2024 IEEE Conference on Artificial Intelligence (CAI) | 979-8-3503-5409-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/CAI59869.2024.00226

Self-Supervised Modular Architecture for Multi-Sensor Anomaly Detection and Localization

Mohammed Ayalew Belay*, Adil Rasheed†, Pierluigi Salvo Rossi*‡
* Department of Electronic Systems, Norwegian University of Science and Technology, Trondheim, Norway
† Department of Engineering Cybernetics, Norwegian University of Science and Technology, Trondheim, Norway
‡ Department of Gas Technology, SINTEF Energy Research, Trondheim, Norway
Emails: mohammed.a.belay@ntnu.no, adil.rasheed@ntnu.no, salvorossi@ieee.org

***Abstract*—In this paper, we propose a novel modular architecture for self-supervised multi-sensor anomaly detection and localization. The framework consists of a spatio-temporal encoder for representation learning, a decoder for latent reconstruction, a predictive memory network for sub-sequence pattern identification, and a denoiser for false-positive reduction. It uniquely combines a reconstruction and latent prediction network and optimizes the modules in an end-to-end mechanism to minimize the combined weighted loss. We demonstrate the flexibility and efficiency of our architecture by introducing different components for each module, showcasing its adaptability and enhanced performance in anomaly detection and localization.**

***Index Terms*—Multi-sensor anomaly detection, unsupervised learning, and anomaly localization**

I. Introduction

Anomaly detection in multi-sensor systems is becoming increasing crucial in various application domains, including industrial monitoring [1], network security [2]–[4], medical applications [5], and autonomous vehicles [6]. The massive collection of multi-sensor data led to the development of several data-driven statistical and machine learning methods for anomaly detection and localization. These tools are implemented in supervised, semi-supervised, or unsupervised learning modes [7]. Multi-sensor anomaly detection faces challenges in supervised and semi-supervised approaches, including labeled data requirement, inability to detect unknown anomalies, data imbalance problem, re-training for dynamic environments, and overfitting problem. Consequently, self-supervised and unsupervised anomaly detection methods are increasingly favored for their ability to overcome these limitations [8]. These methods includes IF [9], OC-SVM [10], USAD [11], DAGMM [12], MAD-GAN [13],TranAD [14], GTA [15].

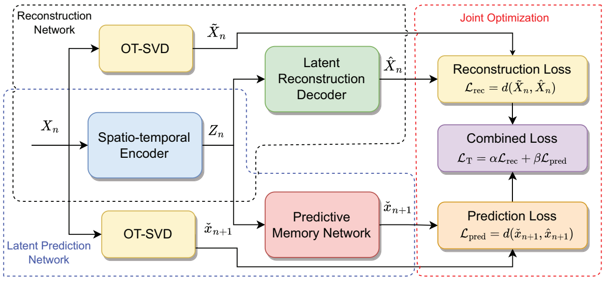

Reconstruction Network
OT-SVD
$\tilde{X}_n$
Latent Reconstruction Decoder
$\hat{X}_n$
Joint Optimization
Reconstruction Loss
$\mathcal{L}_{\text{rec}} = d(\tilde{X}_n, \hat{X}_n)$
X_n
Spatio-temporal Encoder
Z_n
Combined Loss
$\mathcal{L}_{\text{T}} = \alpha\mathcal{L}_{\text{rec}} + \beta\mathcal{L}_{\text{pred}}$
Predictive Memory Network
$\check{x}_{n+1}$
OT-SVD
$\check{x}_{n+1}$
Prediction Loss
$\mathcal{L}_{\text{pred}} = d(\check{x}_{n+1}, \hat{x}_{n+1})$
Latent Prediction Network

Fig. 1: A modular architecture for anomaly detection and localization.

Despite the development of several unsupervised multi-sensor anomaly detection algorithms, their effectiveness and widespread application remain limited due to several challenges. These include the difficulty in detecting diverse types of anomalies (point, contextual, or collective), the failure to remove noise from sensor data leading to false positives, and the need for algorithms to accurately capture spatio-temporal correlations. Moreover, most existing methods relies on single-type anomaly detection and lack integrated end-to-end training with multiple anomaly scoring, which limit their overall performance.

In this work, we introduce a modular architecture designed for unsupervised multi-sensor anomaly detection and localization, as shown in figure 1. This architecture integrates a spatio-temporal encoder to enhance representation learning, coupled with a decoder for latent reconstruction, and a predictive memory network adept at identifying sub-sequence patterns. A crucial addition to our framework is the denoiser module, specifically designed to mitigate false positives, a common issue in anomaly detection systems where noisy sensor data is often mistaken for an anomaly. Another key contribution of our proposed framework is the unique combination of a reconstruction network and a latent prediction network. This dual approach is tailored for the efficient detection of both point and sub-sequence anomalies. Moreover, the framework employs an end-to-end optimization strategy that jointly optimizes the encoder-decoder-memory network. This strategy focuses on minimizing the combined weighted loss, thereby enhancing the robustness of the anomaly detection system.

II. Problem Statement

We consider a multisensor system comprising S sensors with $x_k[n] \in \mathbb{R}$ representing the a measurement from the kth sensor at a specific discrete time point n. We define the *measurement vector* $\boldsymbol{x}[n] = (x_1[n], x_2[n], \ldots, x_S[n])^T \in \mathbb{R}^S$ for all sensors at time n. Consequently, the *measurement matrix* $\boldsymbol{X} = (\boldsymbol{x}[1], \boldsymbol{x}[2], \ldots, \boldsymbol{x}[N]) \in \mathbb{R}^{S \times N}$ represents sensor

This work was partially supported by the Research Council of Norway under the project DIGITAL TWIN within the PETROMAKS2 framework (project nr. 318899).

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00226

readings for N discrete time steps. In unsupervised setting, we utilize a training measurement matrix $\boldsymbol{X}_{\text{train}}$ representing a typical normal system operating conditions. We aim to develop a model $\mathcal{F}(\cdot)$ that accurately represents this normal behavior. For evaluation, we employ a testing measurement matrix $\boldsymbol{X}_{\text{test}} \in \mathbb{R}^{S \times M}$ (where $M \ll N$) encompassing both normal and anomalous conditions. Furthermore, labels for test data are available and given by the label vector $\boldsymbol{y}$, with y_m indicating the presence ($y_m = 1$) or absence ($y_m = 0$) of anomalies at each time step m. The objective of the framework is to produce a representation $\hat{\boldsymbol{y}} = \mathcal{F}(\boldsymbol{X}_{\text{test}})$ that approximates $\boldsymbol{y}$ according to a predefined metric.

III. Proposed Architecture

In the proposed modular architecture, we introduce a generic and adaptable framework for unsupervised multi-sensor anomaly detection and localization. The architecture consists of four key components, each capable of being implemented with different DNN models and computational tools.

1) Spatio-Temporal Encoder: This module captures spatio-temporal patterns within the data. The framework allows for a choice among several models such as MLP, CNN, RNN, GRU and Transformers, enhancing its adaptability to different datasets and anomaly detection requirements. The framework operates on a sliding-window mechanism to output reference signals for reconstruction and prediction tasks. We define the system input $\boldsymbol{X}_n = (\boldsymbol{x}[n], \boldsymbol{x}[n-1], \ldots, \boldsymbol{x}[n-L+1]) \in \mathbb{R}^{S \times L}$ at discrete time n, where L represents the window size. The selected encoder $\mathcal{E}(\cdot)$ then generates a latent representation $\boldsymbol{Z}_n$ to capturing the underlying patterns and dependencies in the data, i.e:

$$\boldsymbol{Z}_n = \mathcal{E}\left(\boldsymbol{X}_n\right) \ . \tag{1}$$

2) Decoder for Latent Reconstruction: The decoder module is responsible for reconstructing the input data from its latent representation. The latent representation $\boldsymbol{Z}_n$ is processed by the decoder $\mathcal{D}(\cdot)$ to output an estimated version of the system input, denoted $\hat{\boldsymbol{X}}_n \in \mathbb{R}^{S \times L}$;

$$\hat{\boldsymbol{X}}_n = \mathcal{D}(\boldsymbol{Z}_n) \ . \tag{2}$$

3) Predictive Memory Network: This module is designed for identifying sub-sequence patterns by predicting future values in the time series data, enhancing the detection of subsequence anomalies. The memory network $\mathcal{M}(\cdot)$ utilize the latent representation $\boldsymbol{Z}_n$ from the encoder to output a one-step-ahead prediction of the measurement vector, denoted $\check{\boldsymbol{x}}_{n+1} \in \mathbb{R}^{S \times 1}$.

$$\check{\boldsymbol{x}}[n+1] = \mathcal{M}(\boldsymbol{Z}_n) \ . \tag{3}$$

4) Denoiser Module (OT-SVD): The framework integrates Optimal Truncated Singular Value Decomposition (OT-SVD) for the purpose of reducing false-positive rates in anomaly detection. This method effectively filters out noise from sensor data that could otherwise be misinterpreted as anomalies. In multi-sensor systems, sensor measurements often exhibit correlation due to physical proximity or similar operating conditions. These dependencies typically result in low-order structures in the data, leading to rank deficiency in the input matrices. This assumption makes a low-rank approximation a suitable strategy for denoising process. For this purpose, we employ a Truncated Singular Value Decomposition (TSVD) low rank estimator. According to the Eckart-Young-Mirsky (EYM) theorem [16], the optimal rank-r approximation ($\tilde{\boldsymbol{X}}_{(r)}$) that minimizes the Frobenius norm is determined by retaining only the first r singular values and their corresponding singular vectors. For an input matrix $\boldsymbol{X}$, its TSVD approximation is calculated as:

$$\tilde{\boldsymbol{X}}_{(r)} = \underset{\hat{\boldsymbol{X}}:\ \operatorname{rank}(\hat{\boldsymbol{X}}) \leq r}{\arg\min} \left\|\boldsymbol{X} - \hat{\boldsymbol{X}}\right\|_F^2 = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T \tag{4}$$

where $\boldsymbol{u}_i$, are the left singular vectors of $\boldsymbol{X}$, $\boldsymbol{v}_i^T$, are the right singular vectors of $\boldsymbol{X}$, σ_i, are the singular values arranged in descending order (i.e., $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(S,L)}$). To determine the optimal rank r, we employ an information-theoretic approach based on random matrix theory [17]. This involves calculating the optimal threshold τ^*, which minimizes the asymptotic mean square error (MSE) between the original matrix and its low-rank approximation, i.e

$$\tau^* = \arg\min_{\tau} \lim_{S \to \infty} \mathbb{E}\left[\left\|\boldsymbol{X} - \tilde{\boldsymbol{X}}_{(r)}\right\|_F^2\right] \ . \tag{5}$$

We consider that the sensor measurements are embedded in additive white Gaussian noise and determine the optimal threshold τ^* for TSVD using:

$$\tau^* = \omega(\rho)\sigma_{\text{med}}, \tag{6}$$

where σ_{med} represents the median of the singular values. $\omega(\rho)$ is dependent on the matrix dimensions and approximated as [18]:

$$\omega(\rho) \approx \begin{cases} 2.858 & S = L \\ 0.56\rho^3 - 0.95\rho^2 + 1.82\rho + 1.43 & S \neq L \end{cases} , \tag{7}$$

Where $\rho = \frac{\min\{S,L\}}{\max\{S,L\}}$. The optimal threshold-dependent rank $r(\tau)$ is thus determined by:

$$r(\tau) = \max\{i : \sigma_i > \tau^*\} \ , \tag{8}$$

During the training process, the module processes the input, $\boldsymbol{X}_n$, to generate a noise-free reference signals for both reconstruction and prediction tasks, denoted $\tilde{\boldsymbol{X}}_n \in \mathbb{R}^{S \times L}$ and $\tilde{\boldsymbol{x}}_{n+1} \in \mathbb{R}^{S \times 1}$, respectively.

A. Loss Functions and End-to-end Optimization

The proposed framework employs a dual-network system for loss computation: a reconstruction network and a latent prediction network. These networks are co-trained in an end-to-end mechanism. The total loss is calculated as a linear combination of their individual losses, optimizing both the reconstruction and prediction accuracy. This dual approach is pivotal for effectively detecting different types of anomalies

Algorithm 1 Training algorithm

```
Input: Normal Training Dataset $\boldsymbol{X} = (\boldsymbol{x}[1], \boldsymbol{x}[2], \dots, \boldsymbol{x}[N])$, window size ($L$), number of epochs ($e$), batch size ($M$), hyperparameters ($\alpha, \beta, \epsilon, lr$)
Output: Trained module parameters (encoder ($\mathcal{E}_\mathrm{w}$), Decoder ($\mathcal{D}_\mathrm{w}$), Memory Network ($\mathcal{M}_\mathrm{w}$))
Data pre-processing, re-sampling, and scaling.
$\mathcal{E}_\mathrm{w}, \mathcal{D}_\mathrm{w}, \mathcal{M}_\mathrm{w} \leftarrow$ initialize model parameters
$k \leftarrow 1$
repeat
    for $j \leftarrow 1$ to $b = N/M$ do
        $Z_n = \mathcal{E}(\boldsymbol{X}_n)$                                   ▷ Eq. 1
        $\hat{\boldsymbol{X}}_n = \mathcal{D}(\boldsymbol{Z}_n)$                                   ▷ Eq. 2
        $\hat{\boldsymbol{x}}[n+1] = \mathcal{M}(\boldsymbol{Z}_n)$                                   ▷ Eq. 3
        $\boldsymbol{U}, \boldsymbol{\Sigma}, \boldsymbol{V} \leftarrow SVD(\boldsymbol{X}_n)$
        $\sigma_\mathrm{med} \leftarrow \mathrm{median}(\mathrm{diag}(\boldsymbol{\Sigma}))$
        $\tau^* = \omega(\rho)\sigma_\mathrm{med}$                                   ▷ Eq. 5
        $r(\tau) \leftarrow \max\{i : \sigma_i > \tau^*\}$                                   ▷ Eq. 8
        $\tilde{\boldsymbol{X}}_n \leftarrow \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T$                                   ▷ Eq. 4
        $\mathcal{L}_{j,\text{rec}} \leftarrow MSE(\tilde{\boldsymbol{X}}_n, \hat{\boldsymbol{X}}_n)$                                   ▷ Eq. 9
        $\mathcal{L}_{j,\text{pred}} \leftarrow MSE(\tilde{\boldsymbol{x}}[n+1], \check{\boldsymbol{x}}[n+1])$                                   ▷ Eq. 10
        $\mathcal{L} \leftarrow \alpha\mathcal{L}_{j,\text{rec}} + \beta\mathcal{L}_{j,\text{pred}}$                                   ▷ Eq. 11
        $\mathcal{E}_\mathrm{w}, \mathcal{D}_\mathrm{w}, \mathcal{M}_\mathrm{w} \leftarrow \mathcal{E}_\mathrm{w}, \mathcal{D}_\mathrm{w}, \mathcal{M}_\mathrm{w} - \mathrm{lr}\nabla\mathcal{L}$
    end for
    $k \leftarrow k + 1$
until $k = e$
```

in sensor measurements. The reconstruction loss is defined as the Mean Squared Error (MSE) between the denoised matrix sequences from OT-SVD $\tilde{\boldsymbol{X}}_n$ and the reconstructed sequences $\hat{\boldsymbol{X}}_n$. For a single sequence, it is given by:

$$\mathcal{L}_\text{rec} = \frac{1}{S}\frac{1}{L}\sum_{k=1}^{S}\sum_{\ell=0}^{L-1}\left(\tilde{x}_k[n-\ell] - \hat{x}_k[n-\ell]\right)^2 . \tag{9}$$

The latent prediction network calculates the loss as the MSE between the denoised vector $\tilde{\boldsymbol{x}}_{n+1}$ and the predicted vector $\check{\boldsymbol{x}}_{n+1}$:

$$\mathcal{L}_\text{pred} = \frac{1}{S}\sum_{k=1}^{S}\left(\tilde{x}_k[n+1]) - \check{x}_k[n+1]\right)^2 . \tag{10}$$

The total loss function combines the two loss types with weights α and β:

$$\mathcal{L} = \alpha\mathcal{L}_\text{rec} + \beta\mathcal{L}_\text{pred}. \tag{11}$$

This approach optimizes the trade-off between reconstruction and prediction accuracy to enable the effective detection of different types of anomalies in sensor measurements. The training process involves updating the combined loss function through back-propagation with mini-batch gradient descent algorithm. The detailed procedure is listed in algorithm 1.

B. Anomaly Detection and Localization

In the inference phase, our framework computes an anomaly score for each time step and each test measurement vector $\boldsymbol{x}[n]$ by summing the prediction and reconstruction errors using the ℓ_2-norm. The anomaly score (s_n) is calculated as follows:

Algorithm 2 Inference algorithm

```
Input: Test dataset containing normal and anomaly data: $\boldsymbol{X} = (\boldsymbol{x}[1], \boldsymbol{x}[2], \dots, \boldsymbol{x}[M])$, window size ($L$), True labels: $\boldsymbol{y} = (y_1, y_2, \cdots, y_M)^T$, Threshold ($\lambda^*$),
Output: Predicted Labels: $\hat{\boldsymbol{y}} = (\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_M)^T$
Data pre-processing, resampling, scaling.
for $m \leftarrow 1$ to $M$ do
    $\hat{\boldsymbol{x}}_m \leftarrow \mathcal{C}(\mathcal{T}(\boldsymbol{X}_n))$
    $\check{\boldsymbol{x}}_m \leftarrow \mathcal{F}(\mathcal{T}(\boldsymbol{X}_n))$
    $s_m \leftarrow \alpha_s\|\boldsymbol{x}_m - \hat{\boldsymbol{x}}_m\|_2 + \beta_s\|\boldsymbol{x}_m - \check{\boldsymbol{x}}_m\|_2$          ▷ Eq. 12
    if $s_m > \lambda^*$ then
        $y_m \leftarrow 1$
    else
        $y_m \leftarrow 0$
    end if                                   ▷ Eq. 15
end for
```

$$s_n = \alpha_s\|\boldsymbol{x}[n] - \hat{\boldsymbol{x}}[n]\|_2 + \beta_s\|\boldsymbol{x}[n] - \check{\boldsymbol{x}}[n]\|_2 , \tag{12}$$

where α_s and β_s are the weights for reconstruction and prediction scores, respectively. We propose a thresholding function based on the training data to detect anomalies in an unsupervised manner. For a training dataset $\boldsymbol{X} = (\boldsymbol{x}[1], \boldsymbol{x}[2], \dots, \boldsymbol{x}[N])$, the anomaly score ($s_n$) for each time step is calculated and thresholding function is defined as:

$$\lambda^* = \frac{1}{N}\sum_{i=1}^{N} s_n + \sqrt{\frac{z^2}{N}\sum_{i=1}^{N}\left(s_n - \frac{1}{N}\sum_{i=1}^{N} s_n\right)^2} \tag{13}$$

Where s_n is calculated by equation (12) and z is a scale factor.

During the inference, our framework not only detects anomalies but also localizes them within the multi-sensor array. Anomaly localization is achieved by examining the reconstruction and prediction errors at the sensor level. For each sensor k, the errors are calculated as:

$$e_k^\text{rec} = \|\tilde{x}_k[n] - \hat{x}_k[n]\|_2 \ , \ e_k^\text{pred} = \|\tilde{x}_k[n] - \check{x}_k[n]\|_2, \tag{14}$$

where e_k^rec and e_k^pred represent the reconstruction and prediction errors for sensor k, respectively. Anomalies are localized by comparing these errors to predefined thresholds θ^rec and θ^pred for reconstruction and prediction errors. The decision rule for anomaly detection and localization is:

$$\hat{y}_m = \begin{cases} 1 & s_m > \lambda^* \text{ and } (e_k^\text{rec} > \theta^\text{rec} \text{ or } e_k^\text{pred} > \theta^\text{pred}) \\ 0 & s_m \le \lambda^* \end{cases} . \tag{15}$$

A sensor measurement vector is classified as an anomaly if the overall score exceeds λ^* and at least one sensor-specific error exceeds its threshold. The inference procedure is summarized in Algorithm 2.

IV. Experiments and Results

A. Multi-sensor Datasets

To evaluate the performance of our proposed framework, we used three real-world multi-sensor datasets. 1) **SWaT** [19], [20]: This dataset is derived from a simulated real-world water treatment system testbed, featuring various network traffic, sensor, and actuator data. 2) **WADI** [21]: It is collected from a testbed that extends the SWaT system and offers a comprehensive view of a network for water treatment, storage, and distribution. We employ two key pre-processing steps (downsampling and feature normalization) on the input data prior to utilizing it in the framework. Downsampling was performed using a median filter with a 1-minute window size and no overlap for both training and test datasets. Labels for the downsampled test data were assigned based on the presence of anomalies within the corresponding window. For feature normalization, we employed min-max scaling to ensure stable training of neural network modules.

B. Baselines and Implementation Details

To evaluate the performance of our proposed architecture, we employ six baseline state-of-the-art anomaly detection methods. These include conventional algorithms such as Isolation Forest (IF) [9], which uses decision trees for unsupervised anomaly detection, and One-Class Support Vector Machines (OC-SVM) [10] that creates a hypersphere around normal data. Deep learning methods are also employed, such as the Multilayer Perceptron Autoencoder (MLP-AE) [22], which identifies anomalies through reconstruction errors, and the UnSupervised Anomaly Detection (USAD) [11], a method relying on adversarially-trained autoencoders. Further, we incorporated Deep Autoencoding Gaussian Mixture Model (DAGMM) [12] and Multivariate Anomaly Detection strategy with GAN (MAD-GAN) [12].

The deep learning frameworks PyTorch and TensorFlow are utilized for model training and evaluation. Data pre-processing was conducted using the Scikit-learn machine learning library. All models were trained in the Google Colaboratory Pro environment using NVIDIA T4 Tensor Core GPU processors.

C. Evaluation Metrics

We evaluate the performance of the proposed framework by considering the problem as a binary classification task using labeled test datasets. Given the data imbalance problem present in anomaly detection (fewer anomalies than normal samples), we prioritize F_1 score and the area under Precision-Recall (AUPR) curves as key performance metrics. Further, we consider the the area under Receiver Operating Characteristic curves (AUROC) for a comprehensive performance comparisons.

D. Results and Discussions

1) Module Variations: The performance analysis of the proposed framework by considering different modules is presented in Tables I and II. We examine five different encoders in combination with two decoders and a fixed memory block. The results indicate a notable superiority of the Transformer encoder in handling complex patterns and dependencies in multi-sensor data for anomaly detection tasks across both SWaT and WADI datasets, particularly when using a CNN decoder and MLP memory networks. This encoder achieves significantly higher F_1, AUC, and $AUPR$ scores compared to conventional recurrent architectures such as RNN, GRU, and LSTM, which, while effective in capturing temporal dependencies, may not effectively model the intricate dynamics in multi-sensor systems. The consistent performance of the CNN encoder across both datasets further highlights the importance of selecting appropriate encoder types to balance performance and efficiency. When utilizing an RNN decoder, the Transformer encoder maintains its lead with a reduced margin. The observed decrease in performance across all encoders for the WADI dataset underscores the inherent complexity of its multi-sensor system, which is partly attributable to nearly twice the number of sensors compared to the SWaT system. The ROC curve and precision-recall curve is presented in Figure 2 for the SWAT dataset. Overall, the results demonstrate the notable impact of encoder and decoder selection in enhancing anomaly detection in different multi-sensor systems.

TABLE I: CNN decoder and MLP memory networks

Encoder	SWaT			WADI		
	F_1	AUC	$AUPR$	F_1	AUC	$AUPR$
CNN	0.3144	0.8609	0.7714	0.0817	0.5633	0.0682
RNN	0.3016	0.8730	0.7852	0.0944	0.6398	0.0878
GRU	0.2965	0.8499	0.6916	0.0987	0.6711	0.0906
LSTM	0.3056	0.8609	0.7640	0.1008	0.7225	0.2257
Transformer	0.6330	0.9166	0.8551	0.1819	0.8313	0.1762

TABLE II: RNN decoder and MLP memory networks

Encoder	SWaT			WADI		
	F_1	AUC	$AUPR$	F_1	AUC	$AUPR$
CNN	0.3291	0.8605	0.7756	0.0871	0.5947	0.1192
RNN	0.3325	0.8761	0.7864	0.0841	0.5926	0.0729
GRU	0.3322	0.8662	0.7653	0.0787	0.5594	0.0790
LSTM	0.3302	0.8693	0.7696	0.0944	0.6665	0.1395
Transformer	0.5270	0.8658	0.7772	0.1123	0.6562	0.0829

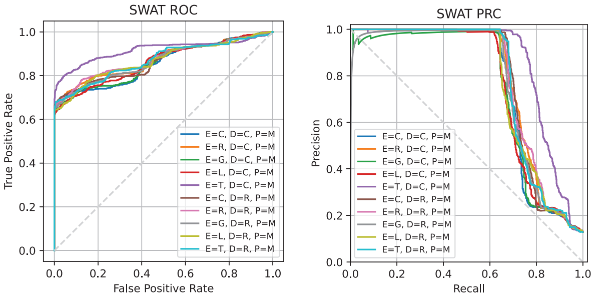


Fig. 2: Module variations performance comparisons (C = CNN, R = RNN, G = GRU, L = LSTM, T = Transformer).

2) Comparison with Baselines: To evaluate the performance of our proposed framework in comparison to the baselines, we selected a module configuration that was demonstrated in our previous evaluations, as detailed in Tables I and II. Specifically, we considered a model configuration integrating a transformer encoder, a convolutional decoder, and an MLP memory module. As summarized in Table III, the proposed model significantly outperforms all baselines across all metrics in the SWaT dataset. The model also maintains a leading performance in the more challenging WADI dataset, achieving the highest F_1 score. This performance indicates the effectiveness of transformer-based models in modeling the complex spatial and temporal relationships in multivariate time series data for anomaly detection, even in challenging and noisy sensor measurements. The results shows the importance of advanced architectures with robust encoder-decoder-memory mechanisms in improving the accuracy of unsupervised anomaly detection in critical multi-sensor systems.

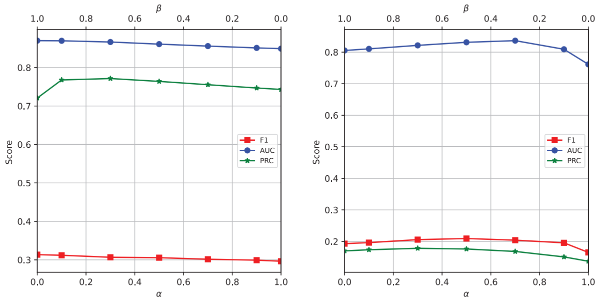
TABLE III: Comparison with baselines

Encoder	SWaT			WADI		
	F_1	AUC	$AUPR$	F_1	AUC	$AUPR$
IF	0.3502	0.8426	0.7577	0.0554	0.7080	0.0987
OC-SVM	0.2932	0.8216	0.7358	0.0965	0.7023	0.1554
MLP-AE	0.3120	0.8263	0.7289	0.0994	0.6708	0.0867
USAD	0.3256	0.8046	0.7031	0.1103	0.6763	0.1196
DAGMM	0.3253	0.8017	0.6917	0.1569	0.7033	0.1359
Our Model	0.6269	0.9166	0.8555	0.1820	0.8313	0.1766

3) Dual-Network Performance: In our dual-network performance analysis, we explored the impact of varying the balance between reconstruction and prediction errors on anomaly detection in multi-sensor systems. This involved fine-tuning the weight coefficients α and β, as detailed in Table IV. The table also includes results from employing single network strategies focusing solely on prediction (with $\alpha = 0, \beta = 1$) and solely on reconstruction (with $\alpha = 1, \beta = 0$). For this analysis, we utilized a model configuration incorporating a transformer and LSTM encoder, a convolutional decoder, and MLP memory. The analysis on the SWaT and WADI datasets using transformer demonstrate that an appropriate mix of reconstruction and prediction capabilities enhances overall performance, compared to focusing solely on either strategy. This is also evident in a result shown in Figure 3 for LSTM encoder. Overall, the result shows the presence of both point and collective anomalies in both datasets. These findings highlight the importance of a balanced and integrated approach in designing effective anomaly detection systems, especially in complex, multi-sensor environments.

TABLE IV: Transformer Encoder performance for different α and β coefficients.

α	β	SWaT			WADI		
		F1	AUC	AUPR	F1	AUC	AUPR
0.0	1.0	0.5470	0.8735	0.7889	0.1931	0.8053	0.1701
0.1	0.9	0.5513	0.8831	0.7972	0.1963	0.8104	0.1737
0.3	0.7	0.5865	0.9017	0.8231	0.2058	0.8215	0.1783
0.5	0.5	0.6269	0.9166	0.8555	0.2093	0.8313	0.1762
0.7	0.3	0.7452	0.9213	0.8629	0.2041	0.8364	0.1684
0.9	0.1	0.8375	0.9213	0.8609	0.1957	0.8092	0.1511
1.0	0.0	0.8347	0.9204	0.8540	0.1653	0.7614	0.1373

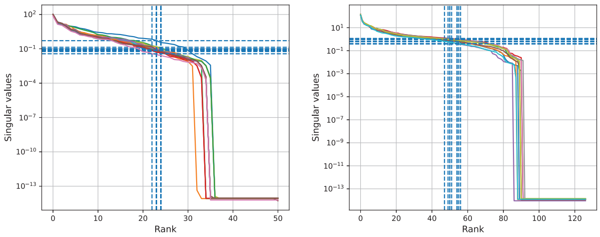


(a) SWaT

(b) WADI

Fig. 3: LSTM Encoder performance for different α and β coefficients.

4) OT-SVD Analysis: In our proposed modular framework, we incorporated an OT-SVD based denoising process to address the challenges posed by noisy and correlated sensor measurements in multi-sensor systems. By applying OT-SVD to both the SWaT and WADI datasets, we aimed to identify the extent of correlation among sensors and the presence of noise in the measurements. The analysis was visualized using scree plots, which depict the singular values against the rank of the input matrix, as shown in figure 4. For the SWaT dataset (containing 51 sensors), with a chosen window length of $L = 500$, the estimated rank varied between 22 and 24. Similarly, for the WADI dataset (containing 117 sensors), with a larger window length of 1000, the estimated rank fluctuated between 47 and 56. This variation in rank and threshold values across different windows underscores the presence of varying noise levels in the measurements, reflecting the complex nature of the industrial processes monitored. These findings suggest that, despite the full-rank nature induced by measurement noise, there is a high degree of correlation and noise present in the sensor measurements. Therefore, the integration of OT-SVD in our framework effectively aids in denoising the data, facilitating a more accurate and reliable anomaly detection in complex multi-sensor systems.



(a) SWaT

(b) WADI

Fig. 4: OT-SVD analysis

E. Ablation Study

For an ablation study, we examine the importance of OT-SVD denoising module within our anomaly detection framework. To achieve this, we assessed the full architecture of our model, which integrates a transformer encoder, a convolutional decoder, and MLP memory, against a variant that excludes the OT-SVD module. The results of this comparison are detailed in Table V. The performance metrics indicate a higher performance of the complete model over its non-denoising counterpart. This finding validates the significance of incorporating a denoising step for a robust multi-sensor anomaly detection.

TABLE V: Performance with and without OT-SVD

Variant	$F1$	AUC	$AUPR$
Without Denoiser	0.5750	0.9161	0.8500
With Denoiser	0.6269	0.9166	0.8555

V. Conclusions

In this study, we propose an adaptive modular framework for unsupervised anomaly detection and localization in multi-sensor systems. It consists of encoders, decoders, memory and denoising blocks. The framework employs a dual network architecture that combines a reconstruction network and a latent prediction network to detect both point and sub-sequence anomalies. It also addresses the challenge of correlated and noisy correlated sensor measurements by employing OT-SVD denoiser. A key strength of our approach lies in its flexible framework, which allows for the integration of diverse neural network architecture for the encoder, decoder, and memory components. This adaptability extends to the denoising process, which can be implemented by different mathematical tools. Overall, our modular framework demonstrates robust performance in anomaly detection, leveraging the strengths of different computational tools to address the complex challenges inherent in unsupervised anomaly detection and localization in multi-sensor environments.

Acknowledgment

This work was partially supported by the Research Council of Norway under the project DIGITAL TWIN within the PETROMAKS2 framework (project nr. 318899).

References

[1] S. Yin, S. X. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 11, pp. 6418–6428, 2014.

[2] P. García-Teodoro, J. Díaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," *Computers & Security*, vol. 28, no. 1-2, pp. 18–28, 6 2009.

[3] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network anomaly detection: Methods, systems and tools," *IEEE Communications Surveys and Tutorials*, vol. 16, no. 1, pp. 303–336, 6 2014.

[4] S. Rajasegarar, C. Leckie, and M. Palaniswami, "Anomaly detection in wireless sensor networks," *IEEE Wireless Communications*, vol. 15, no. 4, pp. 34–40, 6 2008.

[5] A. Ukil, S. Bandyoapdhyay, C. Puri, and A. Pal, "IoT healthcare analytics: The importance of anomaly detection," *Proceedings - International Conference on Advanced Information Networking and Applications, AINA*, vol. 2016-May, pp. 994–997, 6 2016.

[6] F. V. Wyk, Y. Wang, A. Khojandi, and N. Masoud, "Real-time sensor anomaly detection and identification in automated vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 1264–1276, 6 2020.

[7] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys*, vol. 41, no. 3, 2009. [Online]. Available: http://doi.acm.org/10.1145/1541880.1541882

[8] M. A. Belay, S. S. Blakseth, A. Rasheed, and P. Salvo Rossi, "Unsupervised Anomaly Detection for IoT-Based Multivariate Time Series: Existing Solutions, Performance Analysis and Future Directions," *Sensors 2023, Vol. 23, Page 2844*, vol. 23, no. 5, p. 2844, 6 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/5/2844

[9] F. T. Liu, K. M. Ting, and Z. H. Zhou, "Isolation-Based Anomaly Detection," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 6, no. 1, pp. 1–39, 3 2012. [Online]. Available: https://dl.acm.org/doi/10.1145/2133360.2133363

[10] K. L. Li, H. K. Huang, S. F. Tian, and W. Xu, "Improving one-class SVM for anomaly detection," *International Conference on Machine Learning and Cybernetics*, vol. 5, pp. 3077–3081, 2003.

[11] A. Julien, M. Pietro, G. Frédéric, M. Sébastien, and A. Z. Maria, "Usad: Unsupervised anomaly detection on multivariate time series," *dl.acm.org*, vol. 20, pp. 3395–3404, 8 2020. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3394486.3403392

[12] B. Zong, Q. Song, M. Renqiang Min, W. Cheng, C. Lumezanu, D. Cho, and H. Chen, "Deep autoencoding gaussian mixture model for unsupervised anomaly detection," *International conference on learning representations*, 2018. [Online]. Available: https://openreview.net/forum?id=BJJLHbb0-

[13] H. Zhao, Y. Wang, J. Duan, C. Huang, D. Cao, Y. Tong, B. Xu, J. Bai, J. Tong, and Q. Zhang, "Multivariate time-series anomaly detection via graph attention network," *Proceedings - IEEE International Conference on Data Mining, ICDM*, vol. 2020-November, pp. 841–850, 11 2020.

[14] S. Tuli, G. Casale, and N. R. Jennings, "TranAD: Deep Transformer Networks for Anomaly Detection in Multivariate Time Series Data," *Proceedings of the VLDB Endowment*, vol. 15, no. 6, pp. 1201–1214, 1 2022. [Online]. Available: http://arxiv.org/abs/2201.07284

[15] Z. Chen, D. Chen, X. Zhang, Z. Yuan, and X. Cheng, "Learning Graph Structures With Transformer for Multivariate Time-Series Anomaly Detection in IoT," *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 9179–9189, 6 2022.

[16] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, 9 1936. [Online]. Available: https://link.springer.com/article/10.1007/BF02288367

[17] M. Gavish and D. L. Donoho, "Optimal Shrinkage of Singular Values," *IEEE Transactions on Information Theory*, vol. 63, no. 4, pp. 2137–2152, 4 2017.

[18] ——, "The optimal hard threshold for singular values is 4/sqrt(3)," *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 5040–5053, 2014.

[19] J. Goh, S. Adepu, K. N. Junejo, and A. Mathur, "A dataset to support research in the design of secure water treatment systems," *Critical Information Infrastructures Security: 11th International Conference*, pp. 88–99, 2017.

[20] P. M. Aditya and O. T. Nils, "SWaT: A water treatment testbed for research and training on ICS security," *International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater)*, 2016. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7469060/

[21] C. M. Ahmed, V. R. Palleti, and A. P. Mathur, "WADI: A water distribution testbed for research in the design of secure cyber physical systems," *Proceedings - 2017 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks, CySWATER 2017*, pp. 25–28, 4 2017. [Online]. Available: https://dl.acm.org/doi/10.1145/3055366.3055375

[22] M. Sakurada and T. Yairi, "Anomaly detection using autoencoders with nonlinear dimensionality reduction," *ACM International Conference Proceeding Series*, vol. 02-December-2014, pp. 4–11, 6 2014. [Online]. Available: http://dx.doi.org/10.1145/2689746.2689747