advanced the field. These deep learning–based approaches utilize a variety of techniques—including autoencoders (AE) [3, 20, 38], variational autoencoders (VAE)[17, 43], generative adversarial networks (GAN) [13, 15, 24], Transformers [18, 33, 46], and denoising diffusion probabilistic models (DDPM) [7, 32, 52]—to achieve superior anomaly detection performance.

The method employs autoencoders (AE) [38] to reconstruct time series through an encoding–decoding procedure, with anomaly detection based on measuring the reconstruction error between the original and reconstructed data. Building on this foundation, subsequent research has enhanced both model interpretability and performance by decomposing time series. Furthermore, variational autoencoder (VAE)-based approaches [17, 43] not only capture reconstruction errors but also estimate the log-likelihood of the reconstructions, thereby modeling the latent probability distribution of the data and achieving robust anomaly detection. Nonetheless, reconstruction methods that lack proper regularization are susceptible to overfitting and diminished performance , a problem that GANs [13, 15, 24] address by concurrently training a discriminator and a generator to produce high-quality synthetic data. To overcome these issues, Denoising Diffusion Probabilistic Models (DDPM) have been proposed. Leveraging the superior performance of DDPM in generation tasks and its success in time series imputation [28, 45], researchers have applied a DDPM-based imputation method to time series anomaly detection, wherein traditional Gaussian white noise is used in the forward process and the original data values serve as static conditions to guide the generation process [7, 32, 52].

2.2 Diffusion Model

Recent advancements in diffusion models have yielded remarkable success in image processing [11, 16]. Although their application to time series analysis is relatively nascent, researchers have begun exploring their utility in tasks such as imputation [4, 28, 45, 53], forecasting [22, 34], and generation [25]. Time series imputation, which involves restoring missing data, has benefited from diffusion-based approaches that offer flexible assumptions about the underlying data distribution compared to traditional probabilistic methods [26]. The CSDI framework [45] leverages diffusion models for imputation by conditioning the generation process on observed data. Building on this foundation, both DSPD/CSPD [4] and PriSTI [28] similarly guide the diffusion process using observed data. Specifically, DSPD/CSPD defines appropriate noise sources and employs score-matching models for imputation, whereas PriSTI introduces static conditions based on global spatiotemporal correlations and geographical relationships to enhance the diffusion model's performance.

Diffusion models have proven effective for multivariate time series forecasting by leveraging historical interactions among variables to predict future trends. For instance, TimeGrad [34] utilizes an autoregressive diffusion process that generates sequences via a Markov chain originating from white Gaussian noise, thereby ensuring robust and efficient forecasts. In addressing the challenges associated with applying diffusion models to time series data, subsequent research [4, 22] has introduced various novel approaches tailored to these tasks. Furthermore, diffusion models excel in generating time series data, as evidenced by TSGM [25], which employs a score-based diffusion methodology to synthesize data from a Gaussian prior, surpassing baseline methods in both quality and diversity.

Difference: Compared to conventional diffusion models that rely on Gaussian noise, our approach employs custom-tailored noise specifically designed to address trend and seasonal anomaly. Traditional diffusion models are optimized for generative tasks where the objective is to replicate the original distribution accurately. In contrast, anomaly detection requires the generation of a large proportion of normal data to amplify the error associated with anomalies. By customizing the noise to align more closely with the inherent characteristics of the data, our diffusion

model is better equipped to differentiate between typical patterns and anomalies, thereby enhancing detection performance.

3 Method

3.1 Problem Formulation

In this study, we focus on detecting anomalies in multivariate time series (MTS) using an unsupervised approach. A collection of MTS is represented $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$, comprising measurements recorded from timestamps 1 to *N*. The objective is to identify timestamps that significantly deviate from the majority of reference timestamps based on a computed threshold $\mathcal{T}$.

3.2 Framework of GNDC-DM

The overall framework of the model is shown in Fig. 1. The overview of GNDC-DM.. It proposes introduces a three channel anomaly detection architecture to identify trend, seasonal, and mixture anomalies in time series through tailored diffusion models. Specifically: (1) In the diffusion models for trend and season, we design a custom noise function, referred to as gradient noise, which is derived from predominantly normal trend and seasonal patterns. Unlike Gaussian noise, this tailored noise selectively corrupts anomalies in the trend and seasonal signals. During the forward process, gradient noise is incrementally added to the original trend and seasonal data, guiding the resulting noisy data toward normal behavior. This robust prior ensures that the reverse process reconstructs the trend and seasonal components as normal. (2) In the mixture diffusion model, for complex mixture anomalies, the model employs combined noise—encompassing both trend and seasonal aspects—as dynamic conditions to steer the generation process. In this scenario, gradient noise is also added during the forward process. However, in the reverse generation phase, the conditions are dynamically adjusted at each time step. These dynamic conditions, in contrast to static ones, emphasize normal information in the time series, thereby enhancing the generation of normal data.

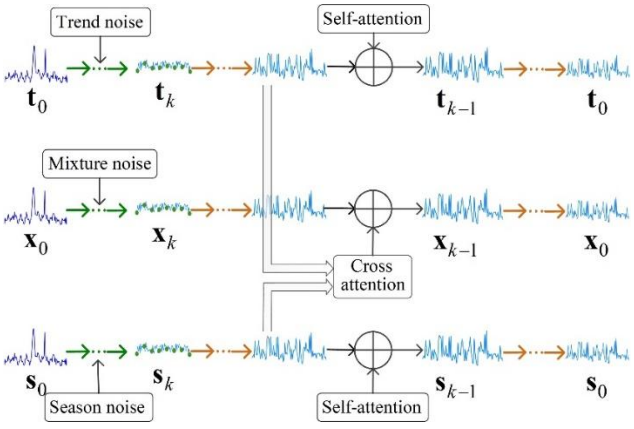

Fig. 1. The overview of GNDC-DM.

3.3 Trend Anomaly Detection

Trend anomalies in time series data are subsequences that induce significant deviations from the overall trend. We employ a diffusion model with gradient noise to reconstruct the trend and subsequently identify anomalies based on the reconstruction error.

To detect trend anomalies, we employ reconstruction-based techniques similar to those in [36, 48], in which both normal and anomalous data points are transformed to reflect normal behavior. Anomalies are subsequently identified by analyzing the reconstruction error, with larger errors indicating a higher likelihood of an anomaly. Inspired by the demonstrated efficacy of diffusion models in related tasks [7, 53], we adopt a diffusion-based approach for reconstruction. In our method, the forward process perturbs the trend data with gradient noise to create a prior distribution, and the reverse diffusion process then refines this prior to generate the reconstructed trend. However, conventional diffusion models typically corrupt the original data using Gaussian white noise, which does not sufficiently attenuate the influence of anomalous points [51]. To address this limitation, we introduce a novel gradient noise, which yields a smoother reconstructed trend and produces larger reconstruction errors for anomalies.

Gradient Noise. Inspired by the successful use of specialized noise in graph data , we propose a novel gradient noise to suppress anomaly-specific information. This noise mechanism shifts the data distribution towards normal patterns, causing anomalous values to be reconstructed as normal while preserving the integrity of the original normal values. Consequently, anomalies are readily identifiable through elevated reconstruction errors.

Anomalies in time series often appear as abrupt fluctuations that deviate from the overall trend, resulting in gradient directions that significantly differ from those of their surrounding data. Noise generated based on gradient direction can enhance the underlying trend of normal points while attenuating the influence of anomalies. To effectively suppress anomalies in the trend, we transform standard Gaussian noise into gradient noise by introducing two key constraints: (i) a directional constraint that aligns the noise with the local gradient, thereby capturing the local dynamics of the time series; and (ii) a stochastic constraint that incorporates Gaussian noise to preserve diversity in the generated samples. Given the trend time series t and Gaussian noise ϵ, the two constraints are expressed as:

$$\epsilon_t = \lambda \cdot \frac{\nabla t}{\| \nabla t \| + \varepsilon} \cdot \| \epsilon \| + \gamma \cdot \epsilon \tag{1}$$

Here, ∇t denotes the gradient of the time series t, and $\| \nabla t \|$ represents its magnitude. The term ε is a small constant introduced to prevent division by zero when $\| \nabla t \|$ approaches zero, thereby ensuring numerical stability. $\| \epsilon \|$ is the magnitude of the Gaussian noise$\epsilon \sim \mathcal{N}(0, I)$. The coefficients λ and γ are tunable hyperparameters that control the relative contributions of the gradient-oriented and random noise components, respectively.

Forward Diffusion Process. Given the gradient noise $\epsilon_{\mathbf{t}}$, we apply the forward diffusion process to inject this noise into the original trend $\mathbf{t}_0$, resulting in a corrupted version that serves as the prior:

$$\mathbf{t}_K = \sqrt{\overline{\alpha}_K}\mathbf{t}_0 + \sqrt{1 - \overline{\alpha}_K}\epsilon_{\mathbf{t}}, \tag{2}$$

where $\overline{\alpha}_T = \prod_{t=1}^{T}(1-\beta_t)$, and $\beta_1, \ldots, \beta_T$ is a fixed variance schedule [16]. Because the gradient trend noise obtained through Eq. (1) encourages the data to converge towards normal data, the abnormal information in the prior is significantly reduced after the forward process.

Denoising Neural Network. To reconstruct trend data, it requires a denoising neural network to remove the noise in the prior and generate clean data. The objective function of this denoising network is to predict the noise from Eq. (2):

$$\mathcal{L}_{tre} = \mathbb{E}_{k,\mathbf{t}_0,\epsilon_\mathbf{t}} \parallel \epsilon_\mathbf{t} - \epsilon_{\theta_\mathbf{t}}(\mathbf{t}_k, k) \parallel_2^2, \tag{3}$$

where $\epsilon_{\theta_\mathbf{t}}(\cdot)$ is the denoising neural network parameterized by $\theta_\mathbf{t}$.

To better learn the normal distribution of the trend, we employ a U-Net based on LDM [36] to implement the denoising network, $\epsilon_{\theta_\mathbf{t}}(\cdot)$. Unlike dynamic conditional denoising networks used in mixed-type anomaly detection—which rely on external conditions such as trend and seasonality—trend anomaly detection focuses on learning the distribution of normal samples without additional conditioning. To align with this objective, we remove the conditional components from the original LDM design and instead incorporate self-attention layers. Selfattention is particularly effective for capturing long-range dependencies within sequences [27, 47], enabling the network to develop a global understanding of the trend structure. Given that normal data typically dominates in anomaly detection tasks [5, 39], leveraging long-range dependencies enhances the network's ability to model normal patterns. This, in turn, amplifies the reconstruction error for anomalous points, thereby improving anomaly detection performance.

Reverse generation process. With the denoising neural network $\epsilon_{\theta_\mathbf{t}}(\cdot)$ and the prior $\mathbf{t}_K$, the reverse generation (diffusion) process iteratively remove noisy data $\mathbf{t}_K$ to generate a clean data $\mathbf{t}_0$:

$$\mathbf{t}_{k-1} = \frac{1}{\sqrt{\alpha_k}}\left(\mathbf{t}_k - \frac{\beta_k}{\sqrt{1-\overline{\alpha}_k}}\epsilon_{\theta_\mathbf{t}}(\mathbf{t}_k, k)\right) + \sqrt{\tilde{\beta}_k}\epsilon_\mathbf{t}, \tag{4}$$

The generation process starts with the noisy data $\mathbf{t}_K$ and generates a clean sample $\mathbf{t}_0$ through iterative denoising, i.e., $\mathbf{t}_K \to \cdots \to \mathbf{t}_k \to \mathbf{t}_{k-1} \cdots \to \mathbf{t}_0$.

3.4 Seasonal Anomaly Detection

For the seasonal component, a similar procedure is employed to reconstruct the data. First, analogous to Eq. (1), the Gaussian noise ϵ' is transformed into the gradient noise $\epsilon_\mathbf{s}$. This tailored noise is then used to perturb the seasonal data $\mathbf{s}_0$ over K steps in the forward diffusion process, resulting in the prior $\mathbf{s}_K$, in a manner corresponding to Eq. (2) for the trend component:

$$\mathbf{s}_k = \sqrt{\overline{\alpha}_k}\mathbf{s}_0 + \sqrt{1-\overline{\alpha}_k}\epsilon_\mathbf{s}, \tag{5}$$

Next, the denoising neural network $\epsilon_{\theta_\mathbf{s}}(\cdot)$ parameterized by $\theta_\mathbf{s}$ is trained to remove the noise in $\mathbf{s}_K$. Finally, the prior $\mathbf{s}_K$ is iteratively removed noisy using $\epsilon_{\theta_\mathbf{s}}(\cdot)$ to generate the clean season data $\mathbf{s}_0$ in the reverse process, like Eq. (4):

$$\mathbf{s}_{k-1} = \frac{1}{\sqrt{\alpha_k}}\left(\mathbf{s}_k - \frac{\beta_k}{\sqrt{1-\overline{\alpha}_k}}\epsilon_{\theta_\mathbf{s}}(\mathbf{s}_k, k)\right) + \sqrt{\tilde{\beta}_k}\epsilon_\mathbf{s}, \tag{6}$$

3.5 Anomaly Detection

Time series often comprise multiple complex components, and in addition to trend and seasonal anomalies, mixture anomalies frequently coexist within the data [23]. To detect such anomalies, we design a time series diffusion model with dynamic conditions. In the forward diffusion process, we replace standard Gaussian noise with the proposed gradient noise to better corrupt mixture anomalies. During the reverse process, dynamic conditions—derived from the decomposed trend and seasonal components—are incorporated to guide the generation of the reconstructed time series. Mixture anomalies are then identified by computing the reconstruction error between the generated time series and the original input, with larger errors indicating a higher likelihood of anomalous behavior.

Denoising Diffusion Probabilistic Models (DDPMs) have shown strong capabilities in accurately modeling data distributions for generative tasks [14]. However, when applied to anomaly detection, existing methods [7, 52] typically rely on static or gradient-based conditioning throughout the generation process. Such static conditioning lacks the flexibility to provide targeted, step-wise guidance during generation, making it challenging to adapt to complex temporal dynamics. As a result, this limitation often leads to suboptimal performance in detecting anomalies within time series data.

To address this limitation, we introduce a dynamic conditioning mechanism that incorporates synchronous noise derived from both the trend and seasonal components during the generation process. This dynamic condition guides the diffusion model to generate time series data that align with normal patterns. Specifically, at each time step in the iterative generation process, we perform trend and seasonal diffusion to obtain trend noise and seasonal noise corresponding to the same time step. These noise components predominantly reflect the characteristics of normal data. By combining them as a dynamic condition, the model is effectively guided to generate time series values that conform to normal behavior. This design ensures that the trend and seasonal signals—both aligned with normal orientations—serve as time-synchronized guidance, not only matching the temporal structure of the data but also directing the generation process toward the normal distribution.

Forward. The diffusion process for time series closely follows the trend forward process described in Eq. (2). Specifically, the time series data $\boldsymbol{x_0}$, sampled from the true distribution, are perturbed over K steps using gradient noise generated in a manner analogous to Eq. (1), resulting in the corrupted representation $\mathbf{x}_K$.

Dynamic Conditions Denoising Network. For the denoising network $\epsilon_{\theta_\mathbf{x}}(\cdot)$ applied to time series data, we incorporate dynamic condition guidance to prioritize learning the distribution of normal data. During the denoising process at each time step, we take the output predicted at the previous step, i.e., $\mathbf{x}_k$, as input. Additionally, we use the corresponding trend $\mathbf{t}_k$ and season $\mathbf{s}_k$ at the same time step as conditioning variables to predict the noise. The predicted noise is then used to generate the time series data for the previous time step, $\mathbf{x}_{k-1}$. The objective function for this process is defined as follows:

$$\mathcal{L}_{mix} = \mathbb{E}_{k,\mathbf{x}_0,\mathbf{t}_k,\mathbf{s}_k,\epsilon}[||\epsilon_{\theta_\mathbf{x}}(\mathbf{x}_k, \mathbf{t}_k, \mathbf{s}_k, k) - \epsilon||_2^2], \tag{7}$$

here, $\epsilon_{\theta_\mathbf{x}}(\mathbf{x}_k, \mathbf{t}_k, \mathbf{s}_k, k)$ denotes the denoising neural network parameterized by $\theta_\mathbf{x}$, where the input is $\mathbf{x}_k$ and the conditions are $\mathbf{t}_k$ and $\mathbf{s}_k$. As a result, the dynamic conditions at each time step change according to the corresponding time step. These dynamic conditions not only ensure consistency across time steps but also incorporate components with normal orientations, effectively guiding the network to reconstruct time series data that aligns with normal patterns.

However, the components in real-world time series are not always quantitative, meaning that using trends and seasons in equal proportions as conditions for time series generation may not be effective for generative tasks. To better align with real-time series data, we preprocess the trend and seasonal components before incorporating them as conditional factors into the denoising network. This ensures that each individual component contributes more effectively to the generation of time series data. Specifically, we control the proportion of trend and season when generating the conditional guidance. Formally, the generation of this conditional guidance is as follows:

$$\mathbf{c} = \delta \mathbf{t}_k + (1 - \delta)\mathbf{s}_k, \tag{8}$$

where δ represents the weight coefficient that controls the proportion of trend and season.

At each k-th time step, the condition $\mathbf{c}$guides the denoising process of the input data $\mathbf{x}_k$ through a cross-attention mechanism. We adopt an LDM-based U-Net architecture [36] as the diffusion network, enhancing it with cross-attention to sequentially integrate the condition $\mathbf{c}$ into the denoising process. To facilitate this integration, we use an embedding layer to transform the condition $\mathbf{c}$ into $\hat{\mathbf{c}} \in \mathbb{R}^{N \times h}$ where the dimension h corresponds to the time series representation $\phi(\mathbf{x}_k)$. The cross-attention mechanism that incorporates the condition is defined as follows:

$$\mathbf{z} = Attention(\mathbf{Q}_{\hat{\mathbf{c}}}, \mathbf{K}_\mathbf{x}, \mathbf{V}_\mathbf{x}) = Softmax(\frac{\mathbf{Q}_{\hat{\mathbf{c}}}\mathbf{K}_\mathbf{x}^T}{\sqrt{d}}) \cdot \mathbf{V}_\mathbf{x} \tag{9}$$

The output $\mathbf{z}$ from the cross-attention mechanism is then used as the input for the subsequent modules in the U-Net architecture. Specifically, the query $\mathbf{Q}_{\hat{\mathbf{c}}} = \mathbf{W}_{\hat{\mathbf{c}}}^q \cdot \hat{\mathbf{c}}$ is derived from the condition, which includes the mixed trend and seasonal noises at the k -th time step. The key and value are given by $\mathbf{K}_\mathbf{x} = \mathbf{W}_\mathbf{x}^k \cdot \phi(\mathbf{x}_k)$ and $\mathbf{V}_\mathbf{x} = \mathbf{W}_\mathbf{x}^v \cdot \phi(\mathbf{x}_k)$, respectively, where $\phi(\mathbf{x}_k) \in \mathbb{R}^{N \times h}$ represents the input time series representation for each cross-attention layer. For the query $\mathbf{Q}_{\hat{\mathbf{c}}}$ from the condition and the key $\mathbf{K}_\mathbf{x}$ from the time series, when both are normal, their similarity will be high, leading to larger attention weights when summing with $\mathbf{V}_\mathbf{x}$. In contrast, when $\mathbf{Q}_{\hat{\mathbf{c}}}$ is normal but $\mathbf{K}_\mathbf{x}$ is abnormal, the similarity decreases, resulting in significantly smaller weights. Consequently, the anomaly information in the time series is effectively attenuated, which helps the network focus on learning the patterns of normal time series.

Reverse. In the generative (reverse diffusion) process at time step k, the prior $\mathbf{x}_k$ is fed into the denoising network $\epsilon_{\theta_\mathbf{x}}(\mathbf{x}_k, \mathbf{t}_k, \mathbf{s}_k, k)$ for denoising:

$$\mathbf{x}_{k-1} = \frac{1}{\sqrt{\alpha_k}}\left(\mathbf{x}_k - \frac{1 - \alpha_k}{\sqrt{1 - \overline{\alpha_k}}}\epsilon_{\theta_\mathbf{x}}(\mathbf{x}_k, \mathbf{t}_k, \mathbf{s}_k, k)\right) + \sqrt{\tilde{\beta}_k}\epsilon_\mathbf{x}, \tag{10}$$

This generative process iteratively refines the distribution until reaching the clean time series data $\mathbf{x}_0$.

3.6 Detection Criterion

Anomalous and normal time series are distinguished based on reconstruction error, with higher errors indicating a greater likelihood of an anomaly. For mixture anomalies in time series, we define them as follows:

$$\mathbf{AS}_{\mathbf{x}}(\mathbf{x}_i) = \sum_{k=1}^{d} | |\mathbf{x}_i^{(k)} - \hat{\mathbf{x}}_i^{(k)}||_2^2 \tag{11}$$

where $\mathbf{x}_i$ is the original time series data and $\hat{\mathbf{x}}_i$ is the generated time series data. Simultaneously, we substitute the time series $\mathbf{x}$ in Eq. (11) with the trend $\mathbf{t}$ and the season $\mathbf{s}$ to separately calculate the scores for trend anomalies $\mathbf{AS}_{\mathbf{t}}(\mathbf{t}_i)$ and seasonal anomalies $\mathbf{AS}_{\mathbf{s}}(\mathbf{s}_i)$. Subsequently, we integrate the computed mixture anomaly scores, trend anomaly scores, and seasonal anomaly scores to derive the overall anomaly score:

$$\mathbf{AS}(\mathbf{x}_i) = \lambda_1 \mathbf{AS}_{\mathbf{x}}(\mathbf{x}_i) + \lambda_2 \mathbf{AS}_{\mathbf{t}}(\mathbf{t}_i) + \lambda_3 \mathbf{AS}_{\mathbf{s}}(\mathbf{s}_i), \tag{12}$$

where λ_1, λ_2 and λ_3 are the balance hyper-parameters.

Following previous works, we calculate the threshold based on the training data. Given the training data $\mathbf{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_N\}$, the corresponding decision threshold is:

$$\mathcal{T} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{AS}\,(\mathbf{x}_i) + \sqrt{\frac{1}{N}\sum_{i=1}^{N}\big(\boldsymbol{AS}(\boldsymbol{x}_i) - \boldsymbol{AVG}(\boldsymbol{X})\big)}, \tag{13}$$

where $\mathbf{AS}(\mathbf{x}_i)$ is the function for calculating anomaly scores, corresponding to Eq. (12), and $\mathbf{AVG}(\mathbf{X})$ is a function for calculating the average of $\mathbf{AS}(\mathbf{x}_i)$. A test sample is considered to be abnormal if $\mathbf{AS}(\mathbf{x}_i) > \mathcal{T}$; otherwise, it is considered normal.

4 EXPERIMENTS

4.1 Benchmark Datasets

We evaluate the effectiveness of our GNDC-DM on four widely used baselines for anomaly identification on time series, as presented in Table 1. Dataset description. (1) **MSL**, collected by NASA, which contains sensor and actuator data from the Mars rover [17]. (2) **SMD**, a five-week collection of resource utilization traces from 28 machines in an internet company's compute cluster [43]. (3) **SWaT**, a 51- dimensional dataset gathered from continuously operating critical infrastructure sensors [29]. (4) **PSM**, compiled internally by eBay, which aggregates metrics from multiple application server nodes [1].

Table 1. Dataset description

Datasets	Applications	# dimension	# point	anomaly(%)
SMD	Server	38	1,416,825	4.2%
MLS	Space	55	132,046	10.5%
SWaT	Water	51	944,919	12.1%
SMD	Server	38	1,416,825	4.20%

4.2 Baselines and Experiment Setup

We perform a thorough evaluation by comparing GNDC-DM with several baseline approaches across different categories: (1) Clustering-based methods (e.g., DeepSVDD [37], THOC [41], and ITAD [42]) group data sequences into clusters and detect anomalies by measuring their distance from the clusters. (2) Prediction-based models (e.g., LSTM [17], CL-MPPCA [44]) train predictive models to estimate future values based on a contextual window of prior observations, flagging anomalies based on discrepancies between predicted and actual values. (3) Reconstruction-based methods (e.g., LSTM-VAE [30], BeatGAN, OmniAnomaly [43], ATransformer, and TFMAE [12]) encode subsequences of normal training time series into a latent space, reconstruct the sequences, and detect anomalies using reconstruction errors. (4) Imputation-based methods (e.g., DiffAD [52] and ImDiffusion [7]) estimate missing values through imputation techniques and identify anomalies based on the magnitude of estimation errors.

Experiments are conducted using PyTorch [31] with a single NVIDIA RTX 24GB GPU. The Adam optimizer [21] is employed with default settings, an initial learning rate of 3×10^{-6}, and a batch size of 16 for all datasets. GNDC-DM is configured with 100 diffusion steps. Hyperparameters and detection thresholds for baseline models are set according to the specifications in their original papers.

4.3 Anomaly Detection Performance

Table 2 presents the precision, recall, and F1-score performance of GNDCDM and baseline methods across various datasets. All results are averaged over multiple runs to ensure robustness. GNDC-DM achieved the highest average F1 scores across the four datasets, surpassing the baseline methods, thereby demonstrating its effectiveness in time series anomaly detection. Specifically, our model obtained F1 scores of 95.27% and 95.34% on the SMD and MSL datasets, respectively. This improvement can be attributed to our decoupled modeling approach for anomaly detection, which effectively captures multiple anomaly types and enhances performance. The results on the industrial datasets further confirm the practical applicability of GNDC-DM. For SWaT and PSM, our approach achieved F1 scores of 97.77% and 98.07%, respectively, setting a new benchmark in the field. GNDC-DM successfully reduces the anomalous components in the time series through gradient noise, while maintaining a favorable precision-recall balance.

Table 2. Performance comparison between GNDC-DM and baselines on the four datasets.

Method	SMD			MSL			SWaT			PSM		
	P	R	F1	P	P	P	P	R	F1	P	R	F1
ITAD	86.22	73.71	79.48	69.44	84.09	76.07	63.13	52.08	57.08	72.80	64.02	68.13
THOC	79.76	90.95	84.99	88.45	90.97	89.69	83.94	86.36	85.13	88.14	90.99	89.54
Deep-SVDD	78.54	79.67	79.10	91.92	76.63	83.58	80.42	84.45	82.39	95.41	86.49	90.73
CL-MPPCA	82.35	76.07	79.08	73.72	88.55	80.45	76.78	81.51	79.07	56.03	99.93	71.81
LSTM	78.56	85.27	81.79	85.44	82.51	83.94	86.15	83.26	84.68	76.93	89.63	82.81
LSTM-VAE	75.76	90.08	82.30	85.49	79.94	82.62	76.00	89.5	82.20	73.62	89.92	80.96
BeatGAN	72.90	84.09	78.10	89.75	85.42	87.53	64.01	87.46	73.92	90.30	93.84	92.04
OmniAnomaly	83.68	86.82	85.22	89.02	86.37	87.67	81.42	84.30	82.83	88.39	74.46	80.83
ATransformer	89.40	95.45	92.33	92.09	95.15	93.59	91.55	96.73	94.07	96.91	98.90	97.89
TFMAE	91.41	91.07	91.24	92.83	95.59	94.19	96.47	97.23	96.85	96.36	98.46	97.40
DiffAD	90.01	95.67	92.75	92.97	95.44	94.19	97.12	96.90	97.01	97.00	98.92	97.95
ImDiffusion	95.20	95.09	94.88	89.30	86.38	87.79	89.88	84.65	87.09	98.11	97.53	97.81
GNDC-DM	94.12	96.45	**95.27**	93.92	96.81	95.34	97.12	98.43	97.77	97.52	98.62	98.07

4.4 Ablation Studies

Subsequently, we present a comprehensive ablation experiment analysis to evaluate the effectiveness of various components within GNDC-DM and to understand how these components contribute to the improvement of anomaly detection performance. Table 3 summarizes the anomaly detection results for different combinations of components across all datasets. It is important to note that all results in the table are reported using the F1-score metric, and the values are averaged over multiple independent runs. The following provides an introduction to each component presented in Table 3: (i) Noise: This component introduces gradient noise to the data during the diffusion process. (ii) Trend: This component decomposes the trend data and applies Trend Reconstruction to detect trend anomalies. (iii) Season: This component separates seasonal data and utilizes Season Reconstruction to detect seasonal anomalies. (iv) Mixture: This component employs Mixture Reconstruction to detect mixed anomalies in the residual data.

Table 3 shows that using only the gradient noise and Mixture Reconstruction components results in minimal performance improvement, indicating that although gradient noise effectively disrupts anomalies, leading to better detection compared to Gaussian noise, the unified model struggles to handle diverse anomaly types. However, incorporating either Trend Reconstruction or Season Reconstruction significantly enhances performance by specifically targeting trend and seasonal anomalies, thus improving overall detection.

Table 3. The F1-scores using different components.

Component				Dataset			
Noise	Trend	Sea-son	Mix-ture	SMD	MSL	SWaT	PSM
✘	✘	✘	✘	89.32	88.88	90.21	90.93
✔	✘	✘	✔	91.87	90.65	92.39	91.76
✘	✔	✔	✔	92.13	91.23	93.46	92.25
✔	✘	✔	✔	94.39	93.52	95.66	95.5
✔	✔	✘	✔	94.65	93.2	96.08	95.9
✔	✔	✔	✘	92.72	91.98	94.11	94.33
✔	✔	✔	✔	**95.27**	**95.34**	**97.77**	**98.07**

4.5 Sensitivity Analysis

Here, we examine the role of the key hyperparameters in the two formulas within GNDC-DM. First, to investigate the impact of δ in Equation 8, we evaluate precision (P), recall (R), and F1-score under varying δ values across four datasets. As shown in Fig. 2, the SMD and PSM datasets exhibit optimal performance at δ =0.6. This is due to the server-collected nature of these datasets, where anomalies primarily manifest as periodic patterns (e.g., daily server load cycles) and sudden deviations (e.g., resource contention spikes). As a result, these datasets require a balanced integration of temporal trends and seasonal components for effective anomaly detection. In contrast, the SWaT dataset, which records persistent attacks in industrial water treatment systems (e.g., pump manipulation over multiple operational cycles), performs better at higher δ values, highlighting the importance of long-term temporal dependencies. Finally, the MSL dataset, which contains transient anomalies caused by random equipment failures (e.g., sensor drifts in Martian environmental data), achieves peak performance at δ =0.4. This lower δ value prioritizes statistical outlier detection over temporal modeling, which aligns with the sporadic nature of the anomalies in this dataset.

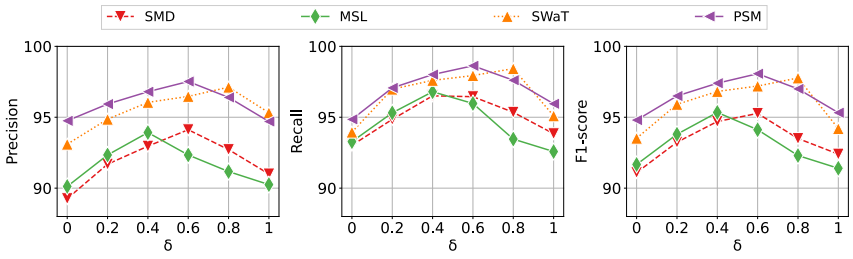

Fig. 2. The impact of δ parameters.

Second, we analyze the sensitivity of two critical hyperparameters in Equation 1: λ and γ. The results across four datasets are summarized in Fig. 3.The impact of λ and γ parameters., with key observations as follows: Impact of λ: Increasing λ from 0.1 to 0.7 enhances detection performance across all datasets. This confirms that gradient-aligned noise effectively suppresses anomalies by strengthening normal trends. However, when λ is too high, performance begins to decrease. This suggests that over-reliance on gradient orientation might distort normal patterns,

leading to a reduction in overall detection accuracy. Impact of γ: Moderate values of γ (with λ ranging from 0.5 to 0.7) strike a balance between diversity and the preservation of normal patterns, leading to optimal results. This indicates that an appropriate level of Gaussian noise helps improve the model's generalization. However, when γ is too high, excessive randomness weakens the contribution of the normal component in gradient noise, reducing the model's ability to effectively detect anomalies.

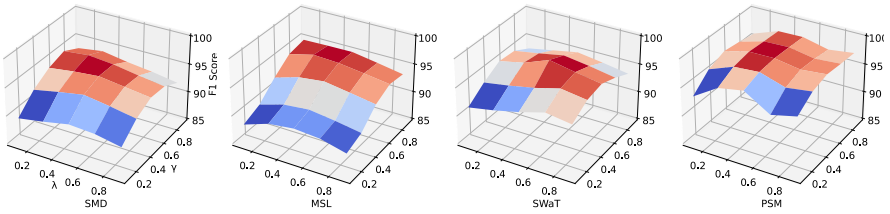

Fig. 3. The impact of λ and γ parameters.

5 Conclusions

In this work, we present GNDC-DM, a novel time series anomaly detection framework based on dynamic conditional diffusion models. By addressing two critical limitations of prior methods—ineffective anomaly suppression from standard Gaussian noise and entangled detection of different anomaly types—our approach offers a tailored solution with three specialized channels for trend, seasonal, and mixed anomalies. The proposed gradient noise integrates gradient based and stochastic elements to distort anomalies while preserving normal structures. In addition, the dynamic conditioning strategy in the mixture channel allows the model to better adapt to complex temporal patterns, significantly improving the identification of mixed anomalies. Extensive experiments across multiple real-world datasets confirm the effectiveness and robustness of GNDCDM, achieving state-of-the-art performance in both detection accuracy and generalizability.

References

1. Abdulaal, A., Liu, Z., Lancewicki, T.: Practical Approach to Asynchronous Multivariate Time Series Anomaly Detection and Localization. Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, pp. 2485–2494. Association for Computing Machinery, Virtual Event, Singapore (2021)
2. Anandakrishnan, A., Kumar, S., Statnikov, A., Faruquie, T., Xu, D.: Anomaly detection in finance: editors' introduction. In: KDD 2017 Workshop on Anomaly Detection in Finance, pp. 1-7. PMLR, (2018)
3. Audibert, J., Michiardi, P., Guyard, F., Marti, S., Zuluaga, M.A.: Usad: Unsupervised anomaly detection on multivariate time series. In: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining, pp. 3395-3404. (2020)
4. Biloš, M., Rasul, K., Schneider, A., Nevmyvaka, Y., Günnemann, S.: Modeling temporal data as continuous functions with process diffusion. (2022)

5. Blázquez-García, A., Conde, A., Mori, U., Lozano, J.A.: A review on outlier/anomaly detection in time series data. ACM computing surveys (CSUR) 54, 1-33 (2021)
6. Breunig, M.M., Kriegel, H.-P., Ng, R.T., Sander, J.: LOF: identifying density-based local outliers. In: Proceedings of the 2000 ACM SIGMOD international conference on Management of data, pp. 93-104. (2000)
7. Chen, Y., Zhang, C., Ma, M., Liu, Y., Ding, R., Li, B., He, S., Rajmohan, S., Lin, Q., Zhang, D.: Imdiffusion: Imputed diffusion models for multivariate time series anomaly detection. arXiv preprint arXiv:2307.00754 (2023)
8. Cook, A.A., Mısırlı, G., Fan, Z.: Anomaly detection for IoT time-series data: A survey. IEEE Internet of Things Journal 7, 6481-6494 (2019)
9. Dai, E., Chen, J.: Graph-augmented normalizing flows for anomaly detection of multiple time series. arXiv preprint arXiv:2202.07857 (2022)
10. Deng, A., Hooi, B.: Graph neural network-based anomaly detection in multivariate time series. In: Proceedings of the AAAI conference on artificial intelligence, pp. 4027-4035. (2021)
11. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. Advances in neural information processing systems 34, 8780-8794 (2021)
12. Fang, Y., Xie, J., Zhao, Y., Chen, L., Gao, Y., Zheng, K.: Temporal-frequency masked autoencoders for time series anomaly detection. In: 2024 IEEE 40th International Conference on Data Engineering (ICDE), pp. 1228-1241. IEEE, (2024)
13. Geiger, A., Liu, D., Alnegheimish, S., Cuesta-Infante, A., Veeramachaneni, K.: Tadgan: Time series anomaly detection using generative adversarial networks. In: 2020 ieee international conference on big data (big data), pp. 33-43. IEEE, (2020)
14. Graham, M.S., Pinaya, W.H., Tudosiu, P.-D., Nachev, P., Ourselin, S., Cardoso, J.: Denoising diffusion models for out-of-distribution detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2948-2957. (2023)
15. Han, C., Rundo, L., Murao, K., Noguchi, T., Shimahara, Y., Milacski, Z.Á., Koshino, S., Sala, E., Nakayama, H., Satoh, S.i.: MADGAN: Unsupervised medical anomaly detection GAN using multiple adjacent brain MRI slice reconstruction. BMC bioinformatics 22, 1-20 (2021)
16. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems 33, 6840-6851 (2020)
17. Hundman, K., Constantinou, V., Laporte, C., Colwell, I., Soderstrom, T.: Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, pp. 387-395. (2018)
18. Jeong, Y., Yang, E., Ryu, J.H., Park, I., Kang, M.: Anomalybert: Self-supervised transformer for time series anomaly detection using data degradation scheme. arXiv preprint arXiv:2305.04468 (2023)
19. Jin, Y., Wei, Y., Cheng, Z., Tai, W., Xiao, C., Zhong, T.: Multi-scale dynamic graph learning for time series anomaly detection (student abstract). In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 23523-23524. (2024)
20. Kim, D., Park, S., Choo, J.: When model meets new normals: test-time adaptation for unsupervised time-series anomaly detection. In: Proceedings of the AAAI conference on artificial intelligence, pp. 13113-13121. (2024)
21. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

22. Kollovieh, M., Ansari, A.F., Bohlke-Schneider, M., Zschiegner, J., Wang, H., Wang, Y.B.: Predict, refine, synthesize: Self-guiding diffusion models for probabilistic time series forecasting. Advances in Neural Information Processing Systems 36, 28341-28364 (2023)
23. Lai, K.-H., Zha, D., Xu, J., Zhao, Y., Wang, G., Hu, X.: Revisiting time series outlier detection: Definitions and benchmarks. In: Thirty-fifth conference on neural information processing systems datasets and benchmarks track (round 1). (2021)
24. Li, D., Chen, D., Jin, B., Shi, L., Goh, J., Ng, S.-K.: MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks. In: International conference on artificial neural networks, pp. 703-716. Springer, (2019)
25. Lim, H., Kim, M., Park, S., Park, N.: Regular time-series generation using SGM. arXiv preprint arXiv:2301.08518 (2023)
26. Lin, L., Li, Z., Li, R., Li, X., Gao, J.: Diffusion models for time-series applications: a survey. Frontiers of Information Technology & Electronic Engineering 25, 19-41 (2024)
27. Lin, Z., Li, M., Zheng, Z., Cheng, Y., Yuan, C.: Self-attention convlstm for spatiotemporal prediction. In: Proceedings of the AAAI conference on artificial intelligence, pp. 11531-11538. (2020)
28. Liu, M., Huang, H., Feng, H., Sun, L., Du, B., Fu, Y.: Pristi: A conditional diffusion framework for spatiotemporal imputation. In: 2023 IEEE 39th International Conference on Data Engineering (ICDE), pp. 1927-1939. IEEE, (2023)
29. Mathur, A.P., Tippenhauer, N.O.: SWaT: A water treatment testbed for research and training on ICS security. In: 2016 international workshop on cyber-physical systems for smart water networks (CySWater), pp. 31-36. IEEE, (2016)
30. Park, D., Hoshi, Y., Kemp, C.C.: A multimodal anomaly detector for robot-assisted feeding using an lstm-based variational autoencoder. IEEE Robotics and Automation Letters 3, 1544-1551 (2018)
31. Paszke, A.: Pytorch: An imperative style, high-performance deep learning library. arXiv preprint arXiv:1912.01703 (2019)
32. Pintilie, I., Manolache, A., Brad, F.: Time series anomaly detection using diffusion-based models. In: 2023 IEEE International Conference on Data Mining Workshops (ICDMW), pp. 570-578. IEEE, (2023)
33. Qin, S., Zhu, J., Wang, D., Ou, L., Gui, H., Tao, G.: Decomposed transformer with frequency attention for multivariate time series anomaly detection. In: 2022 IEEE International Conference on Big Data (Big Data), pp. 1090-1098. IEEE, (2022)
34. Rasul, K., Seward, C., Schuster, I., Vollgraf, R.: Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In: International conference on machine learning, pp. 8857-8868. PMLR, (2021)
35. Ren, H., Xu, B., Wang, Y., Yi, C., Huang, C., Kou, X., Xing, T., Yang, M., Tong, J., Zhang, Q.: Time-series anomaly detection service at microsoft. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, pp. 3009-3017. (2019)
36. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 10684-10695. (2022)
37. Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S.A., Binder, A., Müller, E., Kloft, M.: Deep one-class classification. In: International conference on machine learning, pp. 4393-4402. PMLR, (2018)
38. Sakurada, M., Yairi, T.: Anomaly detection using autoencoders with nonlinear dimensionality reduction. In: Proceedings of the MLSDA 2014 2nd workshop on machine learning for sensory data analysis, pp. 4-11. (2014)

39. Schmidl, S., Wenig, P., Papenbrock, T.: Anomaly detection in time series: a comprehensive evaluation. Proceedings of the VLDB Endowment 15, 1779-1797 (2022)
40. Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., Williamson, R.C.: Estimating the support of a high-dimensional distribution. Neural computation 13, 1443-1471 (2001)
41. Shen, L., Li, Z., Kwok, J.: Timeseries anomaly detection using temporal hierarchical one-class network. Advances in neural information processing systems 33, 13016-13026 (2020)
42. Shin, Y., Lee, S., Tariq, S., Lee, M.S., Jung, O., Chung, D., Woo, S.S.: Itad: integrative tensor-based anomaly detection system for reducing false positives of satellite systems. In: Proceedings of the 29th ACM international conference on information & knowledge management, pp. 2733-2740. (2020)
43. Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W., Pei, D.: 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, pp. 2828-2837. (2019)
44. Tariq, S., Lee, S., Shin, Y., Lee, M.S., Jung, O., Chung, D., Woo, S.S.: Detecting anomalies in space using multivariate convolutional LSTM with mixtures of probabilistic PCA. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, pp. 2123-2133. (2019)
45. Tashiro, Y., Song, J., Song, Y., Ermon, S.: Csdi: Conditional score-based diffusion models for probabilistic time series imputation. Advances in neural information processing systems 34, 24804-24816 (2021)
46. Tuli, S., Casale, G., Jennings, N.R.: Tranad: Deep transformer networks for anomaly detection in multivariate time series data. arXiv preprint arXiv:2201.07284 (2022)
47. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems 30, (2017)
48. Wang, C., Zhuang, Z., Qi, Q., Wang, J., Wang, X., Sun, H., Liao, J.: Drift doesn't matter: dynamic decomposition with diffusion reconstruction for unstable multivariate time series anomaly detection. Advances in Neural Information Processing Systems 36, 10758-10774 (2023)
49. Wang, Z., Xu, X., Zhang, W., Trajcevski, G., Zhong, T., Zhou, F.: Learning latent seasonal-trend representations for time series forecasting. Advances in Neural Information Processing Systems 35, 38775-38787 (2022)
50. Wen, Q., Yang, L., Zhou, T., Sun, L.: Robust time series analysis and applications: An industrial perspective. In: Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining, pp. 4836-4837. (2022)
51. Wyatt, J., Leach, A., Schmon, S.M., Willcocks, C.G.: Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 650-656. (2022)
52. Xiao, C., Gou, Z., Tai, W., Zhang, K., Zhou, F.: Imputation-based time-series anomaly detection with conditional weight-incremental diffusion models. In: Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining, pp. 2742-2751. (2023)
53. Xiao, C., Jiang, X., Du, X., Yang, W., Lu, W., Wang, X., Chetty, K.: Boundary-enhanced time series data imputation with long-term dependency diffusion models. Knowledge-Based Systems 310, 112917 (2025)