TREND-CONDITIONED RESIDUAL DIFFUSION FOR MULTIVARIATE TIME-SERIES ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Unsupervised multivariate time-series anomaly detection is challenging in real-world systems because normal observations often exhibit non-stationary trends, local fluctuations, and evolving inter-variable dependencies. Existing reconstruction-based methods usually learn normal patterns directly in the raw sequence space, which may confuse normal trend variations with anomalies and reduce the discriminability of reconstruction errors. To address this issue, we propose a **Trend-Conditioned Residual Diffusion** framework for multivariate time-series anomaly detection. Instead of recovering the entire sequence from a pure-noise state, the proposed method constructs a smoothed trend state as a structured endpoint and trains a diffusion bridge to recover the fine-grained residual between the original sequence and its trend component. This design shifts the reconstruction target from raw-sequence generation to trend-conditioned residual recovery, thereby providing a more stable normal-pattern baseline for anomaly scoring. To further capture cross-variable normal dependencies, we introduce a channel aggregate-redistribute fusion module that summarizes global channel context and redistributes it to each variable representation. Extensive experiments on five public real-world datasets demonstrate that the proposed method achieves competitive detection performance under commonly used point-adjusted evaluation and remains effective under affiliation-based metrics and detection-delay analysis. Ablation studies further verify the contributions of the diffusion bridge, trend-conditioned residual recovery, and channel-coupling fusion.

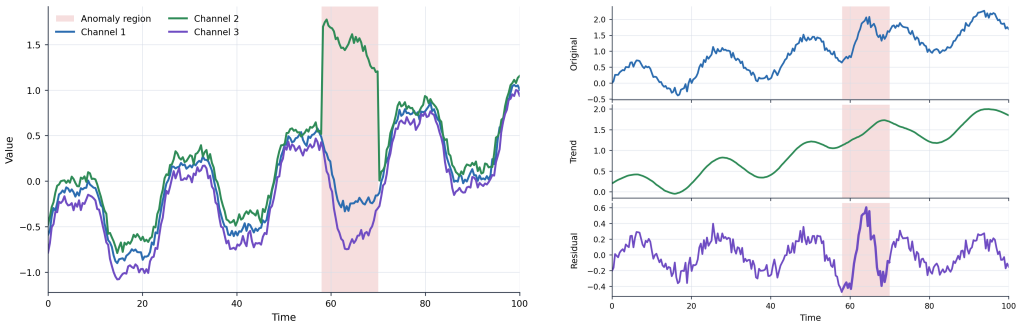
1 INTRODUCTION

Multivariate time-series anomaly detection (MTSAD) supports monitoring in servers, spacecraft, and industrial systems, where labels are scarce and normal behavior is both non-stationary and cross-channel dependent Hundman et al. (2018); Su et al. (2019); Mathur & Tippenhauer (2016). Reconstruction detectors learn from normal data, but raw-space errors can confound benign trend changes with anomalies; expressive models may also reconstruct abnormal local perturbations too well. Moreover, failures can appear as broken relationships between variables rather than isolated amplitude changes Deng & Hooi (2021); Zhao et al. (2020); Xu et al. (2022); Wu et al. (2025).

This setting also exposes a mismatch between conventional diffusion generation and anomaly detection. A standard diffusion process is designed to erase the structure of a clean sample and then regenerate the complete sample from an almost unstructured noisy state Ho et al. (2020). At test time in MTSAD, however, the observed window is already available and may contain abnormal values. Driving that window toward full noise discards not only abnormal fluctuations but also useful evidence about its operating level, phase, and slowly varying trend. The reverse model must consequently recover global evolution and local detail at once. For non-stationary signals, uncertainty in this unnecessarily broad recovery task can itself produce large errors on normal trend changes, while a strongly conditioned reconstructor may still reproduce abnormal details. More corruption is therefore not automatically more useful for anomaly discrimination; the recovery path should preserve trustworthy structure and concentrate its capacity on deviations from that structure.

Figure 1 illustrates the motivation for Trend-Conditioned Residual Diffusion(denoted as TCRD). As shown in Figure 1a, the three correlated channels respond differently within the same anomalous interval, with abnormal deviations being more pronounced in some channels than in others. This

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.



(a) Cross-channel coupling disruption. (b) Trend–residual separation.

Figure 1: Motivation for TCRD. (a) Observations from three channels illustrate their cross-channel coupling behavior; the red shaded region denotes the ground-truth anomalous interval. (b) Trend–residual decomposition of the original series, showing the raw data, its smoothed trend, and the residual obtained by subtracting the smoothed trend from the raw data; the red shaded region denotes the ground-truth anomalous interval.

observation highlights the importance of cross-channel coupling and motivates the use of CCAR-Fusion. As shown in Figure 1b, removing the smoothed trend makes normal residuals more stable while rendering anomalous deviations more prominent. Reconstruction in the residual space therefore helps the diffusion model learn normal patterns.

Unlike diffusion detectors that perform masked imputation or raw-sequence denoising Chen et al. (2023); Xiao et al. (2023), TCRD changes the reconstruction target itself. This distinction is also complementary to decomposition- and multi-resolution-based approaches such as D^3R and MODEM Wang et al. (2023); Zhong et al. (2025). Our contributions are: (i) a trend-conditioned residual diffusion formulation for non-stationary MTSAD; (ii) a deterministic residual-bridge reconstruction process; and (iii) CCAR-Fusion for cross-variable context modeling. Experiments on five public datasets show competitive point-adjusted, affiliation-based, and delay-aware performance, while ablations isolate the effects of all three components.

2 Related Work

2.1 Reconstruction and Dependency Modeling for MTSAD

Classical MTSAD methods estimate normal density or decision boundaries, as in Isolation Forest and one-class learning Liu et al. (2008); Sch"olkopf et al. (2001). Deep detectors instead learn temporal dynamics or reconstruct normal observations. OmniAnomaly uses a stochastic recurrent model, while TranAD and Anomaly Transformer exploit attention to characterize temporal deviations Su et al. (2019); Tuli et al. (2022); Xu et al. (2022). These approaches improve representation capacity, but prediction or reconstruction in the raw space requires a single score to absorb slowly evolving trends, periodic structure, and short abnormal perturbations. Normal regime changes can therefore enlarge the same error that is expected to identify anomalies.

Multivariate detectors further model dependencies because a fault may preserve each channel's marginal range while breaking synchronization across channels. GDN and MTAD-GAT encode inter-variable relations with graphs or attention, whereas DCdetector and CATCH develop contrastive or channel-aware representations Deng & Hooi (2021); Zhao et al. (2020); Yang et al. (2023); Wu et al. (2025). TCRD is complementary to these dependency models: CCAR-Fusion summarizes the current multichannel state into a compact shared core and redistributes it without constructing a fixed graph. More importantly, dependency modeling is coupled to residual recovery, so cross-channel evidence helps determine which fine-scale variations are compatible with the retained operating trend.

2.2 Diffusion-Based Anomaly Detection

Diffusion models learn a sequence of recovery problems and offer stable training for complex time-series distributions Ho et al. (2020). ImDiffusion frames detection as masked imputation and uses intermediate denoising outputs for ensemble inference, while DiffAD performs conditional recovery of missing segments Chen et al. (2023); Xiao et al. (2023). These methods demonstrate that progressive recovery can provide multiple anomaly-sensitive views. Their primary objective nevertheless remains recovery of raw values: under heavy corruption, the model must infer both the global state and the local detail that ultimately contributes to the anomaly score.

Recent diffusion detectors address non-stationarity through decomposition, multi-resolution representations, or memory mechanisms Wang et al. (2023); Zhong et al. (2025); Dou et al. (2025). D^{3}R separates dynamically changing components, MODEM combines low- and high-resolution recovery, and DeAnomaly uses robust decomposition with memory augmentation. These studies establish the value of structural temporal information, but TCRD changes a different part of the formulation: the smoothed trend is the terminal state of the diffusion bridge rather than an auxiliary feature, and the residual is the sole prediction target. The reverse trajectory therefore retains the window's low-frequency operating context and allocates recovery capacity to normal fine-scale behavior. This structured endpoint and residual objective directly target the source of unstable reconstruction errors in non-stationary MTSAD.

3 Method

Given a window $\mathbf{x}_0 \in \mathbb{R}^{K \times L}$, where K and L denote the number of channels and the window length, respectively, TCRD learns normal residual patterns and detects deviations through reconstruction error. Figure 2 is the framework of the method: the following five stages correspond exactly to Steps 1–5 in the figure. Detailed distributions and derivations are deferred to Appendix A.

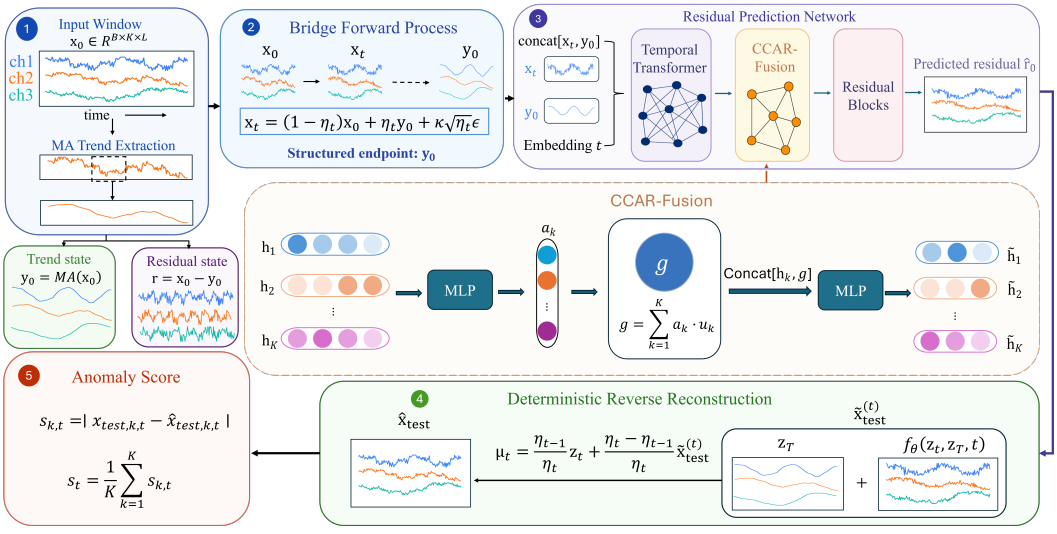


Figure 2: TCRD follows five stages: (1) trend–residual decomposition, (2) bridge forward process, (3) residual prediction network, (4) deterministic reverse reconstruction, and (5) anomaly scoring.

3.1 Step 1: Trend–Residual Decomposition

We first extract a slowly varying trend with a centered moving-average operator MA of window width w and assign the remaining local variation to the residual:

$$\mathbf{y}_0 = MA(\mathbf{x}_0), \qquad \mathbf{r}_0 = \mathbf{x}_0 - \mathbf{y}_0. \tag{1}$$

Boundary values are replicated so that $\mathbf{y}_0$ and $\mathbf{x}_0$ have identical shapes. The trend acts as a structured reference, whereas $\mathbf{r}_0$ is the learning target. This separation prevents the network from spending most of its capacity reproducing benign low-frequency evolution.

3.2 STEP 2: BRIDGE FORWARD PROCESS

Rather than diffusing toward pure noise, TCRD moves the state from the observation toward its trend. For a monotone schedule $0 = \eta_0 < \cdots < \eta_T = 1$, a training state is sampled directly as

$$\mathbf{x}_t = (1 - \eta_t)\mathbf{x}_0 + \eta_t \mathbf{y}_0 + \kappa\sqrt{\eta_t}\,\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{2}$$

Thus, the bridge retains a structured mean throughout the process and ends around $\mathbf{y}_0$. The equivalent Markov transition, endpoint distribution, and posterior are given in Appendix A.1.

3.3 STEP 3: RESIDUAL PREDICTION NETWORK

The predictor receives $\mathbf{x}_t$, the trend condition $\mathbf{y}_0$, and the step embedding t. A Temporal Transformer captures within-channel temporal context Vaswani et al. (2017). CCAR-Fusion then aggregates channel representations $\{\mathbf{h}_k\}_{k=1}^{K}$ into a compact shared core g and redistributes it to every channel:

$$g = \mathrm{Agg}_k(\phi(\mathbf{h}_k)), \qquad \tilde{\mathbf{h}}_k = \psi([\mathbf{h}_k; g]). \tag{3}$$

Here, ϕ and ψ are MLPs and $[\cdot;\cdot]$ denotes concatenation. The fused representation $\tilde{\mathbf{h}}_k$ enriches channel k with global channel context. The fused representations across all channels and temporal positions are passed through subsequent residual blocks and output projections to produce the residual estimate $\hat{\mathbf{r}}_0$. Training minimizes

$$\mathcal{L}_{\mathrm{res}} = \mathbb{E}\left[\frac{1}{K}\sum_{k=1}^{K}\frac{1}{L}\sum_{\ell=1}^{L}\left(\hat{r}_{0,k,\ell} - r_{0,k,\ell}\right)^2\right]. \tag{4}$$

Here, K denotes the number of channels and L denotes the number of temporal positions in an input window. The indices $k \in \{1, \ldots, K\}$ and $\ell \in \{1, \ldots, L\}$ refer to the channel and temporal position, respectively; $r_{0,k,\ell}$ and $\hat{r}_{0,k,\ell}$ are the target and predicted residuals at that location. The dimension-wise channel weights and complete aggregate–redistribute equations appear in Appendix A.2.

3.4 STEP 4: DETERMINISTIC REVERSE RECONSTRUCTION

For a test window $\mathbf{x}_{\mathrm{test}}$, we extract the trend $\mathbf{y}_{\mathrm{test}} = MA(\mathbf{x}_{\mathrm{test}})$ and initialize $\mathbf{z}_T = \mathbf{y}_{\mathrm{test}}$. At each reverse step $t = T, \ldots, 1$, the predictor uses the current state $\mathbf{z}_t$ and fixed trend $\mathbf{z}_T$ to estimate the residual. As shown in Step 4 of Figure 2, the clean estimate and predicted posterior mean are

$$\begin{aligned} \tilde{\mathbf{x}}_{\mathrm{test}}^{(t)} &= \mathbf{z}_T + f_\theta(\mathbf{z}_t, \mathbf{z}_T, t), \\ \boldsymbol{\mu}_t &= \frac{\eta_{t-1}}{\eta_t}\mathbf{z}_t + \frac{\eta_t - \eta_{t-1}}{\eta_t}\tilde{\mathbf{x}}_{\mathrm{test}}^{(t)}. \end{aligned} \tag{5}$$

The update $\mathbf{z}_{t-1} = \boldsymbol{\mu}_t$ combines the current state and clean estimate with nonnegative weights summing to one. No additional Gaussian noise is injected. At $t = 1$, $\eta_0 = 0$ makes the update equal to the clean estimate, giving the final reconstruction $\hat{\mathbf{x}}_{\mathrm{test}} = \mathbf{z}_0$. Appendix A.3 provides the full reconstruction procedure.

3.5 STEP 5: ANOMALY SCORING

For a temporal position t and channel k, we first compute the absolute reconstruction error and then average over all K channels to obtain the anomaly score:

$$\begin{aligned} s_{k,t} &= \left|x_{\mathrm{test},k,t} - \hat{x}_{\mathrm{test},k,t}\right|, \\ s_t &= \frac{1}{K}\sum_{k=1}^{K} s_{k,t}. \end{aligned} \tag{6}$$

Following the ImDiffusion evaluation protocol Chen et al. (2023), we retain ten intermediate reconstructions, threshold their scores using scale-adaptive upper-tail thresholds, and combine the binary decisions by voting. The deterministic trajectory yields repeatable scores for the same input. Appendix A.3 specifies the thresholding and voting procedure.

4 EXPERIMENTS

4.1 DATASETS AND EXPERIMENTAL SETUP

This paper evaluates the proposed method on five public multivariate time-series anomaly detection datasets, namely SMD Su et al. (2019), PSM Abdulaal et al. (2021), SMAP Hundman et al. (2018), MSL Hundman et al. (2018), and SWaT Mathur & Tippenhauer (2016). These datasets cover representative real-world scenarios, including server monitoring, spacecraft telemetry, and industrial control systems. Specifically, SMD and PSM are collected from server monitoring systems, SMAP and MSL are NASA spacecraft telemetry datasets, and SWaT is collected from a secure water treatment industrial control testbed. The statistics of these datasets are summarized in Table 1.

Table 1: Statistics of the datasets used in the experiments.

Dataset	Entities	Dimensions	Train	Test	Anomaly Ratio
SMD	28 machines	38	708,405	708,420	4.16%
PSM	1	25	132,481	87,841	27.76%
SMAP	55 sequences	25	140,825	444,035	12.83%
MSL	27 sequences	55	58,317	73,729	10.48%
SWaT	1	51	495,000	449,919	12.14%

For a comprehensive evaluation, this paper compares TCRD with representative multivariate time-series anomaly detection methods from different categories, including traditional machine learning methods, recurrent models, reconstruction-based models, graph neural network methods, generative models, Transformer-based methods, and diffusion-based methods. The compared methods include IForest Liu et al. (2008), LSTM-AD Malhotra et al. (2015), MSCRED Zhang et al. (2019), GDN Deng & Hooi (2021), MTAD-GAT Zhao et al. (2020), OmniAnomaly Su et al. (2019), InterFusion Li et al. (2021), BeatGAN Zhou et al. (2019), MAD-GAN Li et al. (2019), Anomaly Transformer Xu et al. (2022), TFAD Zhang et al. (2022), TranAD Tuli et al. (2022), NPSR Lai et al. (2023), DiffAD Xiao et al. (2023), ImDiffusion Chen et al. (2023), D^{3}R Wang et al. (2023), and MODEM Zhong et al. (2025).

For the main comparison and ablation studies, we evaluate TCRD using the ensemble-based procedure in the official ImDiffusion implementation Chen et al. (2023), as specified in Eqs. (18)–(20), and report point-adjusted precision, recall, and F1 (PA-F1). We repeat training and evaluation six times with independent random seeds and average each metric across runs. The reported PA-F1 is the mean of the run-level scores, rather than the harmonic mean of the averaged precision and recall. We additionally report affiliation-based precision, recall, and F1 (Aff-P, Aff-R, and Aff-F1), following the evaluation protocol of D^{3}R Wang et al. (2023), to evaluate interval localization.

Unless otherwise specified, we retain the main training configuration used in ImDiffusion Chen et al. (2023) for protocol consistency. The model is trained for 100 epochs with a batch size of 16 and a learning rate of 1.0×10^{-3}. The number of multi-head attention heads is set to 8, and the bridge-step embedding dimension is 128. The number of bridge steps is set to $T = 50$. For TCRD-specific components, a centered moving-average operator with an odd kernel size of 25 is used to construct the smoothed trend state. The bridge perturbation coefficient is set to $\kappa = 1.0$. We set $\eta_0 = 0$, while the nonzero bridge coefficients $\eta_1, \ldots, \eta_T$ are linearly spaced from 10^{-4} to 1.0. The channel-core dimension of CCAR-Fusion is set to 8.

All experiments are conducted in the same computing environment. The model is implemented based on the PyTorch 2.5.1 deep learning framework Paszke et al. (2019) and trained and inferred using a single NVIDIA GeForce RTX 4090 GPU with 24 GB of memory.

4.2 COMPARISON EXPERIMENTS

Table 2 reports the main comparison using point-adjusted metrics. The comparison covers classical reconstruction-based, graph-based, and Transformer-based detectors, as well as recent diffusion-based methods including DiffAD, ImDiffusion, D^{3}R, and MODEM.

Table 2: Comparison using point-adjusted precision (P), recall (R), and F1 (PA-F1), in %. The best and second-best performances are highlighted in bold and underlined, respectively. Arithmetically inconsistent baseline F1 entries are recomputed from their reported P and R as $2PR/(P+R)$. TCRD results are averaged over six independent runs.

Method	SMD			PSM			MSL			SWaT			SMAP		
	P	R	PA-F1	P	R	PA-F1	P	R	PA-F1	P	R	PA-F1	P	R	PA-F1
IForest	20.30	21.30	17.99	66.30	49.19	56.41	60.59	53.28	53.34	97.64	66.50	79.07	28.86	76.71	41.63
LSTM-AD	33.61	32.29	26.39	90.50	77.07	83.13	73.30	57.45	63.78	**99.25**	67.37	80.26	78.41	56.30	65.33
MSCRED	85.67	90.38	84.26	95.55	68.57	79.65	50.08	60.88	48.99	84.23	40.66	54.84	41.07	86.04	55.60
GDN	84.60	78.62	78.65	87.50	83.85	85.64	86.68	80.27	83.35	13.11	5.85	8.08	96.89	54.01	69.36
MTAD-GAT	88.36	83.30	84.63	87.63	87.25	87.44	84.68	82.24	83.44	84.68	82.24	83.44	**97.18**	52.59	68.24
OmniAnomaly	87.51	90.52	87.75	92.55	95.51	94.01	83.21	82.15	82.68	97.49	75.00	84.70	84.07	96.74	89.95
InterFusion	88.15	90.71	87.72	91.28	93.26	92.26	76.88	**94.64**	84.42	96.83	85.30	90.60	87.88	77.04	82.10
BeatGAN	90.13	88.94	87.97	92.04	87.67	89.75	**96.06**	70.20	81.07	96.06	70.20	81.07	89.15	67.81	76.63
MAD-GAN	88.51	90.45	88.03	85.96	88.38	86.98	70.47	78.41	74.23	79.18	54.23	63.85	96.40	54.74	69.82
AnomalyTransformer	87.26	89.70	88.46	92.31	93.29	92.80	86.34	91.08	88.64	80.57	82.34	81.45	85.36	88.03	86.67
TFAD	92.42	91.36	91.89	90.27	97.21	93.61	87.15	89.78	88.44	73.26	84.59	78.52	83.66	86.72	85.16
TranAD	89.06	89.82	87.85	95.06	91.47	92.20	89.51	92.97	**91.15**	70.25	72.66	68.86	82.24	85.02	83.60
NPSR	87.69	90.65	89.15	87.61	94.73	91.02	83.62	85.53	84.56	79.35	83.29	81.23	82.64	86.42	84.49
DiffAD	90.32	95.71	92.94	96.22	97.70	96.95	85.59	88.32	86.93	86.64	88.35	87.49	91.08	87.39	89.19
ImDiffusion	94.53	94.48	94.50	97.72	96.83	97.27	87.74	84.65	86.16	89.88	84.65	87.09	87.71	96.18	91.75
D^3R	86.50	95.34	90.71	88.17	93.25	90.63	78.48	94.33	85.67	80.17	87.40	83.63	86.80	91.76	89.21
MODEM	95.70	96.32	96.01	96.97	**98.35**	97.65	91.28	88.32	89.77	89.42	93.08	91.21	88.50	**97.22**	**92.66**
Ours	**95.81**	**97.36**	**96.58**	**98.75**	97.74	**98.24**	89.37	92.75	90.99	96.24	**94.62**	**95.43**	82.68	96.90	89.23

Overall, TCRD achieves strong and competitive performance across the five benchmark datasets. It obtains the highest F1-score on SMD, PSM, and SWaT, the second-highest F1-score on MSL, while its performance on SMAP remains lower than MODEM and ImDiffusion. These results indicate that trend-conditioned residual recovery is particularly effective on several datasets, although its benefit is dataset-dependent.

4.3 Ablation Study

To isolate the contributions of the diffusion bridge, trend-conditioned residual formulation, and channel-coupling module, we compare the full TCRD model with five degraded variants. *w/o Diffusion Bridge* removes the iterative diffusion-bridge reconstruction process. *Raw-Sequence Diffusion* retains diffusion but changes its target from the trend residual to the raw sequence $\mathbf{x}_0$. *w/o CCAR-Fusion* removes cross-variable aggregate-redistribute fusion, whereas *w/o Bridge-Trend* removes the proposed trend-conditioned bridge mechanism. Finally, *w/o both* removes both Bridge-Trend and CCAR-Fusion. The experiments are performed on four representative datasets, including SMD, PSM, MSL, and SWaT, and the results are reported in Table 3.

Table 3: Ablation study results (PA-F1, %) on representative benchmark datasets. All entries are averaged over six independent runs.

Dataset	w/o Diffusion Bridge	Raw-Sequence Diffusion	w/o CCAR-Fusion	w/o Bridge-Trend	w/o both	Ours
SMD	95.34	95.87	94.88	94.43	94.14	**96.58**
PSM	96.96	96.18	97.97	96.52	96.64	**98.24**
MSL	88.38	80.62	86.17	89.09	86.48	**90.99**
SWaT	92.21	86.98	93.38	84.46	85.40	**95.43**

Table 3 shows that the full TCRD model achieves the highest F1-score on all four datasets. Removing the diffusion bridge consistently lowers F1, most clearly on SWaT (95.43% to 92.21%), showing that iterative bridge reconstruction contributes beyond the underlying prediction network.

Applying diffusion directly to the raw sequence instead of the trend residual sharply reduces F1 on MSL (90.99% to 80.62%) and SWaT (95.43% to 86.98%). This controlled comparison supports recovering residuals around a smoothed trend state in non-stationary settings.

Removing Bridge-Trend also consistently degrades detection performance, with the largest decrease occurring on SWaT, from 95.43% to 84.46%. Removing CCAR-Fusion reduces F1-score on every dataset and has its largest effect on MSL, where performance decreases from 90.99% to 86.17%. The *w/o both* variant remains below the full model across all datasets. Together, these comparisons show that the diffusion bridge and trend-conditioned residual target provide the main reconstruction advantages, while CCAR-Fusion supplies complementary cross-variable information.

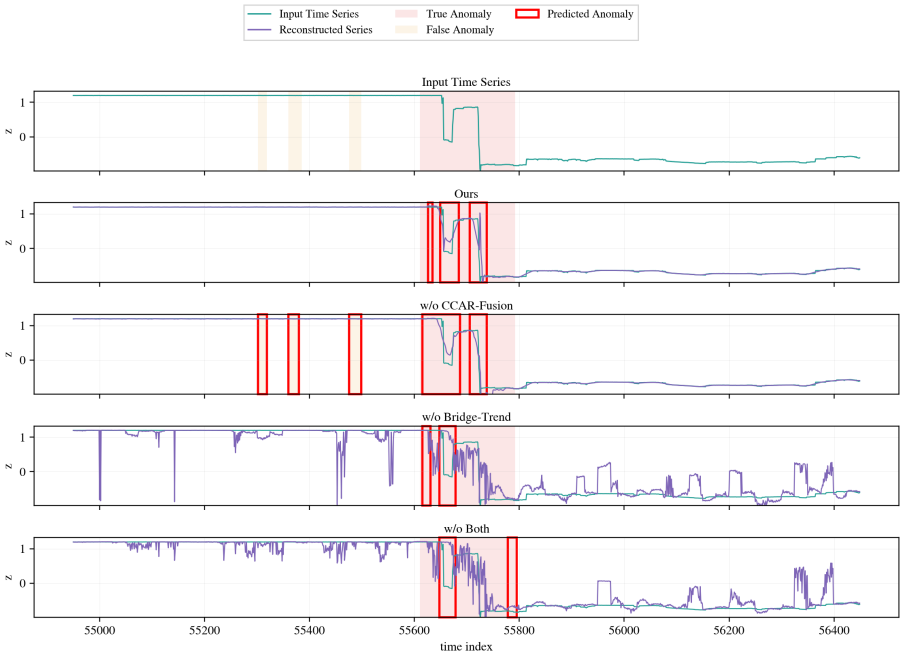


Figure 3: Qualitative ablation on MSL. Red and yellow shading indicate ground-truth and false-positive regions, respectively; boxes mark predicted anomalies. Removing either proposed component introduces less stable reconstruction or additional false alarms.

4.4 Detection Timeliness and Interval Localization

To complement point-adjusted accuracy, we evaluate how promptly anomalies are detected and how closely predicted intervals align with ground truth. Following ImDiffusion Chen et al. (2023), we measure detection timeliness using Average Sequence Detection Delay (ADD). Given S anomalous events, ADD is defined as

$$\mathrm{ADD} = \frac{1}{S}\sum_{i=1}^{S}(T_i - q_i)\,, \tag{7}$$

where q_i is the starting time of anomalous event i, $T_i \geq q_i$ is its corresponding detection time, and S is the total number of anomalous events. A smaller ADD indicates more timely anomaly detection.

For interval localization, we report affiliation-based precision, recall, and F1 (Aff-P, Aff-R, and Aff-F1), following the evaluation protocol of D^3R Wang et al. (2023). These metrics assess the temporal correspondence between predicted and ground-truth anomalous intervals, with Aff-F1 summarizing the balance between Aff-P and Aff-R. Point adjustment credits an entire anomalous interval after a single hit (Appendix A.3); affiliation metrics instead assess interval alignment, complementing ADD's measurement of response speed.

Table 4: Average Sequence Detection Delay (mean ± standard deviation); lower is better. The average uses the four displayed datasets.

Method	SMD	PSM	MSL	SWaT	Average
IForest	90 ± 1	191 ± 17	123 ± 28	539 ± 20	236 ± 10
LSTM-AD	87 ± 1	224 ± 54	115 ± 29	627 ± 4	263 ± 15
MSCRED	32 ± 0	218 ± 35	109 ± 30	1065 ± 339	356 ± 86
GDN	38 ± 1	148 ± 0	106 ± 2	1478 ± 0	443 ± 1
MTAD-GAT	90 ± 100	182 ± 0	96 ± 17	482 ± 80	213 ± 32
OmniAnomaly	26 ± 1	121 ± 11	93 ± 2	550 ± 48	198 ± 12
InterFusion	$\underline{22 \pm 2}$	40 ± 10	$\underline{32 \pm 15}$	454 ± 141	137 ± 36
BeatGAN	38 ± 2	166 ± 11	68 ± 24	607 ± 6	220 ± 7
MAD-GAN	59 ± 57	122 ± 2	88 ± 0	926 ± 337	299 ± 85
TranAD	24 ± 0	127 ± 4	56 ± 12	657 ± 246	216 ± 62
DiffAD	48 ± 3	134 ± 7	74 ± 9	437 ± 50	173 ± 13
ImDiffusion	24 ± 1	28 ± 1	46 ± 4	350 ± 43	112 ± 11
D^3R	65 ± 16	172 ± 6	90 ± 12	643 ± 46	243 ± 13
MODEM	27 ± 2	$\mathbf{26 \pm 0}$	52 ± 5	$\underline{343 \pm 26}$	$\underline{112 \pm 7}$
TCRD	$\mathbf{16 \pm 3}$	$\underline{26 \pm 5}$	$\mathbf{24 \pm 6}$	$\mathbf{249 \pm 37}$	$\mathbf{79 \pm 10}$

Table 5: Comparison using affiliation-based precision (Aff-P), recall (Aff-R), and F1 (Aff-F1), in %, following the evaluation protocol of D^3R Wang et al. (2023). Best and second-best Aff-F1 values are bold and underlined.

Method	PSM			SMD			SWaT		
	Aff-P	Aff-R	Aff-F1	Aff-P	Aff-R	Aff-F1	Aff-P	Aff-R	Aff-F1
Sampling	84.39	51.65	64.08	74.53	31.44	44.22	60.62	84.66	70.65
ECOD	74.60	33.84	46.56	73.98	16.15	26.51	97.61	11.51	20.59
PCA	92.20	37.71	53.53	83.88	40.19	54.34	63.58	72.18	67.61
CBLOF	59.90	98.45	74.49	86.67	33.52	48.34	63.08	70.91	66.77
IForest	100.00	3.35	6.48	100.00	9.37	17.13	61.17	70.14	65.35
LODA	92.66	40.17	56.05	59.02	66.18	62.60	61.77	70.41	65.35
VAE	62.21	87.72	72.80	82.09	43.49	58.66	60.11	73.19	66.02
DeepSVDD	74.05	50.64	60.15	64.98	46.77	54.13	59.11	72.25	65.69
LSTM-AE	75.11	75.86	75.48	84.96	43.49	57.53	60.18	72.19	65.64
MTAD-GAT	79.90	60.14	68.63	85.90	67.69	75.71	65.09	77.51	71.23
TFAD	79.14	71.63	75.20	56.32	97.80	71.49	60.38	71.96	69.53
AnomTrans.	52.01	85.04	64.55	100.00	3.19	6.19	55.41	59.94	57.59
D^3R	62.94	96.19	$\underline{76.09}$	77.15	99.26	**86.82**	69.85	79.51	$\underline{74.36}$
TCRD	73.58	88.47	**80.34**	88.15	81.83	$\underline{84.87}$	67.95	87.08	**76.33**

As reported in Table 4, TCRD obtains the lowest ADD on SMD, MSL, and SWaT, ties MODEM in mean delay on PSM, and reduces the four-dataset average from 112 to 79 relative to both ImDiffusion and MODEM.

TCRD ranks first in Aff-F1 on PSM and SWaT and second on SMD (Table 5). On PSM, TCRD achieves 80.34% Aff-F1, exceeding the runner-up D^3R (76.09%) by 4.25 percentage points. On SWaT, TCRD improves Aff-F1 from 74.36% to 76.33% (+1.97 percentage points) over D^3R, with higher Aff-R (87.08% vs. 79.51%) and slightly lower Aff-P (67.95% vs. 69.85%).

4.5 Qualitative Analysis

Figure 4 illustrates trend–residual separation (Step 1): the smoothed endpoint retains the dominant operating trajectory while local deviations concentrate in the residual. Figure 5 depicts anomaly scoring (Step 5): the complete model produces a focused score peak inside the anomalous interval, whereas removing Bridge-Trend or CCAR-Fusion introduces less selective responses. Together, the two views are consistent with the quantitative ablations.

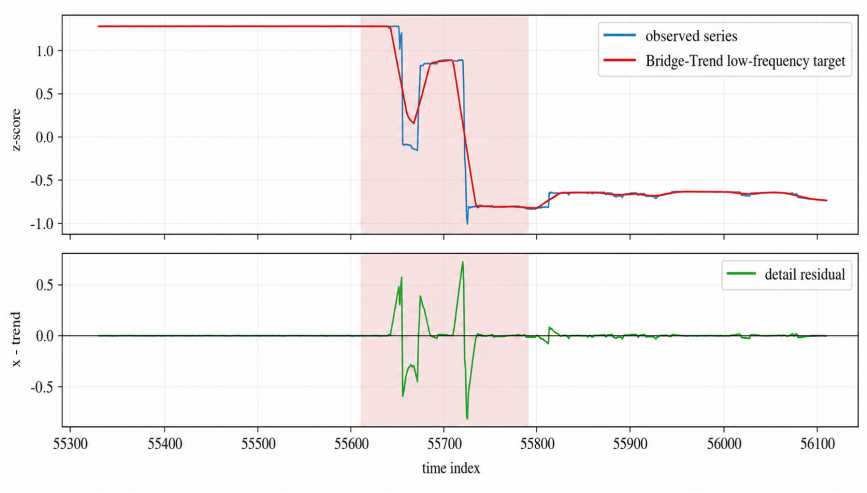


Figure 4: Trend extraction and residual separation on MSL. The shaded region denotes the ground-truth anomalous interval.

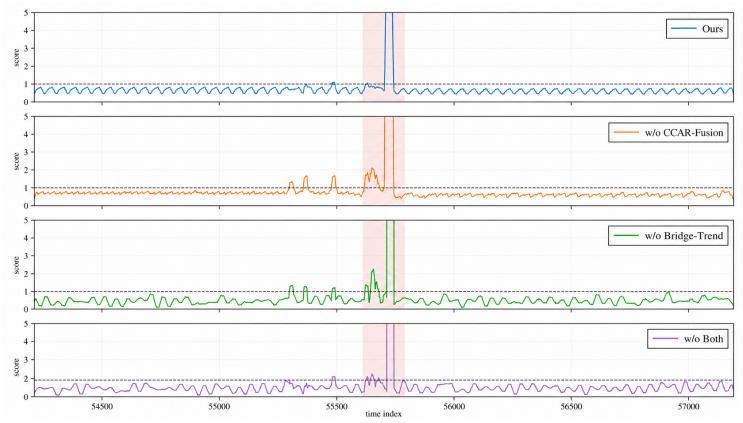


Figure 5: Anomaly-score responses under model variants on MSL. The shaded region denotes the ground-truth anomalous interval, and dashed lines denote detection thresholds.

5 Conclusion

We introduced TCRD, which replaces raw-sequence diffusion with residual recovery conditioned on a smoothed trend and augments the predictor with cross-channel fusion. Results on five public MTSAD benchmarks and targeted ablations support the diffusion bridge, residual objective, and CCAR-Fusion. Current limitations include sensitivity to the fixed smoothing window, reduced precision on long drift-like anomalies, and the practical deployment of iterative diffusion reconstruction in real-time monitoring scenarios. Adaptive trend extraction, calibrated scoring, and accelerated sampling are promising directions.

Reproducibility Statement

To support reproducibility, the submission includes an anonymized reference implementation of the core TCRD architecture, representative configurations, and lightweight verification scripts, while detailed implementation and experimental settings are documented in the appendix. The complete training and evaluation pipeline will be released upon publication.

AI USE STATEMENT

Generative AI tools were used to assist with manuscript polishing, related-work discovery, and limited experimental execution. All scientific content, experimental results, and final claims were reviewed and verified by the authors.

REFERENCES

Ahmed Abdulaal, Zhuanghua Liu, and Tomer Lancewicki. Practical approach to asynchronous multivariate time series anomaly detection and localization. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2485–2494, 2021.

Yuhang Chen, Chaoyun Zhang, Minghua Ma, Yudong Liu, Ruomeng Ding, Bowen Li, Shilin He, Saravan Rajmohan, Qingwei Lin, and Dongmei Zhang. ImDiffusion: Imputed diffusion models for multivariate time series anomaly detection. *arXiv preprint arXiv:2307.00754*, 2023.

Ailin Deng and Bryan Hooi. Graph neural network-based anomaly detection in multivariate time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 4027–4035, 2021.

Hui Dou, Pengcheng Shi, Yiwen Zhang, Pengfei Chen, and Zibin Zheng. DeAnomaly: Anomaly detection for multivariate time series using robust decomposition and memory-augmented diffusion models. *IEEE Transactions on Instrumentation and Measurement*, 74:1–14, 2025. doi: 10.1109/TIM.2025.3570337.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020.

Kyle Hundman, Valentino Constantinou, Christopher Laporte, Ian Colwell, and Tom Soderstrom. Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 387–395, 2018.

Chih-Yu Lai, Fan-Keng Sun, Zhengqi Gao, Jeffrey H. Lang, and Duane S. Boning. Nominality score conditioned time series anomaly detection by point/sequential reconstruction. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Dan Li, Dacheng Chen, Baihong Jin, Lei Shi, Jonathan Goh, and See-Kiong Ng. MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks. In *Artificial Neural Networks and Machine Learning–ICANN 2019*, pp. 703–716. Springer, 2019.

Zhihan Li, Youjian Zhao, Jiaqi Han, Ya Su, Rui Jiao, Xidao Wen, and Dan Pei. Multivariate time series anomaly detection and interpretation using hierarchical inter-metric and temporal embedding. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 3220–3230, 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, 2008.

Pankaj Malhotra, Lovekesh Vig, Gautam Shroff, and Puneet Agarwal. Long short term memory networks for anomaly detection in time series. In *Proceedings of the 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pp. 89–94, 2015.

Aditya P. Mathur and Nils Ole Tippenhauer. SWaT: A water treatment testbed for research and training on ICS security. In *2016 International Workshop on Cyber-Physical Systems for Smart Water Networks*, pp. 31–36, 2016.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Bernhard Sch"olkopf, John C. Platt, John Shawe-Taylor, Alex J. Smola, and Robert C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, 2001.

Ya Su, Youjian Zhao, Chenhao Niu, Rong Liu, Wei Sun, and Dan Pei. Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2828–2837, 2019.

Shreshth Tuli, Giuliano Casale, and Nicholas R. Jennings. TranAD: Deep transformer networks for anomaly detection in multivariate time series data. *Proceedings of the VLDB Endowment*, 15(6): 1201–1214, 2022.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Chengsen Wang, Zirui Zhuang, Qi Qi, Jingyu Wang, Xingyu Wang, Haifeng Sun, and Jianxin Liao. Drift doesn't matter: Dynamic decomposition with diffusion reconstruction for unstable multivariate time series anomaly detection. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

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 *International Conference on Learning Representations*, 2025.

Chunjing Xiao, Zehua Gou, Wenxin Tai, Kunpeng Zhang, and Fan Zhou. Imputation-based time-series anomaly detection with conditional weight-incremental diffusion models. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2742–2751, 2023. doi: 10.1145/3580305.3599391.

Jiehui Xu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Anomaly transformer: Time series anomaly detection with association discrepancy. In *International Conference on Learning Representations*, 2022.

Yiyuan Yang, Chaoli Zhang, Tian Zhou, Qingsong Wen, and Liang Sun. DCdetector: Dual attention contrastive representation learning for time series anomaly detection. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 3033–3045, 2023.

Chaoli Zhang, Tian Zhou, Qingsong Wen, and Liang Sun. TFAD: A decomposition time series anomaly detection architecture with time-frequency analysis. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*, pp. 2497–2507, 2022. doi: 10.1145/3511808.3557470.

Chuxu Zhang, Dongjin Song, Yuncong Chen, Xinyang Feng, Cristian Lumezanu, Wei Cheng, Jingchao Ni, Bo Zong, Haifeng Chen, and Nitesh V. Chawla. A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 1409–1416, 2019.

Hang Zhao, Yujing Wang, Juanyong Duan, Congrui Huang, Defu Cao, Yunhai Tong, Bixiong Xu, Jing Bai, Jie Tong, and Qi Zhang. Multivariate time-series anomaly detection via graph attention network. In *2020 IEEE International Conference on Data Mining*, pp. 841–850, 2020.

Guojin Zhong, Pan Wang, Jin Yuan, Zhiyong Li, and Long Chen. Multi-resolution decomposable diffusion model for non-stationary time series anomaly detection. In *International Conference on Learning Representations*, 2025.

Bin Zhou, Shenghua Liu, Bryan Hooi, Xueqi Cheng, and Jing Ye. BeatGAN: Anomalous rhythm detection using adversarially generated time series. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, pp. 4433–4439, 2019. doi: 10.24963/ijcai.2019/616.

A Detailed Model Formulation

This appendix expands the five-step presentation in Section 3. It provides the complete bridge distribution, posterior, CCAR-Fusion operator, deterministic reverse update, and anomaly score while preserving the notation used in Figure 2.

A.1 Trend decomposition and bridge distribution

For an odd moving-average width w and $h = (w-1)/2$, Step 1 computes

$$y_{0,k,\ell} = \frac{1}{w} \sum_{j=-h}^{h} x_{0,k,\ell+j}, \qquad \mathbf{r}_0 = \mathbf{x}_0 - \mathbf{y}_0, \tag{8}$$

with replicated boundary values. In Step 2, let

$$0 = \eta_0 < \eta_1 < \cdots < \eta_T = 1, \qquad \alpha_t = \eta_t - \eta_{t-1}. \tag{9}$$

The residual-shifting Markov transition is

$$q(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{x}_0, \mathbf{y}_0) = \mathcal{N}(\mathbf{x}_{t-1} - \alpha_t \mathbf{r}_0, \kappa^2 \alpha_t \mathbf{I}). \tag{10}$$

Because $\sum_{i=1}^{t} \alpha_i = \eta_t$, its marginal is

$$q(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{y}_0) = \mathcal{N}\big((1-\eta_t)\mathbf{x}_0 + \eta_t \mathbf{y}_0, \kappa^2 \eta_t \mathbf{I}\big)\,, \tag{11}$$

which directly gives the sampling rule in Eq. (2). At $t = T$, the mean is the structured endpoint $\mathbf{y}_0$ and $q(\mathbf{x}_T \mid \mathbf{x}_0, \mathbf{y}_0) = \mathcal{N}(\mathbf{y}_0, \kappa^2 \mathbf{I})$.

For $t \geq 2$, Gaussian conditioning yields

$$\begin{aligned} q(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}_0, \mathbf{y}_0) &= \mathcal{N}(\tilde{\boldsymbol{\mu}}_t, \tilde{\sigma}_t^2 \mathbf{I}), \\ \tilde{\boldsymbol{\mu}}_t &= \frac{\eta_{t-1}}{\eta_t}\mathbf{x}_t + \frac{\alpha_t}{\eta_t}\mathbf{x}_0, \\ \tilde{\sigma}_t^2 &= \kappa^2 \frac{\eta_{t-1}}{\eta_t}\alpha_t. \end{aligned} \tag{12}$$

Replacing the unknown $\mathbf{x}_0$ with $\hat{\mathbf{x}}_0 = \mathbf{y}_0 + f_\theta(\mathbf{x}_t, \mathbf{y}_0, t)$ parameterizes the reverse distribution. Although posterior matching induces a timestep-dependent coefficient, we use the unweighted residual objective in Eq. (4), which directly optimizes the target used by the network.

A.2 CCAR-Fusion details

At a temporal position, let $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_K] \in \mathbb{R}^{K \times d}$. A two-layer MLP maps every channel to a d_c-dimensional candidate core,

$$\mathbf{u}_k = \mathrm{Linear}_2(\mathrm{GELU}(\mathrm{Linear}_1(\mathbf{h}_k))). \tag{13}$$

CCAR-Fusion normalizes each core dimension over channels and aggregates it:

$$a_{k,j} = \frac{\exp(u_{k,j})}{\sum_{i=1}^{K} \exp(u_{i,j})}, \qquad g_j = \sum_{k=1}^{K} a_{k,j} u_{k,j}, \quad j = 1, \ldots, d_c. \tag{14}$$

The shared core $\mathbf{g} = [g_1, \ldots, g_{d_c}]$ is then redistributed:

$$\tilde{\mathbf{h}}_k = \mathrm{Linear}_4(\mathrm{GELU}(\mathrm{Linear}_3([\mathbf{h}_k; \mathbf{g}]))). \tag{15}$$

This realizes the compact operator in Eq. (3) without explicitly constructing all K^2 pairwise channel interactions.

A.3 Deterministic reconstruction and scoring

For Step 4, initialize $\mathbf{z}_T = \mathbf{y}_{\text{test}}$ and, for $t = T, \dots, 1$, compute

$$\begin{aligned} \hat{\mathbf{r}}^{(t)} &= f_\theta(\mathbf{z}_t, \mathbf{y}_{\text{test}}, t), \\ \tilde{\mathbf{x}}_{\text{test}}^{(t)} &= \mathbf{y}_{\text{test}} + \hat{\mathbf{r}}^{(t)}, \\ \mathbf{z}_{t-1} = \boldsymbol{\mu}_t &= \frac{\eta_{t-1}}{\eta_t}\mathbf{z}_t + \frac{\alpha_t}{\eta_t}\tilde{\mathbf{x}}_{\text{test}}^{(t)}. \end{aligned} \tag{16}$$

No Gaussian term is added, so this trajectory follows the learned posterior mean. The final reconstruction is $\hat{\mathbf{x}}_{\text{test}} = \mathbf{z}_0$. Step 5 then evaluates Eq. (6); explicitly,

$$\begin{aligned} s_{k,t} &= \left| x_{\text{test},k,t} - \hat{x}_{\text{test},k,t} \right|, \\ s_t &= \frac{1}{K}\sum_{k=1}^{K} s_{k,t}. \end{aligned} \tag{17}$$

Here, t indexes temporal positions in the test window, $s_{k,t}$ is the absolute error for channel k, and s_t is its mean over channels.

Ensemble inference. For compatibility with the ImDiffusion evaluation protocol Chen et al. (2023), we retain the ten intermediate reconstructions indexed by $\mathcal{S} = \{0, 3, 6, \dots, 27\}$. Let $e_i(\ell)$ denote the channel-averaged score in Eq. (6) at temporal position $t = \ell$, computed from the reconstruction at selected step $i \in \mathcal{S}$, and let $\bar{E}_i = L^{-1}\sum_{\ell=1}^{L} e_i(\ell)$ be its mean error. Using step 0 as the reference and an upper-tail proportion $\rho_0 = 0.02$, the proportion at step i is adjusted for its error scale as

$$\rho_i = \frac{\bar{E}_0}{\bar{E}_i}\rho_0, \qquad \tau_i = Q_{1-\rho_i}\left(\{e_i(\ell)\}_{\ell=1}^{L}\right), \tag{18}$$

where Q_q is the empirical q-quantile. Equivalently, τ_i is the boundary obtained by ranking the scores and retaining the largest $\max(\lfloor \rho_i L \rfloor, 1)$ values. The step-wise decisions and the accumulated vote are

$$y_{i,\ell} = \mathbb{I}[e_i(\ell) > \tau_i], \qquad v_\ell = \sum_{i \in \mathcal{S}} y_{i,\ell}. \tag{19}$$

Following the official ImDiffusion implementation Chen et al. (2023), we select the voting threshold that maximizes point-adjusted F1 (PA-F1) on the test set, with ground-truth labels $\mathbf{b} \in \{0, 1\}^L$:

$$\xi^* = \arg\max_{\xi \in \{0,\dots,9\}} F_1\left(\mathrm{PA}\left(\{\mathbb{I}[v_\ell > \xi]\}_{\ell=1}^{L}, \mathbf{b}\right), \mathbf{b}\right). \tag{20}$$

Here, PA marks all points in a ground-truth anomalous interval as detected if at least one point in that interval is detected.

B Additional Experimental Results

B.1 Additional Qualitative Example

Fig. 6 presents a qualitative detection example. The predicted anomalous intervals show clear overlap with the ground-truth anomalous regions, while most normal regions maintain relatively low reconstruction errors. This case further illustrates how the proposed residual-recovery framework produces anomaly responses around abnormal segments.

C Extended Discussion and Future Work

In this paper, we propose TCRD, a trend-conditioned residual diffusion framework for unsupervised multivariate time-series anomaly detection. Different from existing reconstruction-based methods that directly model the original non-stationary series, TCRD first extracts a smoothed trend state and then guides the diffusion process to recover fine-grained residual information. In addition, CCAR-Fusion is introduced to enhance cross-channel dependency modeling, enabling the model to

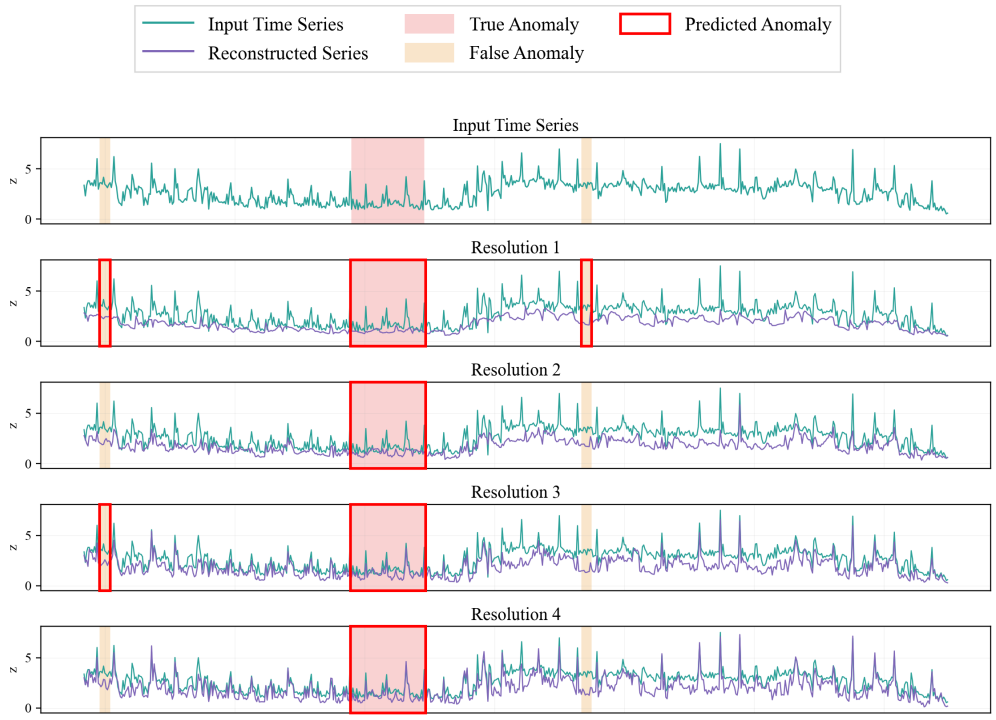


Figure 6: Qualitative detection example of TCRD on a test segment. The shaded regions denote ground-truth anomalous intervals or false-alarm regions, and the predicted anomalous intervals are marked by boxes.

better capture coupling changes among variables. Extensive experiments on multiple public datasets show that TCRD achieves competitive detection performance under point-wise, delay-based, and affiliation-based evaluation protocols. The ablation results show consistent degradation when the diffusion bridge is removed or when diffusion directly targets the raw sequence, thereby supporting both the bridge process and the trend-conditioned residual objective. The ablation and visualization results also verify the complementary role of CCAR-Fusion.

Nevertheless, TCRD still has several limitations. First, although the proposed method achieves strong recall and competitive F1-scores, its precision can still be limited on datasets with long-duration anomalous trends, such as SMAP. This is because long-lasting drift-like anomalies may be partially absorbed by the smoothed trend state, making the residual abnormal patterns less distinguishable. Second, the current trend extraction strategy relies on a predefined moving-average window. Although this design is simple and effective, a fixed window may not be optimal for all datasets, especially when anomalies have long durations or when the normal trend changes rapidly. Third, the effect of CCAR-Fusion can be dataset-dependent, suggesting that cross-channel coupling modeling still needs to be more adaptive to different multivariate systems. Finally, the practical deployment of iterative diffusion reconstruction in real-time monitoring scenarios remains to be investigated.

In future work, we plan to improve TCRD from the following directions. First, we will explore adaptive or learnable trend decomposition strategies to replace the fixed moving-average target, so that the model can better handle different types of non-stationary patterns. Second, we will investigate more effective threshold selection and anomaly score calibration strategies to reduce false alarms and improve precision, especially for datasets with long-duration drift-like anomalies. Third, we will further optimize the diffusion sampling process through acceleration or distillation techniques to make inference more practical while maintaining detection accuracy. Finally, we will extend the proposed framework to more real-world multivariate monitoring scenarios and investigate more robust channel-dependency modeling methods for complex dynamic systems.