

# Fault Detection for Process Industries via Temporal CapsNet Encoder-Assisted One-Class Classifier

Sheng Wang<sup>ID</sup>, Qiang Zhao<sup>ID</sup>, Yinghua Han<sup>ID</sup>, and Jinkuan Wang<sup>ID</sup>, *Member, IEEE*

**Abstract**—Fault detection plays a pivotal role in ensuring safety and efficiency in process industries. Subspace learning-based fault detection methods have gained recognition for their effective data structure characterization and noise mitigation. However, harnessing the benefits of subspace learning for fault detection tasks and ensuring significant discrimination between normal and fault feature representations to enhance classifier accuracy remain underexplored areas. In this study, we introduce a temporal capsule network (CapsNet) encoder-assisted one-class classifier (TceOne) methodology that enables joint optimization of subspace learning and fault detection. We modify the CapsNet to preserve temporal correlations of multivariate time series in the subspace, thereby enhancing the discriminability between normal and fault subspace representations. Normal subspace representations are confined to a compact region by minimizing the discriminative hypersphere radius of the one-class classifier, leaving fault features sparsely distributed outside the hypersphere. We then establish a specific subspace distance metric that draws normal data closer to the center and distances fault data from it. This metric accounts for the properties of CapsNet instantiation parameters, integrating the variations in both direction and magnitude of subspace representation. We demonstrate the effectiveness and superiority of our proposed methodology through experiments conducted on the Tennessee Eastman (TE) process.

**Index Terms**—Capsule network (CapsNet), fault detection, one-class classifier, process industry, subspace learning.

## I. INTRODUCTION

PROCESS industries encompass diverse sectors, including chemical, petrochemical, food and beverage, and pharmaceuticals, which rely on intricate and interconnected systems for their secure, profitable operation. Yet, these

Manuscript received 1 June 2023; revised 8 July 2023; accepted 27 July 2023. Date of publication 7 August 2023; date of current version 17 August 2023. This work was supported in part by the National Natural Science Foundation of China under Grant U21A20475, in part by the Natural Science Foundation of Hebei Province of China under Grant E2022501017 and Grant F2020501040, in part by the Research Fund from the State Key Laboratory of Rolling and Automation, Northeastern University under Grant 2021RALKFKT007, and in part by the Fundamental Research Funds for the Central Universities under Grant N2223001. The Associate Editor coordinating the review process was Dr. Siliang Lu. (*Corresponding author: Qiang Zhao*)

Sheng Wang and Jinkuan Wang are with the College of Information Science and Engineering, Northeastern University, Shenyang 110819, China (e-mail: 2010325@stu.neu.edu.cn; wjk@neu.edu.cn).

Qiang Zhao is with the School of Control Engineering and the Hebei Key Laboratory of Micro-Nano Precision Optical Sensing and Measurement Technology, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China (e-mail: zhaoqiang@neu.edu.cn).

Yinghua Han is with the School of Computer and Communication Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China (e-mail: yhan@neu.edu.cn).

Digital Object Identifier 10.1109/TIM.2023.3302346

systems are prone to various challenges, such as feedstock property variations, fluctuations in operating conditions, and equipment degradation. These issues can induce faults, leading to considerable production losses, safety hazards, and quality concerns. Consequently, fault detection in processes is pivotal to ensuring reliable and safe operation of these systems. One specific application scenario is the monitoring and control of chemical processes. Chemical plants, characterized by their complex and interconnected systems, can benefit from the proposed fault detection method by continuously monitoring real-time process data. Through the analysis of this data, operators can detect abnormal patterns indicative of faults or anomalies. Prompt corrective measures, such as adjusting process parameters or initiating maintenance activities, can then be undertaken to minimize risks and avert accidents, equipment damage, and production losses.

The advancement in sensor and computing technologies has facilitated the accumulation of massive volumes of process data, paving the way for data-driven techniques in fault detection. A prevalent strategy is the application of dimensionality reduction techniques, projecting high-dimensional data onto a lower dimensional subspace. This subspace characterization aids in identifying data structures and mitigating noise [1], [2], thereby enabling the classification of normal and faulty samples based on subspace feature representations. These subspace learning methods include statistical dimensionality reduction algorithms such as principal component analysis (PCA) [3], [4], kernel PCA [5], and so on [6], [7], [8]. Deep neural networks (DNNs)-based methods have also been valued such as convolutional neural network (CNN) [9], [10], long-short term memory network (LSTM) [11], [12], auto-encoder (AE) [13], [14], and their variants [15], [16]. DNNs can be trained end-to-end, enabling the model to learn task-specific representations directly from the data. This arrangement optimizes the subspace learning concurrently with the downstream fault detection task. Although these methods have demonstrated robust feature expression capabilities, the challenge lies in how to extract sufficiently discriminative features from multivariate time series. Ensuring significant differences between the subspace features of normal and faulty data is paramount for accurate fault detection.

In industrial environments, manufacturing processes are meticulously engineered to meet specific production targets and operate under well-defined mechanisms [17]. These mechanisms often exhibit inherent correlations that adhere to stringent chronological constraints during normal operations.

For instance, fluctuations in raw materials may impact controlled parameters, subsequently affecting measured parameters. However, the emergence of a fault can disrupt these temporal correlations, with changes in these correlations before and after the fault being pivotal for investigating the discriminative feature subspace. This observation suggests that focusing on changes in data temporal correlation can facilitate fault detection. Despite several attempts to enhance fault detection by adequately projecting features into the subspace, the perspective of learning subspace features from temporal correlation is seldom considered.

In this study, our objective is to learn temporal correlations by enhancing the capsule network (CapsNet) in a novel multivariate time series encoder. Although conventional architectures such as CNN and LSTM prove reliable in various scenarios, they can overlook subtle interdependencies within the data due to their reliance on max-pooling and sequential processing [18], [19], respectively. CapsNet, however, offers a unique approach to processing these kinds of data. It encapsulates the combinations of features in capsules, where the direction of a capsule vector represents the type of the feature, and its magnitude (bounded to 1) indicates the probability of the feature's existence [20], [21], [22]. Furthermore, capsules can encode various feature properties, such as value, trend, and duration. In industrial processes, the data change in the value, trend, and duration before and after the fault can be captured by the CapsNet, resulting in alterations in capsule magnitude and direction. This ability of CapsNet to retain high-resolution details and capture complex interdependencies conserves the input data's hierarchical temporal correlations, forming the bedrock of our design choice.

Drawing upon this foundation, it is acknowledged that the emergence of a fault disrupts these temporal correlations, leading to deviations from normal patterns. Theoretically, enhancing the discriminability of normal and fault subspace features is expected to improve the accuracy of fault detection [23]. Recent research has focused on constructing a one-class classifier within the subspace [24], aiming to encapsulate normal features within a compact hypersphere, thereby achieving effective discrimination of fault features outside the hypersphere. While the one-class classifier proves to be an effective approach, it is not suitable for a CapsNet subspace with capsule representations. Because these methods minimize the average Euclidean distance to constrain normal data features within a compact hypersphere, ignoring specific properties of capsules. This approach weakens the effectiveness of CapsNet to extract temporal correlations and may even reduce it to a regular neural network. To address this, our modified one-class classifier accounts for variations in capsule direction and magnitude, ensuring that the classifier can construct a well-fit and discriminative hypersphere, further enhancing the discriminability of subspace features between normal and faulty states.

To effectively address the challenges identified in process industry fault detection, we propose a novel method known as the temporal Capsnet encoder-assisted one-class classifier (TceOne). The method hinges on two primary components:

a temporal Capsnet encoder (TCE) for feature subspace construction and an enhanced one-class classifier. The TCE forms the core of our method, acting as the primary tool for extracting features from process data. It comprises two feature extractors working in tandem. The first employs a separable multiscale temporal CNN (SMTCN) to extract temporal features from individual time series as primary capsules. The second extractor utilizes a multihead self-attention (MSA) mechanism to transform these primary capsules into digital capsules, thereby enabling the extraction of temporal correlations from multivariate time series. The second component of TceOne is an enhanced one-class classifier. It leverages the unique properties of capsules, taking into account both the direction and magnitude of capsule vectors. We propose a suitable distance metric for TCE subspaces, which measures the distance between a test sample and the center of the subspace relative to its boundary. This facilitates the differentiation between normal and faulty conditions in process data. TceOne has been validated using the Tennessee Eastman (TE) process and offers the following contributions:

- 1) A novel fault detection method for process industries is proposed, known as TceOne, which allows for joint optimization of subspace feature extraction and one-class classification.
- 2) TCE is capable of extracting temporal correlations from multivariate time series, effectively maximizing the discrimination of normal and fault features in the subspace, thus easing the task of finding the hypersphere for the one-class classifier.
- 3) The newly proposed subspace distance metric combines the direction and magnitude of capsule vectors, offering a more accurate discriminative hypersphere for the one-class classifier and thereby improving the effectiveness of fault detection.

The remainder of this article is organized as follows. In Section II, we review the one-class classifier for fault detection. In Section III, we revisit the DSVDD. In Section IV, we present the overall framework and provide a detailed presentation of the proposed methods. In Section V, we present a case study and corresponding experimental results. Finally, in Section VI, we provide the conclusion of this work.

## II. RELATED WORK

Fault detection is commonly regarded as a classification task that aims to differentiate between normal and faulty conditions [25]. Fault detection methods based on one-class classifier are particularly useful in scenarios where there is an abundance of data representing normal conditions of process industries but few or no samples of faulty conditions. These methods aim to identify samples that deviate significantly from what the subspace learning model deems normal.

Local Outlier Factor (LOF) is a popular outlier detection method used for fault detection [26], [27]. LOF computes the local density deviation of each data point with respect to its neighbors, denoting the degree of abnormality or outlierness of each data point. However, it can be computationally

expensive as the number of data points increases. Also, it requires appropriate selection of parameters, like the number of neighbors to consider, which can be challenging [28], [29]. Similarly, the isolation forest (IF) isolates anomalies and is used for fault detection by randomly partitioning data and measuring the number of partitions required to isolate an instance from other instances [30], [31], [32]. However, its drawbacks include instability and sensitivity to outliers. The random partitioning approach can lead to inconsistent results, and outliers can distort the partitioning process, affecting fault detection accuracy. When applying LOF and IF for fault detection, the feature selection step is also important. This step involves analyzing the data and selecting the relevant features that have the greatest discriminative power to distinguish between normal and faulty samples.

Support vector machine (SVM)-based methods, i.e., one-class support vector machines (OC-SVM) and support vector data description (SVDD), have gained popularity for fault detection based on subspace learning. OC-SVM works by finding an optimal hyperplane in a high-dimensional feature space that maximally discriminates between normal and faulty samples, enabling accurate classification and detection of faults [33], [34]. The performance of OC-SVM can be sensitive to the choice of kernel function. Selecting the most suitable kernel for a given dataset is not always straightforward, which severely limits the generalization of OC-SVM [35]. However, SVDD aims to find a hypersphere in the feature space that encloses the majority of the data points representing the target class [36], [37]. This makes it more robust to outliers and suitable for various applications where the target class may exhibit different shapes or distributions.

Subsequently, deep support vector data description (DSVDD) is an extension of SVDD that incorporates DNNs to learn more expressive representations of the subspace feature. Many studies have attempted to enhance the performance of fault detection tasks by integrating DSVDD with a DNN encoder, aiming to utilize DSVDD to refine the subspaces learned by the encoder [38], [39], [40]. However, these DNNs based on CNNs and LSTMs are limited in their ability to handle temporal correlations and construct discriminative feature subspaces. The unique capabilities of CapsNet, on the other hand, make it a promising candidate for enhancing the discriminative power of feature subspaces in fault detection tasks, as it can retain high-resolution details and capture complex interdependencies. Despite the potential of CapsNet, there has been limited research on effectively combining it with one-class classifiers to augment the discriminative power of feature subspaces. The integration of CapsNet and one-class classifiers has the potential to provide a more nuanced and sensitive approach to fault identification in process industry operations. This combination can help reduce the occurrence of undetected faults, minimizing production losses and safety hazards. By leveraging the strengths of both CapsNet and one-class classifiers, we aim to address the limitations of existing methods and achieve more accurate and reliable fault detection in industrial processes.

### III. PRELIMINARIES

#### A. DSVDD

DSVDD learns the mapping from the original space  $\mathcal{X} \subseteq \mathbb{R}^n$  to the subspace  $\mathcal{F} \subseteq \mathbb{R}^m$  by building an encoder  $\phi(\cdot; \mathcal{W}) : \mathcal{X} \rightarrow \mathcal{F}$ , where  $\mathcal{W} = \{\mathbf{W}^1, \dots, \mathbf{W}^L\}$  is the weight set of neural network with  $L$  layers. For a given input  $\mathbf{x}$ ,  $\phi(\mathbf{x}; \mathcal{W})$  represents the feature representation in the subspace. The encoder can be trained to minimize the volume of a data-enclosing sphere in subspace  $F$  that is characterized by center  $\mathbf{c}$ , by minimizing the mean distance from the subspace representation of each sample to the center. In fault detection tasks, the simplified objective function of DSVDD is written as follows:

$$\min \frac{1}{p} \sum_{i=1}^p \|\phi(\mathbf{x}_i; \mathcal{W}) - \mathbf{c}\|^2 + \frac{\lambda}{2} \sum_{\ell=1}^L \|\mathbf{W}^\ell\|_F^2 \quad (1)$$

where the first term is a quadratic loss for penalizing the distance of every network representation  $\phi(\mathbf{x}_i; \mathcal{W})$  to  $\mathbf{c}$ , and  $\mathbf{c}$  is the mean output of the encoder after being pretrained. The second term is a network weight decay regularizer with hyperparameter  $\lambda > 0$ , and  $\|\cdot\|_F$  is the Frobenius norm.

For a given test input  $\mathbf{x}_{\text{new}}$ , the fault score can be naturally defined using the distance from the feature representation to the center of the hypersphere, i.e.,

$$s(\mathbf{x}_{\text{new}}) = \|\phi(\mathbf{x}_{\text{new}}; \mathcal{W}^*) - \mathbf{c}\|_2 \quad (2)$$

where  $\mathcal{W}^*$  represents the parameters after training. The further the subspace feature of the test data is from the center, the larger the anomaly score, and the more likely it is to be inferred to be the fault sample.

### IV. TCE-ASSISTED ONE-CLASS CLASSIFIER

The proposed TceOne-based fault detection approach, tailored specifically for the process industry, is depicted in Fig. 1. It begins by processing raw multivariate time series data through the SMTCN module. This stage independently extracts temporal features from each series and converts them into primary capsules. Next, the MSA mechanism is employed to discern interdependencies between the primary capsules, resulting in digital capsules. These digital capsules are decoded using a specialized decoder to generate reconstructed multivariate time series samples. This reconstruction forms the basis for the initial model training, where the focus is on minimizing reconstruction loss. Concurrently, the digital capsules, serving as subspace feature representations, facilitate the fine-tuning of the TCE and the classifier through minimization of the distance between the subspace features and the center. The distribution of the features of normal samples, coupled with a predetermined significance level, informs the determination of the classifier threshold. Finally, fault detection is performed by determining whether the distance between the center and the subspace mapping of the input (represented by digital capsules) exceeds the set threshold. Further details about the proposed method will be provided in the sections that follow.

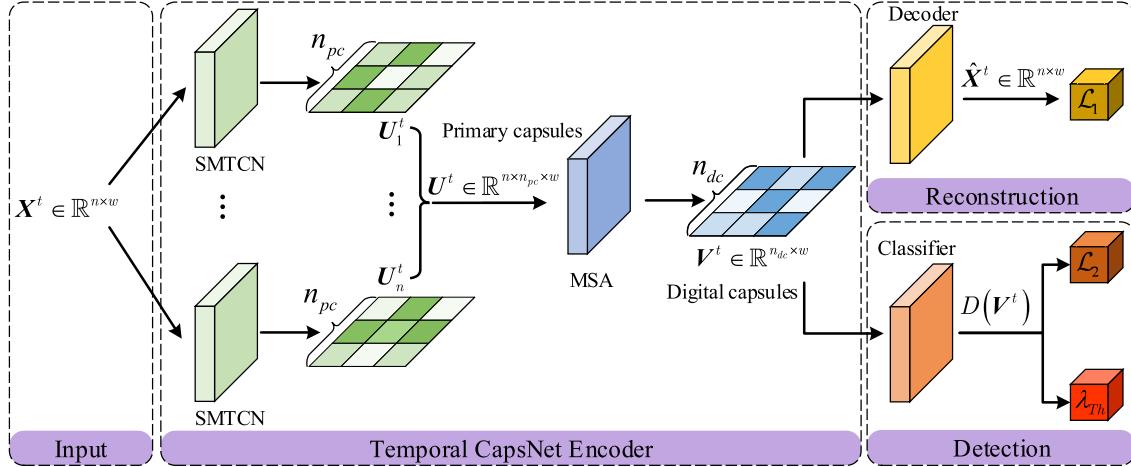


Fig. 1. Schematic of the TceOne-based fault detection model. The model uses SMTCN to process the input  $X^t$  and generate primary capsules  $U^t$ . Temporal correlations within  $U^t$  are discerned using the MSA, producing digital capsules  $V^t$ . These are decoded into reconstructions  $\hat{X}^t$ , aiding model pretraining. The distance between  $V^t$  and their center, i.e.,  $D(V^t)$ , assists in the fine-tuning of the classifier. The model performs fault detection by comparing  $D(V^t)$  with a threshold,  $\lambda_{Th}$ .

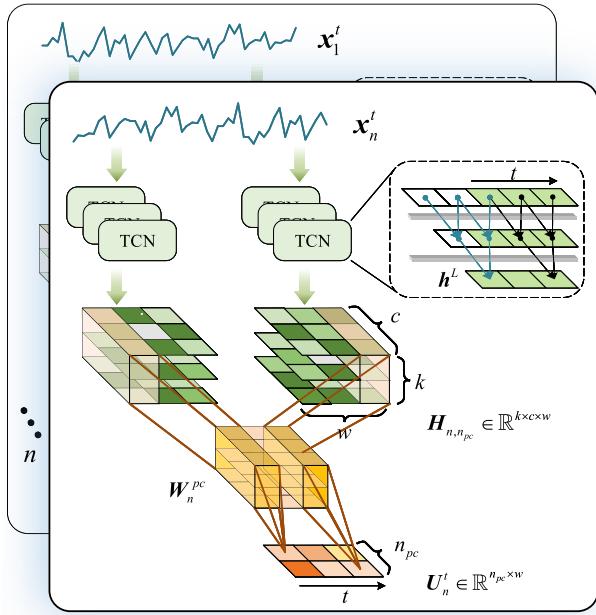


Fig. 2. Details of the SMTCN module. The input of SMTCN module is the multivariate time series data, i.e.,  $[x_1^t, \dots, x_n^t] \in \mathbb{R}^{n \times w}$  and the output is the feature maps of primary capsules, i.e.,  $[U_1^t, \dots, U_n^t] \in \mathbb{R}^{n \times n_{pc} \times w}$ .

#### A. Temporal CapsNet Encoder

Let  $X = \{X^t = [x_1^t, \dots, x_n^t] | t = 1, \dots, T\} \in \mathbb{R}^{n \times T}$  be the samples continuously observed from  $T$  time points in the process containing  $n$  time series under normal conditions, where  $x_i^t = [x_i^{t-w+1}, \dots, x_i^t]$  and  $w$  is the time window size. As shown in Fig. 2, SMTCN consists of  $n$  concurrent modules, each of which independently extracts local features of a time series  $x_i^t$  in terms of primary capsules. Firstly,  $x_i^t$  passes through  $n_{pc}$  groups of TCN blocks with  $L$  layers,  $k$  kernel sizes, and  $c$  output channels. The output of the first and hidden layers can be formalized as follows:

$$\mathbf{h}^1 = \sigma(\mathbf{w}^1 \otimes x_i^t) \quad (3)$$

$$\mathbf{h}^\ell = \sigma(\mathbf{h}^{\ell-1} + \mathbf{w}^\ell \otimes \mathbf{h}^{\ell-1}) \quad (4)$$

where  $\mathbf{w}^\ell$  is the weight of the  $\ell$ th layer,  $\ell = 2, \dots, L$ , the notation  $\otimes$  denotes dilated convolutional calculation, and  $\sigma(\cdot)$  indicates a nonlinear activation function. Referring to (4), the layerwise identity mappings, i.e., the first  $\mathbf{h}^{\ell-1}$ , are employed to prevent gradient dispersion in the deep network caused by weights less than 1. Due to the presence of  $\mathbf{h}^{\ell-1}$ , the gradient of the upper level can be directly propagated to the lower level in the backward updating. Notably, the inclusion of bias terms should be purposely avoided, as it can lead the network to learn a constant function that directly maps to the center of the hypersphere. In effect, this will lead to the collapse of the hypersphere [24]. In the subsequent networks, the bias terms are also discarded. The output of the TCN blocks merged from different kernel sizes and channels, i.e.,  $\mathbf{H}_{i,j} \in \mathbb{R}^{k \times c \times w}$ ,  $j = 1, \dots, n_{pc}$ , are compressed by

$$U_i^t = \{\mathbf{u}_{i,j}^t = \text{squash}(W_i^{pc} \cdot \mathbf{H}_{i,j}) | j = 1, \dots, n_{pc}\} \quad (5)$$

where notation  $\cdot$  denotes dot product, and  $U_i^t \in \mathbb{R}^{n_{pc} \times w}$ ,  $i = 1, \dots, n$  represents the feature maps extracted through the SMTCN module. A nonlinear activation function  $\text{squash}(\mathbf{x}) = (\|\mathbf{x}\|^2 / (1 + \|\mathbf{x}\|^2))(\mathbf{x} / \|\mathbf{x}\|)$  is applied to capsules with vector properties.

The activation function  $\text{squash}(\mathbf{x})$  not only augments the network's capacity to manage nonlinear dependencies but also normalizes each capsule vector's magnitude between 0 and 1. The design of  $\text{squash}(\mathbf{x})$  ensures the preservation of the capsule vectors' direction and relative proportions even as their magnitudes are downscaled [18], [19]. This property renders CapsNet invariant to changes in the overall length of input vectors. If the length of the input vectors is increased or decreased, it does not significantly impact the output. This property ensures that CapsNet focuses more on the arrangement and relationships between features rather than their absolute magnitudes, making it more robust to variations in input data. By normalizing the magnitude of the capsule vector, the activation function provides a measure of confidence or activation level associated with the corresponding feature. Rather than focusing exclusively on the absolute

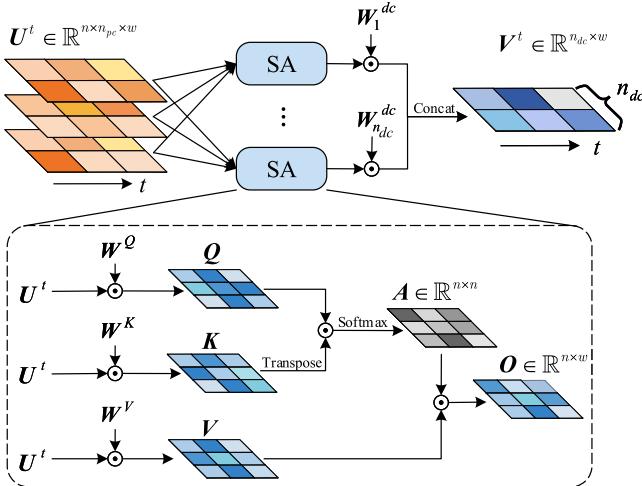


Fig. 3. Details of the MSA module. The input of MSA module is the primary capsules, i.e.,  $\mathbf{U}^t \in \mathbb{R}^{n \times n_{pc} \times w}$  and the output is the feature maps of digital capsules  $\mathbf{V}^t \in \mathbb{R}^{n_{dc} \times w}$ .

magnitudes of features, the network emphasizes their existence and arrangement within the input data. By preserving the capsules' direction and relative proportions, CapsNet learns temporal correlations and configurations of features, leading to robust and meaningful representations.

Let  $\mathbf{u}_{i,j}^t \in \mathbb{R}^w$  denote the  $j$ th primary capsule of  $\mathbf{x}_i^t$ , in which the element  $u_{i,j}^{t(\tau)} \in \mathbf{u}_{i,j}^t$ ,  $\tau = t-w+1, \dots, t$  represents the feature corresponding to the RF time steps before time  $\tau$ , ensuring that the encoding of the primary capsules preserves temporal information. RF is the receptive field of TCN block

$$\text{RF}_L = \text{RF}_{L-1} + d^{L-1}(K-1) \prod_{\ell < L-1} S_\ell \quad (6)$$

where  $d$  is the dilated rate,  $K$  is the kernel size, and  $\prod_{\ell < L-1} S_\ell$  is the multiplication of all strides in the front  $L-1$  layers. SMTCN employs multiscale kernels to extract local dynamic features of time series under different receptive fields [41].

Next, the features of  $n$  time series are concatenated to form the primary capsules of SMTCN, i.e.,  $\mathbf{U}^t = \{\mathbf{U}_i^t | i = 1, \dots, n\} \in \mathbb{R}^{n \times n_{pc} \times w}$ , which are then fed into the MSA module to discover the intercorrelations between the primary capsules. As shown in Fig. 3, for each SA block, the input is projected into three metrics  $\mathbf{Q}$ ,  $\mathbf{K}$ , and  $\mathbf{V}$  via linear projections. The correlation of each primary capsule is measured by the scale dot product of  $\mathbf{Q}$  and  $\mathbf{K}$  [42], which is further used as the attention weights to participate in the weighted summation of  $\mathbf{V}$ . The scenario of SA can be formalized as follows:

$$\mathbf{O} = \text{softmax}\left(\frac{\mathbf{W}^Q \mathbf{U}^t \cdot (\mathbf{W}^K \mathbf{U}^t)'}{\sqrt{n}}\right)(\mathbf{W}^V \mathbf{U}^t) \quad (7)$$

where  $\mathbf{W}^*$  are the projection weights. The softmax function is utilized to ensure that the sum of each row of attention weights equals to 1, which represents the degree of association between each capsule and others. The output of each SA, i.e.,  $\mathbf{O}_i \in \mathbb{R}^{n \times w}$ , constitutes a set of temporal correlation features which are compressed and merged into digit

capsules by

$$\mathbf{V}^t = \{\mathbf{v}_i^t = \text{Squash}(\mathbf{W}_i^{\text{dc}} \cdot \mathbf{O}_i) | i = 1, \dots, n_{dc}\} \quad (8)$$

where  $\mathbf{V}^t \in \mathbb{R}^{w \times n_{dc}}$  preserves the temporal information of original time series as well, specifically,  $v_i^{t(\tau)} \in \mathbf{v}_i^t$  is calculated as the weighted sum of the features at time  $\tau$  within  $\mathbf{U}^t$ .

To mitigate the complexity of TCE, we propose the utilization of the MSA as a substitute of the dynamic routing mechanism of CapsNet [22]. MSA is capable of identifying important contextual information that might be disregarded by dynamic routing, particularly in tasks that necessitate a holistic understanding of the input data, such as preserving temporal correlations among multivariate time series. Moreover, MSA offers enhanced computational efficiency compared to dynamic routing. Specifically, the complexity of MSA, according to (7), can be approximated as  $O(nwn_{pc}^2)$ . However, the complexity of the dynamic routing is approximated as  $O(rnwn_{pc}n_{dc})$ , where  $r$  represents the number of routing iterations. The complexity of dynamic routing is higher because the condition  $n_{dc} > n_{pc}$  holds as CapsNet requires higher dimensions to represent more complex temporal correlation features. And the dynamic routing involves iterative computations between capsules, so its complexity increases with  $r$  in multiples. Especially, when the input dimension  $n$  increases,  $n_{pc}$  and  $n_{dc}$  also increase. As a result, the complexity of dynamic routing grows faster due to its dependence on these increasing variables.

In the decoder, a separable deconvolution network is employed to transform subspace features back into time series data, which splits features into multiple channels and performs independent deconvolutional operations on each channel. This approach can effectively prevent the confusion between channels and improve the quality of reconstruction, which can be formulated as follows:

$$\hat{\mathbf{X}}^t = \sigma(\mathbf{W}_L^t \sigma(\dots \sigma(\mathbf{W}_1^t \mathbf{V}^t))) \quad (9)$$

where  $\hat{\mathbf{X}}^t \in \mathbb{R}^{n \times w}$  is the reconstruction of  $\mathbf{X}^t \in \mathbb{R}^{n \times w}$ ,  $\mathbf{W}_\ell^t$  is the weight of the transpose convolution kernel,  $\ell = 1, \dots, L'$ , and  $L'$  is the number of deconvolution layers.

The analysis above clearly shows that the encoding and decoding of multivariate time series in our proposed model strictly adheres to chronological constraints. Specifically, the TCE extracts features at each time step into the primary capsules  $\mathbf{U}^t$  from the raw data that occurred prior to the respective time step. This imparts the extracted features with temporal attributes consistent with the underlying time series. Similarly, the digit capsules  $\mathbf{V}^t$  preserve this temporal information by avoiding any disruption to the chronological order. In essence, the features of time  $\tau$  within  $\mathbf{V}^t$  result from the intercorrelations of the corresponding primary capsules extracted before time  $\tau$ . Moreover, the reconstructions of time  $\tau$  produced by the decoder are based solely on the deconvolution output of the digit capsules at time  $\tau$ . Thus, the encoding and decoding processes within our model ensure the preservation of the temporal information of the multivariate time series throughout the subspace learning process. Clearly, the TCE is designed with a unique perspective that enforces strict temporal constraints on the correlations between time series.

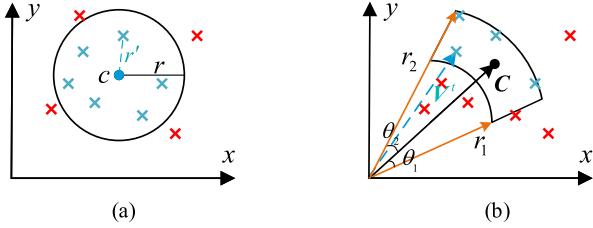


Fig. 4. Comparison of DSVDD and TceOne. In (a), the boundary is decided by the radii between normal features and center. In (b), the boundary is decided by the differences in direction and magnitude between normal features and center. (a) DSVDD. (b) TceOne.

This ensures the exactitude of temporal correlation modeling for multivariate time series.

### B. Subspace Distance Metric

The fault detection task based on DSVDD can be visually represented as Fig. 4(a), where the distance between the subspace representation and center are calculated by the average Euclidean distance, as show in (2). This approach neglects the direction and magnitude of subspace representations that are in the form of capsules, resulting in a potential decline in the effectiveness of fault detection. To address this issue, we propose a more suitable subspace distance metric by taking into account the differences in direction and magnitude between the subspace feature representation and the center, it can be formalized as follows:

$$D(\mathbf{V}^t) = \sum_{i=1}^{n_{dc}} \alpha \left( e^{-\frac{\mathbf{v}_i^t \cdot \mathbf{c}_i}{\|\mathbf{v}_i^t\| \|\mathbf{c}_i\|}} - e^{-1} \right) + \beta (\|\mathbf{v}_i^t\| - \|\mathbf{c}_i\|)^2 \quad (10)$$

where the first term calculates the cosine angle between the capsule  $\mathbf{v}_i^t$  and the center vector  $\mathbf{c}_i$ . As the angle between the two directions decreases, the value of this term approaches 0, indicating greater consistency between the directions. The second term measures the difference between the magnitude of  $\mathbf{v}_i^t$  and  $\mathbf{c}_i$ . When the magnitude is closer to each other, the value of this term approaches 0. The coefficients  $\alpha$  and  $\beta$  are proportional and sum to 1, controlling the weights of the two terms. A smaller value of  $D(\mathbf{V}^t) \geq 0$  indicates a closer distance between  $\mathbf{V}^t$  and  $\mathbf{C}$  in the subspace.

Referring to (10), the subspace feature representations of normal samples should maintain proximity to the center in both direction and magnitude. Conversely, fault samples may exhibit deviations from the center in either direction or magnitude. For instance, in the 2-D plane depicted in Fig. 4(b), normal features are confined to the direction interval  $[\theta_c - \theta_1, \theta_c + \theta_2]$  and the magnitude interval  $[r_1, r_2]$ , where  $\theta_1 = \theta_2$  and  $r_c = (1/2)(r_1 + r_2)$ . Here,  $\theta_c$  and  $r_c$ , respectively, denote the direction and magnitude of center  $\mathbf{C}$ . Thus, the discrimination boundary of TceOne is an irregular hypersphere, which accommodates a more compact boundary compared with DSVDD and leads to improved discrimination between normal and fault samples.

### C. Transfer Learning for TceOne

During the pretraining process, both the TCE and the decoder parameters are simultaneously optimized to learn the normal patterns. The objective function in this reconstruction

task can be expressed as follows:

$$\min \sum_{t=1}^T \| \mathbf{X}^t - \phi(\mathbf{X}^t; \mathcal{W}_{\text{total}}) \|_2^2. \quad (11)$$

The pretrained TCE parameters  $\mathcal{W} \subset \mathcal{W}_{\text{total}}$  are transferred into the TceOne model, and the center  $\mathbf{C} = (1/T) \sum_{t=1}^T \phi(\mathbf{X}^t; \mathcal{W})$  is kept fixed during the fine-tuning process. The objective function for the TceOne fine-tuning is formulated as follows:

$$\min \frac{1}{T} \sum_{t=1}^T D(\phi(\mathbf{X}^t; \mathcal{W})) + \frac{\lambda}{2} \sum_{\mathbf{W} \in \mathcal{W}} \|\mathbf{W}\|_F^2 \quad (12)$$

where  $\mathbf{W} \subset \mathcal{W}$  are the parameters of the pretrained TCE model that need to be further fine-tuned. The first term encourages the optimization process to find the optimal values of the parameters that minimize the distance between the subspace features and the center. Hence, the boundary of normal pattern are compact enough to improve the discrimination between normal and fault samples. The second term is a weight decay term that regularizes the parameters in  $\mathcal{W}$  by adding a penalty proportional to the squared Frobenius norm of the weight matrices. This term discourages the parameters from becoming too large and helps to prevent overfitting by reducing the model's capacity to fit noise in the data [38].

The enhancements made to the one-class classifier aim to enhance the discriminability between the representation of normal and fault subspace, achieved by minimizing the radius of the discriminative hypersphere for normal data and pushing the fault features further away from the subspace center. These enhancements utilize the properties of the TCE instantiation parameters, which synthesize diverse directions and magnitudes of subspace representation. Specifically, the normal subspace representations are confined to a compact region, while the fault features are distributed outside this region. As a result, the normal data are brought closer to the center, facilitating a clearer differentiation from the fault data.

To clarify, the distance between the subspace representation of the new sample  $\mathbf{X}^{\text{new}}$  and  $\mathbf{C}$  is computed as  $D(\phi(\mathbf{X}^{\text{new}}; \mathcal{W}^*))$ , where  $\mathcal{W}^*$  denotes the fine-tuned TCE parameters. If the distance is larger than a predefined threshold  $\lambda_{\text{th}}$ , the sample is classified as faulty; otherwise, it is classified as normal. The value of  $\lambda_{\text{th}}$  is determined by the training set and the desired significance level  $\text{th}\%$ . Specifically,  $\lambda_{\text{th}}$  is set to the  $\text{th}\%$  quantile of the distances between the feature representations of the training samples and  $\mathbf{C}$ . In other words,  $\lambda_{\text{th}}$  is the value such that  $\text{th}\%$  of the training samples have distances lower than  $\lambda_{\text{th}}$ , i.e., the  $\lambda_{\text{th}} = \text{th}\% \text{ quantile of } \{D(\phi(\mathbf{X}^t; \mathcal{W}^*)) | t = 1, \dots, T\}$ .

## V. EXPERIMENT AND DISCUSSION

The reliability and accuracy of the proposed TceOne fault detection method are validated through TE benchmark process with various fault types. Furthermore, the study examines the differences in feature subspace construction among various methods, revealing the superior feature extraction capabilities of TCE.

### A. Dataset and Settings

The TE process is a widely studied industrial recycle reactor process [43], often used as a benchmark for fault detection tasks. It involves 52 variables, including 19 composition measurement variables, 22 continuous process variables, and 11 control variables, with 20 different faults. In our study, we partitioned the data of each fault types into three distinct sets: a training set, a validation set, and a test set. The training set, comprising a subset of 2000 samples collected under normal operating conditions, is primarily used for adjusting the parameters (weights and biases) during training. The validation set, composed of a subset of 1000 samples also gathered under normal conditions, is utilized during the training phase to tune hyperparameters and assess the model's performance. This provides a means to combat overfitting and ensure that the model is capable of generalizing well to unseen data. The test set, completely independent of the training and validation data, comprises 1400 samples, with an initial 500 samples collected under normal conditions and the remaining 900 samples collected under various fault conditions. The test set is instrumental in providing an unbiased evaluation of the model's ability to correctly distinguish between normal and faulty states.

Due to the high cost associated with missing the detection of fault samples, the fault detection rate (FDR) is chosen as the metric to evaluate the performance of the proposed method. FDR is a widely accepted measure in fault detection and process monitoring, typically calculated as the ratio of true positive (TP) detections to the total number of actual faults present in the system. It can be expressed using the following formula:  $FDR = TP / (TP + FN)$ , where TPs represent the number of correctly detected faults; false negatives (FNs) represent the number of undetected faults. FDR measures the percentage of true process faults that are detected correctly, with higher FDR indicating a lower likelihood of missing an alarming process fault.

During the experimental phase, we juxtapose the results yielded by our method against a spectrum of comparable methodologies across 20 distinct datasets, each corresponding to a different process fault type in TE process. Moreover, we utilize the mean value derived from a stratified tenfold cross-validation for each fault result. This rigorous procedure corroborates the generalization capacity of our fault detection approach, demonstrating its consistent performance across varied datasets and fault types. The proposed TceOne method is evaluated against several advanced fault detection methods, including kernel principal component analysis (KPCA) [44], variable selection-canonicalvariate analysis-Kullback Leibler divergence (VS-CVA-KLD) [6], adversarial autoencoder (AAE) [14], least squares support vector machine (LS-SVM) [45], squirrel search algorithm and support vector data description (SSA-SVDD) [36], kernel principal component analysis-based support vector data description (KPCA-SVDD) [37], and variational autoencoder-based support vector data description (VAE-SVDD) [40], which have all been shown to be effective for fault detection tasks. Additionally, two ablation models, i.e., TceOne-01 and TceOne-10, are incorporated to evaluate the impact of the TCE module and the proposed subspace distance metric on the performance of

fault detection. The model TceOne-01 substitutes the TCE module with a conventional convolutional module, thereby disregarding the temporal dimension of the time series data. This substitution facilitates an analysis of the TCE module's role in harnessing temporal dependencies. In contrast, the TceOne-10 employs Euclidean distance instead of our novel subspace distance metric, thus enabling us to probe its impact on enhancing the compactness of the decision boundary. By juxtaposing the results gleaned from these ablation models against those from the complete TceOne-based approach, we can discern the effectiveness of the proposed modules and their influence on the fault detection performance.

The layer configurations of the TCE and decoder were determined based on empirical experimental validation and related works in the field of fault detection. The proposed model was trained using the Adam optimizer with a learning rate of 0.001, which decays at a rate of 0.1 per 5 epochs. The batch size was set to 128, and the maximum number of epochs was set to 200. The iterative training termination condition was triggered when the training loss failed to improve for 10 consecutive epochs. Other hyperparameters such as  $w = 5$ ,  $n_{pc} = 10$ ,  $n_{dc} = 10$ ,  $th = 99$ ,  $\alpha = 0.6$ , and  $\beta = 0.4$  were set, and the remaining hyperparameters were optimized using Bayesian optimization to improve the model's reconstruction task. The experiments were implemented using Python 3.6.8 and torch-1.10.0, and executed on a PC equipped with an NVIDIA GeForce RTX 3090 GPU and 128 GB RAM. For the purpose of a rigorous and unbiased comparison, we have diligently tuned the parameters of each comparative model to optimize their performance and yield the most favorable outcomes.

### B. Overall Performance of TceOne

The experimental results demonstrate that the proposed TceOne algorithm outperforms all the compared fault detection methods in the majority of the TE fault datasets. Specifically, TceOne achieves the best FDR values in 15 out of 20 comparisons, as shown in Table I, where the bolded numbers represent the highest FDR values for each fault. Furthermore, TceOne also achieves the best result in the Avg. metrics comparison, indicating its effectiveness in fault detection and competitiveness with other compared methods. It exhibited a robust capability to detect a variety of faults under diverse circumstances, showing its strong generalization performance. These results demonstrate that the proposed method is capable of accurately detecting faults in TE processes and outperforms other state-of-the-art methods.

One of the challenges in fault detection is the difficulty in detecting faults that occur in the presence of feedback control systems. This is demonstrated in the fault detection charts for fault 10, shown in Fig. 5. The feedback controller in this case tries to return to the set point while reducing the influence of the fault, making fault detection challenging. The detection index stops rising and begins to oscillate between intervals, as shown in the figure. However, it should be emphasized that the system is still faulty at this time and will not immediately return to normal. Nearly two hours after the failure occurred, the comparison method was difficult to detect, and TceOne performed better. These results demonstrate the effectiveness

TABLE I  
FDRS FOR ALL FAULTS IN THE TE PROCESS

Fault No.	KPCA	VS-CVA-KLD	AAE	LS-SVM	SSA-SVDD	KPCA-SVDD	VAE-SVDD	TceOne-01	TceOne-10	TceOne
1	0.988	0.996	0.997	0.997	0.967	<b>0.998</b>	0.995	0.994	<b>0.998</b>	<b>0.998</b>
2	0.981	0.988	<b>0.992</b>	0.938	0.935	0.958	0.988	0.982	0.982	0.989
3	0.258	0.252	0.329	0.314	0.429	0.328	0.274	0.231	0.305	<b>0.492</b>
4	<b>1.000</b>	0.969	0.975	<b>1.000</b>	0.834	<b>1.000</b>	0.992	0.785	<b>1.000</b>	<b>1.000</b>
5	0.218	<b>0.447</b>	0.310	0.103	<b>0.447</b>	0.141	0.275	0.266	0.283	0.415
6	0.990	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	0.913	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
7	<b>1.000</b>	0.996	0.997	0.999	0.926	0.998	0.993	0.997	<b>1.000</b>	<b>1.000</b>
8	0.938	0.952	0.961	0.788	0.869	0.851	0.950	0.953	0.957	<b>0.974</b>
9	0.226	0.447	0.361	0.270	0.331	0.247	0.265	0.226	0.322	<b>0.699</b>
10	0.205	0.572	0.713	0.315	0.527	0.363	0.684	0.743	0.763	<b>0.970</b>
11	0.321	0.208	0.794	0.788	0.794	0.849	0.693	0.641	0.785	<b>0.980</b>
12	0.964	0.704	0.935	0.775	0.912	0.878	0.973	0.925	<b>0.981</b>	0.898
13	0.970	0.972	0.960	0.913	0.924	0.905	0.974	0.970	0.971	<b>0.980</b>
14	0.939	0.639	0.982	0.975	0.915	0.975	0.994	<b>0.999</b>	<b>0.999</b>	0.998
15	0.275	0.167	0.274	0.316	<b>0.481</b>	0.316	0.276	0.258	0.237	0.364
16	0.135	0.308	0.381	0.175	0.348	0.158	0.269	0.378	0.385	<b>0.525</b>
17	0.735	0.629	0.926	0.935	0.927	0.957	0.945	0.883	0.933	<b>0.986</b>
18	0.950	0.916	0.935	0.967	0.966	<b>0.986</b>	0.981	0.983	0.985	0.933
19	0.193	0.355	0.378	0.113	0.377	0.171	0.142	0.287	0.270	<b>0.987</b>
20	0.238	0.307	0.312	0.459	0.539	0.463	0.498	0.513	0.585	<b>0.960</b>
Avg.	0.626	0.641	0.716	0.657	0.718	0.677	0.708	0.701	0.737	<b>0.857</b>

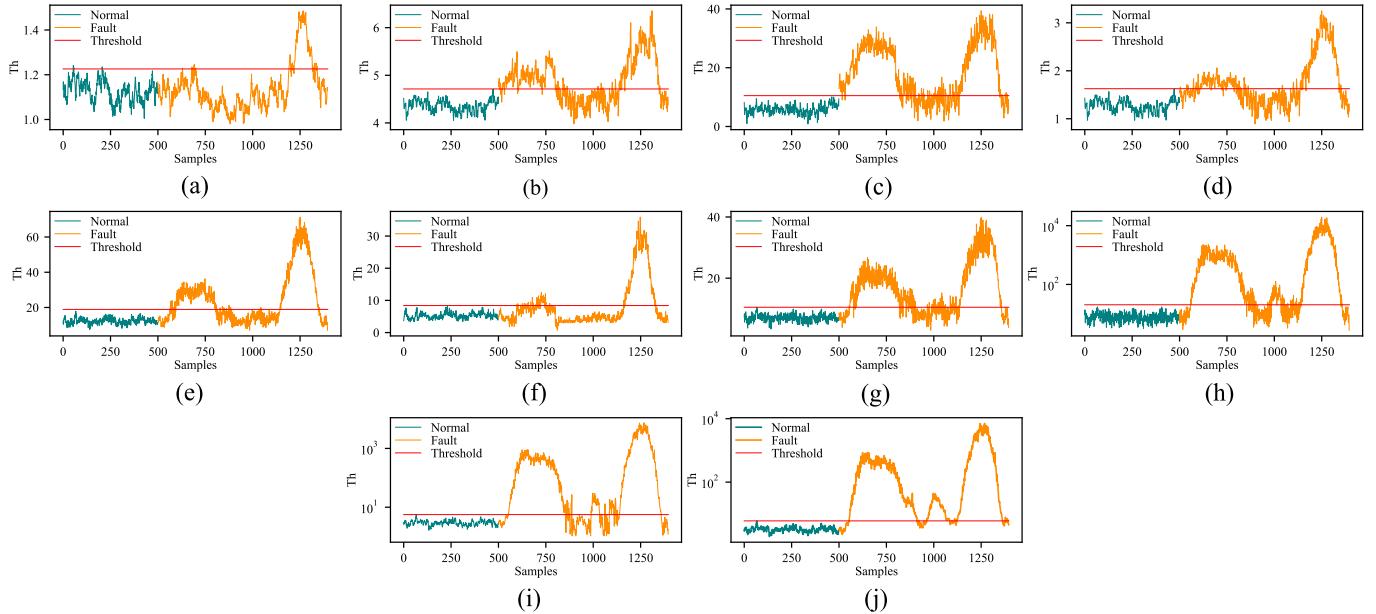


Fig. 5. Fault detection for fault 10 in the TE process, the x-axis and y-axis represent sample timestamps and detection metric, respectively. (a) KPCA. (b) VS-CVA-KLD. (c) AAE. (d) LS-SVM. (e) SSA-SVDD. (f) KPCA-SVDD. (g) VAE-SVDD. (h) Ablation model TceOne-01. (i) Ablation model TceOne-10. (j) Proposed TceOne.

of TceOne in detecting faults even in the presence of feedback control systems.

The TceOne approach is designed with an emphasis on optimizing the FDR. It is a critical measure in the process industry where undetected faults could lead to significant downtime and potential safety risks. This design consideration places TceOne as a practical and effective solution for real-world industrial applications, offering an advantage over other methods that may prioritize different performance metrics. The comparison methods such as KPCA, LS-SVM, SSA-SVDD, and KPCA-SVDD may struggle to accurately encapsulate complex intercorrelations within the multivariate time series. DNN-based methods like AAE and VAE-SVDD, despite their proficiency in modeling nonlinearity, can over-

look critical temporal information. The experimental results also show that TceOne consistently outperforms TceOne-01 and TceOne-10 in fault detection scenarios across most different types of faults. This demonstrates the significance of both the TCE module and the proposed subspace distance metric in achieving high fault detection performance.

TceOne leverages CapsNet to effectively extract temporal features while preserving the chronological order of time series data. This unique capability enables TceOne to accurately capture temporal relationships and configurations of features, leading to robust and meaningful representations. CapsNets emerge as a particularly promising approach due to their capacity to preserve hierarchical spatial and temporal correlations. Nonetheless, they bring about increased complexity.

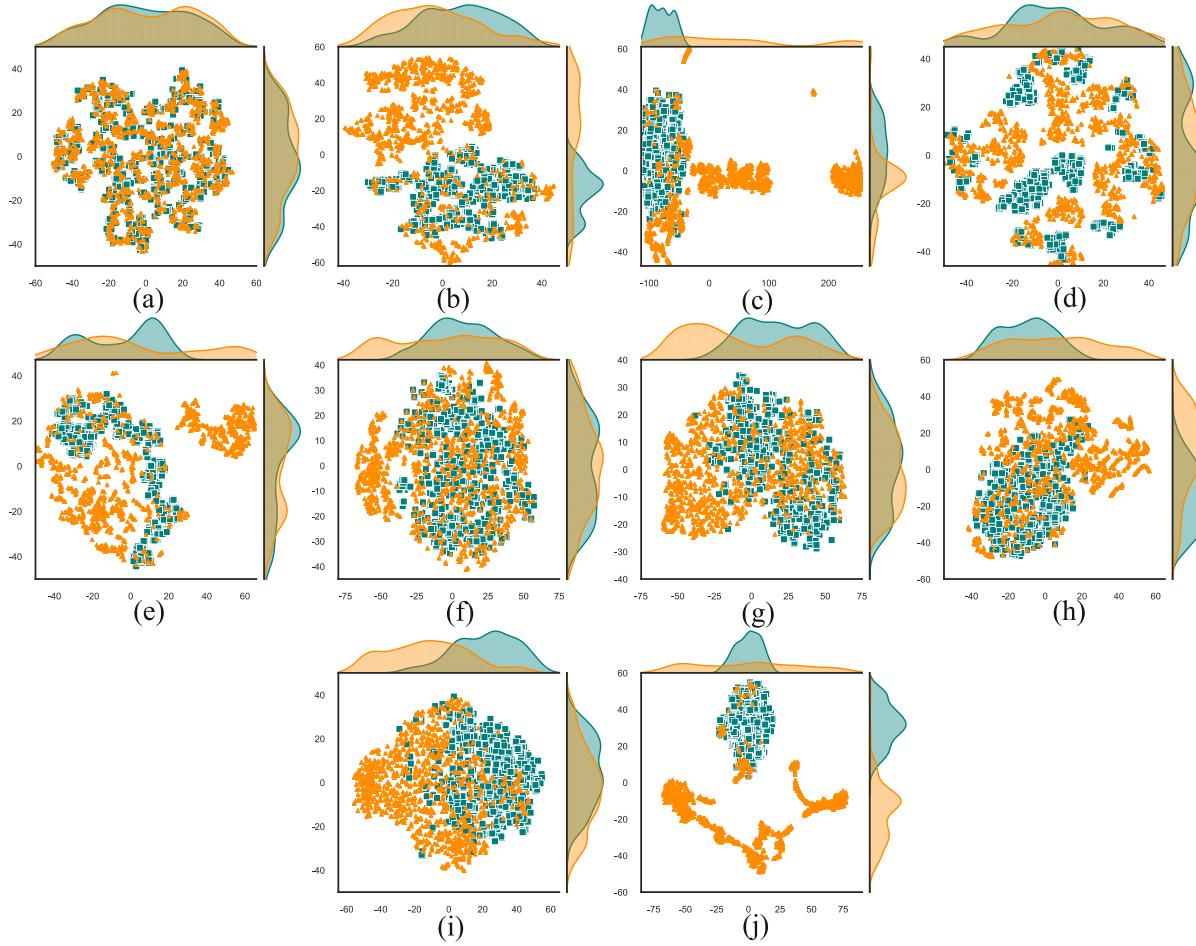


Fig. 6. t-SNE visualization of the subspace feature for fault 10 in the TE process. The teal and orange dots represent the subspace features of the normal and faulty samples after dimensionality reduction, respectively. (a) KPCA. (b) VS-CVA-KLD. (c) AAE. (d) LS-SVM. (e) SSA-SVDD. (f) KPCA-SVDD. (g) VAE-SVDD. (h) Ablation model TceOne-01. (i) Ablation model TceOne-10. (j) Proposed TceOne.

It is worth noting that this primarily manifests as increased off-line training costs and has negligible impact on online detection efficiency. In light of the enhanced fault detection performance, the ability to capture intricate details of temporal correlations, and overall robustness in the face of varied and complex industrial processes, the benefits of CapsNets outweigh the added complexity. Moreover, these comparison methods like LS-SVM, SSA-SVDD, KPCA-SVDD, and VAE-SVDD employ the Euclidean distance metric, which may fall short in accurately capturing the underlying data distribution, particularly in high-dimensional space. To address this, TceOne introduces a new subspace distance metric that melds the cosine angle and magnitude of capsule vectors, providing a superior fit for the discriminative hypersphere and thereby bolstering the effectiveness of fault detection.

Our proposed TceOne approach employs a unique projection strategy that situates normal samples within a compact region in the subspace while dispersing fault samples outside of this boundary. This strategy improves the separation between normal and fault samples. The robustness of TceOne's boundary compactness and its enhanced discrimination ability compared to DSVDD is highlighted by the concentration of normal samples and their distinct separation from

abnormal samples within the subspace. Nonetheless, given the high-dimensional nature of the subspace features in the context of multivariate time series in industrial processes, it is practically challenging to directly observe these features. Consequently, we resort to indirect methods to visualize and validate these subspace features. We utilize t-distributed stochastic neighbor embedding (t-SNE) to visualize these high-dimensional subspace features in our experiments. The t-SNE is a prevalent technique for dimensionality reduction and visualizing of high-dimensional feature spaces [23], [46]. It effectively maps data to a lower dimensional space, preserving the local structure and relationships between data points. Applying t-SNE to the subspace features allows us to depict the sample distribution and relationships in a 2-D space, thereby providing valuable insight into the discriminative power and effectiveness of the TceOne method.

To illustrate the corresponding feature visualization learned by different methods, we selected the TE process dataset of fault 10. We apply t-SNE to reduce the feature representations of the testing data to a 2-D space, providing a comparison of the discrimination of subspace features projected by TceOne and other compared methods. As depicted in Fig. 6, the proposed TceOne method shows a marked ability to confine

normal data within a more compact area while dispersing fault data to more scattered locations. This provides a clearer demarcation between the regions corresponding to normal and fault data, demonstrating superior discriminative abilities compared to the compared methods.

Within the t-SNE visualization, subspace features derived from methods, such as KPCA, LS-SVM, often exhibit a degree of scattering, indicating a certain level of misclassification or less-than-ideal distinction between normal and faulty states. KPCA and KPCA-SVDD are a nonlinear dimensionality reduction technique. Despite their ability to handle nonlinear patterns, they struggle with multivariate time series that bear complex temporal dependencies. In addition, the effectiveness of KPCA is contingent on the choice of kernel and its parameters. VS-CVA-KLD assumes a Gaussian distribution of data, a supposition that may not be consistently met in real-world situations. AAE, LS-SVM, and SSA-SVDD may also fall short in adequately modeling temporal correlations among different time steps within multivariate time series. The features derived from ablation models, i.e., TceOne-01 and TceOne-10, appear more overlapped and less distinguishable, signifying less discriminative power. TceOne-01 does not utilize the TCE, which means that important temporal patterns in the data could be overlooked. This lack of temporal context might result in a less effective separation between normal and faulty states. TceOne-10 uses a traditional Euclidean distance metric rather than the proposed subspace distance measure. The absence of this component may not fully discern the subtle differences between normal and faulty states, leading to reduced discriminative power.

Conversely, the subspace features produced by TceOne form two distinct clusters, clearly separated from one another. TceOne effectively confines normal data to a compact region while mapping fault data to more dispersed locations, thereby exhibiting superior discriminative abilities. This corroborates our proposition that the fusion of TCE and the enhanced subspace distance metric yields more discriminative subspace representations, advantageous for fault detection. TCE adeptly extracts and encapsulates temporal features, providing TceOne an edge over traditional methods in rendering complex temporal relationships within time series. Furthermore, MSA discerns interdependencies among primary capsules, capturing significant contextual information that might be overlooked by alternative methods. The subspace distance metric, combining the cosine angle and magnitude of capsule vectors. This further enhances the capability of TceOne to learn discriminative features, resulting in a more effective separation between normal and faulty states in the feature subspace.

Overall, the experimental findings attest to the efficacy of the proposed TceOne algorithm in identifying faults within TE processes. In the majority of fault detection scenarios, TceOne surpasses the performance of other contemporary methods and maintains its fault detection capabilities effectively, even under the influence of feedback control systems. The visualization of feature representations further endorses the discriminative prowess of TceOne and its proficiency in deriving compact and discriminative subspace representations, crucial for fault detection. The implications of these results carry substantial

weight within industrial contexts, where the criticality of fault detection for upholding process safety and reliability cannot be overstated.

## VI. CONCLUSION

The findings of this study demonstrate the efficacy of a novel fault detection approach, based on TceOne, within process industries. Distinct from traditional feature learning methodologies, the proposed TCE integrated with subspace distance metric allows for comprehensive learning of temporal correlation features embedded within multivariate time series. Consequently, this enriches the discriminability between normal and fault samples within the subspace feature representation. Empirical results derived from the TE process validate the superiority of TceOne over alternative strategies in terms of FDR. The enhanced effectiveness is further corroborated by the feature visualization outcomes. Our method confines normal data within a more compact region while dispersing fault data across a wider area. As such, TceOne boasts improved discriminative ability relative to other comparative methodologies. Overall, our proposed TceOne-based fault detection method, offers a promising solution for detecting faults in process industries. It empowers real-time monitoring of process data in chemical, power generation, and manufacturing industries. Through timely identification of anomalous patterns, operators can swiftly act to mitigate risk, forestall accidents, reduce downtime, and optimize production efficiency. The implementation of our method can substantially boost safety measures, enhance operational efficiency, and promote heightened productivity and profitability within process industries.

In future research, two areas hold promise for advancing fault detection in process industries. First, exploring the combination of multiple one-class classifiers could bolster detection performance by leveraging specialized classifiers for diverse fault patterns. Secondly, addressing the dynamic nature of process industries is crucial. Developing adaptive subspace learning models capable of adapting to evolving systems and changing operating conditions represents an important research direction.

## REFERENCES

- [1] C. Liu, K. Wang, Y. Wang, and X. Yuan, "Learning deep multimanifold structure feature representation for quality prediction with an industrial application," *IEEE Trans. Ind. Informat.*, vol. 18, no. 9, pp. 5849–5858, Sep. 2022.
- [2] D. Liu, Y. Wang, C. Liu, X. Yuan, C. Yang, and W. Gui, "Data mode related interpretable transformer network for predictive modeling and key sample analysis in industrial processes," *IEEE Trans. Ind. Informat.*, vol. 19, no. 9, pp. 9325–9336, Sep. 2023, doi: [10.1109/TII.2022.3227731](https://doi.org/10.1109/TII.2022.3227731).
- [3] Y. Tao, H. Shi, B. Song, and S. Tan, "A novel dynamic weight principal component analysis method and hierarchical monitoring strategy for process fault detection and diagnosis," *IEEE Trans. Ind. Electron.*, vol. 67, no. 9, pp. 7994–8004, Sep. 2020.
- [4] M. Wang, D. Zhou, and M. Chen, "Recursive hybrid variable monitoring for fault detection in nonstationary industrial processes," *IEEE Trans. Ind. Informat.*, vol. 18, no. 10, pp. 7296–7304, Oct. 2022.
- [5] P. Wu, R. M. G. Ferrari, Y. Liu, and J.-W. van Wingerden, "Data-driven incipient fault detection via canonical variate dissimilarity and mixed kernel principal component analysis," *IEEE Trans. Ind. Informat.*, vol. 17, no. 8, pp. 5380–5390, Aug. 2021.

- [6] J. Dong, L. Jiang, C. Zhang, and K. Peng, "A novel quality-related incipient fault detection method based on canonical variate analysis and Kullback–Leibler divergence for large-scale industrial processes," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–10, 2022.
- [7] Y. Si, Y. Wang, and D. Zhou, "Key-performance-indicator-related process monitoring based on improved kernel partial least squares," *IEEE Trans. Ind. Electron.*, vol. 68, no. 3, pp. 2626–2636, Mar. 2021.
- [8] P. Zhou, R. Zhang, J. Xie, J. Liu, H. Wang, and T. Chai, "Data-driven monitoring and diagnosing of abnormal furnace conditions in blast furnace ironmaking: An integrated PCA-ICA method," *IEEE Trans. Ind. Electron.*, vol. 68, no. 1, pp. 622–631, Jan. 2021.
- [9] X. Cheng and S. Cheng, "Infrared thermographic fault detection using machine vision with convolutional neural network for blast furnace chute," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–9, 2022.
- [10] Z. Zhi, L. Liu, D. Liu, and C. Hu, "Fault detection of the harmonic reducer based on CNN-LSTM with a novel denoising algorithm," *IEEE Sensors J.*, vol. 22, no. 3, pp. 2572–2581, Feb. 2022.
- [11] C. Wang, Z. Wang, W. Liu, Y. Shen, and H. Dong, "A novel deep offline-to-online transfer learning framework for pipeline leakage detection with small samples," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–13, 2023.
- [12] W. Yu, C. Zhao, and B. Huang, "MoniNet with concurrent analytics of temporal and spatial information for fault detection in industrial processes," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 8340–8351, Aug. 2022.
- [13] X. Kong, X. Li, Q. Zhou, Z. Hu, and C. Shi, "Attention recurrent autoencoder hybrid model for early fault diagnosis of rotating machinery," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021.
- [14] K. Jang, S. Hong, M. Kim, J. Na, and I. Moon, "Adversarial autoencoder based feature learning for fault detection in industrial processes," *IEEE Trans. Ind. Informat.*, vol. 18, no. 2, pp. 827–834, Feb. 2022.
- [15] A. Maged, C. F. Lui, S. Hardiy, and M. Xie, "Variational autoencoders-LSTM based fault detection of time-dependent high dimensional processes," *Int. J. Prod. Res.*, pp. 1–16, Feb. 2023, doi: [10.1080/00207543.2023.2175591](https://doi.org/10.1080/00207543.2023.2175591).
- [16] J. Yu, X. Liu, and L. Ye, "Convolutional long short-term memory autoencoder-based feature learning for fault detection in industrial processes," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–15, 2021.
- [17] C. Tian and C. Zhao, "Single model-based analysis of relative causal changes for root-cause diagnosis in complex industrial processes," *Ind. Eng. Chem. Res.*, vol. 60, no. 34, pp. 12602–12613, Sep. 2021.
- [18] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 3856–3866.
- [19] G. E. Hinton, A. Krizhevsky, and S. D. Wang, "Transforming auto-encoders," in *Proc. 21st Int. Conf. Artif. Neural Netw.* New York, NY, USA: Springer-Verlag, 2011, pp. 44–51.
- [20] G. Yang, H. Tao, R. Du, and Y. Zhong, "Compound fault diagnosis of harmonic drives using deep capsule graph convolutional network," *IEEE Trans. Ind. Electron.*, vol. 70, no. 4, pp. 4186–4195, Apr. 2023.
- [21] R. Huang, J. Li, Y. Liao, J. Chen, Z. Wang, and W. Li, "Deep adversarial capsule network for compound fault diagnosis of machinery toward multidomain generalization task," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.
- [22] D. Zhao, S. Liu, T. Zhang, H. Zhang, and Z. Miao, "Subdomain adaptation capsule network for unsupervised mechanical fault diagnosis," *Inf. Sci.*, vol. 611, pp. 301–316, Sep. 2022.
- [23] D. Li, Q. Tao, J. Liu, and H. Wang, "Center-aware adversarial autoencoder for anomaly detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 6, pp. 2480–2493, Jun. 2022.
- [24] L. Ruff et al., "Deep one-class classification," in *Proc. 35th Int. Conf. Mach. Learn.*, 2018, pp. 4393–4402.
- [25] T. Ergen and S. S. Kozat, "Unsupervised anomaly detection with LSTM neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 8, pp. 3127–3141, Aug. 2020.
- [26] Q. Xie, G. Tao, C. Xie, and Z. Wen, "Abnormal data detection based on adaptive sliding window and weighted multiscale local outlier factor for machinery health monitoring," *IEEE Trans. Ind. Electron.*, vol. 70, no. 11, pp. 11725–11734, Nov. 2023.
- [27] C. Zhang, D. Hu, and T. Yang, "Anomaly detection and diagnosis for wind turbines using long short-term memory-based stacked denoising autoencoders and XGBoost," *Rel. Eng. Syst. Saf.*, vol. 222, Jun. 2022, Art. no. 108445.
- [28] L. Meneghetti, M. Terzi, S. Del Favero, G. A. Susto, and C. Cobelli, "Data-driven anomaly recognition for unsupervised model-free fault detection in artificial pancreas," *IEEE Trans. Control Syst. Technol.*, vol. 28, no. 1, pp. 33–47, Jan. 2020.
- [29] M. E. Basiri et al., "Improving sentiment polarity detection through target identification," *IEEE Trans. Computat. Social Syst.*, vol. 7, no. 1, pp. 113–128, Feb. 2020.
- [30] S. Chen, R. Yang, M. Zhong, X. Xi, and C. Liu, "A random forest and model-based hybrid method of fault diagnosis for satellite attitude control systems," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–13, 2023.
- [31] W. Du, Z. Guo, C. Li, X. Gong, and Z. Pu, "From anomaly detection to novel fault discrimination for wind turbine gearboxes with a sparse isolation encoding forest," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–10, 2022.
- [32] H. Wang, W. Jiang, X. Deng, and J. Geng, "A new method for fault detection of aero-engine based on isolation forest," *Measurement*, vol. 185, Nov. 2021, Art. no. 110064.
- [33] S. Fong and S. Narasimhan, "An unsupervised Bayesian OC-SVM approach for early degradation detection, thresholding, and fault prediction in machinery monitoring," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.
- [34] Y. Zhu, C. Du, Z. Liu, Y. Chen, and Y. Zhao, "A turboshaft aeroengine fault detection method based on one-class support vector machine and transfer learning," *IEEE Trans. Semicond. Manuf.*, vol. 35, no. 3, pp. 457–469, Aug. 2022.
- [35] C. I. Lang et al., "One class process anomaly detection using kernel density estimation methods," *J. Aerosp. Eng.*, vol. 35, no. 6, pp. 457–469, Nov. 2022.
- [36] J. A. Navarro-Acosta, I. D. García-Calvillo, and E. O. Reséndiz-Flores, "Fault detection based on squirrel search algorithm and support vector data description for industrial processes," *Soft Comput.*, vol. 26, no. 24, pp. 13639–13650, Jul. 2022.
- [37] L. Cai, H. Yin, J. Lin, H. Zhou, and D. Zhao, "A relevant variable selection and SVDD-based fault detection method for process monitoring," *IEEE Trans. Autom. Sci. Eng.*, early access, Aug. 17, 2022, doi: [10.1109/TASE.2022.3198668](https://doi.org/10.1109/TASE.2022.3198668).
- [38] Q. Wu, W. Lu, and X. Yan, "Process monitoring of nonlinear uncertain systems based on part interval stacked autoencoder and support vector data description," *Appl. Soft Comput.*, vol. 129, Nov. 2022, Art. no. 109570.
- [39] W. Mao, J. Chen, X. Liang, and X. Zhang, "A new online detection approach for rolling bearing incipient fault via self-adaptive deep feature matching," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 2, pp. 443–456, Feb. 2020.
- [40] Y. Zhou, X. Liang, W. Zhang, L. Zhang, and X. Song, "VAE-based deep SVDD for anomaly detection," *Neurocomputing*, vol. 453, pp. 131–140, Sep. 2021.
- [41] W. Chen and K. Shi, "Multi-scale attention convolutional neural network for time series classification," *Neural Netw.*, vol. 136, pp. 126–140, Apr. 2021.
- [42] Y. Zhou, K. Xu, and F. He, "Root cause diagnosis in multivariate time series based on modified temporal convolution and multi-head self-attention," *J. Process Control*, vol. 117, pp. 14–25, Sep. 2022.
- [43] A. Bathelt, N. L. Ricker, and M. Jelali, "Revision of the Tennessee Eastman Process model," *IFAC-PapersOnLine*, vol. 48, no. 8, pp. 309–314, 2015.
- [44] D. Zheng, L. Zhou, and Z. Song, "Kernel generalization of multi-rate probabilistic principal component analysis for fault detection in nonlinear process," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 8, pp. 1465–1476, Aug. 2021.
- [45] G. Kaur, P. Chanak, and M. Bhattacharya, "Obstacle-aware intelligent fault detection scheme for industrial wireless sensor networks," *IEEE Trans. Ind. Informat.*, vol. 18, no. 10, pp. 6876–6886, Oct. 2022.
- [46] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 86, pp. 2579–2605, Nov. 2008.



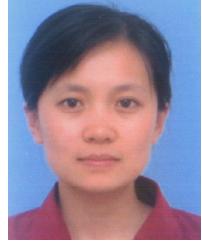
**Sheng Wang** received the B.S. degree in automation and the M.S. degree in control engineering from the College of Information Science and Engineering, Northeastern University, Shenyang, China, in 2017 and 2020, respectively, where he is currently pursuing the Ph.D. degree in control science and engineering.

His research interests include process control, fault detection and diagnosis, and causality discovery.



**Qiang Zhao** received the Ph.D. degree in navigation guidance and control from the College of Information Science and Engineering, Northeastern University, Shenyang, China, in 2017.

He is currently an Associate Professor with the School of Control Engineering, Northeastern University at Qinhuangdao, Qinhuangdao, China. His research interests include fault detection and diagnosis, industry big data analysis, and data-driven quality monitoring.



**Yinghua Han** received the B.S., M.S., and Ph.D. degrees in navigation guidance and control from the College of Information Science and Engineering, Northeastern University, Shenyang, China, in 2003, 2005, and 2008, respectively.

She is currently a Professor with the School of Computer and Communication Engineering, Northeastern University at Qinhuangdao, Qinhuangdao, China. Her research interests include industry big data analysis and optimal operation of smart grid.



**Jinkuan Wang** (Member, IEEE) received the M.S. degree in automation from Northeastern University, Shenyang, China, in 1985, and the Ph.D. degree in information and communication from the University of Electro-Communications, Chofu, Japan, in 1993.

In 1990, he joined the Institute of Space Astronautical Science, Sagamihara, Japan, as a Special Member. Since 1994, he has been an Engineer with the Research Department, COSEL, Toyama, Japan. Since 1998, he has been a Professor with the College of Information Science and Engineering, Northeastern University. His research interests include industry big data analysis, intelligent control, wireless sensor networks, adaptive array signal processing, and optimal operation of smart grid.