ANOMSEER: REINFORCING MULTIMODAL LLMS TO REASON FOR TIME-SERIES ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Time-series anomaly detection (TSAD) with multimodal large language models (MLLMs) is an emerging area, yet a persistent challenge remains: MLLMs rely on coarse time-series heuristics but struggle with multi-dimensional, detailed reasoning, which is vital for understanding complex time-series data. We present ANOMSEER to address this by reinforcing the model to ground its reasoning in precise, structural details of time series, unifying anomaly classification, localization, and explanation. At its core, an expert chain-of-thought trace is generated to provide a verifiable, fine-grained reasoning from classical analyses (e.g., statistical measures, frequency transforms). Building on this, we propose a novel time-series grounded policy optimization (TimerPO) that incorporates two additional components beyond standard reinforcement learning: a time-series grounded advantage based on optimal transport and an orthogonal projection to ensure this auxiliary granular signal does not interfere with the primary detection objective. Across diverse anomaly scenarios, ANOMSEER, with Qwen2.5-VL-3B/7B-Instruct, outperforms larger commercial baselines (e.g., GPT-4o) in classification and localization accuracy, particularly on point- and frequency-driven exceptions. Moreover, it produces plausible time-series reasoning traces that support its conclusions.

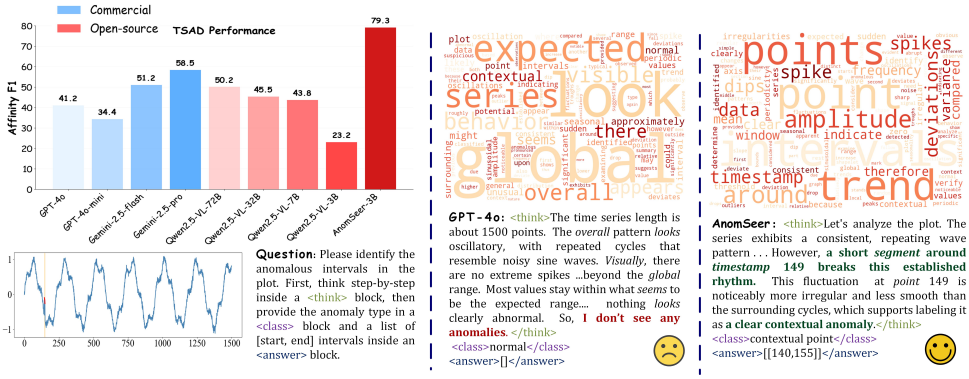


Figure 1: Comparison of model performance and time-series reasoning quality. **Left**: Affinity F1 (%) of different models on TSAD benchmarks. **Middle**: GPT-4o results, including word frequency distributions in reasoning (top) and its coarse-grained answer (bottom). **Right**: ANOMSEER results, including word frequency distributions in reasoning (top) and its fine-grained answer (bottom).

1 INTRODUCTION

Recent advances in large language models (LLMs) have opened new opportunities for time-series anomaly detection (TSAD) (Xu et al., 2021). Building on this progress, we focus on a practical yet underexplored setting, *time-series reasoning for anomalies* (Yang et al., 2025; Kong et al., 2025), where the goal goes beyond flagging abnormal segments: models must also produce coherent, linguistically grounded explanations. Emerging studies (Zhou & Yu, 2024; Xu et al., 2025; He et al., 2025) have revealed that LLMs exhibit stronger zero-shot robustness when reasoning over visual renderings of time series (e.g., line plots) rather than raw numeric sequences. This advantage arises

from human-like pattern perception and greater token efficiency enabled by compact, semantically rich images (He et al., 2025; Liu et al., 2024). These insights naturally motivate multimodal LLMs (MLLMs) as the backbone for advancing TSAD in an *reasoning-centric* manner, i.e., detecting, attributing, and justifying anomalies with structured natural language grounded in visual cues.

Despite these strengths, MLLMs fundamentally lack built-in time-series priors, and their reasoning often resorts to coarse time-series heuristics and struggles with detailed time-series analysis (Figure 1 (Middle)), thereby leading to suboptimal performance. While reinforcement learning (RL) (Sutton & Barto, 2018) has proven more effective than supervised fine-tuning (SFT) (Zhang et al., 2025b; Liu et al., 2025b; Luo et al., 2025; Tan et al., 2025) at incentivizing the emergent reasoning of LLMs in other domains (Guo et al., 2025; Wei et al., 2025; Feng et al., 2025), its reliance on globally verifiable rule-based goals may be ill-suited for the model to capture subtle, fine-grained time-series patterns. Consequently, even well-trained MLLMs may only excel at salient, out-of-range anomalies yet struggle to articulate nuanced shifts (e.g., small trend drifts) with faithful textual evidence. This discrepancy raises a central question for MLLM in TSAD:

> *Can we incentivize MLLMs to ground their time-series reasoning in fine-grained, multi-dimensional evidence, ensuring faithful and verifiable anomaly interpretations?*

To address this challenge, we propose AnomSeer, a novel time-series MLLM post-training approach that not only detects anomalies but also produces structured, evidence-based explanations to support its decisions. Our core idea is to fuse the analytical rigor of classical numerical TSAD with the holistic visual intuition of MLLMs through two components: *(i) expert chain-of-thought (ExpCoT)* trace, which encodes structured reasoning inspired by classical TSAD workflows, and *(ii) time-series grounded policy optimization (TimerPO)*, a novel temporal-aware RL algorithm that softly aligns the model's reasoning with ExpCoT trajectories. Instead of merely correcting outputs, AnomSeer utilizes the analytical rigor of traditional TSAD methods, such as residual inspection (Hyndman & Athanasopoulos, 2018) and wavelet-based drift detection (Thill et al., 2017), and embeds it into the MLLM's learning process. TimerPO operationalizes this integration by measuring the semantic deviation from an ExpCoT using optimal transport (Caffarelli & McCann, 2010; Bonneel et al., 2011) and transforms this distance into a refinement advantage signal. This signal is then orthogonally projected, ensuring it acts as non-interfering auxiliary guidance of the main RL objective. Consequently, TimerPO enhances the model's fine-grained temporal-aware reasoning capabilities (Figure 1 (Right)) without perturbing its global visual understanding or the primary optimization objective. We summarize our key contributions as follows:

- We explore a pivotal challenge hindering the effectiveness of MLLMs for TSAD: the tendency of MLLMs to rely on coarse visual "eyeballing" rather than engaging in fine-grained numerical reasoning. We introduce AnomSeer, a novel approach that bridges this gap by transferring classical, detailed TSAD priors into the time-series reasoning process of MLLMs during training.
- We propose TimerPO, a new RL algorithm designed for time-series reasoning in TSAD. TimerPO guides fine-grained, numerical time-series knowledge into the model's reasoning. It leverages optimal transport to create auxiliary advantage signals and applies them as non-interfering corrective guidance for RL training via orthogonal projection.
- Extensive experiments across diverse TSAD tasks demonstrate that AnomSeer consistently outperforms strong MLLM baselines (e.g., GPT-4o) in detection accuracy and localization precision, unifying detection, categorization, and reasoning. Critically, it produces fine-grained, plausible reasoning traces grounded in detailed time-series evidence, achieving faithful and verifiable interpretations in time-series anomaly detection.

2 Related Work

Time series anomaly detection is a critical task in domains like healthcare and cybersecurity, aiming to detect deviations from normal temporal patterns (Wu et al., 2025; Shentu et al., 2024). Traditional methods rely on statistical techniques and machine learning methods (e.g., Z-score (Bhatnagar et al., 2021), Isolation Forest (Liu et al., 2008) and One-Class SVM (Schölkopf et al., 1999)), while recent advances leverage deep models such as Autoencoders (Zong et al., 2018; Park et al., 2018) for reconstruction- or prediction-based detection. Despite their effectiveness, these models struggle in industrial settings due to the scarcity of anomaly data, limiting generalization. To address this,

recent efforts explore pre-trained (Zhou et al., 2023; Zhang et al., 2025a) and time-series foundation models (Goswami et al., 2024; Gao et al., 2024) for zero- and few-shot detection. However, these approaches are primarily optimized for accuracy, lacking the ability to analyze anomaly types, reason about temporal patterns, and explain why a given sample is anomalous.

Time-series reasoning with LLMs is an emerging research frontier (Kong et al., 2025). To enable LLMs to perform time-series analysis, researchers have primarily explored two input strategies: prompting with numerical data (Alnegheimish et al., 2024) or visual representations (Zhuang et al., 2024; He et al., 2025; Xu et al., 2025; Zhou & Yu, 2024). While the visual approach, feeding plots into MLLMs such as GPT-4o, is often more token-efficient, its effectiveness is limited by the fact that these models are not explicitly trained on time-series visualizations. To instill temporal understanding, recent works have primarily relied on integrating classical modules (Chen et al., 2025; Liu et al., 2025a), employing auxiliary techniques (He et al., 2025; Zhuang et al., 2024), or large-scale SFT (Yang et al., 2025). An alternative and promising path involves RL to promote structured problem-solving, as seen in DeepSeek-R1 (Guo et al., 2025). Building on this, recent work such as TimeMaster (Zhang et al., 2025b) trains MLLMs for classification tasks by combining SFT with RL to enable interpretable temporal reasoning over visualized series. Nevertheless, RL for enhancing anomaly detection in MLLMs remains underexplored. In this paper, we show that vanilla RL struggles to detect subtle anomalies and propose a new method to mitigate this limitation.

3 PRELIMINARY

Time-series anomaly detection. Time-series anomaly detection (TSAD) aims to identify abnormal patterns within temporal data. Following standard practice (Zhou & Yu, 2024), we use $\mathbf{X} = \{\mathbf{x}_t\}_{t=1}^{T}$ to denote a univariate time series of length T, where each observation $\mathbf{x}_t \in \mathbb{R}$ sampled at regular intervals and may correspond to either normal or anomalous behavior. Anomalies are defined as continuous intervals of data points that deviate significantly from the expected pattern. They can be categorized into point-wise anomalies (contextual point and global point) and range-wise anomalies (trend, shapelet, and seasonal), resulting in five types in total. Formally, let $\mathcal{A} = \{(t_s^{(i)}, t_e^{(i)})\}_{i=1}^{k}$ denote the set of anomalous intervals, where $1 \le t_s^{(i)} \le t_e^{(i)} \le T$. Each tuple $(t_s^{(i)}, t_e^{(i)})$ specifies the start and end indices of the i-th anomalous segment; in particular, $t_s^{(i)} = t_e^{(i)}$ denotes a single-point anomaly. The primary goal of TSAD is to infer the set $\mathcal{A}$ with high accuracy.

Multimodal time-series formulation. To enable MLLMs to perform time-series anomaly detection, the input of MLLM consists of the time-series input $\mathbf{X}$ and context prompt $\mathbf{c}$ that encode domain knowledge, natural-language instructions, or task-specific queries to guide the model's reasoning process. To enable multimodal processing, we follow the *visualization input strategy* (Liu et al., 2024; Xu et al., 2025; Zhang et al., 2025b), rendering the raw time series into a line-plot image $\mathbf{X} \to \mathbf{I}$ and then feeding it to the MLLM's vision encoder. This approach allows the model to leverage its pre-trained visual reasoning abilities on a representation that is both compact and semantically rich (Xu et al., 2025; Xie et al., 2024).

Multimodal LLM inference. We define a time-series MLLM π_θ (parameterized by θ) that specifies a conditional distribution over an output sequence $\mathbf{y} = \{y_1, y_2, \ldots, y_N\}$, where each token y_n may correspond to an anomaly label, an interval boundary, or a natural-language reasoning. Given the rendered time-series data $\mathbf{I}$ and textual context $\mathbf{c}$, the model generates outputs autoregressively: $\pi_\theta(\mathbf{y} \mid \mathbf{I}, \mathbf{c}) = \prod_{n=1}^{N} \pi_\theta(y_n \mid y_{<n}, \mathbf{I}, \mathbf{c})$. This formulation unifies reasoning, explanation and detection in a single generative process, allowing the model to produce structured outputs that are both context-aware and interpretable.

4 METHODOLOGY

Time-series MLLMs often rely on coarse visual heuristics and fail to produce numerically grounded, fine-grained reasoning for TSAD. This weakness limits their ability to detect subtle anomalies such as frequency shifts or small trend drifts in complex time-series data. To address this, we introduce ANOMSEER, a novel MLLM post-training approach for TSAD that couples classical time-series statistical rigor with the expressive reasoning ability of MLLMs. ANOMSEER is trained with two key components: (1) *expert chain-of-thought (ExpCoT)*, which generates structured, expert-like reason-

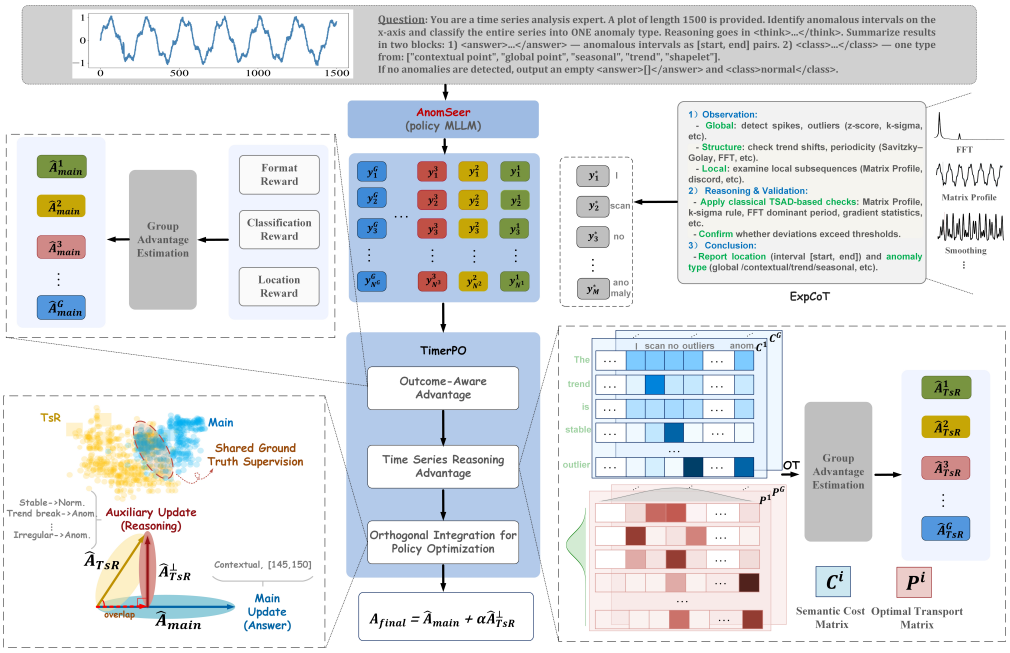


Figure 2: The overall framework of ANOMSEER. ANOMSEER first generates ExpCoT reasoning traces $\mathbf{y}^*$ from the time-series data based on classical TSAD techniques (e.g., FFT). TimerPO then computes the outcome-aware advantage and leverages optimal transport to compute the time-series reasoning advantage, which is orthogonally integrated into policy optimization to ensure stable training and improved reasoning quality.

ing traces from ground-truth time series using statistical diagnostics (e.g., histogram-based outlier scores, FFT, matrix profile); and (2) *time-series grounded policy optimization (TimerPO)*, a new RL algorithm that leverages ExpCoT to establish the corrective, orthogonal advantages to refine reasoning without overriding the detection objective. Figure 2 presents an overview of ANOMSEER. In the remainder of this section, we will detail the design of ExpCoT (Section 4.1) and the TimerPO optimization algorithm (Section 4.2), and discuss how they jointly enable accurate, interpretable, and numerically faithful anomaly detection.

4.1 EXPERT CHAIN-OF-THOUGHT GENERATION

To ground the reasoning of time-series MLLM with classical time-series detailed analysis for TSAD, we introduce the *expert chain-of-thought (ExpCoT)* trace, a structured reasoning that mirrors the stepwise detection of a human analyst. Unlike an LLM-generated CoT, which may rely on heuristic pattern matching, ExpCoT is grounded in systematically derived, quantitatively verifiable evidence. ExpCoT is generated per instance, starting from ground-truth annotations. We apply classical statistical and signal-processing techniques to extract descriptive statistics, candidate anomaly categories, and precise temporal localization. This trace delivers rich, multi-dimensional guidance that goes beyond a simple correct/incorrect signal, encouraging fine-grained and interpretable reasoning.

Crucially, ExpCoT adheres to a disciplined *three-stage* reasoning path (*Observation* → *Reasoning & Validation* → *Conclusion*), closely mirroring the stepwise process of human analytical reasoning.

Observation: systematically progress from global patterns to local irregularities across multiple views, surfacing candidate anomalies prior to formal testing.

The *"Observation"* stage performs a hierarchical scan of the time series $\mathbf{X}$ to extract preliminary statistical features. (**1**) *Global Scan:* We first assess extreme values by examining the global data distribution via a histogram-based outlier score (Goldstein & Dengel, 2012). (**2**) *Structural Scan:* If no global outliers are present, we analyze fundamental properties such as trend stability using smoothed gradients (Thill et al., 2017) and periodicity via FFT-based frequency analysis (Ren et al.,

2019). (**3**) *Local Scan:* If the series appears structurally stable, we perform a localized search for dissimilar subsequences (discords) using the Matrix Profile (Yeh et al., 2016). This fine-grained scan provides the key statistical features that guide the subsequent detection process.

Reasoning & Validation: use classical TSAD techniques to formalize ground truth as fine-grained, testable claims, supported by targeted quantitative time-series analysis.

The *"Reasoning & Validation"* stage establishes a causal link between preliminary observations and formal statistical evidence of anomalies. First, it leverages the ground-truth anomaly type to align statistical markers with visual patterns (e.g., "A sharp spike around $t \approx 150$ deviates significantly from the rest of the data, suggesting a contextual anomaly"). This classification then guides the selection of a targeted statistical method for validation; for example, a suspected trend shift is validated using gradient analysis (Thill et al., 2017), while the aforementioned contextual anomaly is confirmed by its Matrix Profile score (Yeh et al., 2016). The numerical outcome is translated into a natural language explanation (e.g., "The discord's z-score of 4.2 at timestamp 145 exceeds the 3-sigma threshold, confirming a significant pattern deviation").

Conclusion: integrate multi-dimensional insights and fine-grained evidence into a precise, defensible anomaly judgment.

The final *"Conclusion"* stage synthesizes the findings into a conclusive summary. It integrates the multi-dimensional understanding from the *"Observation"* stage with the detailed, quantitative evidence from the *"Reasoning & Validation"* stage to deliver a definitive judgment, e.g., "Therefore, the detected anomaly is a contextual point, located in the interval [145, 150]".

In summary, as shown in Figure 2, ExpCoT provides a structured reasoning trace that embeds analytical rigor and numerically grounded logic. This makes it particularly effective for identifying subtle anomalies and offers fine-grained, informed guidance for subsequent MLLM training. See examples of ExpCoT in Appendix C.2.

4.2 Time-Series Grounded Policy Optimization

To leverage ExpCoT and enable the reasoning of MLLM grounded in fine-grained time-series analysis, we introduce TimerPO, a novel RL method building upon Group Relative Policy Optimization (GRPO) (Shao et al., 2024). We begin with the vanilla GRPO formulation. Given the rendered time-series instance $\mathbf{I}$ and textual context $\mathbf{c}$, the model produces a *group* of candidate responses $\mathcal{G} = \{\mathbf{y}^1, \mathbf{y}^2, ..., \mathbf{y}^G\}$ where G denotes the group size. This group-based generation enables pairwise relative reward comparisons, which are subsequently used to compute group-aware advantages.

Outcome-Aware Advantage. For each generated response $\mathbf{y}^i \in \mathcal{G}$, the task reward is a weighted sum of (i) a format reward $r^{\text{fmt},\,i} \in \{0, 1\}$ that checks if the predefined output format of time-series MLLM is valid, (ii) a classification reward $r^{\text{cls},\,i}$ for anomaly type accuracy and (iii) a detection location reward $r^{\text{loc},,i}$, which integrates common anomaly-detection metrics (Zhou & Yu, 2024):

$$r^i = \lambda^{\text{fmt}} r^{\text{fmt},\,i} \;+\; \lambda^{\text{cls}} r^{\text{cls},\,i} \;+\; \lambda^{\text{loc}} r^{\text{loc},\,i}. \tag{1}$$

where $\lambda^{\text{fmt}}, \lambda^{\text{cls}}, \lambda^{\text{loc}}$ are tunable weights. To stabilize optimization, rewards are normalized within each group, yielding the main advantage:

$$\widehat{A}^i_{\text{main}} = \frac{r^i - \mu_r}{\sigma_r + \varepsilon}, \quad \mu_r = \frac{1}{G}\sum_{i=1}^{G} r^i, \quad \sigma_r^2 = \frac{1}{G}\sum_{i=1}^{G} \left(r^i - \mu_r\right)^2, \tag{2}$$

with the vectorized form $\widehat{A}_{\text{main}} = (\widehat{A}^1_{\text{main}}, \ldots, \widehat{A}^G_{\text{main}})^\top \in \mathbb{R}^G$ serving as the normalized baseline signal for subsequent policy updates. However, such outcome-aware advantages risk encouraging coarse, heuristic reasoning for time series data (e.g., detecting only obvious outliers while ignoring subtle but meaningful temporal patterns).

Time-Series Reasoning Advantage. To explicitly encourage fine-grained reasoning, TimerPO leverages the Optimal Transport (OT) (Villani et al., 2008; Li et al., 2024) to quantify the semantic alignment between a model's reasoning trace $\mathbf{y}^i = \{y^i_1, \ldots, y^i_{N^i}\}$ and the corresponding ExpCoT's reasoning trace $\mathbf{y}^\star = \{y^\star_1, \ldots, y^\star_M\}$ where N^i and M are their lengths. Given $\mathbf{y}^i$ and $\mathbf{y}^\star$, we extract

the final-layer embeddings from the MLLM π_θ, obtaining embedding vectors $\mathbf{e}^i$ for $\mathbf{y}^i$ and $\mathbf{e}^\star$ for $\mathbf{y}^\star$. We then construct a semantic cost matrix $\mathbf{C}^i \in \mathbb{R}^{N^i \times M}$ whose (n, m)-th entry measures the cosine distance between token embeddings:

$$C^i_{nm} = 1 - \frac{\mathbf{e}^i_n \cdot \mathbf{e}^\star_m}{\|\mathbf{e}^i_n\| \, \|\mathbf{e}^\star_m\|}, \quad n = 0, \ldots, N^i, \quad m = 0, \ldots, M. \tag{3}$$

Let $\mathbf{u}^i \in \Delta^{N^i-1}$ and $\mathbf{v} \in \Delta^{M-1}$ denote the marginal distributions over token positions for the model and the corresponding ExpCoT trace, obtained by normalizing their generation probabilities. The OT distance for response $\mathbf{y}^i$ is defined by

$$W^i = \min_{\mathbf{P}^i \in \Pi(\mathbf{u}^i, \mathbf{v})} \langle \mathbf{P}^i, \mathbf{C}^i \rangle_F, \quad \Pi(\mathbf{u}^i, \mathbf{v}) = \left\{ \mathbf{P}^i \geq 0 \;\middle|\; \mathbf{P}^i \mathbb{1}_M = \mathbf{u}^i, \; (\mathbf{P}^i)^\top \mathbb{1}_{N^i} = \mathbf{v} \right\}, \tag{4}$$

where $\langle \cdot, \cdot \rangle_F$ is the Frobenius product, and W^i measures the minimal semantic effort required to transform the model's reasoning distribution into the ExpCoT distribution. In practice, we approximate the solution of Equation (4) with the entropic-regularized Sinkhorn–Knopp (Cuturi, 2013) for efficiency and smoothness. Then, we use $r^i_{\mathrm{TsR}} = \exp(-W^i/\tau)$ as the reasoning reward and obtain the *time-series reasoning advantage*:

$$\widehat{A}^i_{\mathrm{TsR}} = \frac{r^i_{\mathrm{TsR}} - \mu_{\mathrm{TsR}}}{\sigma_{\mathrm{TsR}} + \varepsilon}, \quad \mu_{\mathrm{TsR}} = \frac{1}{G} \sum\nolimits_{i=1}^{G} r^i_{\mathrm{TsR}}, \quad \sigma^2_{\mathrm{TsR}} = \frac{1}{G} \sum\nolimits_{i=1}^{G} (r^i_{\mathrm{TsR}} - \mu_{\mathrm{TsR}})^2. \tag{5}$$

Collecting across the group $\mathcal{G}$ yields $\widehat{A}_{\mathrm{TsR}} = (\widehat{A}^1_{\mathrm{TsR}}, \ldots, \widehat{A}^G_{\mathrm{TsR}})^\top \in \mathbb{R}^G$ which serves as a relative measure of reasoning quality.

Orthogonal Integration for Policy Optimization. A naive combination of task and reasoning rewards risks interference, as ExpCoT guidance may overlap with the primary detection objective under shared ground truth supervision. To avoid this, TimerPO orthogonalizes the time-series grounded advantage with respect to the main advantage, retaining only the complementary part:

$$\widehat{A}^\perp_{\mathrm{TsR}} = \widehat{A}_{\mathrm{TsR}} - \frac{\langle \widehat{A}_{\mathrm{TsR}}, \, \widehat{A}_{\mathrm{main}} \rangle}{\|\widehat{A}_{\mathrm{main}}\|_2^2 + \varepsilon} \, \widehat{A}_{\mathrm{main}} \tag{6}$$

We then compose the final advantage for each response by

$$A^i_{\mathrm{final}} = \widehat{A}^i_{\mathrm{main}} + \alpha \left(\widehat{A}^\perp_{\mathrm{TsR}} \right)^i, \qquad i = 1, \ldots, G, \tag{7}$$

where α is a hyperparameter controlling the strength of the reasoning refinement. This composite advantage, A^i_{final}, then drives the policy update by replacing the standard normalized advantage in the clipped objective function:

$$\mathcal{L}(\theta) = \frac{1}{G} \sum_{i=1}^{G} \frac{1}{|\mathbf{y}^i|} \sum_{n=1}^{|\mathbf{y}^i|} \min\Big(\rho^i_n \, A^i_{\mathrm{final}}, \; \mathrm{clip}(\rho^i_n, \, 1-\epsilon, \, 1+\epsilon) \, A^i_{\mathrm{final}} \Big) - \beta \, \mathrm{KL}\big[\pi_\theta \,\|\, \pi_{\mathrm{ref}}\big], \tag{8}$$

where ρ^i_n is the importance ratio for the n-th token of response $\mathbf{y}^i$, and ϵ, β are the PPO clipping and KL weights, respectively. By operating at the advantage level, TimerPO offers a stable mechanism to instill ExpCoT reasoning, enhancing the model's analytical precision while keeping the primary detection update direction unchanged.

Overall. AnomSeer employs the pure RL training strategy to enhance MLLMs without SFT as a cold-start or any modifications to the model architecture. During training, we first construct ExpCoT using the analytical rigor of traditional TSAD methods, and subsequently refine the model's policy using orthogonalized time-series reasoning advantages through TimerPO. This simple yet effective integrated design efficiently instills expert knowledge into the pre-trained model within a single reinforcement learning phase. At inference time, AnomSeer operates in a fully end-to-end manner, requiring no external components or incurring any additional token overhead. As shown in Figure 3, the trained AnomSeer receives the question and produces outputs that include step-by-step analysis, anomaly type classification, and precise interval localization. Appendix B provides the pseudocode for the overall AnomSeer procedure.

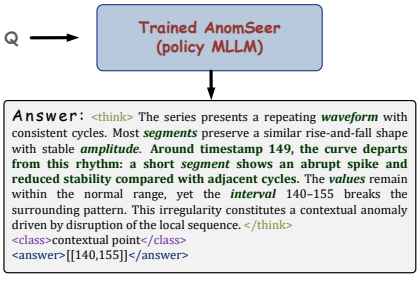


Figure 3: An example of TSAD reasoning produced by AnomSeer.

Table 1: Performance comparison on the AnomLLM test dataset. Results are reported as the mean and standard deviation over three runs for anomaly classification accuracy (%) and location detection accuracy metrics (%): Affinity-Precision (P), Affinity-Recall (R), and Affinity-F1 (F1).

Modality	Type	Method	Classification	Location												Avg F1
			Accuracy	Frequency			Trend			Range			Point			
				P	R	F1	P	R	F1	P	R	F1	P	R	F1	
Commercial																
Image+Text	Prompting	GPT-4o	$17.2_{\pm 1.1}$	$11.1_{\pm 0.2}$	$10.8_{\pm 0.2}$	$10.9_{\pm 0.2}$	$40.3_{\pm 0.4}$	$48.4_{\pm 0.1}$	$43.5_{\pm 0.2}$	$55.0_{\pm 0.5}$	$61.8_{\pm 0.5}$	$57.0_{\pm 0.5}$	$51.5_{\pm 0.2}$	$58.8_{\pm 0.2}$	$53.4_{\pm 0.2}$	41.2
Image+Text	Prompting	GPT-4o-mini	$17.8_{\pm 1.2}$	$10.3_{\pm 0.1}$	$10.1_{\pm 0.1}$	$10.2_{\pm 0.1}$	$19.4_{\pm 0.2}$	$29.4_{\pm 0.2}$	$23.2_{\pm 0.2}$	$48.0_{\pm 0.1}$	$58.6_{\pm 0.1}$	$51.3_{\pm 0.1}$	$51.3_{\pm 0.1}$	$58.2_{\pm 0.1}$	$52.7_{\pm 0.1}$	34.4
Image+Text	Prompting	Gemini-2.5-Flash	$10.0_{\pm 0.5}$	$21.4_{\pm 0.1}$	$16.6_{\pm 0.1}$	$17.9_{\pm 0.1}$	$34.6_{\pm 0.1}$	$36.0_{\pm 0.1}$	$35.2_{\pm 0.1}$	$76.0_{\pm 0.1}$	$78.9_{\pm 0.1}$	$76.7_{\pm 0.1}$	$76.4_{\pm 0.5}$	$74.9_{\pm 0.5}$	$74.9_{\pm 0.5}$	51.2
Image+Text	Prompting	Gemini-2.5-Pro	$12.6_{\pm 0.1}$	$17.4_{\pm 0.1}$	$22.0_{\pm 0.1}$	$19.1_{\pm 0.3}$	$58.8_{\pm 0.5}$	$60.0_{\pm 0.1}$	$59.0_{\pm 0.2}$	$79.4_{\pm 0.1}$	$83.2_{\pm 0.5}$	$81.3_{\pm 0.4}$	$76.1_{\pm 0.5}$	$74.4_{\pm 0.4}$	$74.5_{\pm 0.5}$	58.5
Numerical+Text	Prompting	SigLLM (GPT-3.5)	\	$16.9_{\pm 0.5}$	$14.9_{\pm 0.5}$	$15.8_{\pm 0.5}$	$20.3_{\pm 0.5}$	$20.5_{\pm 0.5}$	$19.6_{\pm 0.1}$	$67.8_{\pm 0.1}$	$67.5_{\pm 0.1}$	$67.7_{\pm 0.1}$	$34.4_{\pm 0.1}$	$38.6_{\pm 0.1}$	$36.2_{\pm 0.1}$	34.8
Open-source																
Image+Text	Prompting	Qwen2.5-VL-72B-Instruct	$14.6_{\pm 0.5}$	$40.2_{\pm 0.1}$	$28.3_{\pm 0.1}$	$31.4_{\pm 0.1}$	$30.8_{\pm 0.2}$	$33.8_{\pm 0.1}$	$32.1_{\pm 0.7}$	$76.8_{\pm 0.4}$	$73.9_{\pm 0.1}$	$74.6_{\pm 0.1}$	$63.2_{\pm 0.5}$	$64.6_{\pm 0.3}$	$62.7_{\pm 0.1}$	50.2
Image+Text	Prompting	Qwen2.5-VL-32B-Instruct	$10.2_{\pm 0.5}$	$19.3_{\pm 0.4}$	$20.2_{\pm 0.5}$	$18.9_{\pm 0.2}$	$34.3_{\pm 0.5}$	$37.2_{\pm 0.2}$	$35.5_{\pm 0.5}$	$72.0_{\pm 0.5}$	$70.3_{\pm 0.5}$	$70.7_{\pm 0.5}$	$55.9_{\pm 0.5}$	$59.3_{\pm 0.5}$	$56.7_{\pm 0.5}$	45.5
Image+Text	Prompting	Qwen2.5-VL-7B-Instruct	$25.3_{\pm 0.2}$	$18.5_{\pm 0.4}$	$16.4_{\pm 0.1}$	$16.8_{\pm 0.1}$	$52.7_{\pm 0.1}$	$53.8_{\pm 0.2}$	$53.1_{\pm 0.3}$	$48.6_{\pm 0.4}$	$45.4_{\pm 0.1}$	$46.4_{\pm 0.1}$	$61.9_{\pm 0.1}$	$58.6_{\pm 0.1}$	$59.0_{\pm 0.1}$	43.8
Image+Text	Prompting	Qwen2.5-VL-3B-Instruct	$11.4_{\pm 0.1}$	$7.1_{\pm 0.1}$	$9.3_{\pm 0.2}$	$7.9_{\pm 0.4}$	$17.4_{\pm 0.5}$	$22.0_{\pm 0.5}$	$19.1_{\pm 0.1}$	$29.7_{\pm 0.1}$	$31.3_{\pm 0.1}$	$29.7_{\pm 0.1}$	$34.3_{\pm 0.5}$	$40.6_{\pm 0.5}$	$36.2_{\pm 0.1}$	23.2
Image+Text	Training	Qwen2.5-VL-3B-SFT3.2K	$29.7_{\pm 0.1}$	$19.0_{\pm 0.1}$	$24.4_{\pm 0.5}$	$19.7_{\pm 0.1}$	$29.7_{\pm 0.2}$	$34.8_{\pm 0.1}$	$30.2_{\pm 0.2}$	$40.1_{\pm 0.1}$	$48.1_{\pm 0.4}$	$40.8_{\pm 0.1}$	$40.4_{\pm 0.5}$	$49.3_{\pm 0.1}$	$40.6_{\pm 0.2}$	32.8
Image+Text	Training	Qwen2.5-VL-3B-SFT32K	$35.6_{\pm 0.5}$	$12.0_{\pm 0.5}$	$14.3_{\pm 0.1}$	$12.4_{\pm 0.1}$	$57.6_{\pm 0.5}$	$57.5_{\pm 0.2}$	$57.4_{\pm 0.5}$	$40.4_{\pm 0.1}$	$52.1_{\pm 0.1}$	$41.3_{\pm 0.2}$	$44.3_{\pm 0.5}$	$58.3_{\pm 0.2}$	$46.3_{\pm 0.5}$	39.4
Image+Text	Training	TimeMaster-3B	$57.9_{\pm 0.6}$	$57.3_{\pm 0.5}$	$50.3_{\pm 0.1}$	$51.4_{\pm 0.2}$	$76.0_{\pm 0.5}$	$77.3_{\pm 0.1}$	$76.6_{\pm 0.5}$	$77.8_{\pm 0.5}$	$83.5_{\pm 0.1}$	$80.1_{\pm 0.1}$	$77.7_{\pm 0.5}$	$82.1_{\pm 0.1}$	$79.6_{\pm 0.5}$	71.9
Image+Text	Training	AnomSeer-3B (Ours)	$62.8_{\pm 0.5}$	$63.7_{\pm 0.5}$	$58.4_{\pm 0.5}$	$58.9_{\pm 0.5}$	$84.2_{\pm 0.2}$	$85.9_{\pm 0.1}$	$84.9_{\pm 0.1}$	$83.3_{\pm 0.3}$	$89.2_{\pm 0.1}$	$85.6_{\pm 0.1}$	$86.0_{\pm 0.1}$	$90.3_{\pm 0.1}$	$87.8_{\pm 0.1}$	79.3
Image+Text	Training	AnomSeer-7B (Ours)	$65.0_{\pm 0.5}$	$68.3_{\pm 0.5}$	$59.4_{\pm 0.2}$	$60.8_{\pm 0.2}$	$86.6_{\pm 0.5}$	$89.0_{\pm 0.5}$	$87.7_{\pm 0.5}$	$91.6_{\pm 0.1}$	$97.8_{\pm 0.4}$	$94.3_{\pm 0.1}$	$93.4_{\pm 0.1}$	$96.9_{\pm 0.5}$	$94.9_{\pm 0.1}$	84.4

5 Experiments

Benchmarks. To evaluate the performance and generalization ability of AnomSeer, we consider three diverse TSAD benchmarks: (**1**) *AnomLLM* (Zhou & Yu, 2024), a synthetic dataset containing frequency, trend, out-of-range and point anomalies[1]; (**2**) *VisualTimeAnomaly* (Xu et al., 2025), a mixed synthetic–real, image-based benchmark covering a broader spectrum of anomaly types[2]; and (**3**) *TSB-UAD* (Paparrizos et al., 2022; Qiu et al., 2025), a real-world univariate collection from domains such as ECG and web traffic, with diverse anomaly types, ratios, and sequence lengths. Training is conducted solely on the synthetic AnomLLM benchmark (3,200 instances), ensuring clean, high-fidelity ExpCoT supervision. Evaluation is then performed on the test sets of AnomLLM, the mixed real-world VisualTimeAnomaly, and TSB-UAD, providing a rigorous test of generalization to diverse, previously unseen anomalies.

Baselines. We compare against both commercial (GPT-4o, GPT-4o-mini, Gemini-2.5-Pro, Gemini-2.5-Flash) and open-source MLLMs (Qwen2.5-VL-72B/32B/7B/3B-Instruct), as well as two representative LLM-based temporal reasoning baselines: *SigLLM* (GPT-3.5-based) (Alnegheimish et al., 2024) and *TimeMaster* (Qwen2.5-VL-3B-based, trained with SFT and GRPO) (Zhang et al., 2025b). We further compare against SFT baselines: Qwen2.5-VL-3B-SFT3.2k, fine-tuned on 3,200 instances, and Qwen2.5-VL-3B-SFT32k, fine-tuned on 32,000 instances.

Metrics. We report both anomaly-type classification accuracy and label-based metrics for localization performance, including Affinity-Precision (P), Affinity-Recall (R), and Affinity-F1 (F1), following the definitions in Huet et al. (2022). These metrics are suitable because LLMs generate discrete anomalous intervals, which can be converted into binary labels rather than continuous scores, and they better capture the temporal consistency of anomaly detection (Zhou & Yu, 2024; Xu et al., 2025).

Hyperparameters. We build AnomSeer on Qwen2.5-VL-3B/7B-Instruct (Bai et al., 2025). Following Zhang et al. (2025b), we set the group size $G = 5$ and the PPO clipping $\epsilon = 0.2$. The reward weights are empirically chosen as $\lambda^{\text{fmt}} = 0.1$, $\lambda^{\text{cls}} = 0.2$, and $\lambda^{\text{loc}} = 0.7$. TimerPO's reasoning advantage weight is fixed at $\alpha = 0.3$. More experimental details are provided in Appendix D.

5.1 Main Results

As shown in Table 1, AnomSeer consistently achieves state-of-the-art results across all anomaly detection tasks on the AnomLLM benchmark. Remarkably, even at a lightweight 3B scale, our model substantially outperforms much larger and more resource-intensive MLLMs such as GPT-4o and Gemini-2.5-Pro in both anomaly type classification and Affinity-F1 metrics, and its performance

[1] In AnomLLM, *frequency*, *trend*, and contextual *point* anomalies are harder as they require contextual awareness, while *range* anomalies are easier since they can be detected from obvious global point deviations.

[2] In VisualTimeAnomaly, *range-wise* anomalies (shapelet, seasonal, and trend) are generally easier, while *point-wise contextual* and *global* anomalies, which manifest as subtle and dispersed single points, are harder.

further improves with the 7B variant. We also observe that simply increasing the amount of SFT data yields only marginal gains, even with $10\times$ more SFT data (32k instances), performance still falls short of AnomSeer. One possible reason is that SFT emphasizes only positive reasoning paths while neglecting negative ones, leading the model to develop only a shallow understanding rather than genuinely learning. Notably, for numerically subtle anomalies such as frequency shifts, AnomSeer maintains a clear advantage, whereas GRPO-trained MLLMs like TimeMaster continue to lag behind. This result suggests that globally verifiable RL objectives alone are insufficient for modeling fine-grained temporal variations, whereas our AnomSeer explicitly encourages fine-grained temporal reasoning that leads to more accurate anomaly detection.

5.2 Ablation Study and Hyperparameter Analysis

We next conduct a detailed ablation study together with a hyperparameter sensitivity analysis. Table 2 provides several key takeaways. First, we replace ExpCoT with CoT generated by GPT-4o, which leads to a marked degradation, particularly on challenging frequency anomalies. This demonstrates that generic CoT supervision imparts mere surface-level fluency rather than in-depth temporal reasoning. It further highlights a crucial insight: the analytical rigor of classical methods is not obsolete, but rather a valuable resource for shaping the next generation of truly capable time-series MLLMs. Second, removing the orthogonalization mechanism causes a moderate drop in performance, underscoring its crucial role in mitigating spurious correlations between reasoning quality and task success. Third, eliminating all components reduces the method to a vanilla GRPO setup and yields the worst average performance, confirming that outcome-based rewards alone are insufficient to foster the fine-grained anomaly detection skills required for complex TSAD.

Table 2: Ablation study on different components of AnomSeer-3B using Affinity F1 score (%).

Components			Anomaly Scenarios			
ExpCoT	$\widehat{A}^{\perp}_{\mathrm{TsR}}$	Orth	Frequency	Trend	Range	Point
✗	✓	✓	49.8	79.5	84.4	86.1
✓	✓	✗	53.5	81.1	83.5	85.4
✗	✗	✗	50.4	77.8	81.8	80.6
✓	✓	✓	**58.9**	**84.9**	**85.6**	**87.8**

Figure 4 presents the effect of varying the temporal reasoning weight α in our TimerPO objective. Across all anomaly types, AnomSeer consistently maintains a substantial margin over the GPT-4o baseline (grey dashed line), showing that even under suboptimal α values, the integration of structured temporal reasoning signals offers clear benefits. The model remains relatively robust within the range $\alpha \in [0.3, 0.7]$, where performance is stable and near-optimal for frequency, trend, range, and point anomalies alike. This highlights the importance of balancing outcome-level and reasoning-level rewards: too small a weight diminishes the impact of explicit reasoning supervision, while too large a weight can overshadow task-level alignment, leading to slight degradation. In practice, $\alpha = 0.3$ works well as a default, though careful dataset-specific tuning may yield further gains.

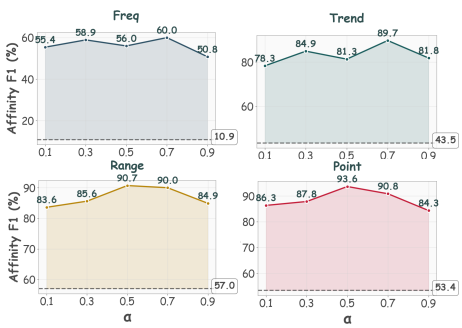
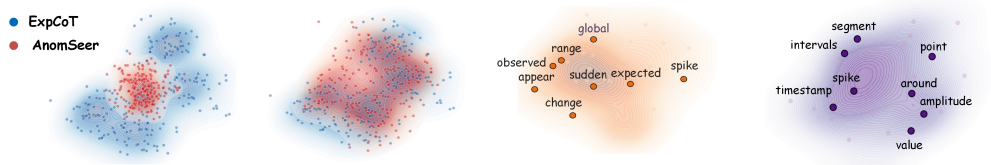


Figure 4: Hyperparameter sensitivity analysis on α, comparing our method with the GPT-4o baseline (grey dashed line).

5.3 Effect of TimerPO on Reasoning Pattern

To show that AnomSeer enables time-series MLLMs reasoning grounded in fine-grained statistics, we analyze the effect of TimerPO on distributional alignment and linguistic usage before and after RL training, as shown in Figure 5. Panels (a)-(b) illustrate that, prior to TimerPO, ExpCoT (blue) and AnomSeer outputs (red) occupy noticeably divergent regions in the representation space, with the latter exhibiting a relatively narrow distribution. This mismatch highlights that the model's reasoning is overly global and lacks diversity. A similar trend is observed in token usage. In the pre-training stage (c), top words are generic and coarse-grained (e.g., global, sudden, change), reflecting surface-level anomaly descriptions. After TimerPO (d), the vocabulary shifts toward finer-grained and temporally grounded tokens (e.g., timestamp, intervals, amplitude), which better capture struc-



(a) Pre-training Alignment (b) TimerPO Alignment (c) Pre-training Top Words (d) TimerPO Top Words

Figure 5: Comparison of distribution alignment between ExpCoT (blue) and AnomSeer (red) outputs, as well as token usage before and after applying TimerPO.

tured reasoning over time. Therefore, these results demonstrate that TimerPO not only improves distributional alignment with expert reasoning but also enriches the semantic granularity of reasoning traces, moving from broad anomaly descriptors to precise temporal markers. We also compare GRPO and our TimerPO-trained models in Appendix E.6, which further confirms the effectiveness of our method in enhancing temporal reasoning.

5.4 Generalization Performance

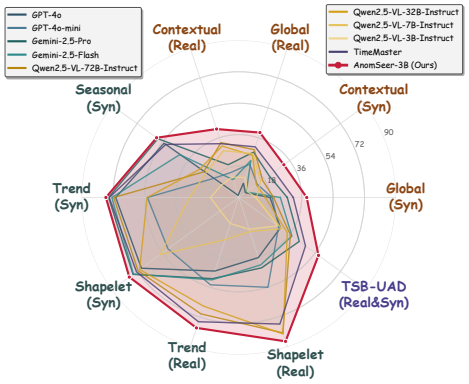


Figure 6: Comparison of model generalization performance (Affinity F1%) across point-wise tasks, range-wise tasks, and the real-world TSB-UAD benchmark.

At last, we evaluate the generalization ability of AnomSeer. We test the model (trained on the synthetic AnomLLM) on two distinct and more challenging benchmarks: VisualTimeAnomaly (a hybrid synthetic-real dataset with richer anomaly types) and TSB-UAD (a real-world univariate collection). Importantly, shapelet anomalies represent a completely new category absent during training. Despite this, as shown in Figure 6, our method demonstrates strong accuracy on such cases. This ability to detect and explain shapelet anomalies shows that the model is not restricted to pattern memorization, but can generalize to qualitatively novel anomaly behaviors. Moreover, on point-wise contextual anomalies, a notably harder task that requires fine-grained discrimination, AnomSeer delivers clear gains over baseline MLLMs, underscoring its ability to move beyond surface-level visual cues. Finally, on the TSB-UAD collection of real-world datasets, which spans diverse domains, AnomSeer sustains its advantage and confirms that the improvements extend beyond synthetic benchmarks to practical anomaly detection scenarios. Overall, these results verify that our approach achieves not only high in-domain accuracy but also robust generalization to unseen and real-world anomalies.

6 Conclusions and Limitations

In this paper, we introduced AnomSeer, an RL post-training method that enables multimodal LLMs to detect and reason about time-series anomalies in a fine-grained and accurate manner. By grounding MLLMs' reasoning in the fine-grained, multi-dimensional evidence of classical TSAD, AnomSeer attains state-of-the-art performance across diverse benchmarks. Beyond surpassing strong baselines such as GPT-4o in detection accuracy and localization, it delivers verifiable, detailed time-series explanations, elevating MLLMs from coarse visual heuristics to principled, testable analysis. Nevertheless, AnomSeer was developed primarily on univariate time-series data in TSAD, and extending it to more complex multivariate scenarios remains an open direction. A potential solution is to reframe each variable as an image-like subrepresentation and then reason over its joint structure, enabling the model to capture both localized temporal patterns and cross-variable dependencies in a coherent manner. Another direction may be to explore how to incorporate external knowledge to better account for real-world events that drive anomaly dynamics.

Ethics Statement

This paper has no potential ethics issues.

Reproducibility Statement

To facilitate the reproducibility of our results, we provide the anonymized code at the repository link at https://anonymous.4open.science/r/AnomSeer_sub-C286, along with detailed hyperparameter settings in Appendix D.4. All experimental data are obtained from publicly available download sources, as described in Appendix D.1.

References

Sarah Alnegheimish, Linh Nguyen, Laure Berti-Equille, and Kalyan Veeramachaneni. Large language models can be zero-shot anomaly detectors for time series? *arXiv preprint arXiv:2405.14755*, 2024.

Arip Asadulaev, Rostislav Korst, Aleksandr Korotin, Vage Egiazarian, Andrey Filchenkov, and Evgeny Burnaev. Rethinking optimal transport in offline reinforcement learning. *Advances in Neural Information Processing Systems*, 37:123592–123607, 2024.

Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, et al. Qwen2. 5-vl technical report. *arXiv preprint arXiv:2502.13923*, 2025.

Aadyot Bhatnagar, Paul Kassianik, Chenghao Liu, Tian Lan, Wenzhuo Yang, Rowan Cassius, Doyen Sahoo, Devansh Arpit, Sri Subramanian, Gerald Woo, Amrita Saha, Arun Kumar Jagota, Gokulakrishnan Gopalakrishnan, Manpreet Singh, K C Krithika, Sukumar Maddineni, Daeki Cho, Bo Zong, Yingbo Zhou, Caiming Xiong, Silvio Savarese, Steven Hoi, and Huan Wang. Merlion: A machine learning library for time series. 2021.

Nicolas Bonneel, Michiel Van De Panne, Sylvain Paris, and Wolfgang Heidrich. Displacement interpolation using lagrangian mass transport. In *Proceedings of the 2011 SIGGRAPH Asia conference*, pp. 1–12, 2011.

Luis A Caffarelli and Robert J McCann. Free boundaries in optimal transport and monge-ampere obstacle problems. *Annals of mathematics*, pp. 673–730, 2010.

Feiyi Chen, Leilei Zhang, Guansong Pang, Roger Zimmermann, and Shuiguang Deng. Synergizing large language models and task-specific models for time series anomaly detection. *arXiv preprint arXiv:2501.05675*, 2025.

Liqun Chen, Ke Bai, Chenyang Tao, Yizhe Zhang, Guoyin Wang, Wenlin Wang, Ricardo Henao, and Lawrence Carin. Sequence generation with optimal-transport-enhanced reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 7512–7520, 2020.

Xu Chu, Zhixin Zhang, Tianyu Jia, and Yujie Jin. Stackelberg self-annotation: A robust approach to data-efficient llm alignment. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.

Jean-Antoine Désidéri. Multiple-gradient descent algorithm (mgda) for multiobjective optimization. *Comptes Rendus Mathematique*, 350(5-6):313–318, 2012.

Lang Feng, Zhenghai Xue, Tingcong Liu, and Bo An. Group-in-group policy optimization for llm agent training. *arXiv preprint arXiv:2505.10978*, 2025.

Shanghua Gao, Teddy Koker, Owen Queen, Tom Hartvigsen, Theodoros Tsiligkaridis, and Marinka Zitnik. Units: A unified multi-task time series model. *Advances in Neural Information Processing Systems*, 37:140589–140631, 2024.

Markus Goldstein and Andreas Dengel. Histogram-based outlier score (hbos): A fast unsupervised anomaly detection algorithm. *KI-2012: poster and demo track*, 1:59–63, 2012.

Mononito Goswami, Konrad Szafer, Arjun Choudhry, Yifu Cai, Shuo Li, and Artur Dubrawski. Moment: A family of open time-series foundation models. *arXiv preprint arXiv:2402.03885*, 2024.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

Zelin He, Sarah Alnegheimish, and Matthew Reimherr. Harnessing vision-language models for time series anomaly detection. *arXiv preprint arXiv:2506.06836*, 2025.

Alexis Huet, Jose Manuel Navarro, and Dario Rossi. Local evaluation of time series anomaly detection algorithms. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 635–645, 2022.

Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.

Pascal Klink, Haoyi Yang, Carlo D'Eramo, Jan Peters, and Joni Pajarinen. Curriculum reinforcement learning via constrained optimal transport. In *International Conference on Machine Learning*, pp. 11341–11358. PMLR, 2022.

Yaxuan Kong, Yiyuan Yang, Shiyu Wang, Chenghao Liu, Yuxuan Liang, Ming Jin, Stefan Zohren, Dan Pei, Yan Liu, and Qingsong Wen. Position: Empowering time series reasoning with multimodal llms. *arXiv preprint arXiv:2502.01477*, 2025.

Meng Li, Guangda Huzhang, Haibo Zhang, Xiting Wang, and Anxiang Zeng. Optimal transport-based token weighting scheme for enhanced preference optimization. *arXiv preprint arXiv:2505.18720*, 2025a.

Xuhong Li, Jiamin Chen, Yekun Chai, and Haoyi Xiong. Gilot: Interpreting generative language models via optimal transport. In *Forty-first International Conference on Machine Learning*, 2024.

Zhuo Li, Yuege Feng, Dandan Guo, Jinpeng Hu, Anningzhe Gao, and Xiang Wan. Aplot: Robust reward modeling via adaptive preference learning with optimal transport. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 5524–5538, 2025b.

Bo Liu, Xingchao Liu, Xiaojie Jin, Peter Stone, and Qiang Liu. Conflict-averse gradient descent for multi-task learning. *Advances in Neural Information Processing Systems*, 34:18878–18890, 2021.

Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *2008 eighth ieee international conference on data mining*, pp. 413–422. IEEE, 2008.

Haoxin Liu, Chenghao Liu, and B Aditya Prakash. A picture is worth a thousand numbers: Enabling llms reason about time series via visualization. *arXiv preprint arXiv:2411.06018*, 2024.

Jun Liu, Chaoyun Zhang, Jiaxu Qian, Minghua Ma, Si Qin, Chetan Bansal, Qingwei Lin, Saravan Rajmohan, and Dongmei Zhang. Large language models can deliver accurate and interpretable time series anomaly detection. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pp. 4623–4634, 2025a.

Zijia Liu, Peixuan Han, Haofei Yu, Haoru Li, and Jiaxuan You. Time-r1: Towards comprehensive temporal reasoning in llms. *arXiv preprint arXiv:2505.13508*, 2025b.

Yucong Luo, Yitong Zhou, Mingyue Cheng, Jiahao Wang, Daoyu Wang, Tingyue Pan, and Jintao Zhang. Time series forecasting as reasoning: A slow-thinking approach with reinforced llms. *arXiv preprint arXiv:2506.10630*, 2025.

Igor Melnyk, Youssef Mroueh, Brian Belgodere, Mattia Rigotti, Apoorva Nitsure, Mikhail Yurochkin, Kristjan Greenewald, Jiri Navratil, and Jarret Ross. Distributional preference alignment of llms via optimal transport. *Advances in Neural Information Processing Systems*, 37: 104412–104442, 2024.

John Paparrizos, Yuhao Kang, Paul Boniol, Ruey S Tsay, Themis Palpanas, and Michael J Franklin. Tsb-uad: an end-to-end benchmark suite for univariate time-series anomaly detection. *Proceedings of the VLDB Endowment*, 15(8):1697–1711, 2022.

Daehyung Park, Yuuna Hoshi, and Charles C Kemp. A multimodal anomaly detector for robot-assisted feeding using an lstm-based variational autoencoder. *IEEE Robotics and Automation Letters*, 3(3):1544–1551, 2018.

Xiangfei Qiu, Zhe Li, Wanghui Qiu, Shiyan Hu, Lekui Zhou, Xingjian Wu, Zhengyu Li, Chenjuan Guo, Aoying Zhou, Zhenli Sheng, et al. Tab: Unified benchmarking of time series anomaly detection methods. *arXiv preprint arXiv:2506.18046*, 2025.

Hansheng Ren, Bixiong Xu, Yujing Wang, Chao Yi, Congrui Huang, Xiaoyu Kou, Tony Xing, Mao Yang, Jie Tong, and Qi Zhang. 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.

Bernhard Schölkopf, Robert C Williamson, Alex Smola, John Shawe-Taylor, and John Platt. Support vector method for novelty detection. *Advances in neural information processing systems*, 12, 1999.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Y Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

Qichao Shentu, Beibu Li, Kai Zhao, Yang Shu, Zhongwen Rao, Lujia Pan, Bin Yang, and Chenjuan Guo. Towards a general time series anomaly detector with adaptive bottlenecks and dual adversarial decoders. *arXiv preprint arXiv:2405.15273*, 2024.

Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

Mingtian Tan, Mike A Merrill, Zack Gottesman, Tim Althoff, David Evans, and Tom Hartvigsen. Inferring events from time series using language models. *arXiv preprint arXiv:2503.14190*, 2025.

Markus Thill, Wolfgang Konen, and Thomas Bäck. Time series anomaly detection with discrete wavelet transforms and maximum likelihood estimation. In *Intern. Conference on Time Series (ITISE)*, volume 2, pp. 11–23, 2017.

Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2008.

Yake Wei and Di Hu. Mmpareto: Boosting multimodal learning with innocent unimodal assistance. *arXiv preprint arXiv:2405.17730*, 2024.

Yuxiang Wei, Olivier Duchenne, Jade Copet, Quentin Carbonneaux, Lingming Zhang, Daniel Fried, Gabriel Synnaeve, Rishabh Singh, and Sida I Wang. Swe-rl: Advancing llm reasoning via reinforcement learning on open software evolution. *arXiv preprint arXiv:2502.18449*, 2025.

Xingjian Wu, Xiangfei Qiu, Zhengyu Li, Yihang Wang, Jilin Hu, Chenjuan Guo, Hui Xiong, and Bin Yang. CATCH: Channel-aware multivariate time series anomaly detection via frequency patching. In *ICLR*, 2025.

Zhe Xie, Zeyan Li, Xiao He, Longlong Xu, Xidao Wen, Tieying Zhang, Jianjun Chen, Rui Shi, and Dan Pei. Chatts: Aligning time series with llms via synthetic data for enhanced understanding and reasoning. *arXiv preprint arXiv:2412.03104*, 2024.

Jiehui Xu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Anomaly transformer: Time series anomaly detection with association discrepancy. *arXiv preprint arXiv:2110.02642*, 2021.

Xiongxiao Xu, Haoran Wang, Yueqing Liang, Philip S Yu, Yue Zhao, and Kai Shu. Can multimodal llms perform time series anomaly detection? *arXiv preprint arXiv:2502.17812*, 2025.

Yiyuan Yang, Zichuan Liu, Lei Song, Kai Ying, Zhiguang Wang, Tom Bamford, Svitlana Vyetrenko, Jiang Bian, and Qingsong Wen. Time-ra: Towards time series reasoning for anomaly with llm feedback. *arXiv preprint arXiv:2507.15066*, 2025.

Chin-Chia Michael Yeh, Yan Zhu, Liudmila Ulanova, Nurjahan Begum, Yifei Ding, Hoang Anh Dau, Diego Furtado Silva, Abdullah Mueen, and Eamonn Keogh. Matrix profile i: all pairs similarity joins for time series: a unifying view that includes motifs, discords and shapelets. In *2016 IEEE 16th international conference on data mining (ICDM)*, pp. 1317–1322. Ieee, 2016.

Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. *Advances in neural information processing systems*, 33: 5824–5836, 2020.

Haoran Zhang, Yong Liu, Yunzhong Qiu, Haixuan Liu, Zhongyi Pei, Jianmin Wang, and Mingsheng Long. Timesbert: A bert-style foundation model for time series understanding. *arXiv preprint arXiv:2502.21245*, 2025a.

Junru Zhang, Lang Feng, Xu Guo, Yuhan Wu, Yabo Dong, and Duanqing Xu. Timemaster: Training time-series multimodal llms to reason via reinforcement learning. *arXiv preprint arXiv:2506.13705*, 2025b.

Tian Zhou, Peisong Niu, Liang Sun, Rong Jin, et al. One fits all: Power general time series analysis by pretrained lm. *Advances in neural information processing systems*, 36:43322–43355, 2023.

Zihao Zhou and Rose Yu. Can llms understand time series anomalies? *arXiv preprint arXiv:2410.05440*, 2024.

Jiaxin Zhuang, Leon Yan, Zhenwei Zhang, Ruiqi Wang, Jiawei Zhang, and Yuantao Gu. See it, think it, sorted: Large multimodal models are few-shot time series anomaly analyzers. *arXiv preprint arXiv:2411.02465*, 2024.

Bo Zong, Qi Song, Martin Renqiang Min, Wei Cheng, Cristian Lumezanu, Daeki Cho, and Haifeng Chen. Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In *ICLR*, 2018.

A THE USE OF LLMS

The core methodology of this paper is fully original and developed by the authors without the use of LLMs.

B PSEUDO CODE

The training pipeline of ANOMSEER is provided as follows:

Algorithm 1 Training Time-Series MLLMs with ANOMSEER

Require: Initial policy $\pi_{\theta_{\text{old}}}$, task distribution $p(\mathbf{X})$, discount factor γ, clipping parameter ϵ, KL penalty β, group size G, ExpCoT generator, TimerPO hyperparameter α
for each training iteration **do**
  Update old policy: $\theta_{\text{old}} \leftarrow \theta$
  // Data preparation phase
  Sample time-series $\mathbf{X} \sim p(\mathbf{X})$ and render visualization I
  Generate expert chain-of-thought $\mathbf{y}^\star \leftarrow \text{ExpCoT}(\mathbf{X})$
  Construct input $(\mathbf{I}, \mathbf{c})$
  // Advantage computation
  Sample group of responses $\mathcal{G} = \{\mathbf{y}^i \sim \pi_{\theta_{\text{old}}}(\cdot|\mathbf{I}, \mathbf{c})\}_{i=1}^{G}$
  for each $\mathbf{y}^i \in \mathcal{G}$ **do**
    Compute outcome reward: $r^i = \lambda^{\text{fmt}} r^{\text{fmt},\,i} + \lambda^{\text{cls}} r^{\text{cls},\,i} + \lambda^{\text{loc}} r^{\text{loc},\,i}$
    Normalize to obtain outcome-aware advantage $\widehat{A}^i_{\text{main}}$ via Eq. (2)
    Compute semantic OT distance W^i between $\mathbf{y}^i$ and $\mathbf{y}^\star$ via Eq. (4)
    Derive reasoning reward $r_i^{\text{TsR}} = \exp(-W^i/\tau)$ and normalize to $\widehat{A}^i_{\text{TsR}}$ via Eq. (5)
  end for
  // Orthogonal integration of advantages
  Compute orthogonalized reasoning advantage:

$$\widehat{A}^{\perp}_{\text{TsR}} = \widehat{A}_{\text{TsR}} - \frac{\langle \widehat{A}_{\text{TsR}},\, \widehat{A}_{\text{main}} \rangle}{\|\widehat{A}_{\text{main}}\|_2^2 + \varepsilon}\, \widehat{A}_{\text{main}}$$

  Final advantage: $A^i_{\text{final}} = \widehat{A}^i_{\text{main}} + \alpha \left(\widehat{A}^{\perp}_{\text{TsR}}\right)^i$
  // Policy update
  Update θ by maximizing the TimerPO objective:

$$\mathcal{L}(\theta) = \frac{1}{G}\sum_{i=1}^{G} \frac{1}{|\mathbf{y}^i|} \sum_{n=1}^{|\mathbf{y}^i|} \min\Big(\rho_n^i\, A^i_{\text{final}},\ \text{clip}(\rho_n^i,\, 1-\epsilon,\, 1+\epsilon)\, A^i_{\text{final}}\Big) - \beta\, \text{KL}\big[\pi_\theta \,\|\, \pi_{\text{ref}}\big],$$

end for

Algorithm 2 Inference with ANOMSEER

Require: Trained policy π_θ, input time series $\mathbf{X}$, instruction prompt $\mathbf{c}$
Render visualization: $\mathbf{I} \leftarrow \mathcal{R}(\mathbf{X})$
Construct model input: $(\mathbf{I}, \mathbf{c})$
// Forward inference
Generate model response: $\mathbf{y} \sim \pi_\theta(\cdot \mid \mathbf{I}, \mathbf{c})$
Get output $\mathbf{y}$ including anomaly type, location, and reasoning
return anomaly prediction results

C MORE DETAILS OF ANOMSEER

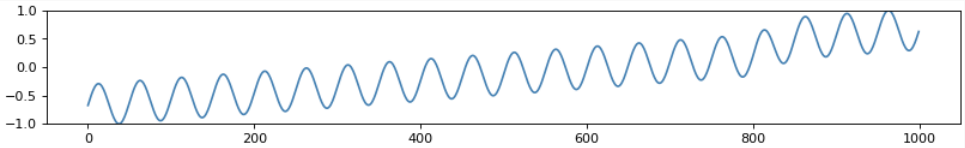
C.1 STRUCTURED OUTPUT FOR REASONING.

A key objective of ANOMSEER is to elicit *textual reasoning* that illuminates the model's analysis process. To achieve this, we enforce a structured output format to decouple the reasoning steps from

the final prediction. The model is prompted to first articulate its analytical process within <think> </think> tags, provide the predicted anomaly category (e.g., trend, global, contextual) within <class> </class> tags, and present the specific anomalous interval(s) within <answer> </answer> tags. This structured prompting strategy bridges low-level visual cues with high-level, human-interpretable reasoning in a unified framework. To illustrate this design, we present our full TSAD prompt in Fig. 7.

Prompt for Anomaly Detection (Image Input)

Input Image:



Prompt: You are a time series analysis expert. A time series plot of length 1000 is provided. Your task is to identify anomalous intervals along the x-axis.
The plot shows temporal patterns and typical value ranges based on both the x and y axes and may exhibit one type of anomaly from the set ["contextual point", "global point", "seasonal", "trend", "shapelet"]. Carefully examine the image to detect intervals that deviate from the expected behavior and infer the anomaly type.
Begin your detailed reasoning inside a <think>...</think> block. In this block, analyze the plot, identify suspicious intervals, and explain why they are considered anomalies and the type.
After reasoning, summarize your findings in the following format inside an <answer>...</answer> block: Each anomaly should be written as a [start, end] pair using x-axis positions. Then choose the type inside <class>...</class> from the set ["contextual point", "global point", "seasonal", "trend", "shapelet"].
Examples: <answer> [[120, 150], [320, 350]] </answer> <class>trend</class> or <answer> [[120, 150]] </answer> <class>global point</class> If no anomalies are found, return: <answer>[]</answer> <class>normal</class>

Figure 7: Prompt definition for time-series anomaly detection

C.2 Details on ExpCoT.

We adopt the common anomaly taxonomy (Qiu et al., 2025) with five categories: (i) **Out-of-Range / Global Point**, (ii) **Contextual Point**, (iii) **Trend Shift**, (iv) **Seasonal/Frequency Deviation**, and (v) **Shapelet/Subsequence**. For each category, we pair characteristic signatures with classical, quantitatively verifiable analyses. ExpCoT is instantiated *per instance* from the ground-truth (GT) anomaly type and temporal annotation, and follows a disciplined three-stage path: Observation → Reasoning & Validation → Conclusion. In *Observation*, we perform a unified hierarchical scan of the series: starting with global distributions (e.g., extreme values), then examining structural properties (e.g., trend and periodicity), and finally analyzing localized patterns (e.g., subsequence dissimilarity) to surface candidate anomalies. *Reasoning & Validation* aligns the GT type and location with a targeted statistical probe and reports the resulting numerical evidence. *Conclusion* integrates these findings into a precise, GT-consistent statement of anomaly type and localization. Figures 9–11 illustrate some cases, and we provide the detailed pipeline for each anomaly type below.

(i) Out-of-Range / Global Point. Observation: Apply the defined global–structural–local scan to find salient deviations as candidates for anomaly detection. Reasoning & Validation: Apply a k-sigma envelope $[\mu - k\sigma, \mu + k\sigma]$ to formalize range departures; aggregate excursions into contiguous intervals and summarize (μ, σ) and the implied bounds. Conclusion: Retain the GT interval(s) as the definitive localization; envelope breaches serve as corroborating evidence.

(ii) Contextual Point. OBSERVATION: Apply the defined global–structural–local scan to find salient deviations as candidates for anomaly detection. REASONING & VALIDATION: Examine fixed-length, z-normalized subsequences using the Matrix Profile: let $d(i)$ be the discord distance and $i^* = \arg\max_i d(i)$. Standardize $\{d(i)\}$ to $z(i)$; if $z(i^*) > \tau$ (e.g., τ=3.5), the subsequence $[i^*, i^*+m)$ constitutes strong evidence of a contextual departure. CONCLUSION: State the GT contextual-point interval(s) as final, summarizing the dominant discord and its standardized magnitude as quantitative support.

(iii) Trend Shift. OBSERVATION: Apply the defined global–structural–local scan to find salient deviations as candidates for anomaly detection. REASONING & VALIDATION: Smooth the series and analyze the gradient g_t; highlight segments where $|g_t - \bar{g}|$ exceeds a multiple of the empirical dispersion of $\{g_t\}$, and merge adjacent exceedances into candidate intervals indicating a shift in slope or level. CONCLUSION: Present the GT trend-shift span(s) as the conclusive localization, together with the gradient summary (center, dispersion, and threshold) as supporting evidence.

(iv) Seasonal/Frequency Deviation. OBSERVATION: Apply the defined global–structural–local scan to find salient deviations as candidates for anomaly detection. REASONING & VALIDATION: Estimate the dominant period over sliding windows (FFT-based periodogram) and identify windows whose periods deviate beyond a robust tolerance around the typical period (e.g., median $\pm k \times 1.4826 \cdot \mathrm{MAD}$). Map these window-level deviations back to the time axis and merge them into intervals. CONCLUSION: Declare the GT seasonal/frequency interval(s) as final, reporting the typical period, its robust dispersion, and the deviation range as quantitative support.

(v) Shapelet/Subsequence. OBSERVATION: Apply the defined global–structural–local scan to find salient deviations as candidates for anomaly detection. REASONING & VALIDATION: Use a subsequence dissimilarity scan (e.g., Matrix Profile), prioritizing the most pronounced discord(s) and, when desired, assessing cross-scale stability across nearby window lengths to strengthen evidence. CONCLUSION: When GT specifies a shapelet/subsequence anomaly, return the GT interval(s) as the definitive localization and include the strongest dissimilar segment(s) as auxiliary evidence.

Instantiation with Ground Truth. For every instance, ExpCoT is generated from the GT class and temporal annotation: OBSERVATION anchors on the GT interval(s) and applies the unified scan (global $\rightarrow$ structural $\rightarrow$ local); REASONING & VALIDATION then selects the analysis matched to the GT type and reports concrete numerical evidence (global envelope deviation, standardized discord magnitude, smoothed-gradient exceedance, or dominant-period drift); CONCLUSION integrates these results and retains the GT interval(s) as the final localization, yielding a faithful, interpretable trace for supervising MLLM training. In practice, these traces are first generated automatically by code to provide quantified validation, and are subsequently refined by human experts for greater fluency and high-fidelity interpretability.

Table 3: Comparison of AnomLLM, VisualTimeAnomaly and TSB-UAD.

Category	**AnomLLM**	**VisualTimeAnomaly**	**TSB-UAD**
Synthetic	✓	✓	✓
Real-world	✗	✓	✓
Length range	1,000	≈1,000	1,570–230,400
Avg. anomaly rate	∼5%	∼5–15%	∼0.2–9.8%
Anomaly Types			
Global point	✓	✓	✓
Contextual point	✓	✓	✓
Trend	✓	✓	✓
Seasonal	✓	✓	✓
Shapelet	✗	✓	✓

Contextual Point Localization (GT [897,902])

Observation — *Global Scan:* The series exhibits moderate variation with mean μ=0.123 and standard deviation σ=0.087. The maximum standardized deviation $\max|z|$=1.41 suggests no extreme global outliers are present. *Structural Scan:* Gradient analysis reveals low trend instability (σ_{grad}=0.0034), indicating stable baseline behavior. A clear dominant period appears at ∼47.2 time units, revealing regular structural patterns. *Pattern Scan:* Discord search identifies the most dissimilar subsequence at t≈899 with discord score =0.156, suggesting potential local anomalies.
Reasoning & Validation — I observed that the subsequences within $[897, 902]$ are globally plausible but locally inconsistent, indicating a contextual anomaly. Matrix Profile analysis (window m=50) reveals strongest discord at t≈899 with z-score=4.2. Since $z > 3.5$, this confirms significant contextual deviation.
Conclusion — Integrating hierarchical scan insights with targeted contextual point validation, the evidence converges on anomalous behavior in interval $[897, 902]$. Therefore, the detected anomaly is classified as contextual point, precisely localized to $[897, 902]$.

Figure 8: Example of ExpCoT reasoning trace for contextual point anomaly detection.

Global Point Localization (GT [1245,1248])

Observation — *Global Scan:* The series centers near zero with mean μ=0.002 and moderate spread σ=0.098. However, the maximum standardized deviation $\max|z|$=6.73 indicates the presence of extreme values that deviate substantially from the typical range. *Structural Scan:* The series demonstrates high trend stability with minimal gradient variation (σ_{grad}=0.0012), suggesting consistent baseline behavior. No dominant periodic structure is evident. *Pattern Scan:* Discord analysis identifies the most unusual subsequence at t≈1246 with notable discord score =0.234, highlighting potential anomalous regions.
Reasoning & Validation — I observed that the values within $[1245, 1248]$ exhibit out-of-range behavior, representing clear global point anomalies. Applying k-σ envelope $[-0.292, 0.296]$ validation: 4 points at $t \approx 1250$ exceed boundaries, confirming significant global deviation from normal range.
Conclusion — Integrating hierarchical scan insights with targeted global point validation, the evidence converges on anomalous behavior in interval $[1245, 1248]$. Therefore, the detected anomaly is classified as global point, precisely localized to $[1245, 1248]$.

Figure 9: ExpCoT reasoning trace for global point (out-of-range) anomaly detection.

D Experimental Details

D.1 Dataset Statistics

We evaluate three public resources to assess models' performance and generalizability across various TSAD scenarios. The detailed dataset statistics and anomaly coverage are summarized in Table 3.

1) **AnomLLM** (Zhou & Yu, 2024) provides controlled synthetic time-series anomaly detection benchmarks. Following the default generation settings (https://github.com/rose-stl-lab/anomllm), we generate eight anomaly types: out-of-range, point, frequency, trend, flat-trend, noisy-point, noisy-freq, and noisy-trend. They can be grouped into four categories: range, point, freq, and trend. For nomenclature consistency in this paper, we map the original task names to our taxonomy as follows: Range → Global point, Point → Contextual point, Freq → Seasonal, and Trend → Trend. Given this synthetic generation process, global (out-of-range) anomalies are typically the easiest to detect, whereas contextual point, trend, and seasonal anomalies are more difficult due to their reliance on local context, regime changes, and frequency shifts, respectively.

2) **VisualTimeAnomaly** (Xu et al., 2025) converts numerical time series into images across various scenarios; in our study, we focus on the univariate setting and adhere to the default synthetic

Trend Shift Localization (GT [8934,9078])

Observation — *Global Scan:* The series shows slight negative bias with mean μ=−0.045 and considerable variability σ=0.512. The moderate maximum standardized deviation $\max|z|$=2.18 suggests no extreme global outliers are present. *Structural Scan:* Gradient analysis reveals moderate trend instability (σ_{grad}=0.0078), indicating some underlying directional changes. No dominant periodic structure is evident. *Pattern Scan:* Discord search locates the most dissimilar pattern at t≈8967 with discord score =0.189, pointing to regions of structural change.
Reasoning & Validation — I observed that the segment within $[8934, 9078]$ shows a clear trend shift, with the long-term slope undergoing a marked change. Gradient analysis on smoothed series (window=21) shows baseline slope $\mu_{\text{grad}}=-0.0002 \pm 0.0078$. 47 points at $t \approx 8950$ exceed k-σ gradient threshold ($\pm$0.0234), confirming abnormal trend changes.
Conclusion — Integrating hierarchical scan insights with targeted trend shift validation, the evidence converges on anomalous behavior in interval $[8934, 9078]$. Therefore, the detected anomaly is classified as trend shift, precisely localized to $[8934, 9078]$.

Figure 10: ExpCoT reasoning trace for trend shift anomaly detection.

Frequency Deviation Localization (GT [2167,2189])

Observation — *Global Scan:* The series maintains near-zero centering with mean μ=0.014 and high variability σ=0.712. The moderate maximum standardized deviation $\max|z|$=1.89 indicates well-contained global fluctuations. *Structural Scan:* Gradient analysis shows moderate trend variation (σ_{grad}=0.0234), consistent with oscillatory behavior. A strong dominant period emerges at ∼50.0 time units, suggesting regular cyclical patterns. *Pattern Scan:* Discord analysis identifies the most atypical subsequence at t≈2178 with discord score =0.142, indicating potential frequency irregularities.
Reasoning & Validation — I observed that the oscillations within $[2167, 2189]$ exhibit frequency deviation patterns, where the periodic structure changes relative to the expected seasonal baseline. Sliding period analysis (W=120) reveals typical period μ=50.2 with robust σ=4.8. 8 windows fall outside robust range $[35.8, 64.6]$, confirming frequency drift.
Conclusion — Integrating hierarchical scan insights with targeted frequency deviation validation, the evidence converges on anomalous behavior in interval $[2167, 2189]$. Therefore, the detected anomaly is classified as frequency deviation, precisely localized to $[2167, 2189]$.

Figure 11: ExpCoT reasoning trace for frequency deviation anomaly detection.

workflow (https://github.com/mllm-ts/VisualTimeAnomaly). The benchmark includes point-wise (global/contextual) and range-wise (trend/seasonal/shapelet) anomalies for univariate series. Within this dataset, point-wise anomalies are the hardest to localize visually, whereas range-wise anomalies are comparatively easier due to their salient coarse-grained patterns.

3) **TSB-UAD** (Qiu et al., 2025) unifies 1,635 univariate series from the original TSB-UAD (Paparrizos et al., 2022) by filtering out low-quality series (e.g., those without anomalies or with an anomaly ratio >10%), resulting in a high-quality collection that includes both real-world and synthetic datasets. We adopt the official defaults and taxonomy from the repository (https://github.com/decisionintelligence/TAB). The TAB-UAD dataset covers both univariate and multivariate settings (treating each multivariate dataset as multiple univariate time series and evaluating them individually). The anomaly coverage includes point (global/contextual) and subsequence (trend/shapelet/seasonal) categories, as well as mixed types. The collected series span diverse domains such as industrial sensors, medical signals, finance, and web traffic, making the benchmark both comprehensive and representative of real-world anomaly detection challenges.

D.2 BASELINES

For each benchmark, we evaluate three groups of models. For the closed-source MLLMs, we access commercial APIs, including MaaS_GP_4o_20241120, MaaS_GP_4o_mini_20240718, MaaS_Ge_2.5_pro_20250617, and MaaS_Ge_2.5_flash_lite_20250722. For the open-source counterparts, we rely on HuggingFace checkpoints such as Qwen/Qwen2.5-VL-72b-Instruct and its smaller variants (e.g., 32B/7B/3B). We further compare against supervised fine-tuned baselines, including Qwen2.5-VL-3B-SFT3.2k, fine-tuned on 3,200 instances, and Qwen2.5-VL-3B-SFT32k, fine-tuned on 32,000 instances.

In addition, we include two representative LLM-based temporal reasoning baselines. SIGLLM (Alnegheimish et al., 2024) is a GPT-3.5-based detector for anomaly identification. We evaluate SigLLM under the default settings provided in its official repository (https://github.com/rose-stl-lab/sigllm), using the original prompts and raw numerical inputs. TIMEMASTER (Zhang et al., 2025b), which builds on Qwen2.5-VL-3B with supervised fine-tuning (SFT) and GRPO and adopts image inputs, is also trained under its default public release[3]. For all models except SIGLLM, we use the same prompt templates (see Figure 7) to ensure consistency and fairness.

D.3 METRICS

We evaluate detection quality using the affiliation-based metrics introduced by Huet et al. (2022)[4], namely *Affi_Precision*, *Affi_Recall*, and their harmonic mean *Affi_F1*. These affiliation-based metrics can be viewed as event-level extensions of the classical precision/recall/F1-score to time-series anomaly detection (Huet et al., 2022). Affi_Precision and Affi_Recall evaluate each ground-truth event locally, and are parameter-free. Moreover, their construction via comparison to a random reference predictor makes the resulting scores both theoretically principled and practically useful for TSAD, especially in LLM-based TSAD settings (Zhou & Yu, 2024; Liu et al., 2024; Xu et al., 2025). Below, we provide their detailed definitions.

Setup. Recall that a univariate time series of length T is denoted by $\mathbf{X} = \{\mathbf{x}_t\}_{t=1}^{T}$. Ground-truth anomaly intervals are given by

$$\mathcal{A} = \{(t_s^{(i)}, t_e^{(i)})\}_{i=1}^{k}, \qquad 1 \le t_s^{(i)} \le t_e^{(i)} \le T,$$

where each interval $(t_s^{(i)}, t_e^{(i)})$ denotes the i-th anomalous segment (with $t_s^{(i)} = t_e^{(i)}$ corresponding to a single-point anomaly). We assume these intervals are pairwise disjoint. For convenience, we identify each interval with the corresponding set of time indices,

$$J_i = \{t \in \{1, \dots, T\} : t_s^{(i)} \le t \le t_e^{(i)}\}.$$

Thus the collection of ground-truth events is

$$\mathcal{J} = \{J_j\}_{j=1}^{n},$$

where $n = k$ and the J_j are pairwise disjoint subsets of $\{1, \dots, T\}$.

Similarly, we denote the predicted anomaly intervals by

$$\widehat{\mathcal{A}} = \{(\hat{t}_s^{(i)}, \hat{t}_e^{(i)})\}_{i=1}^{\hat{k}},$$

and write

$$\widehat{J}_i = \{t \in \{1, \dots, T\} : \hat{t}_s^{(i)} \le t \le \hat{t}_e^{(i)}\}, \qquad \widehat{\mathcal{J}} = \{\widehat{J}_i\}_{i=1}^{\hat{k}}.$$

All sets J_j and $\widehat{J}_i$ are subsets of the index set $\mathcal{T} = \{1, \dots, T\}$. For any $A \subseteq \mathcal{T}$, we write $|A|$ for its cardinality. For $t \in \mathcal{T}$ and $Y \subseteq \mathcal{T}$, we define

$$\mathrm{dist}(t, Y) = \min_{y \in Y} |t - y|$$

as the distance (in time indices) from t to the set Y, with the convention that $\mathrm{dist}(t, \varnothing) = +\infty$.

[3] https://github.com/langfengQ/TimeMaster

[4] The official implementation of these metrics is publicly available at https://github.com/ahstat/affiliation-metrics-py/.

Affiliation regions. Following (Huet et al., 2022), we partition the time index set $\mathcal{T}$ into *affiliation regions* $\{E_j\}_{j=1}^n$, one for each ground-truth event J_j:

$$E_j \;=\; \big\{t \in \mathcal{T} : j = \arg\min_{k \in \{1,\dots,n\}} \mathrm{dist}(t, J_k)\big\},$$

with ties broken arbitrarily so that $\{E_j\}_{j=1}^n$ forms a partition of $\mathcal{T}$, i.e. $\mathcal{T} = \biguplus_{j=1}^n E_j$ and $E_j \cap E_k = \varnothing$ for $j \neq k$. For each j, we denote by

$$\widetilde{P}_j \;=\; \Big(\bigcup_{i=1}^{\hat{k}} \widehat{J}_i\Big) \cap E_j$$

the subset of predicted anomalous time indices that fall inside the affiliation region E_j.

Random reference predictor. For each $j \in \{1, \dots, n\}$, we define a random reference predictor by drawing a time index

$$X_j \sim \mathrm{Unif}(E_j),$$

uniformly at random from E_j. The *precision-side baseline distance* is

$$D_j^{\mathrm{prec}} \;=\; \mathrm{dist}(X_j, J_j),$$

and its survival function (complementary CDF) is

$$\overline{F}_j^{\mathrm{prec}}(d) \;=\; \mathbb{P}\big(D_j^{\mathrm{prec}} \geq d\big), \qquad d \geq 0.$$

Intuitively, $\overline{F}_j^{\mathrm{prec}}(d)$ measures how likely a random prediction in E_j lies at distance at least d from the true event J_j.

For the recall side, for each j and each time index $t \in J_j$, we define

$$D_{j,t}^{\mathrm{rec}} \;=\; \mathrm{dist}(t, X_j),$$

and the corresponding survival function

$$\overline{F}_{j,t}^{\mathrm{rec}}(d) \;=\; \mathbb{P}\big(D_{j,t}^{\mathrm{rec}} \geq d\big), \qquad d \geq 0.$$

Affi_Precision. For a fixed ground-truth event J_j, the *local affiliation-precision score* $P_{\mathrm{prec}}(j)$ compares the actual predictions in E_j to the random baseline:

$$P_{\mathrm{prec}}(j) \;=\; \begin{cases} \dfrac{1}{|\widetilde{P}_j|} \displaystyle\sum_{t \in \widetilde{P}_j} \overline{F}_j^{\mathrm{prec}}\big(\mathrm{dist}(t, J_j)\big), & \text{if } |\widetilde{P}_j| > 0, \\ \text{(ignored)}, & \text{if } |\widetilde{P}_j| = 0. \end{cases}$$

Only those events with $|\widetilde{P}_j| > 0$ contribute to the global precision. Let

$$S \;=\; \big\{j \in \{1, \dots, n\} : |\widetilde{P}_j| > 0\big\}$$

be the set of ground-truth events for which at least some prediction mass falls into E_j. The global *Affi_Precision* is defined as

$$\text{Affi\_Precision} \;=\; \begin{cases} \dfrac{1}{|S|} \displaystyle\sum_{j \in S} P_{\mathrm{prec}}(j), & \text{if } |S| > 0, \\ 0, & \text{if } |S| = 0. \end{cases}$$

Affi_Recall. For the recall side, each ground-truth event J_j defines a local score $P_{\mathrm{rec}}(j)$ by averaging, over all time indices $t \in J_j$, how much better the prediction $\widetilde{P}_j$ is than the random baseline:

$$P_{\mathrm{rec}}(j) \;=\; \frac{1}{|J_j|} \sum_{t \in J_j} \overline{F}_{j,t}^{\mathrm{rec}}\big(\mathrm{dist}(t, \widetilde{P}_j)\big),$$

where

$$\mathrm{dist}(t, \widetilde{P}_j) \;=\; \min_{z \in \widetilde{P}_j} |t - z|,$$

with the convention that if $\widetilde{P}_j = \varnothing$, then $\mathrm{dist}(t, \widetilde{P}_j) = +\infty$ and $\overline{F}^{\mathrm{rec}}_{j,t}(\mathrm{dist}(t, \widetilde{P}_j)) = 0$.

The global *Affi_Recall* is obtained by averaging $P_{\mathrm{rec}}(j)$ over all ground-truth events:

$$\mathrm{Affi\_Recall} \;=\; \frac{1}{n} \sum_{j=1}^{n} P_{\mathrm{rec}}(j).$$

Affi_F1. Finally, the *Affi_F1* score is defined as the harmonic mean of Affi_Precision and Affi_Recall. Let

$$P \;=\; \mathrm{Affi\_Precision}, \qquad R \;=\; \mathrm{Affi\_Recall},$$

then

$$\mathrm{Affi\_F1} \;=\; \begin{cases} 0, & \text{if } P + R = 0, \\ \dfrac{2PR}{P + R}, & \text{otherwise.} \end{cases}$$

By construction, Affi_Precision, Affi_Recall, and Affi_F1 all take values in the interval $[0, 1]$.

D.4 Implementation Details

D.4.1 Time-Series Image Input

We follow the common plotting conventions used in prior work on MLLMs (Xu et al., 2025; Zhou & Yu, 2024; Zhang et al., 2025b) to ensure fairness. The line plots do not include shaded or highlighted regions, and anomalous intervals are not explicitly marked. Each time-series image is rendered at a resolution of 805×124 pixels.

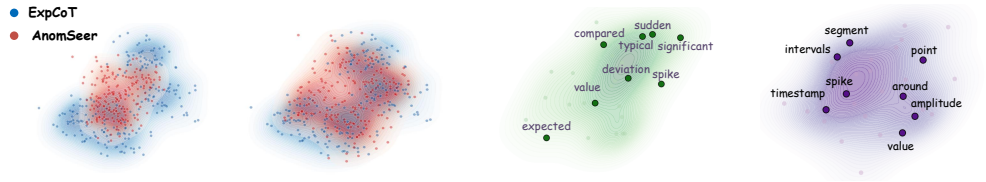
Table 4: Hyperparameter settings.

Algorithm	Hyperparameter	Value	Algorithm	Hyperparameter	Value
GRPO	Max response length	1024	TimerPO	r^{fmt}	0.1
	Batch size	128		r^{cls}	0.2
	Mini-batch size	128		r^{loc}	0.7
	KL loss coefficient	0.001		α	0.3
	Group size	5			
	Learning rate	1e-6			

D.4.2 Training Setup

We initialize our backbone with the publicly available Qwen2.5-VL-3B-Instruct[5] and Qwen2.5-VL-7B-Instruct[6] checkpoints. Our overall training pipeline only includes a TimerPO stage based purely on reinforcement learning. We build our implementation on the public RL training library[7] and the temporal reasoning training framework[8]. We summarize our hyperparameter settings in Table 4, where the GRPO configuration follows TimeMaster for fairness. The models are trained on 3,200 synthetic samples from AnomLLM and evaluated on the AnomLLM synthetic test set, VisualTimeAnomaly, and TSB-UAD, which cover broader anomaly types and varying sequence lengths to assess generalization to unseen real-world scenarios.

[5] https://huggingface.co/Qwen/Qwen2.5-VL-3B-Instruct
[6] https://huggingface.co/Qwen/Qwen2.5-VL-7B-Instruct
[7] https://github.com/volcengine/verl
[8] https://github.com/langfengQ/TimeMaster



(a) GRPO Alignment (b) TimerPO Alignment (d) GRPO Top Words (d) TimerPO Top Words

Figure 12: Comparison of distributional alignment between ExpCoT (blue) and AnomSeer (red) outputs, along with token usage under GRPO and TimerPO training.

Method	Classification	Location												
	Accuracy	Frequency			Trend			Range			Point			Avg F1
		P	R	F1	P	R	F1	P	R	F1	P	R	F1	
TimeMaster-3B	57.90±1.49	57.30±1.24	50.30±0.25	51.40±0.50	76.00±1.24	77.30±0.25	76.60±1.24	77.80±1.24	83.50±0.25	80.10±0.25	77.70±1.24	82.10±0.25	79.60±1.24	71.92±0.81
ANOMSEER-3B	62.80±1.24	63.70±1.24	58.40±1.24	58.90±1.24	84.20±0.50	85.90±0.25	84.90±0.25	83.30±0.75	89.20±0.25	85.60±0.25	86.00±0.25	90.30±0.25	87.80±0.25	79.30±0.50
ANOMSEER-7B	65.00±1.24	68.30±1.24	59.40±0.50	60.80±0.50	86.60±1.24	89.00±1.24	87.70±1.24	91.60±0.25	97.80±0.99	94.30±0.25	93.40±0.25	96.90±1.24	94.90±0.25	84.42±0.56

Table 5: Mean ± 95% confidence interval half-width over 3 seeds.

D.5 System Configuration

All experiments were conducted on a computing setup equipped with 4 NVIDIA A100-SXM4 GPUs (80 GB each) and 4 NVIDIA RTX A6000 GPUs (48 GB each) for Qwen-3B, and 4 NVIDIA H100-SXM4 GPUs (96 GB each) for Qwen-7B.

E More Experimental Results

E.1 Confidence Intervals and Computational Cost

To complement the results in the main paper, we provide the complete set of performance metrics corresponding to Table 1, including mean values over three runs together with their 95% confidence intervals. As shown in Table 5, the consistently small intervals support the robustness of our findings and indicate that AnomSeer performs stably across repeated trials.

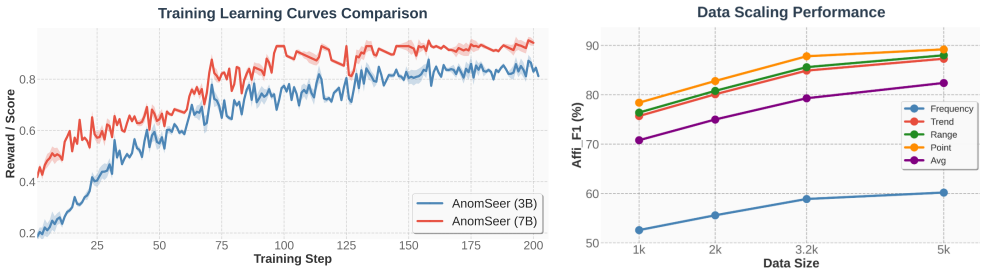
We also report the computational profile of AnomSeer (3B) trained on NVIDIA RTX A6000 GPUs with 48 GB of memory. The training phase requires 12.4 hours of wall-clock time using four GPUs in parallel. For inference, the model operates on a single GPU, utilizing approximately 7 GB of memory and achieving an average latency of 4.8 seconds per time-series sample. These computational characteristics fall within acceptable limits for practical deployment in TSAD scenarios.

E.2 Learning Curves & Data Scaling

We present the learning curves and data-scaling results in Figure 13. We first observe that the learning curves for both the 3B and 7B models exhibit stable and monotonic improvement, with performance rising rapidly during the initial 50-100 training steps before gradually stabilizing. In addition, scaling the training set from 1k to 5k examples yields consistent gains across all four tasks. The average Affinity F1 score continues to improve as the dataset grows, with no clear signs of saturation. These results suggest that the current data regime remains in a growth phase, and further increasing the amount of training data is likely to yield additional performance gains.

E.3 Optimization and Alignment Ablation

Advantage-level orthogonalization vs. gradient-level projection. We compared TimerPO to two multi-objective optimization baselines: (i) a weighted-sum objective (no projection) and (ii) PCGrad-style gradient orthogonalization (Yu et al., 2020). Table 6 summarizes the results. TimerPO consistently outperforms both weighted-sum and gradient-level projection across all anomaly types. Orthogonalizing auxiliary signals at the advantage level promotes complementary contributions



(a) (b)

Figure 13: (a) Learning curves of training score versus training steps for the 3B and 7B models, and (b) data-scaling performance for the 3B model evaluated from 1k to 5k training examples.

prior to gradient computation, whereas PCGrad only modifies gradients when explicit conflicts are detected. This reduces partial interference between objectives and results in smoother, lower-variance optimization trajectories.

Method	Freq.	Trend	Range	Point	Avg
Orthogonalization strategies					
TimerPO (ours)	58.9	84.9	85.6	87.8	**79.3**
Weighted-sum (no proj.)	53.5	81.1	83.5	85.4	75.9
PCGrad (gradient level)	54.2	80.2	84.5	86.4	76.3
Alignment objectives					
TimerPO (ours)	58.9	84.9	85.6	87.8	**79.3**
Cosine similarity	42.2	73.8	84.1	86.8	71.7
CLIP-style similarity	48.5	74.1	84.3	86.7	73.4

Table 6: Comparison of orthogonalization strategies (top) and alignment objectives (bottom).

Replacing OT with cosine or contrastive similarity. To isolate the contribution of OT-based alignment, the OT module in TimerPO was replaced with two alternatives: (i) token-wise cosine similarity and (ii) a CLIP-style InfoNCE objective (temperature $= 0.07$). As shown in Table 6, OT yields a 6.8% improvement in average F1 relative to cosine and contrastive similarity. OT provides structure-aware alignment by modeling semantic distances between reasoning tokens rather than treating tokens independently. These findings indicate that OT geometry plays an essential role in aligning model reasoning with temporally structured anomaly patterns.

E.4 Extensibility

We now discuss the extensibility of the proposed method, with results summarized in Tab. 7. **1) Multivariate time series.** Although the main experiments focus on univariate data, the framework is not limited to this setting. Multivariate inputs can be converted into a unified visual representation by rendering each variable as a subplot within a single image. Empirical results on a multivariate benchmark demonstrate that the method generalizes effectively beyond the univariate setting. **2) Short-term and boundary anomalies.** Short-duration or boundary anomalies are typically under-represented in existing datasets and therefore challenging to detect reliably. A simple targeted augmentation strategy yields notable improvements on a dedicated evaluation set of such cases. These findings indicate that lightweight preprocessing and sampling adjustments can enhance robustness in challenging anomaly scenarios.

E.5 Comparison with Traditional TSAD Methods

To provide a unified view of classical time-series anomaly detection (TSAD) methods and our framework, Table 8 summarizes representative baselines across four anomaly types. Traditional

Setting	GPT-4o	Gemini-2.5	Qwen2-VL	Ours (7B)
Multivariate (synthetic / real-world)	62.7 / 54.2	77.3 / 65.2	45.0 / 24.5	**83.5 / 72.4**
Short / boundary anomalies	Original: 56.2		**+ Augmentation: 75.8**	

Table 7: Results for multivariate evaluation (top) and short/boundary anomaly robustness (bottom).

approaches such as FFT, Matrix Profile, gradient-based detection, ARIMA, and statistical thresholding operate directly on raw signals and typically produce detection outputs only. While they can perform well in specific scenarios, they often rely on careful parameter tuning (e.g., window selection or differencing) and exhibit limited robustness across diverse anomaly patterns.

In contrast, AnomSeer outperforms traditional approaches across all anomaly types and, more importantly, supports detection, classification, and natural language reasoning within a single model. This broader output capability enables interpretability and generalization across diverse anomaly patterns, rather than optimizing for a single metric or domain-specific signal property.

Method	Capability	Freq.	Trend	Range	Point	Avg
FFT	Location only	65.9	18.0	28.5	27.8	35.1
Matrix Profile	Location only	11.4	29.4	67.2	87.4	48.9
Gradient	Location only	57.1	58.5	55.8	65.9	59.3
Ensemble (voting)	Location only	59.0	17.4	69.2	92.4	59.5
ARIMA	Location only	62.6	4.9	67.3	74.2	52.3
Thresholding	Location only	57.6	25.2	60.0	47.6	47.6
AnomSeer (3B)	Loc. + Cls. + Reasoning	58.9	84.9	85.6	87.8	**79.3**
AnomSeer (7B)	Loc. + Cls. + Reasoning	60.8	87.7	94.3	94.9	**84.4**

Table 8: Comparison with classical TSAD baselines and the proposed AnomSeer.

E.6 Details on Effect of TimerPO

To highlight the advantage of TimerPO over vanilla GRPO in temporal reasoning, we further compare their behaviors in Figure 12. While GRPO narrows the distributional gap to some extent, the model outputs remain relatively constrained and still exhibit a clear mismatch compared to expert reasoning. A similar pattern is evident in token usage: GRPO-trained outputs are dominated by outcome-oriented words such as compared and expected, whereas TimerPO encourages the use of more fine-grained, temporally grounded terms like timestamp, intervals, and amplitude, which anchor reasoning to concrete temporal structures. These findings confirm that TimerPO provides a more principled enhancement over GRPO, enabling models to move beyond surface outcome alignment toward genuine temporal reasoning.

E.7 More Case Studies on Reasoning

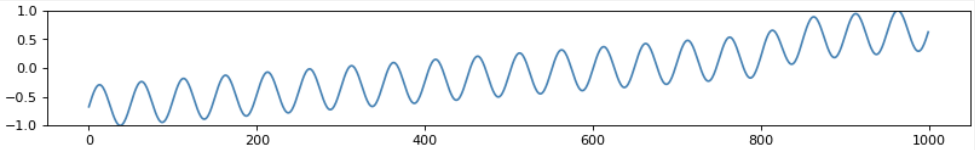
We provide several case studies illustrating our model's complete reasoning process on corresponding data visualizations. These examples show our approach's effectiveness in focusing on specific segments and timestamps for fine-grained analysis. We also present a failure case in Fig. 14: a short anomaly within the interval [998, 1000] at the sequence's very end goes undetected by AnomSeer, which incorrectly classifies it as 'normal'. This highlights the need for future work to improve the sensitivity of MLLMs to such boundary-case anomalies.

F Extended Related Work

LLM-based time series anomaly detection (TSAD) is an emerging area, with several exploratory methods recently proposed. These approaches vary in modalities, backbones, and integration strategies. For example, SigLLM (Alnegheimish et al., 2024) and CoLLaTe (Chen et al., 2025) use numeric–text inputs with GPT-3.5 or GPT-4, relying on prompting with external post-processing or

Response of AnomSeer on the ground-truth trend anomaly [[800, 850]].

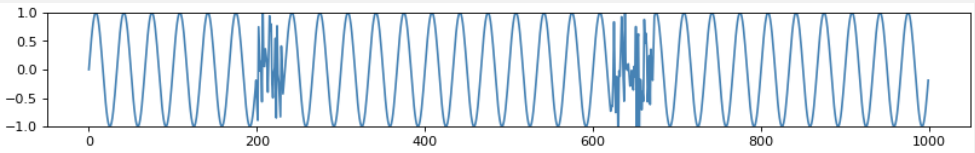
Input Image:



AnomSeer Response: <think> The plot shows a general wave-like pattern with intervals where the amplitude becomes slightly elevated. Such behavior could hint at a potential trend change, but the deviations are not very pronounced. The global structure remains smooth, with no sudden spikes or sharp dips that would qualify as clear anomalies. However, I did notice a small irregularity forming around timestep 850, where the local pattern begins to diverge slightly from the surrounding trend. The fluctuation is modest in amplitude, with peaks rising only marginally higher and troughs dipping slightly deeper than the preceding cycles. This subtle change suggests a weak signal of trend variation rather than a strong anomaly. </think>
<answer> [[800, 900]] </answer>
<class> trend </class>

Response of AnomSeer on the contextual point anomaly [[200,231],[622,673]]

Input Image:



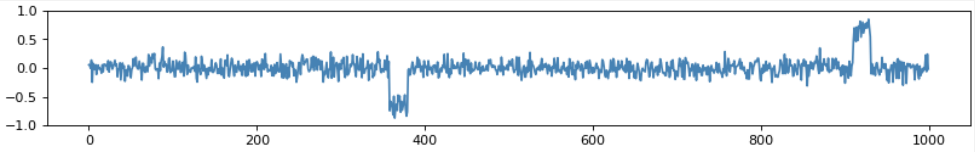
AnomSeer Response: <think> The time series exhibits two noticeable intervals that deviate from the surrounding pattern. The first occurs between roughly 200 and 250, while the second spans around 600 to 700. Both segments rise sharply compared with the smoother baseline, resembling local spikes rather than gradual changes. These departures stand out in their immediate neighborhoods but do not exceed the overall global range of the series. Because these deviations break the contextual continuity of the waveform, they are better characterized as contextual point anomalies. </think>
<answer> [[200, 250], [600, 700]] </answer>
<class> contextual point </class>

task-specific TSAD modules, but without reasoning ability. LLMAD (Liu et al., 2025a) augments GPT-4-turbo with retrieval-based domain knowledge to support anomaly classification and localization, though it requires an external database for prompting. On the multimodal side, TAMA (Zhuang et al., 2024) and VLM4TS (He et al., 2025) employ image–text inputs with GPT-4o, together with post-processing or ViT-based components. More recently, Time-RA (Yang et al., 2025) applies large-scale SFT on Qwen2.5-VL-7B, but its coverage remains incomplete, particularly in anomaly localization. In contrast, our method uses a compact open-source backbone, Qwen2.5-VL-3B/7B, and reinforcement learning to directly equip the model with anomaly classification, localization, and reasoning, without external modules or proprietary APIs.

Optimal Transport (OT) offers a principled geometric framework for aligning probability distributions and has seen increasing adoption in both reinforcement learning (RL) and large language model (LLM) alignment. In RL, OT has been leveraged to structure learning signals and align task or policy distributions (Klink et al., 2022; Asadulaev et al., 2024; Chen et al., 2020), notably in curriculum design (Klink et al., 2022) and as a regularizer for offline policy learning (Asadulaev et al.,

Response of AnomSeer on the global point anomaly [[358,380],[910,931]]

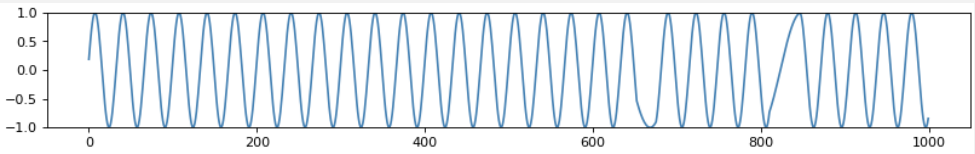
Input Image:



AnomSeer Response: <think> The plot contains several sharp, discontinuous excursions in amplitude. By scanning the sequence for abrupt changes that break this continuity, I identified three distinct intervals: one near 350, another around 880, and a third close to 900. Each spike pops outside the local band, then quickly snaps back. These intervals deviate noticeably from the surrounding baseline, suggesting global point anomalies in the series. </think>
<answer> [[350, 450], [880, 950]] </answer>
<class> global point </class>

Response of AnomSeer on the freq anomaly [[653,676],[811,846]]

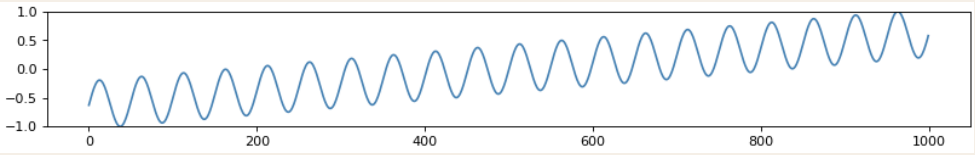
Input Image:



AnomSeer Response: <think> To identify anomalies, I first examine the trend and periodic structure of the plot. The early section displays a regular sinusoidal rhythm with no unusual deviations. Around the 650th timestep, however, the frequency begins to distort, showing dips and irregular oscillations that break the expected cycle. In addition, I observed another disturbance near the 850th mark, where the pattern again deviates from the baseline. This suggests a change in the underlying dynamics, with the presence of a seasonal anomaly disrupting the expected pattern. </think>
<answer> [[580, 700], [850, 900]] </answer>
<class> seasonal </class>

Failure case where AnomSeer fails to detect the ground-truth trend anomaly at [998,1000]

Input Image:



AnomSeer Response: <think> No anomalies detected in the plot. </think>
<answer> [] </answer>
<class> normal </class>

Figure 14: Failure case

2024). Within LLM, OT has been used to support preference modeling (Li et al., 2025a; Melnyk

et al., 2024; Désidéri, 2012; Chu et al.; Li et al., 2025b), by aligning full reward distributions (Melnyk et al., 2024) or applying token-level weighting schemes to highlight semantically important regions (Li et al., 2025a). Most of these approaches focus on final outcome alignment, operating over entire sequences or aggregated behaviors. But our work applies OT at the reasoning-token level, aligning the model's intermediate reasoning steps with structured ExpCoT traces derived from classical TSAD primitives. This enables process-level supervision, enhancing the model's temporal reasoning capabilities rather than merely refining output preferences.

Multi-objective optimization methods (Désidéri, 2012; Yu et al., 2020; Liu et al., 2021; Wei & Hu, 2024) aim to stabilize training across competing tasks by projecting conflicting gradients into compatible directions. For example, PCGrad (Yu et al., 2020) explicitly projects one task's gradient onto the normal plane of another when conflicts arise. In contrast, our TimerPO introduces orthogonal projection in the advantage space, not to resolve inter-task interference, but to preserve the independent contribution of an auxiliary reasoning advantage. Since this auxiliary signal reflects structured supervision rather than a separate objective, our projection design allows it to complement the main anomaly detection reward without disruption. To the best of our knowledge, this is the first approach to combine token-level OT alignment with advantage-space disentanglement to enhance temporal reasoning in multimodal LLMs.