This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/TII.2023.3275701, IEEE Transactions on Industrial Informatics

Subdomain-Alignment Data Augmentation for Pipeline Fault Diagnosis: An Adversarial Self-Attention Network

Chuang Wang, Zidong Wang, *Fellow, IEEE*, Lifeng Ma, *Member, IEEE*, Hongli Dong, *Senior Member, IEEE*, and Weiguo Sheng, *Member, IEEE*

***Abstract*—Data augmentation (DA) has the potential to address the issue of imbalanced and insufficient datasets (I&ID) in pipeline fault diagnosis. However, the majority of existing DA methods for time series are inspired by computer vision techniques, ignoring the temporal dynamic properties and fine-grained fault features, which leads to limited performance of the augmentation. To tackle this problem, we introduce a novel DA approach called the subdomain-alignment adversarial self-attention network (SA-ASN), which takes into account both temporal association and semantic correlation. Our approach features a novel temporal association learning (TAL) mechanism, which transfers temporal information from the discriminator to the generator via a customized knowledge-sharing structure, improving the reliability of synthetic long-range associations. Additionally, we introduce a prototype-assisted subdomain alignment (PASA) strategy that forms a hierarchical structure in the synthetic dataset by incorporating local semantic correlation into the model training. With the support of TAL and PASA, our SA-ASN algorithm enhances the authenticity of temporal structure at the instance level and improves the discriminability of fault features at the category level. Our experimental results show that the SA-ASN algorithm provides a more diverse and accurate augmentation of pipeline data. The effectiveness of our SA-ASN algorithm encourages the use of data-driven diagnostic models in complex real-world oilfield pipeline networks.**

***Index Terms*—Adversarial learning, data augmentation, multi-head self-attention mechanism, pipeline fault diagnosis, subdomain alignment.**

I. Introduction

Fault diagnosis plays a crucial role in reducing maintenance costs and enhancing the reliability of oil and gas transportation systems. In recent years, data-driven intelligent approaches have made significant advancements in industrial fault diagnosis [6], [14], [24]. However, the success of these approaches is heavily dependent on balanced data categories and a sufficient amount of labeled data. This is not always the case in pipeline fault diagnosis, as some fault categories are rarely available due to normal operation and obtaining labeled data is expensive, time-consuming, and challenging. As a result, the performance of existing data-driven approaches decreases when applied to pipeline fault diagnosis. As illustrated in Fig. 1, a diagnostic model trained on imbalanced and insufficient datasets (I&IDs) fails to capture discriminative features, resulting in a high rate of false alarms and missed alarms. Hence, there is an urgent need to address the issue of I&ID to ensure the efficacy of the diagnostic model.

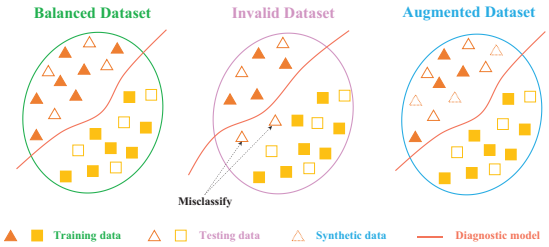


Fig. 1: Intelligent data-driven diagnostic models trained on balanced dataset, invalid dataset, and augmented dataset, respectively.

To address the challenging I&ID issue, much research effort has been made to increase the size of the original dataset by generating synthetic data, also known as data augmentation (DA) [4]. Currently, this line of research can be broadly categorized into three categories: random transformation, pattern mixing, and generative models [7]. Random transformation-based DA methods primarily focus on adding random noise, flipping or cropping, time-domain warping, and frequency-domain warping. For instance, a deep learning approach for rotating machinery fault diagnosis has been proposed in [13], where DA techniques were used to generate additional valid samples for model training. In [8], a dynamic time-warping method has been developed to augment run-to-fail data to predict the remaining useful life of machinery. However, these random transformation methods have limited effectiveness

This work was supported in part by the National Natural Science Foundation of China under Grants 61933007, U21A2019, and 62273180, the Hainan Province Science and Technology Special Fund of China under Grant ZDYF2022SHFZ105, the "Pioneer" and "Leading Goose" R&D Program of Zhejiang Province of China under Grant 2023C01022, and the Alexander von Humboldt Foundation of Germany. *(Corresponding author: Hongli Dong.)*

Chuang Wang and Hongli Dong are with Sanya Offshore Oil & Gas Research Institute, Northeast Petroleum University, Sanya 572025, China, also with the Artificial Intelligence Energy Research Institute, Northeast Petroleum University, Daqing 163318, China, and also with the Heilongjiang Provincial Key Laboratory of Networking and Intelligent Control, Northeast Petroleum University, Daqing 163318, China. (E-mails: wangchuang64@126.com, shiningdhl@vip.126.com)

Zidong Wang is with the Department of Computer Science, Brunel University London, Uxbridge, Middlesex, UB8 3PH, United Kingdom. (Email: Zidong.Wang@brunel.ac.uk)

Lifeng Ma is with the School of Automation, Nanjing University of Science and Technology, Nanjing 210094, China. (E-mail: malifeng@njust.edu.cn)

Weiguo Sheng is with School of Information Science and Engineering, Hangzhou Normal University, Hangzhou 311121, China. (Email: weiguouk@hotmail.com)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/TII.2023.3275701, IEEE Transactions on Industrial Informatics

in improving diagnostic performance, as they only alter the shape of original samples rather than generating new ones. To overcome this limitation, pattern mixing-based DA methods have been developed that combine two or more patterns to generate a new one. In this regard, one widely used augmentation method is the synthetic minority oversampling technique (SMOTE) [2], which generates a new sample by interpolating between an existing sample and its K-nearest neighbors. Nevertheless, data interpolation is susceptible to hard negative samples, which often leads to ambiguous sample features and reduces classification performance.

Generative models, on the other hand, explore untapped real-world distributions to produce new samples. This method not only increases the diversity of features but also ensures that the samples are valid. The most well-known generative model is the generative adversarial network (GAN) [5], which has become a popular research topic in DA due to its powerful feature extraction and representation capabilities [18], [26], [29]. For instance, an approach combining the GAN algorithm and Gaussian mixture model has been proposed in [18] to solve the I&ID problem. Furthermore, a multi-scale progressive GAN algorithm has been put forward in [29] to synthesize the surface defect images. Moreover, a multi-granularity GAN algorithm has been developed in [26] to improve the quantity and quality of wafer maps for accurate defect identification.

Thus far, most existing GAN-based time-series data augmentation methods have taken inspiration from visual DA techniques, and these methods may not perform optimally when applied to pipeline fault diagnosis for the following reasons. Firstly, these methods are designed to capture pixel-level features which are not suitable for learning time-series temporal dynamics. These results in the generator having difficulty in capturing the variation trend of the time-series, leading to unreliable long-range associations in the synthetic series. Additionally, previous methods primarily focus on aligning the global distribution between the real and synthetic datasets. As a result, synthetic data from different health states may be blended and deviate from real data under the corresponding health state, causing ambiguous decision boundaries. To address these limitations, the objective of this paper is to propose an advanced time-series generative model.

The key contributions of this paper are as follows.

1) We propose a new DA approach, referred to as the subdomain-alignment adversarial self-attention network (SA-ASN), which explicitly takes into account both temporal association and semantic correlation.
2) We design a temporal association learning (TAL) mechanism to transfer the temporal dynamic information from the discriminator to the generator through a customized knowledge-sharing structure, thereby enhancing the reliability of the synthetic time series.
3) We implement a prototype-assisted subdomain alignment (PASA) strategy to incorporate semantic correlation into data generation, resulting in maximized intra-class compactness and inter-class separability.
4) Through qualitative visualization and quantitative evaluation, our SA-ASN algorithm is compared to state-of-the-art DA methods. The results demonstrate the effectiveness of our approach in addressing the I&ID issue.

The structure of this paper is as follows: In Section II, related works including intelligent fault diagnosis and attention mechanism are discussed. Section III provides a detailed description of the proposed SA-ASN algorithm. The experimental results and analysis are presented in Section IV. Finally, in Section V, the conclusions of this paper are drawn.

II. Related Work

A. Intelligent fault diagnosis

With the advancement of deep learning technologies, data-driven intelligent fault diagnosis has seen remarkable success in various industrial applications [27]. For instance, [1] introduces a classifier based on Long Short-Term Memory (LSTM) that learns the temporal association and predicts the health state of the pipeline. Another example is the ensemble learning framework in [21], which combines a Sparse Autoencoder Network and an improved Support Vector Machine to enhance the accuracy of pipeline fault detection. Despite these successes, the majority of these intelligent methods depend heavily on balanced and sufficient datasets that cover all health states. As a result, ensuring the quality and quantity of the data is crucial to improve the accuracy of pipeline fault diagnosis. In recent years, data augmentation methods are developed to address data scarcity and imbalance by augmenting imperfect data with satisfactory quality and desired diversity. For example, [6] puts forth a Trinetworks-form based GAN algorithm for pipeline leakage detection in the presence of incomplete sensor data. Additionally, [28] proposes a mixed-GANs algorithm to provide additional data for training a high-accuracy model in terms of pipeline leakage detection.

Notably, the above GAN-based data augmentation methods borrow ideas directly from computer vision without considering temporal dynamic properties and fine-grained fault features, resulting in limited augmentation performance. To address this problem, some researchers have attempted to construct a GAN framework by utilizing a network structure suitable for learning time series. For example, [14] proposes a hybrid GAN framework based on bidirectional LSTM for fault diagnosis in machine health monitoring. Moreover, [24] develops a method combining GAN and LSTM to predict pipeline leakage. Nevertheless, LSTM-based generators and discriminators have difficulty in obtaining global optimal solutions due to the vanishing gradient problem and limited computational resources. Therefore, this paper proposes an SA-ASN algorithm that integrates a multi-head attention mechanism to recover complex temporal structures and learn fine-grained features.

B. Attention mechanism in fault diagnosis

In recent years, attention mechanism has gained widespread popularity in the field of time-series data processing for fault diagnosis due to its ability to model internal dependencies and global representations effectively [10], [11], [20], [23] and the references therein. For instance, a fault-attention generative

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/TII.2023.3275701, IEEE Transactions on Industrial Informatics

probabilistic adversarial autoencoder has been proposed in [23] for anomaly detection by discovering the low-dimensional manifold embedded in the high-dimensional space of the signal, where the fault-attention abnormal state indicator has been constructed using the distribution probability of low-dimensional features and reconstruction errors. Furthermore, an attention recurrent autoencoder has been presented in [10] for early fault diagnosis and severity detection of rotating machinery. Moreover, a joint attention feature transfer network has been developed in [11] to deal with the problem of data imbalance by transferring generalized representations obtained from other classes to the class with scarce samples, where the attention module has been used to maintain the discriminability between the features of different classes.

III. A Novel SA-ASN Algorithm

In this section, we will provide a comprehensive overview of our proposed SA-ASN algorithm. Subsequently, the core components of the SA-ASN algorithm, such as the TAL mechanism and the PASA strategy, will be described in detail. Finally, we will present a summary of the training process for the SA-ASN algorithm in the form of a pseudo-code algorithm.

A. Overview of SA-ASN algorithm

A new SA-ASN algorithm, as shown in Fig. 2, is proposed to improve the diagnosis accuracy of pipeline faults by addressing the I&ID problem. The algorithm starts by utilizing the TAL mechanism to learn long-range temporal associations from the discriminator through a customized knowledge-sharing structure. The PASA strategy is then introduced to draw together synthetic and real data from the same underlying category, while separating data from different categories, as demonstrated in Fig. 3. With the support of the TAL mechanism and the PASA strategy, the SA-ASN algorithm is able to recover the instance-level variation tendencies and construct the category-level hierarchical structure. Finally, the algorithm is updated using an alternating iterative technique, following the training approach proposed in [5].

B. Temporal association learning (TAL) mechanism

In the domain of time series analysis, the complexity of the features poses a challenge when compared to image data. Directly applying GAN algorithms from computer vision may result in invalid synthetic temporal structures. To overcome this challenge, we propose the TAL mechanism to learn and recover the complex dynamic properties in real data.

The discriminator in a GAN identifies real and synthetic data by comparing the most discriminative regions. Based on this principle, it is meaningful to recover these distinguishable features contained in the temporal structure to ensure the reliability of the synthetic series. The self-attention mechanism, as proposed in [20], is an effective tool for modeling global dependencies of series. Its multi-head version, which connects feature information from different subspaces, is even more effective in highlighting discriminative features. Accordingly, in the TAL mechanism, we first introduce multi-head self-attention blocks into the discriminator to capture long-range associations and highlight discriminative features from real data. Next, we employ a well-designed knowledge-sharing structure to transfer this information from the discriminator to the generator, leading to the production of effective synthetic temporal structures.

Here, we provide a detailed description of the TAL mechanism as follows, where the definitions of relevant functions (e.g. LayerNorm$(\cdot)$, LeakyRelu$(\cdot)$, Softmax$(\cdot)$, Concat$(\cdot)$) can be found in [20].

The discriminator and the generator in the SA-ASN algorithm are constructed by stacking one-dimensional convolution blocks and multi-head self-attention blocks alternately. Formally, the overall equations of the discriminator at the l-th layer are described as follows:

$$\mathcal{F}_D^l = \mathrm{LayerNorm}\left(\mathrm{MA}\left(\mathcal{X}^{l-1}\right) + \mathcal{X}^{l-1}\right), \tag{1}$$

$$\mathcal{X}^l = \mathrm{LeakyRelu}\left(\mathrm{CB}\left(\mathcal{F}_D^l\right)\right), \tag{2}$$

where $\mathcal{X}^{l-1} \in \mathbb{R}^{N\times 1\times J}$, $l \in \{1, 2, \ldots, L\}$ is the input of the l-th layer with length J, N denotes input volume, $\mathcal{F}_D^l$ is the hidden representation, $\mathrm{CB}(\cdot)$ denotes the convolution block, and $\mathrm{MA}(\cdot)$ represents the multi-head self-attention block. Note that the transferable knowledge obtained by $\mathrm{MA}(\cdot)$ in the discriminator is shown as follows:

$$\mathcal{S}_{D,i}^l = \mathrm{Softmax}\left(\frac{\mathcal{Q}_{D,i}^l {\mathcal{K}_{D,i}^l}^T}{\sqrt{d_i^l}}\right), \tag{3}$$

$$\mathcal{H}_{D,i}^l = \mathcal{S}_{D,i}^l \mathcal{V}_{D,i}^l, \tag{4}$$

where $\mathcal{Q}_{D,i}^l = \mathcal{X}_i^{l-1} W_{\mathcal{Q},D,i}^l$, $\mathcal{K}_{D,i}^l = \mathcal{X}_i^{l-1} W_{\mathcal{K},D,i}^l$, and $\mathcal{V}_{D,i}^l = \mathcal{X}_i^{l-1} W_{\mathcal{V},D,i}^l \in \mathbb{R}^{N\times 1\times \frac{d_i^l}{h}}$ represent the query, key, value of i-th head of MA, respectively; h denotes the number of head; $W_{\mathcal{Q},D,i}^l$, $W_{\mathcal{K},D,i}^l$, and $W_{\mathcal{V},D,i}^l \in \mathbb{R}^{J\times \frac{d_i^l}{h}}$ are the trainable parameter matrices; d^l is the dimension of parameter matrices; $\mathcal{S}_{D,i}^l$ denotes the self-attention map (i.e., the series-associations); and $\mathcal{H}_{D,i}^l$ represents the hidden representation after the self-attention in the l-th layer. Therefore, the output of $\mathrm{MA}\left(\mathcal{X}^{l-1}\right)$ can be presented as follows:

$$\mathcal{MH}_D^l(\mathcal{Q}, \mathcal{K}, \mathcal{V}) = \mathrm{Concat}(\mathcal{H}_{D,1}^l, \mathcal{H}_{D,2}^l, \ldots, \mathcal{H}_{D,h}^l) W_{\mathcal{O},D}^l, \tag{5}$$

where $W_{\mathcal{O},D}^l \in \mathbb{R}^{d^l\times d^l}$ is the parameter matrix. To model the temporal association in the sample space, the discriminator is first optimized to distinguish between the real and synthetic series, which can be formulated as follows:

$$\begin{aligned}\max_{\theta_D} \mathcal{L}_D =& \mathbb{E}_{\mathcal{X}\sim P_S(\mathcal{X})}\left[\log D\left(\mathcal{X}\right)\right] \\ &+ \mathbb{E}_{\mathcal{Z}\sim P_L(\mathcal{Z})}\left[\log\left(1 - D\left(G\left(\mathcal{Z}\right)\right)\right)\right],\end{aligned} \tag{6}$$

where $\mathcal{X}$ denotes the real series obtained from sample space $P_S(\mathcal{X})$, $\mathcal{Z}$ represents the latent variable sampled from latent space $P_L(\mathcal{Z})$, $G\left(\mathcal{Z}\right)$ is the data generated by the generator, θ_D is the network parameter of the discriminator, and $D(\cdot)$ is the probability that the discriminator outputs. Then, we freeze

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/TII.2023.3275701, IEEE Transactions on Industrial Informatics

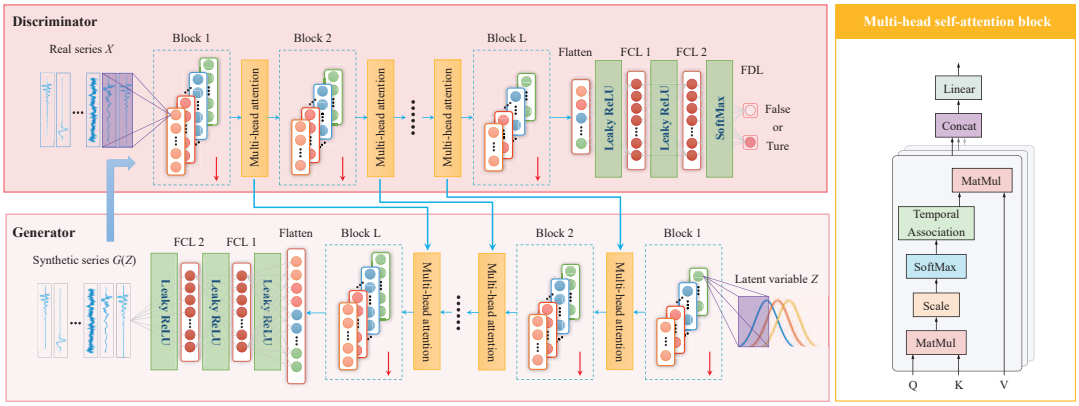


Fig. 2: Architecture of the proposed SA-ASN algorithm.

the discriminator and transfer its self-attention maps into the generator, which can be formalized as follows:

$$\mathcal{H}_{G,i}^{l} = \mathcal{S}_{D,i}^{L-l+1} V_{G,i}^{l}, \tag{7}$$

$$\mathcal{MH}_{G}^{l}(\mathcal{Q},\mathcal{K},\mathcal{V}) = \mathrm{Concat}(\mathcal{H}_{G,1}^{l},\mathcal{H}_{G,2}^{l},\ldots,\mathcal{H}_{G,h}^{l}) W_{\mathcal{O},G}^{l}, \tag{8}$$

where $\mathcal{S}_{D,i}^{L-l+1}$ denotes the self-attention map of the $(L-l+1)$-th layer in the sample space; $\mathcal{V}_{G,i}^{l} = \mathcal{Z}_{i}^{l-1} W_{\mathcal{V},G,i}^{l}$ is the value matrix. Finally, we train the generator to deceive the discriminator by mapping the latent variables into the synthetic data that is similar to the real data, which can be represented as follows:

$$\min_{\theta_G} \mathcal{L}_G = \mathbb{E}_{\mathcal{Z}\sim P_L(\mathcal{Z})} \left[\log\left(1 - D\left(G\left(\mathcal{Z}\right)\right)\right)\right], \tag{9}$$

where θ_G is the network parameter of the generator.

C. Prototype-assisted subdomain alignment (PASA) strategy

Conventional GAN algorithms only strengthen the global similarity between the real and synthetic distributions via the adversarial mechanism without considering the semantic correlations between the class-conditional distributions. However, discriminative features of the same category are not always identical. In this case, even if the temporal structure of the synthetic data is close to that of the real data, its features representing the health state may not necessarily align with those of the real data. In other words, the synthetic data of different categories may be mixed up with each other and deviate from the real data of corresponding category. To overcome this weakness, we propose a PASA strategy to align the local distributions between the real and synthetic datasets.

As shown in Fig. 3, we exploit the prototype that characterizes the representative embedding of a set of semantically similar instances in the feature space to adjust the subdomain distributions of the synthetic data. More specifically, the PASA strategy first maximizes the correlation between real prototype and synthetic data in the same category. Concurrently, the PASA strategy keeps features away from the decision boundaries by minimizing the correlation between real prototypes and different categories of synthetic data. With the

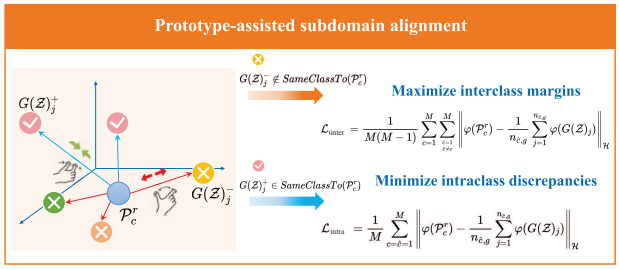


Fig. 3: Prototype-assisted subdomain alignment via bidirectional self-supervised loss.

PASA strategy, the SA-ASN algorithm explicitly incorporates semantic information regarding the health states into the generation process, and thus improves the distinguishability of the synthetic data. In this study, the inter-class and intra-class correlations are jointly optimized by a bidirectional self-supervised loss, which can be formulated as follows:

$$\begin{aligned}\mathcal{L}_{\mathrm{BS}} =& \frac{1}{M}\sum_{c=\hat{c}=1}^{M}\left\|\varphi\left(\mathcal{P}_c^r\right) - \frac{1}{n_{\hat{c},g}}\sum_{j=1}^{n_{\hat{c},g}}\varphi\left(G\left(\mathcal{Z}\right)_j\right)\right\|_{\mathcal{H}} \\ &- \gamma\frac{1}{M\left(M-1\right)}\sum_{c=1}^{M}\sum_{\substack{\hat{c}=1\\ \hat{c}\neq c}}^{M}\left\|\varphi\left(\mathcal{P}_c^r\right)\right. \\ &\left. - \frac{1}{n_{\hat{c},g}}\sum_{j=1}^{n_{\hat{c},g}}\varphi\left(G\left(\mathcal{Z}\right)_j\right)\right\|_{\mathcal{H}},\end{aligned} \tag{10}$$

where $\gamma > 0$ is a balancing parameter, $\mathcal{H}$ denotes the reproducing kernel Hilbert space (RKHS) and $\varphi(\cdot)$ is the nonlinear mapping from the feature space to the RKHS. M represents the number of sample categories. $G\left(\mathcal{Z}\right)_j$ denotes the j-th synthetic data generated by the SA-ASN algorithm. $n_{\hat{c},g}$ is the number of samples in $\hat{c}$ category. $\mathcal{P}_c^r$ denotes the prototype of the c-th category in the sample space, which can be calculated as follows:

$$\mathcal{P}_c^r = \frac{1}{n_{c,r}}\sum_{k=1}^{n_{c,r}} F_\theta\left(\mathcal{X}_k\right), \tag{11}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/TII.2023.3275701, IEEE Transactions on Industrial Informatics

where $\mathcal{P}_c^r$ denotes the mean vector of embedded support data belonging to the c-th class. $F_\theta(\cdot)$ is an embedding function via deep neural networks. $n_{c,r}$ denotes the number of the real data in the c-th class. $\mathcal{X}_k$ is the k-th real data.

D. Overall objective function of the SA-ASN algorithm

In this paper, we jointly optimize the adversarial loss and the bidirectional self-supervised loss to produce high-fidelity and diversified synthetic data. By incorporating (6), (9), and (10), the overall objective function of the SA-ASN algorithm is summarized as follows:

$$\mathcal{L} = \min_{\theta_G} \max_{\theta_D} \mathcal{L}_{\text{adv}} + \alpha \min_{\theta_G} \mathcal{L}_{\text{BS}}, \tag{12}$$

where α is a balancing parameter; θ_G and θ_D are the parameters of generator and discriminator, respectively. Moreover, Algorithm 1 displays the training procedure of our SA-ASN algorithm.

Algorithm 1 The SA-ASN Algorithm

Input: Real data $\{\mathcal{X}_k\}_{k=1}^{n_{c,r}}$; Latent variable $\{\mathcal{Z}_j\}_{j=1}^{n_{\hat{c},g}}$;
Output: Synthetic data $G(\mathcal{Z})$;

for number of training epoch **do**
  for number of discriminator iterations **do**
    Sample mini-batch of m real data from distribution $\mathcal{X} \sim P_S(\mathcal{X})$;
    Sample mini-batch of m latent variables from distribution $\mathcal{Z} \sim P_L(\mathcal{Z})$;
    Compute discriminator loss $\mathcal{L}_D$ via (6);
    Update discriminator via $\mathcal{L}_D$;
  end for
  for number of generator iterations **do**
    Sample mini-batch of m latent variables from distribution $\mathcal{Z} \sim P_L(\mathcal{Z})$;
    Compute generator loss $\mathcal{L}_G$ via (9);
    Compute bidirectional self-supervised loss $\mathcal{L}_{\text{BS}}$ via (10);
    Update generator by $\mathcal{L}_G$ and $\mathcal{L}_{\text{BS}}$;
  end for
end for

IV. Experiments and Analysis

A. Setups

1) Dataset description: The SA-ASN algorithm is evaluated on the pipeline dataset obtained from the simulation platform. Specifically, the pipeline length is 180m, the operating pressure is 0.5MPa, and the flow rate is $10\text{m}^3/\text{h}$. The voltage signals are collected at a sampling frequency of 1024Hz under four different health states, i.e., large leak (LL), medium leak (ML), small leak (SL), and normal condition (NC). Details of the pipeline data for each health state are shown in Fig. 4.

2) Baselines: To assess the proposed SA-ASN algorithm comprehensively, the synthetic data of SA-ASN are compared with those of seven other state-of-the-art baselines:

- *Generative adversarial nets [5]:* It proposes two objective functions for updating the generator, namely saturating and non-saturating objective functions.

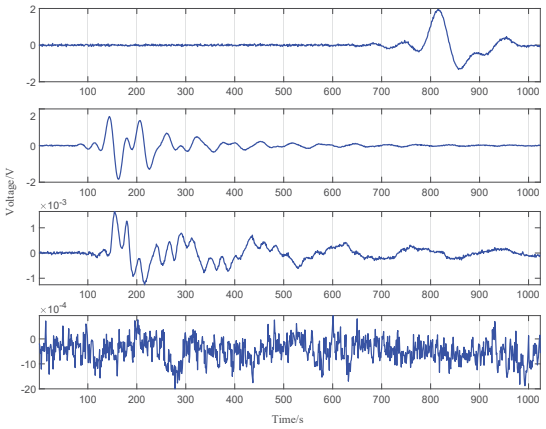


Fig. 4: Details of the pipeline data for each health state (from top to bottom in this figure, the health states are, in order, large leak, medium leak, small leak, and normal condition).

- *Least squares GAN (LSGAN) [15]:* It leverages least square loss to reduce the discrepancy between the synthetic and real dataset.
- *Conditional GAN (CGAN) [16]:* It employs additional information such as category vectors as constraints to guide data generation.
- *InfoGAN [3]:* It puts forward a generative model that could control synthetic data patterns by incorporating latent representations into the generator.
- *Auxiliary classifier GAN (ACGAN) [17]:* It develops a semi-supervised generative model to simultaneously generate synthetic data for multiple categories.
- *Synthetic minority over-sampling technique (SMOTE) [2]:* It creates synthetic data by randomly interpolating between minority class and its K-nearest neighbors.
- *Variational AutoEncoder (VAE) [9]:* It aims to construct a potential probability distribution for each sample by pushing the encoder input closer to the decoder output.

3) Model configuration: We employ the 1-D convolution architecture and the 1-D transposed convolution architecture to construct the discriminator and the generator, respectively. The configurations of the SA-ASN algorithm are presented in Table I. Particularly, all deep neural network-based algorithms share the same network parameters. The hyperparameters are fine-tuned and selected by cross-validation. During training, we utilize the Adam optimizer with a learning rate of 1e-4 to update the weights and biases of the SA-ASN algorithm. All the experiments are conducted with PyTorch using NVIDIA GEFORCE RTX 3090 GPU, Intel(R) Core(TM) i9-10900k, 3.70-GHz CPU.

B. Qualitative visualization

First, the synthetic data generated by the proposed SA-ASN algorithm is presented along with the real data, as shown in Fig. 5, in which the real data is shown in blue dotted line and the synthetic data is shown in orange solid line. Each column represents one health state. It can be observed that the pipeline data generated by the SA-ASN algorithm is extremely consistent with the characteristics of the real data.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/TII.2023.3275701, IEEE Transactions on Industrial Informatics

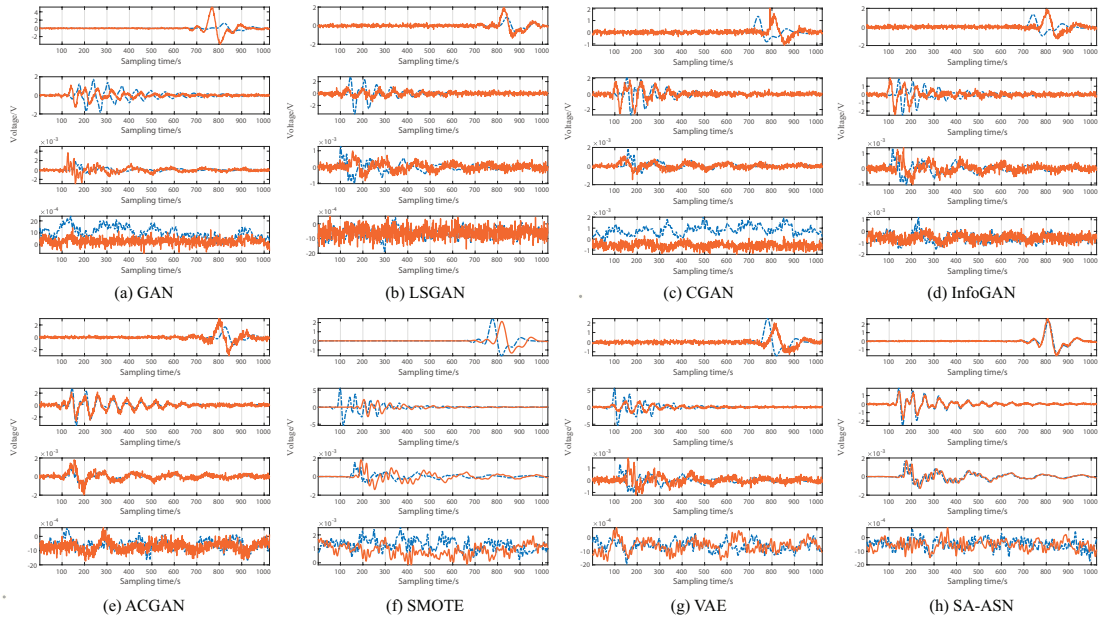


Fig. 6: Real data and synthetic data. (a) GAN. (b) LSGAN. (c) CGAN. (d) InfoGAN. (e) ACGAN. (f) SMOTE. (g) VAE. (h) SA-ASN (ours).

TABLE II: Quantitative evaluation of the proposed SA-ASN and all baselines by ED, CD, MMD, KL, EMD, Cosine, and PCC. (Best results are shown in **bold**.)

Model	Health state	ED ↓	CD ↓	MMD ↓	KL ↓	EMD ↓	Cosine ↑	PCC ↑
GAN	Large leak	5.8259	1.0000	2.45-03	0.1908	0.1148	0.9123	0.3792
	Medium leak	5.2397	0.7275	2.02E-03	0.0955	0.1962	0.9543	0.3453
	Small leak	3.3422	0.4950	2.59E-03	0.0315	8.06E-05	0.9781	0.6551
	Normal condition	7.4409	0.7314	2.49E-03	**0.1419**	1.90E-04	0.9044	0.0974
LSGAN	Large leak	5.9141	0.8307	2.74E-03	0.1541	0.0905	0.9218	0.4090
	Medium leak	3.8426	0.4368	2.11E-03	0.0249	0.2052	0.9858	0.6515
	Small leak	5.7607	0.7475	1.59E-03	0.0888	6.79E-05	0.9365	0.3013
	Normal condition	7.3484	0.7614	2.80E-03	0.1510	1.82E-04	0.9109	0.0500
CGAN	Large leak	5.6606	1.0000	2.21E-03	0.1781	0.0892	0.9190	0.3535
	Medium leak	2.7552	0.3782	1.59E-03	0.0147	0.1638	0.9905	0.7748
	Small leak	5.5668	0.7160	1.84E-03	0.0898	9.21E-05	0.9366	0.1346
	Normal condition	7.1029	0.7613	2.67E-03	0.1502	1.35E-04	**0.9132**	0.0672
InfoGAN	Large leak	2.3412	0.3690	2.22E-03	0.0180	0.1133	0.9872	0.1041
	Medium leak	5.5612	0.8718	2.18E-03	0.0945	0.1761	0.9485	0.6977
	Small leak	3.3184	0.5624	1.77E-03	0.0435	3.75E-05	0.9778	0.6502
	Normal condition	7.4060	0.7917	2.66E-03	0.1450	2.20E-04	0.9105	0.0409
ACGAN	Large leak	2.2632	0.4087	1.15E-03	0.0170	0.1078	0.9892	0.8464
	Medium leak	5.6679	0.8106	1.69E-03	0.0917	0.0855	0.9555	0.2266
	Small leak	7.2984	0.6938	4.88E-05	0.1125	1.17E-04	0.9384	0.1401
	Normal condition	7.3893	0.7551	6.03E-04	0.1517	4.50E-05	0.9090	0.0038
SMOTE	Large leak	2.4679	0.4376	1.13E-03	0.0384	0.0135	0.9819	0.7749
	Medium leak	3.0388	0.4999	**5.76E-05**	0.0252	**0.0228**	0.9870	0.5999
	Small leak	4.0798	0.5776	3.05E-04	0.0875	3.20E-05	0.9599	0.4534
	Normal condition	8.6473	0.7983	2.39E-03	0.1978	1.38E-03	0.8780	0.2027
VAE	Large leak	2.1520	0.3654	8.45E-04	0.0377	0.1015	0.9863	0.8160
	Medium leak	5.2304	0.7243	3.44E-04	0.0904	0.0936	0.9633	0.2419
	Small leak	4.2955	0.8951	5.61E-04	0.1136	2.94E-05	0.9537	0.3310
	Normal condition	8.1732	0.7548	4.17E-04	0.1872	1.48E-04	0.8873	0.0389
SA-ASN	Large leak	**0.5844**	**0.0865**	**1.43E-04**	**0.0022**	**0.0629**	**0.9992**	**0.9887**
	Medium leak	**1.3848**	**0.2091**	1.25E-04	**0.0039**	0.0532	**0.9983**	**0.9478**
	Small leak	**1.2103**	**0.2425**	**1.54E-05**	**0.0060**	**7.31E-06**	**0.9976**	**0.9577**
	Normal condition	**7.0452**	**0.6675**	**2.61E-05**	0.1460	**3.14E-05**	0.9105	**0.2846**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/TII.2023.3275701, IEEE Transactions on Industrial Informatics

TABLE III: Performance evaluation of pipeline fault diagnosis via FAR and MAR. (Accuracy %)

Models	FAR					MAR				
	LL	ML	SL	NC	Average	LL	ML	SL	NC	Average
GAN	0.00	1.83	31.42	4.31	9.39	68.97	11.09	0.00	31.98	29.01
LSGAN	11.95	9.20	0.00	0.83	5.49	17.41	0.00	47.77	0.00	16.29
CGAN	0.00	17.51	0.61	0.00	4.53	14.45	0.00	39.17	0.34	13.49
InfoGAN	0.11	0.06	6.74	0.00	1.73	12.44	5.30	0.32	2.92	5.24
ACGAN	2.08	12.31	2.07	1.06	4.38	27.72	3.18	0.00	20.36	12.82
SMOTE	0.00	5.65	42.62	15.18	15.86	92.14	93.13	0.17	3.61	47.26
VAE	2.45	4.15	30.61	2.82	10.00	88.32	19.96	0.00	10.79	29.76
SA-ASN	0.00	0.00	0.04	0.00	0.02	0.02	0.03	0.00	0.06	0.03

TABLE IV: Performance evaluation of pipeline fault diagnosis via FDR and ACC. (Accuracy %)

Models	FDR					ACC				
	LL	ML	SL	NC	Average	LL	ML	SL	NC	Average
GAN	31.03	88.91	100.00	68.02	71.99	82.50	95.88	76.50	88.88	85.94
LSGAN	82.59	100.00	52.23	100.00	83.71	86.67	93.08	87.96	99.38	91.77
CGAN	85.55	100.00	60.83	99.66	86.51	96.33	86.83	89.67	99.99	93.19
InfoGAN	87.56	94.70	99.68	97.08	94.76	96.92	98.67	94.92	99.25	97.44
ACGAN	72.27	96.81	100.00	79.63	87.18	91.25	89.95	98.41	94.04	93.41
SMOTE	7.85	6.86	99.82	96.38	52.73	76.04	72.58	67.45	87.75	75.96
VAE	11.67	80.03	100.00	89.21	70.23	75.79	91.91	77.08	95.20	84.99
SA-ASN	97.83	96.93	99.83	93.82	97.10	99.50	99.20	97.12	98.33	98.91

TABLE I: Configurations of the SA-ASN algorithm.

Network parameters	Values
Number of convolution blocks	3
Number of transposed convolution blocks	3
Number of convolution kernels	3
Number of transposed convolution kernels	3
Balancing parameter	0.01
Size of convolution kernels	1×2
Size of transposed convolution kernels	1×2
Stride	2
Length of real time series	1024
Training epochs	2000
Number of heads in the PASA	8
Hyperparameter α	0.01
Hyperparameter γ	1

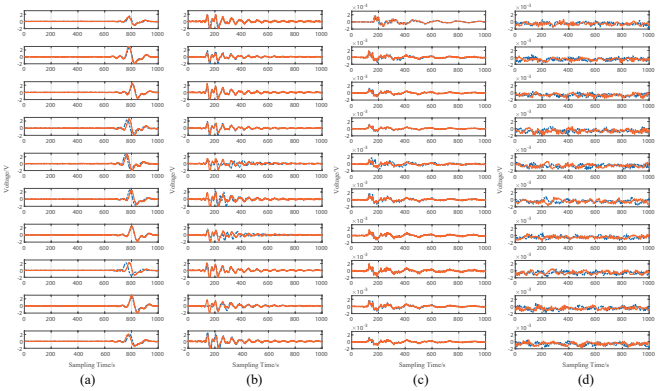


Fig. 5: Real data and synthetic data generated by the SA-ASN algorithm. (a) Pipeline signal changes under LL. (b) Pipeline signal changes under ML. (c) Pipeline signal changes under SL. (d) Pipeline signal changes under NC.

We are now in a position to comprehensively evaluate the effectiveness of our SA-ASN algorithm by comparing its performance with that of some state-of-the-art baselines. As shown in Fig. 6(a)-(e), the temporal properties associated with the health states in the synthetic data are significantly degraded by the off-the-shelf GAN framework. Subsequently, it can be seen that the synthetic data from the SMOTE and VAE cannot recover the subtle amplitude variations, which inevitably makes the follow-up fault diagnosis task more challenging. Differently, our SA-ASN algorithm well learns the temporal association and semantic representation from the sample space, which enables the synthetic data more natural and smooth in the whole structure. We believe that the above satisfactory visual results mainly benefit from the cooperation between the TAL mechanism and the PASA strategy. In particular, the former helps the generator to learn global context information from the real data, while the latter contributes to aligning discriminative features between the real and synthetic distributions.

C. Quantitative evaluation

We further assess our SA-ASN algorithm by seven quantitative metrics including Euclidean distance (ED), Chebyshev distance (CD), maximum mean discrepancy (MMD), Kullback-Leibler divergence (KL), Earth-Mover distance (EMD), cosine similarity (Cosine), and Pearson correlation coefficient (PCC). Note that the first five metrics measure the difference between the real and generated distributions, and thus their values should be small (denoted by $\downarrow$). On the contrary, Cosine and PCC should have large values, as

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/TII.2023.3275701, IEEE Transactions on Industrial Informatics

they reveal the correlation between the given data (denoted by ↑). Table II reports the statistical results under the four health states. It is obvious that our SA-ASN algorithm achieves superior performance surpassing or comparable to other baselines. For example, in the LL scenario, our algorithm outperforms the second-best algorithm in all metrics with significant superiority. These results suggest that our SA-ASN algorithm is more effective in generating time series due to the consideration of temporal association and semantic correlation.

D. Ablation study

To verify the contribution of each component in the SA-ASN algorithm, we carry out incremental experiments. As illustrated in Section III, the SA-ASN algorithm consists of two key components: the TAL mechanism and the PASA strategy. Therefore, we take the standard GAN as the baseline and add the TAL and the PASA respectively. Compared variants are listed as follows: standard GAN, standard GAN + TAL, standard GAN + PASA, and SA-ASN.

Fig. 7 shows the results of the distribution visualization via t-SNE. According to Fig. 7(a) and Fig. 7(b), it can be found that the TAL mechanism helps to recover the time-structured features of the real data in the generated data. However, the standard GAN+TAL has no ability to learn the discriminative features of each health state, resulting in invalid decision boundaries. Furthermore, Fig. 7(c) presents the favorable properties of intra-class compactness and inter-class separability. Nevertheless, due to the loss of temporal dynamic information, standard GAN + PASA has the problem of misclassification. On the contrary, the proposed SA-ASN algorithm obtains the best instance and distribution characteristics, including high-quality temporal association and distinct decision boundaries. In summary, each component is important and indispensable for achieving the best performance of the proposed SA-ASN algorithm.

E. Pipeline fault diagnosis

1) Implementation details: In this subsection, the effectiveness of our algorithm is investigated via pipeline fault diagnosis. Experimental cases with different augmentation (AM) rates are carried out to evaluate the performance of the SA-ASN algorithm and all baselines in these cases. The AM rate is defined as follows.

$$\mathrm{AM} = N_{\mathrm{normal}}/N_{\mathrm{fault}}, \tag{13}$$

where N_{normal} represents the sample number of normal condition and N_{fault} denotes the sample number of all fault conditions. In order to comprehensively analyze the diagnostic reliability, four metrics are used in the experiment, including false-alarm rate (FAR), missing-alarm rate (MAR), fault-detection rate (FDR), and total accuracy (ACC). More specifically, the lower the FAR and MAR, the more accurate the classification results (the opposite is true for FDR and ACC). Note that the N_{normal} in this experiment is set to 2000.

2) Experimental results: Following the above experimental setup, we conduct the fault diagnosis on the pipeline datasets with six AM rates, including AM rate=100:1, AM rate=20:1, AM rate=10:1, AM rate=5:1, AM rate=2:1, and AM rate=1:1. Fig. 8 shows the effect of different AM rates on the ACC. It is observed that our SA-ASN algorithm outperforms the other baselines with increasing AM rate. To be specific, the SA-ASN algorithm brings an improvement up to 1.47% in ACC compared to the second-best algorithm when AM rate=1:1. Moreover, the other three metrics in the case of AM rate = 1:1

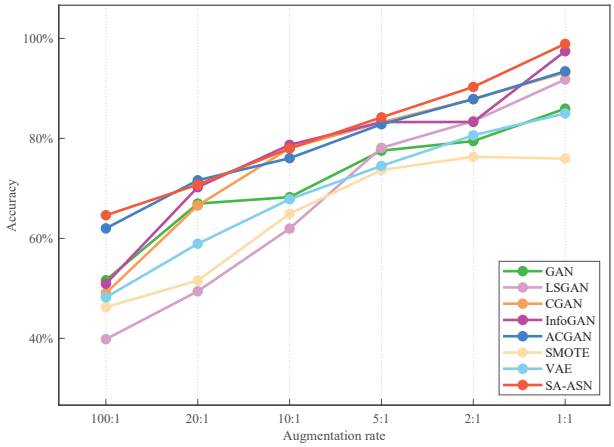


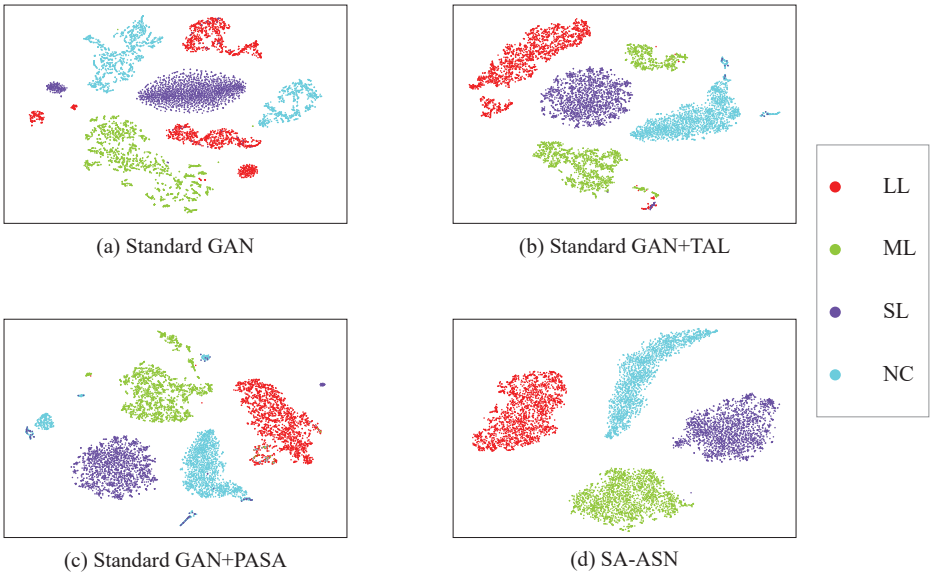
Fig. 8: Diagnostic accuracy at different AM rates.

are shown in Tables III-IV. Compared with the baselines, our SA-ASN algorithm decreases the averaged FAR and averaged MAR by 1.71% and 5.2%, respectively. Besides, the proposed SA-ASN algorithm significantly outperforms all baselines in FDR and ACC with improvements of 2.34% and 1.47% over the second-best algorithm. All the above results suggest that the SA-ASN algorithm effectively improves the reliability of temporal structure and the discriminability of features, thereby overcoming the I&ID issue.

V. Conclusion

In this paper, a novel subdomain-alignment adversarial self-attention network (SA-ASN) has been proposed to solve the I&ID problem. Different from the traditional GAN framework, our SA-ASN algorithm has explicitly considered temporal association and semantic correlation during data generation. More specifically, we have put forward a TAL mechanism to model the complex dynamics of real data in the discriminator and transfer the attention maps to the generator via a customized knowledge-sharing structure. Furthermore, we have developed a PASA strategy to align the local class-conditional distributions between the real and synthetic datasets. Experimental results have demonstrated that the performance improvement of the data-driven diagnostic model significantly benefits from the augmented dataset. Therefore, we can conclude that the proposed SA-ASN algorithm is able to learn the underlying distributions of real-world data, while effectively preserving the temporal features and semantic representations of the real data in the synthetic data. In the future, we will focus on corresponding research directions, including but not

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/TII.2023.3275701, IEEE Transactions on Industrial Informatics



(a) Standard GAN (b) Standard GAN+TAL

(c) Standard GAN+PASA (d) SA-ASN

Fig. 7: Distribution visualization of t-SNE embedding.

limited to designing advanced control strategies [12], [22], [30] and developing new DA techniques [19], [25].

REFERENCES

[1] Y. An, X. Wang, R. Chu, B. Yue, L. Wu, J. Cui and Z. Qu, Event classification for natural gas pipeline safety monitoring based on long short-term memory network and Adam algorithm, *Structural Health Monitoring*, vol. 19, no. 4, pp. 1151-1159, 2020.

[2] N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, SMOTE: Synthetic minority over-sampling technique, *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[3] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever and P. Abbeel, InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets, In: *Advances in Neural Information Processing Systems (NIPS)*, Barcelona, Spain, 2016.

[4] J. Dou, G. Wei, Y. Song, D. Zhou and M. Li, Switching triple-weight-SMOTE in empirical feature space for imbalanced and incomplete data, *IEEE Transactions on Automation Science and Engineering*, in press, DOI: 10.1109/TASE.2023.3240759.

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, Generative adversarial nets, In: *Advances in Neural Information Processing Systems (NIPS)*, Red Hook, NY, USA: Curran, 2014, pp. 2672–2680.

[6] X. Hu, H. Zhang, D. Ma and R. Wang, A tnGAN-based leak detection method for pipeline network considering incomplete sensor data, *IEEE Transactions on Instrumentation and Measurement*, vol. 70, art. no. 3510610, 2020.

[7] B. K. Iwana and S. Uchida, An empirical survey of data augmentation for time series classification with neural networks, *PLoS One*, vol. 16, art. no. e0254841, 2021.

[8] S. Kim, N. H. Kim and J. H. Choi, Prediction of remaining useful life by data augmentation technique based on dynamic time warping, *Mechanical Systems and Signal Processing*, vol. 136, art. no. 106486, 2020.

[9] D. P. Kingma and M. Welling, Auto-encoding variational bayes, *arXiv preprint arXiv:1312.6114*, 2013.

[10] X. Kong, X. Li, Q. Zhou, Z. Hu and C. Shi, Attention recurrent autoencoder hybrid model for early fault diagnosis of rotating machinery, *IEEE Transactions on Instrumentation and Measurement*, in press, DOI: 10.1109/TIM.2021.3051948.

[11] B. Li, B. Tang, L. Deng and J. Wei, Joint attention feature transfer network for gearbox fault diagnosis with imbalanced data, *Mechanical Systems and Signal Processing*, vol. 176, art. no. 109146, 2022.

[12] H. Li, P. Wu, N. Zeng, Y. Liu and F. E. Alsaadi, A survey on parameter identification, state estimation and data analytics for lateral flow immunoassay: from systems science perspective, *International Journal of Systems Science*, vol. 53, no. 16, pp. 3556-3576, 2022.

[13] X. Li, W. Zhang, Q. Ding and J. Q. Sun, Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation, *Journal of Intelligent Manufacturing*, vol. 31, pp. 433–452, 2020.

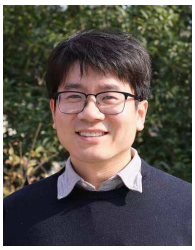
[14] H. Liu, H. Zhao, J. Wang, S. Yuan and W. Feng, LSTM-GAN-AE: A promising approach for fault diagnosis in machine health monitoring, *IEEE Transactions on Instrumentation and Measurement*, vol. 71, art. no. 3503113, 2021.
[15] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang and S. P. Smolley, Least squares generative adversarial networks, In: *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, 2017, pp. 2794–2802.
[16] M. Mirza and S. Osindero, Conditional generative adversarial nets, *arXiv preprint arXiv:1411.1784*, 2014.
[17] A. Odena, C. Olah and J. Shlens, Conditional image synthesis with auxiliary classifier GANs, In: *Proceedings of the International Conference on Machine Learning*, Sydney, Australia, 2017, pp. 2642–2651.
[18] X. Ren, W. Lin, X. Yang, X. Yu and H. Gao, Data augmentation in defect detection of sanitary ceramics in small and non-i.i.d datasets, *IEEE Transactions on Neural Networks and Learning Systems*, in press, DOI: 10.1109/TNNLS.2022.3152245.
[19] F. M. Shakiba, M. Shojaee, S. M. Azizi and M. Zhou, Real-time sensing and fault diagnosis for transmission lines, *International Journal of Network Dynamics and Intelligence*, vol. 1, no. 1, pp. 36-47, Dec. 2022.
[20] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, Attention is all your need, In: *Advances in Neural Information Processing Systems (NIPS)*, California, America, 2017.
[21] C. Wang, F. Han, Y. Zhang and J. Lu, An SAE-based resampling SVM ensemble learning paradigm for pipeline leakage detection, *Neurocomputing*, vol. 403, pp. 237–246, 2020.
[22] X. Wang, Y. Sun and D. Ding, Adaptive dynamic programming for networked control systems under communication constraints: a survey of trends and techniques, *International Journal of Network Dynamics and Intelligence*, vol. 1, no. 1, pp. 85-98, Dec. 2022.
[23] J. Wu, Z. Zhao, C. Sun, R. Yan and X. Chen, Fault-attention generative probabilistic adversarial autoencoder for machine anomaly detection, *IEEE Transactions on Industrial Informatics*, vol. 16, no. 12, pp. 7479–7488, 2020.
[24] P. Xu, R. Du and Z. Zhang, Predicting pipeline leakage in petrochemical system through GAN and LSTM, *Knowledge-Based Systems*, vol. 175, pp. 50–61, 2019.
[25] B. Yang, Z. Liu, G. Duan and J. Tan, Mask2Defect: A prior knowledge-based data augmentation method for metal surface defect inspection, *IEEE Transactions on Industrial Informatics*, vol. 18, no. 10, pp. 6743–6755, 2022.
[26] J. Yu and J. Liu, Multiple granularities generative adversarial network for recognition of wafer map defects, *IEEE Transactions on Industrial Informatics*, vol. 18, pp. 1674–1683, 2022.
[27] Y. Yuan, G. Ma, C. Cheng, B. Zhou, H. Zhao, H.-T. Zhang and H. Ding, A general end-to-end diagnosis framework for manufacturing systems, *National Science Review*, vol. 7, no. 2, pp. 418-429, 2020.
[28] H. Zhang, X. Hu, D. Ma, R. Wang and X. Xie, Insufficient data generative model for pipeline network leak detection using generative adversarial networks, *IEEE Transactions on Cybernetics*, vol. 52, no. 7, pp. 7107–7120, 2022.
[29] H. Zhang, D. Pan, J. Liu and Z. Jiang, A novel MAS-GAN-based data synthesis method for object surface defect detection, *Neurocomputing*, vol. 499, pp. 106–114, 2022.
[30] X. M. Zhang, Q.-L. Han and X. Ge, The construction of augmented Lyapunov-Krasovskii functionals and the estimation of their derivatives in stability analysis of time-delay systems: a survey, *International Journal of Systems Science*, vol. 53, no. 12, pp. 2480-2495, 2022.

Chuang Wang received the B.Sc. degree in Automation from Northeast Petroleum University, Daqing, China, in 2017. He is currently pursuing the Ph.D. degree in Petroleum and Natural Gas Engineering at the Northeast Petroleum University, Daqing, China. His research interests include evolutionary calculation and deep learning techniques.

Zidong Wang (Fellow, IEEE) received the B.Sc. degree in mathematics in 1986 from Suzhou University, Suzhou, China, and the M.Sc. degree in applied mathematics in 1990 and the Ph.D. degree in electrical engineering in 1994, both from Nanjing University of Science and Technology, Nanjing, China.

He is currently Professor of Dynamical Systems and Computing in the Department of Computer Science, Brunel University London, U.K. From 1990 to 2002, he held teaching and research appointments in universities in China, Germany and the U.K. Prof. Wang's research interests include dynamical systems, signal processing, bioinformatics, control theory and applications. He has published more than 700 papers in international journals. He is a holder of the Alexander von Humboldt Research Fellowship of Germany, the JSPS Research Fellowship of Japan, William Mong Visiting Research Fellowship of Hong Kong.

Prof. Wang serves (or has served) as the Editor-in-Chief for *International Journal of Systems Science*, the Editor-in-Chief for *Neurocomputing*, the Editor-in-Chief for *Systems Science & Control Engineering*, and an Associate Editor for 12 international journals including IEEE TRANSACTIONS ON AUTOMATIC CONTROL, IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY, IEEE TRANSACTIONS ON NEURAL NETWORKS, IEEE TRANSACTIONS ON SIGNAL PROCESSING, and IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS-PART C. He is a Member of the Academia Europaea, a Member of the European Academy of Sciences and Arts, an Academician of the International Academy for Systems and Cybernetic Sciences, a Fellow of the IEEE, a Fellow of the Royal Statistical Society and a member of program committee for many international conferences.



Lifeng Ma received the B.Sc. degree in Automation from Jiangsu University, Zhenjiang, China, in 2004 and the Ph.D. degree in Control Science and Engineering from Nanjing University of Science and Technology, Nanjing, China, in 2010. From August 2008 to February 2009, he was a Visiting Ph.D. Student in the Department of Information Systems and Computing, Brunel University London, U.K. From January 2010 to April 2010 and May 2011 to September 2011, he was a Research Associate in the Department of Mechanical Engineering, the University of Hong Kong. From March 2015 to Februry 2017, he was a Visiting Research Fellow at the King's College London, U.K.

He is currently a Professor in the School of Automation, Nanjing University of Science and Technology, Nanjing, China. His current research interests include control and signal processing, machine learning and deep learning. He has published more than 50 papers in refereed international journals. He serves as an editor for *Neurocomputing* and *International Journal of Systems Science*.

Hongli Dong (Senior Member, IEEE) received the Ph.D. degree in control science and engineering from the Harbin Institute of Technology, Harbin, China, in 2012.

From 2009 to 2010, she was a Research Assistant with the Department of Applied Mathematics, City University of Hong Kong, Hong Kong. From 2010 to 2011, she was a Research Assistant with the Department of Mechanical Engineering, The University of Hong Kong, Hong Kong. From 2011 to 2012, she was a Visiting Scholar with the Department of Information Systems and Computing, Brunel University London, London, U.K. From 2012 to 2014, she was an Alexander von Humboldt Research Fellow with the University of Duisburg-Essen, Duisburg, Germany. She is currently a Professor with the Artificial Intelligence Energy Research Institute, Northeast Petroleum University, Daqing, China. She is also the Director of the Heilongjiang Provincial Key Laboratory of Networking and Intelligent Control, Daqing. Her current research interests include robust control and networked control systems.

Dr. Dong is a very active reviewer for many international journals.

Weiguo Sheng received the M.Sc. degree in information technology from the University of Nottingham, U.K., in 2002 and the Ph.D. degree in computer science from Brunel University, U.K., in 2005. Then, he worked as a Researcher at the University of Kent, U.K. and Royal Holloway, University of London, U.K. He is currently a Professor at Hangzhou Normal University. His research interests include evolutionary computation, data mining/clustering, pattern recognition and machine learning.