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.2023.3337364, IEEE Transactions on Industrial Informatics

Support-Sample-Assisted Domain Generalization via Attacks and Defenses: Concepts, Algorithms and Applications to Pipeline Fault Diagnosis

Chuang Wang, Zidong Wang, *Fellow, IEEE*, Qinyuan Liu, Hongli Dong, *Senior Member, IEEE*, and Weiguo Sheng, *Member, IEEE*

***Abstract*—This paper is concerned with domain generalization (DG), a practical yet challenging scenario in transfer learning where the target data are not available in advance. The key insight of DG is focused on learning a robust model that can generalize to the unseen domain by leveraging knowledge from the source domain. To this end, we propose a novel algorithm known as Support-Sample-Assisted Adversarial Attacks (SSAA) for DG. In the SSAA algorithm, an attack-defense strategy is deployed to enhance the target model's generalizability and transferability. This strategy includes a non-targeted attack stage, during which attack samples are generated to form pseudo-target domains with near-realistic covariate shifts. Subsequently, in the model defense stage, a bi-classifier structure is used to distinguish support samples from the generated attack samples. These support samples form a new decision boundary encompassing all unseen samples, prompting an extension of the existing decision boundary to meet these samples. Experimental results on cross-domain fault diagnosis tasks suggest that SSAA outperforms current state-of-the-art DG methods, indicating a promising avenue for further DG development.**

***Index Terms*—Domain generalization, attack-defense strategy, support sample, transfer learning, domain adaptation**

I. Introduction

Transfer learning (TL) techniques have recently attracted considerable attention in various fields such as pipeline fault diagnosis [6], [25], battery status monitoring [13], [14], and machinery safety assessment [18], [30], due to their potential in tackling the domain shift issue. Domain adaptation (DA), a typical scenario in TL, has proven effective in transferring learned knowledge from a well-labeled source domain to an unlabeled target domain. Most existing DA methods are designed to extract discriminative domain-invariant features in high-level space using two main strategies [21]. One strategy is to reduce the distribution discrepancies between the source and target domains by matching statistical moments, and the other is to generate domain-confused features by leveraging an additional domain discriminator.

Despite the noteworthy achievements of DA, existing methods, including distribution matching and adversarial training, all assume the prior availability of a labeled source domain and an unlabeled target domain for model training. Unfortunately, such an assumption may be unrealistic in practical scenarios due to changing working conditions and the rarity of target data. Consider, for example, the fault diagnosis task in pipeline operation and maintenance. Variables such as geography, internal medium, and atmospheric environment result in varying pipeline data distributions. Moreover, collecting pipeline failure data requires the pipeline to operate continuously under failure or near-failure conditions, an approach unsuitable for ensuring energy security. Consequently, obtaining target samples before model deployment presents a considerable challenge for implementing DA, given the unpredictable nature of influencing factors. Existing DA methods fail to extract domain-invariant features when domain discrepancies are unknown. To minimize dependency on target data, domain generalization (DG) has surfaced as an alternative approach, allowing the training of robust models without any target information. Currently, DG is the most promising tool for managing new instances and unseen categories.

The fundamental concept of DG revolves around mining domain-invariant features that are sensitive to category differentiation but insensitive to domain shifts [2], [17], [30]. For instance, a well-designed DG network that balances multi-source domain invariance and specificity has been proposed in [30] to undertake real-time cross-domain classification tasks. Furthermore, a conditional contrastive DG method developed in [17] aims to maximize the intra-class similarity and inter-class separability of realistic pairs from multi-source domains. In [2], a deep hybrid DG approach has been formulated to learn domain-invariant and discriminative features from multi-source domains, facilitating knowledge generalization across different workloads and machines. From these reviews, it e-

This work was supported in part by the National Natural Science Foundation of China under Grants 61933007, U21A2019, and 62222312, the Hainan Province Science and Technology Special Fund of China under Grant ZDYF2022SHFZ105, the Fundamental Research Funds for the Central Universities, 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)

Qinyuan Liu is with the Department of Computer Science and Technology, Tongji University, Shanghai 201804, China, also with the Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai 200092, China, and also with the Shanghai Artificial Intelligence Laboratory, Shanghai, China. (E-mail: liuqy@tongji.edu.cn)

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

merges that multi-source DG (MultiDG) is the most frequently used setting, striving to enhance the model's generalizability to the target domain by reducing the model's sensitivity to distribution shifts among multi-source domains. However, the practical applicability of MultiDG is limited due to two key issues: 1) the inconsistent number of source domains across different tasks negatively impacts the model's generalizability, and 2) the process of collecting and labeling data from multiple source domains is costly.

A practical yet often overlooked setting is single-source DG (SingleDG), where only one source domain is available during training. However, such a single source domain hampers the model's ability to learn domain-invariant features, as it fails to provide any information about domain variation. Consequently, the model tends to overfit to domain-specific signals. Domain expansion (DE), the most common method for addressing SingleDG, seeks to broaden the decision boundary of a single source domain by generating pseudo-target domains. The primary objective of DE is to augment the diversity of domain distributions while firmly maintaining the original semantics. In pursuit of this goal, in this paper, an attack-defense strategy is introduced, and the efficacy of the attack-defense model (ADM) in managing SingleDG is demonstrated. We will revisit and contrast the concepts of SingleDG and the attack-defense strategy, culminating in two key conclusions on the functioning of the ADM.

Firstly, a classifier trained on the source domain tends to misclassify target samples, as these share the same semantic information as the source samples but exhibit different data distributions. This mirrors the adversarial attack field, where successful attacks occur when original samples are augmented with minor perturbations that alter data distribution, leading the model to confidently produce incorrect outputs. From this, the first conclusion is that both the attack strategy and the target domain violate the assumption of data being independent and identically distributed (i.i.d). Moreover, the goal of SingleDG is to eliminate domain discrepancies, enabling the application of the source classifier to the target domain with satisfactory results. In terms of the defense strategy, the aim is to achieve a robust model capable of providing the desired output, even when subjected to attacks. Thus, the second conclusion is that both the defense strategy and SingleDG share a common ultimate goal: enhancing the model's generalizability and transferability to handle i.i.d-violated samples. Therefore, we propose that all misclassified target samples can be viewed as infinitely occurring attack samples during the testing stage, also known as out-of-distribution (OOD) samples. Concurrently, SingleDG essentially combats a natural generator that incessantly attacks the target model by introducing domain perturbations.

Following the above discussions, it is argued that the attack-defense strategy can provide valuable guidance for implementing DE. Specifically, enhancing generalizability and transferability depends on the creation of pseudo-target domains through continuous attacks and the extension of the decision boundary via effective defenses. The complexities of this strategy lie in: 1) generating imperceptible and diverse attack samples to effectively explore unknown domains, and 2) extending decision boundaries using these attack samples. With these challenges in mind, this paper proposes a novel **S**upport-**S**ample-assisted **A**dversarial **A**ttacks (SSAA) algorithm for DE, aimed at addressing the SingleDG problem. Firstly, an adversarial framework composed of a generator and a discriminator is designed to generate pseudo-target samples (henceforth referred to as attack samples) in the direction of gradient ascent, with the aim of continually attacking the model and exploring the unseen target domain's decision boundary. Subsequently, a bi-classifier structure, encompassing an auxiliary classifier and a target classifier, is created to identify support samples capable of forming a new decision boundary that includes all unseen samples. Finally, the SSAA is designed to learn discriminative and transferable features from these support samples, thereby extending the classification boundary and offering a robust classifier for the target samples.

The core contributions of this paper are highlighted as follows.

1) An adversarial framework, comprised of a generator and a discriminator, is introduced to generate diverse and smooth attack instances, while strongly maintaining the original semantic information, thereby assisting in exploring the distribution boundary of the unseen target domain.
2) A well-crafted bi-classifier structure is employed to identify support samples, which are then utilized to form a new decision boundary, thereby enhancing the generalizability of the cross-domain model.
3) Extensive experiments conducted on cross-domain fault diagnosis tasks reveal that the proposed SSAA outperforms several existing state-of-the-art DG algorithms, and this substantiates the effectiveness and potential applicability of the SSAA algorithm.

The rest of this paper is structured as follows. Section II discusses related works on DA, DG, and adversarial attacks. Section III provides a detailed description of the novel SSAA algorithm. Experimental results and their respective analysis are presented in Section IV. Finally, Section V draws conclusions from the study.

II. Related Work

A. Domain Adaptation

Deep Neural Networks (DNNs), trained on large-scale labeled datasets, have made significant advancements in various practical applications such as industrial fault diagnosis [11], [29], image identification [10], and object detection [24], [28]. However, despite these successes, conventional DNNs trained on labeled datasets often struggle to generalize to unlabeled test data with differing distributions, an issue known as domain shift. To mitigate this limitation, DA has been developed, enabling the learning of an adaptive classifier for the target domain by transferring knowledge from the source domain [21]. Existing DA methods have been designed to extract discriminative domain-invariant features in high-level spaces through distribution matching or adversarial learning. Distribution matching specifically seeks to eliminate domain discrepancies by matching all statistical moments, including Maximum Mean Discrepancy (MMD), Deep Correlation Alignment (CORAL) [20], and Central Moment Discrepancy (CMD) [27]. Adversarial learning, on the other hand, draws inspiration from Generative Adversarial Networks (GAN) [4], employing an additional domain discriminator to create domain-confused features through a two-player min-max game.

B. Domain Generalization

While DA methods have recently shown promising results in tackling the domain shift problem, such success heavily relies on an idealistic assumption that target data can be accessed prior to deploying a deep model. In many fields, due to stringent data privacy regulations, target data is generally unavailable. To address this constraint, Domain Generalization (DG), which involves DA without using any target information, has been proposed for classification or regression tasks on unseen data. Initially, considerable efforts were focused on MultiDG, as seen in [2], [9], [23] and references therein. Prior works on MultiDG can be categorized into three groups: 1) *data augmentation*, which generates virtual data to assist in learning general representations; 2) *domain-invariant representation learning*, which employs statistical moment matching or adversarial learning to extract domain-invariant representations from multiple source domains; and 3) *learning strategy*, which uses various techniques such as self-supervised learning, meta-learning, and ensemble learning to enhance generalization ability. However, the collection and labeling of data from multiple domains is costly, rendering MultiDG less practical in real-world scenarios. In this paper, we examine a more challenging and realistic setting, known as SingleDG. SingleDG aims to enhance the generalizability and transferability of the target model by generating pseudo-target domains that differ from the source domain.

C. Adversarial Attacks

In networked systems, signals transmitted over the network are susceptible to attacks, given that sensors are interconnected via a shared network medium. Consequently, the security of networked control systems has garnered substantial research interest over past decades, significantly advancing the development of attack-defense strategies, as seen in [16], [26]. Likewise, the behavior of network attacks on DNNs has drawn considerable attention due to their potential severe impacts across various real-world applications. These include semantic image segmentation in computer vision, network intrusion detection in cyber security, and road sign recognition in the physical world.

Adversarial attack is a prevalent form of attack, where adversarial samples are designed to be misclassified with high confidence by introducing imperceptible perturbations to the original input. To bolster the model's robustness against such attacks, numerous researchers have recommended generating adversarial samples to directly perturb and disrupt the target model during training. Consequently, a defense strategy to resist these attacks, referred to as adversarial training, is developed. For instance, in [5], a Fast Gradient Sign Method (FGSM) has been proposed to generate perturbations along the gradient of the objective function. Subsequently, various variants of FGSM, including Projected Gradient Descent (PGD) [15], Iterative FGSM (I-FGSM) [7], and Momentum Iterative FGSM (MI-FGSM) [1], have been developed to further enhance the target model's robustness. Unlike these studies, this paper addresses samples with significant perturbations, i.e., Out-Of-Distribution (OOD) samples. This challenge is more akin to cross-domain knowledge transfer rather than pure adversarial attacks.

III. An SSAA Algorithm

A. Problem Definition

Without loss of generality, this paper centers its attention on cross-domain classification tasks on SingleDG. The goal of SingleDG is to learn valuable knowledge (domain-invariant feature representations) from a single source domain and apply it to a new unseen target domain.

Let $\{x_1, x_2, \ldots, x_n\}$ and $\{y_1, y_2, \ldots, y_n\}$ denote the samples and corresponding labels, where n is the number of samples/labels. We define that a domain $\mathbb{D} = \{X, Y\}$ consists of domain samples X and corresponding category labels Y. In the SSAA, we have a labeled source domain $\mathbb{D}_s = \{X_s, Y_s\}$ supported by n_s source samples $\{x_{s,1}, x_{s,2}, \ldots, x_{s,n_s}\}$ and n_s source labels $\{y_{s,1}, y_{s,2}, \ldots, y_{s,n_s}\}$. Similarly, the target domain that is not available during the training stage is denoted as $\mathbb{D}_t = \{X_t, Y_t\}$ with n_t target samples $\{x_{t,1}, x_{t,2}, \ldots, x_{t,n_t}\}$ and n_t target labels $\{y_{t,1}, y_{t,2}, \ldots, y_{t,n_t}\}$.

Due to the time-varying working conditions, the marginal distribution of the target domain is different from that of the source domain, i.e., $P(X_s) \neq P(X_t)$. In this paper, the source domain, pseudo-target domain, and target domain share the same label space, which comprises H discrete labels $\{1, 2, \ldots, H\}$. Therefore, we aim to train a robust model with the help of source and pseudo-target domains to predict the data labels in the target domain with minimum prediction error.

B. Key Ideas

The objective of this paper is to learn a model robust to the fluctuations of the input. The key idea is to exploit DE and adversarial training under the guidance of the attack-defense strategy. Specifically, the DE is abstracted as an attack process, aiming to generate attack samples with invariant semantic information utilizing the constraint of adversarial attack. Furthermore, adversarial training is formalized as a defense strategy that endeavors to accurately identify attack samples by extending the decision boundary of the target model.

The iterative procedure of the SSAA algorithm consists of three fundamental steps.

1) **Adversarial attack implementation.** In this paper, attack samples are defined as samples that are similar to the original samples but misclassified by the target model. Formally, for the original samples x_s that can be classified correctly, i.e., $C(F(x_s)) = y_s$ (C is classifier and F is feature extractor), adversarial samples $\hat{x}_s$ tend to be classified in the wrong category, i.e., $C(F(\hat{x}_s)) \neq$

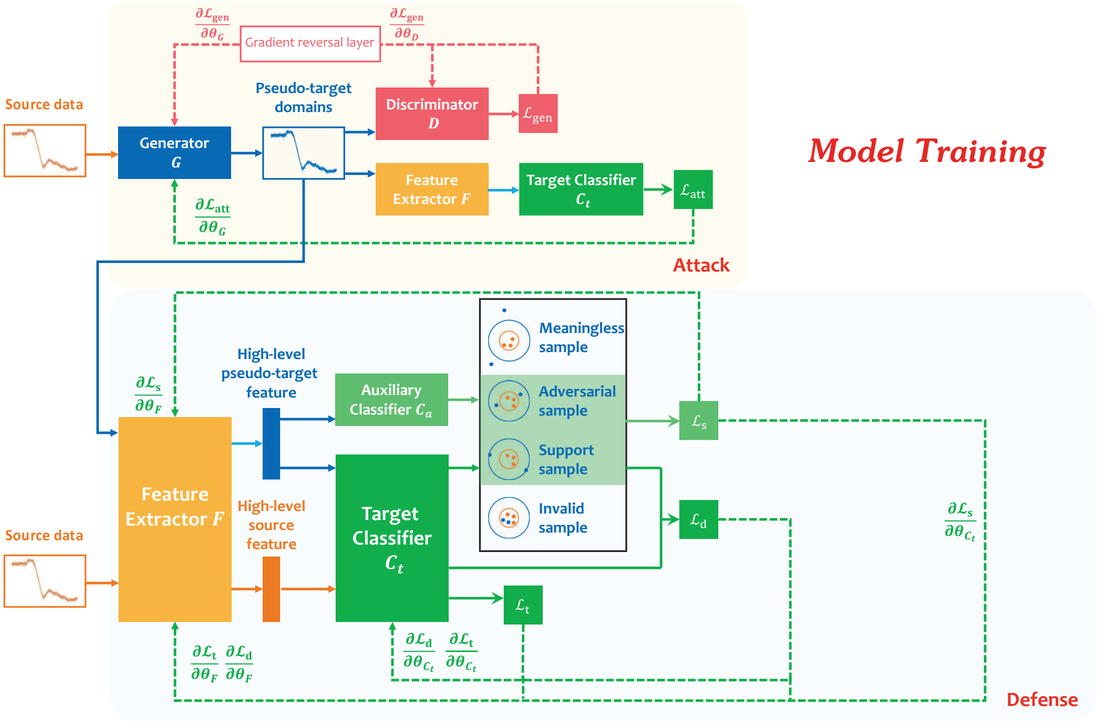

Model Training
$\frac{\partial \mathcal{L}_{gen}}{\partial \theta_G}$
Gradient reversal layer
$\frac{\partial \mathcal{L}_{gen}}{\partial \theta_D}$
Source data
Generator G
Pseudo-target domains
Discriminator D
$\mathcal{L}_{gen}$
Feature Extractor F
Target Classifier C_t
$\mathcal{L}_{att}$
$\frac{\partial \mathcal{L}_{att}}{\partial \theta_G}$
Attack
$\frac{\partial \mathcal{L}_s}{\partial \theta_F}$
High-level pseudo-target feature
Auxiliary Classifier C_a
Meaningless sample
Adversarial sample
Support sample
Invalid sample
$\mathcal{L}_s$
$\mathcal{L}_d$
$\frac{\partial \mathcal{L}_s}{\partial \theta_{C_t}}$
Feature Extractor F
High-level source feature
Target Classifier C_t
Source data
$\mathcal{L}_t$
$\frac{\partial \mathcal{L}_t}{\partial \theta_F}$ $\frac{\partial \mathcal{L}_d}{\partial \theta_F}$
$\frac{\partial \mathcal{L}_d}{\partial \theta_{C_t}}$ $\frac{\partial \mathcal{L}_t}{\partial \theta_{C_t}}$
Defense

Fig. 1: Overview of the proposed SSAA model. It can be observed that the training of the SSAA model consists of two stages: the attack stage and the defense stage. The purpose of the attack stage is to train the generator and discriminator to produce attack samples with imperceptible perturbations. The purpose of the defense stage is to train the feature extractor and classifier to accurately identify attack samples.

y_s. Therefore, attack samples are generated to disrupt the optimization direction of the target model while not altering the semantic representations. Note that the attacks used in this paper are non-targeted.

2) **Support sample exploration.** The key to defending against attacks is to improve the ability of the target model to generalize outside the source domain, which can be achieved by extending the decision boundary. However, this is a challenging task as the magnitude of covariate shifts is priori unknown. To tackle this problem, a bi-classifier structure is devised to find support samples that are natural and effective tools for locating the decision boundary.

3) **Decision boundary extension.** For the feature extractor, we encourage it to extend the decision boundary of the target model by accurately classifying the support samples. With repeated two-stage training, the target model can progressively extend the decision boundary until it contains all new unseen samples.

C. Overall Framework

Based on the above definitions, we formulate the SSAA as a generator G that produces attack samples to deceive the target model, a discriminator D that assists in the generation of attack samples, a feature extractor F that maps the input samples to a high-dimensional feature space, and dual classifiers C_t and C_a that output predicted source domain labels and find support samples. More specifically, the backbone networks of G and F are one-dimensional convolutional neural networks (1D-CNN). C_t and C_a consist of fully connected (FC) layers, activation functions ReLU, and LogSoftmax layers. D is achieved through the structure of FC→ReLU→FC→ReLU→FC→ReLU→FC→Sigmoid. The overview of the proposed SSAA model is presented in Fig. 1, and network structures are shown in Fig. 2.

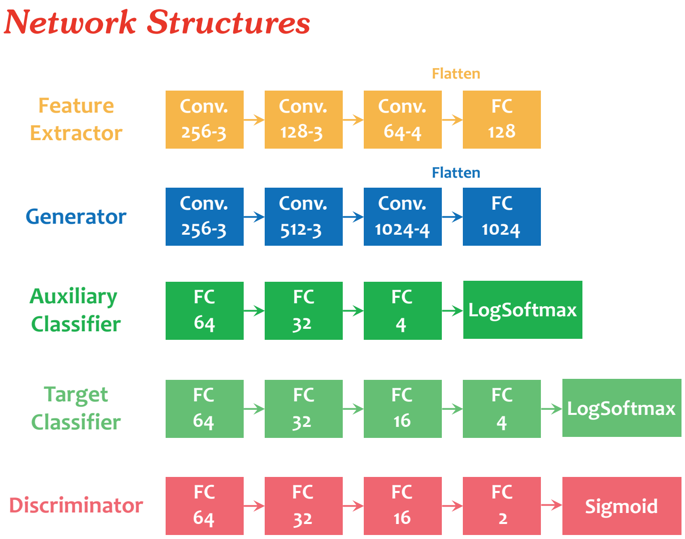


Fig. 2: Network structures of feature extractor, generator, auxiliary classifier, target classifier, and discriminator.

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.2023.3337364, IEEE Transactions on Industrial Informatics

D. Adversarial Attack Implementation

As discussed in Section III-B, the important principle for the construction of attack samples is to maintain their imperceptibility, which indicates that the attack samples should have the same semantic information as the original samples. In response to this rule, a discriminator is introduced into the SSAA with the aim of influencing the target model with satisfying imperceptibility. Specifically, qualified perturbations can make the discriminator fail to distinguish between attack samples and original samples. Therefore, the generator and discriminator can be updated by:

$$\begin{aligned}\min_{\theta_G}\max_{\theta_D}\mathcal{L}_{\text{gen}} =& \frac{1}{n_s}\sum_{i=1}^{n_s}\left[\log\left(D\left(x_{s,i}\right)\right)\right.\\ &\left.+\log\left(1-D\left(G\left(x_{s,i}\right)\right)\right)\right], \end{aligned} \tag{1}$$

where the attack sample is defined by $\hat{x}_{s,i} \triangleq G\left(x_{s,i}\right)$, θ_G and θ_D represent the parameters of G and D, respectively.

In the attack stage, we expect the attack samples to exert negative impacts on the target model by altering its original inputs, which can be achieved by optimizing the following objective:

$$\min_{\theta_G}\mathcal{L}_{\text{att}} = -\,\mathcal{L}_{\text{cla}}\left(f\left(\hat{x}_s\right), y_s\right), \tag{2}$$

where $f = F \circ C$ is the DNN and $\circ$ represents composite mapping. $\mathcal{L}_{\text{cla}}$ is the cross-entropy loss, which is formulated as follows:

$$\mathcal{L}_{\text{cla}} = -\frac{1}{n_s}\sum_{i=1}^{n_s}\sum_{h=1}^{H}\mathbf{1}_{[h=y_{s,i}]}\log\left(p\left(y_{s,i}=h \mid f\left(\hat{x}_{s,i}\right)\right)\right), \tag{3}$$

where n_s denotes the number of attack samples, H represents the number of categories, and $\mathbf{1}_{[\cdot]}$ is the indicator function. $\mathbf{1}_{[h=y_{s,i}]} = 1$ if $\hat{x}_{s,i}$ belongs to category h, otherwise $\mathbf{1}_{[h=y_{s,i}]} = 0$.

E. Defend Against Attacks

1) Support Sample Exploration: The key idea of defense strategy lies in extending the decision boundary to encompass all unseen samples, which generally requires the L_∞ norm of covariate shift to remain below a reasonable threshold ε as $\left\|\hat{x}_s - x_s\right\|_\infty \leq \varepsilon$. In fact, finding a suitable threshold ε is difficult or even impossible in real-world scenarios, which severely limits the practical application of SingleDG. To address this issue, the proposed SSAA relaxes the threshold restriction, which means that the magnitude of the covariate shift is no longer limited. Under such circumstances, four types of attack samples would naturally emerge from the generation process.

- The first type is *meaningless sample*, which has inconsistent semantic information with the source domain sample. The meaningless sample can achieve a high attack success rate, but violates the DE principle, which means that such samples may degenerate the performance of the target model.
- The second type is the *adversarial sample*, which essentially belongs to the same category as the neighboring original sample but is misclassified by the target model. The adversarial sample is highly smooth and diverse, and therefore effective in improving the generalizability and transferability of the target model.
- The third type is the *support sample*, which is defined as the feature vector closest to the decision boundary. The support sample is the most appropriate tool to delimit the valid boundary for pseudo-target domains.
- The fourth type is *invalid sample*, which is unable to influence the target model.

In this paper, we adopt dual classifiers C_t and C_a to identify the four types of samples described above. Formally, the characteristics of meaningless samples, adversarial samples, support samples, and invalid samples are summarized as follows.

- Meaningless sample: $C_t\left(F\left(\hat{x}_{s,h}\right)\right) \neq y_{s,h}$ and $C_a\left(F\left(\hat{x}_{s,h}\right)\right) \neq y_{s,h}$, where $\hat{x}_{s,h}$ is the attack sample associated with category h and $y_{s,h}$ is the corresponding real label.
- Adversarial sample: $C_t\left(F\left(\hat{x}_{s,h}\right)\right) = y_{s,h}$ and $C_a\left(F\left(\hat{x}_{s,h}\right)\right) \neq y_{s,h}$, or vice versa.
- Support sample is a special form of adversarial sample. Therefore, supposing that the adversarial sample characteristics are matched, we can identify the support sample by finding $\operatorname{argmax}\left(\mathcal{L}_{\text{cla}}\left(f\left(\tilde{x}_s\right), y_s\right)\right)$.
- Invalid sample: $C_t\left(F\left(\hat{x}_{s,h}\right)\right) = C_a\left(F\left(\hat{x}_{s,h}\right)\right) = y_{s,h}$.

Remark 1: In summary, OOD samples describe samples that are different from the source domain distribution, pseudo-target domain samples are generated OOD samples, and attack samples are pseudo-target domain samples that are generated by attack-defense strategy, including adversarial samples and support samples. Therefore, in this paper, the support samples are not generated by the pseudo-target domain, but by the adversarial attacks. Additionally, the support samples belong to the pseudo-target domain.

2) Decision Boundary Extension: First, we optimize θ_F and θ_{C_t} simultaneously by minimizing the $\mathcal{L}_{\text{t}}$ on the labeled source data, which can be formulated as:

$$\min_{\theta_F,\theta_{C_t}}\mathcal{L}_{\text{t}} = -\frac{1}{n_s}\sum_{i=1}^{n_s}\sum_{h=1}^{H}\mathbf{1}_{[h=y_{s,i}]}\log\left(p\left(y_{s,i}=h \mid f(x_{s,i})\right)\right), \tag{4}$$

where θ_F and θ_{C_t} are the parameters of F and C_t respectively.

Additionally, to defend against imperceptible perturbations, the target model is learned with a conditional maximum mean discrepancy (CMMD) loss [21] that penalizes the discrepancy between the adversarial sample and the source sample. The CMMD loss can be described as:

$$\begin{aligned}&\min_{\theta_F,\theta_{C_t}}\mathcal{L}_{\text{d}} = \frac{1}{H}\sum_{h=1}^{H}\\ &\left\|\frac{1}{n_{s,h}}\sum_{i=1}^{n_{s,h}}\varphi\left(f\left(x_{s,h,i}\right)\right) - \frac{1}{n'_{s,h}}\sum_{i=1}^{n'_{s,h}}\varphi\left(f\left(\tilde{x}_{s,h,i}\right)\right)\right\|_{\mathcal{H}}^2, \end{aligned} \tag{5}$$

where φ is a nonlinear mapping, $\mathcal{H}$ denotes the reproducing kernel Hilbert space (RKHS), $\tilde{x}_s$ denotes adversarial sample,

n'_s is the number of adversarial samples including support samples.

To efficiently and effectively resist attacks, a weighting mechanism is employed in the optimization process of the target model. Specifically, the support sample is given a larger weight to extend the decision boundary and the adversarial sample is assigned a smaller weight to defend against attacks. This process is formulated as:

$$\min_{\theta_F,\theta_{C_t}} \mathcal{L}_\mathrm{s} = -\frac{1}{n'_s}\sum_{i=1}^{n'_s}\sum_{h=1}^{H} w_i \mathbf{1}_{[h=y_{s,i}]}\log\left(p\left(y_{s,i}=h \mid f(\tilde{x}_{s,i})\right)\right), \tag{6}$$

where w_i is the sample weight calculated by:

$$w_i = \frac{e^{\mathcal{L}_\mathrm{cla}(f(\tilde{x}_s),y_s)_i}}{\sum_{i=1}^{n'_s} e^{\mathcal{L}_\mathrm{cla}(f(\tilde{x}_s),y_s)_i}}. \tag{7}$$

The training process of the proposed SSAA is shown in Algorithm 1.

Algorithm 1: The training process of the SSAA

```
Input: Labeled source domain $\{(x_{s,i}, y_{s,i})\}_{i=1}^{n_s}$.
Input: The number of total training epoch $N_0$; attack
       epoch $N_1$; defense epoch $N_2$
Output: Feature extractor weights $\theta_F$, target classifier
        weights $\theta_{C_t}$; balancing hyperparameters $\lambda_a$, $\lambda_d$,
        and $\lambda_s$.
Initialize $\theta_G$, $\theta_D$, $\theta_F$, $\theta_{C_a}$, and $\theta_{C_t}$.
for $i_0 \leq N_0$ do
    for $i_1 \leq N_1$ do
        Update the discriminator $D$ by ascending
        stochastic gradient, i.e., $\nabla_{\theta_D}\mathcal{L}_\mathrm{gen}$;
    end
    Update the generator $G$ by descending stochastic
    gradient, which is formulated by:
        $\nabla_{\theta_G}\frac{1}{n_s}\sum_{i=1}^{n_s}\log\left(1-D\left(G\left(x_{s,i}\right)\right)\right)+\lambda_a\mathcal{L}_\mathrm{att}$;
                                        ▷ Stage 1: Non-targeted attack
    for $i_2 \leq N_2$ do
        Identify adversarial samples;
        Update the feature extractor $F$ and dual
        classifiers $\theta_{C_t}, \theta_{C_a}$ by descending stochastic
        gradient, that is, $\nabla_{\theta_F,\theta_{C_t},\theta_{C_a}}\mathcal{L}_\mathrm{t}+\lambda_d\mathcal{L}_\mathrm{d}+\lambda_s\mathcal{L}_\mathrm{s}$.
    end
                                        ▷ Stage 2: Model defense
end
```

IV. Simulation Experiments

This paper utilizes pipeline fault diagnosis as the chosen testbed. Due to their economic, safety, and stability advantages, pipelines have become a prevalent feature in modern oil and gas transportation systems. However, during pipeline operation, incidents of leakage often occur, predominantly caused by pipeline deterioration or corrosion. These incidents pose significant threats to both property and life, rendering them among the most hazardous accidents. The urgency, therefore, lies in developing a model capable of achieving high accuracy within target distributions [6], [21]. In response to this pressing issue, this section will evaluate and compare the performance of the proposed SSAA algorithm with several existing state-of-the-art DG methods on pipeline fault diagnosis tasks.

A. Data Description

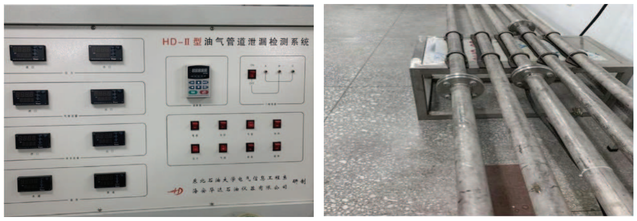
In this paper, two kinds of pipeline datasets collected by different platforms are employed for comparison experiments. The details are described as follows.

The first kind of dataset is the negative pressure wave dataset (NPWD) collected by the ZJ-CSGD type simulation platform, as shown in Fig. 3. The parameters of the pipeline are set as follows: the length is $180.2\mathrm{m}$, the sampling frequency is $1024\mathrm{Hz}$, and the flow rate is $10\mathrm{m}^3/\mathrm{h}$. In NPWD, the working pressure of the pipeline is set as different values to produce multiple source domains for performance evaluation. Specifically, three pressure conditions are included in the NPWD, namely high pressure (HP), medium pressure (MP), and low pressure (LP).

(a) (b)

Fig. 3: Console and platform of the negative pressure wave dataset.

The second kind of dataset is the acoustic wave dataset (AWD) collected by the HD-II type simulation platform, as shown in Fig. 4. The parameters of the pipeline are set as follows: the length is $160\mathrm{m}$, the sampling frequency is $5000\mathrm{Hz}$, and the flow rate is $60\mathrm{m}^3/\mathrm{h}$. It should be noted that acoustic signals in the AWD are collected under a fixed pressure (FP) condition $0.5\mathrm{MPa}$.



(a) (b)

Fig. 4: Console and platform of the acoustic wave dataset.

In the NPWD and the AWD, the pipeline data under each pressure condition can be divided into four categories, where failure signals including large leakage, medium leakage, and small leakage (LL, ML, and SL) are collected by switching valves and normal signals (NS) are collected in healthy condition. The data are downsampled from $5000\mathrm{Hz}$ to $1024\mathrm{Hz}$ because a high sampling rate would increase the complexity of data processing and reduce the speed of modeling operations. Furthermore, in order to mitigate the impact of noise on model training, an improved VMD method proposed in [22] is used in this paper to perform data denoising.

B. Implementation Details

In this paper, the training set, validation set, and test set are all zero-mean normalized by the mean and standard deviation of the training set. The split ratio of the training/validation/test set is 0.7:0.1:0.2.

For all experiments, the stochastic gradient descent (SGD) optimizer is employed to optimize the generator G and discriminator D with an initial learning rate of 10^{-4}. The Adam optimizer is used to optimize the feature extractor F, the target classifier C_t, and the auxiliary classifier C_a with an initial learning rate of 10^{-3}. The total number of the SSAA training epochs is set to 35, the number of attack epochs is set to 5, and the number of defense epochs is set to 10. The training process of the SSAA is early stopped within 3 epochs. The batch size is set to 32. Additionally, the impacts of balancing hyperparameters on model performance are investigated through sensitivity analysis.

For fairness, we re-implement six advanced comparison algorithms, including

- Empirical Risk Minimization (ERM) aims to optimize the model parameters by minimizing the expectation of the loss function values on the training dataset.
- Domain-Adversarial Neural Networks (DANN) [3] was proposed by Ganin et al. in 2016. DANN promotes the model to learn domain-invariant feature representations by introducing a domain-adversarial loss.
- DeepCORAL [20] was proposed by Sun et al. in 2016. The objective of DeepCORAL is to achieve domain adaptation by minimizing the difference in the covariance matrix between two domains.
- Group Distributionally Robust Optimization (GroupDRO) [19] was proposed by Sagawa et al. in 2020. The primary principle of the GroupDRO algorithm is to improve the generalization performance of the model by dividing the training data into multiple groups and optimizing the empirical worst-group risk.
- Meta-Learning for Domain Generalization (MLDG) [8] was proposed by Li et al. in 2018. The key idea of MLDG is to improve the generalization performance of the model through multilevel representation learning.
- Domain-Invariant Feature Exploration (DIFEX) [12] was proposed by Lu et al. in 2022. DIFEX improves the generalization performance of the model by learning both internally-invariant and mutually-invariant features.

For comparison methods, all hyperparameters are selected through the grid search method using the validation set. All methods, including the SSAA and baselines, are repeated five times, implemented by PyTorch, and trained on a single NVIDIA GEFORCE RTX 3090 GPU with 24GB memory.

C. Evaluation Metric

Accuracy is employed as an evaluation metric in this paper, which can be formulated as follows:

$$\text{Accuracy} = \frac{n_{\text{tp}} + n_{\text{tn}}}{n_{\text{tp}} + n_{\text{fn}} + n_{\text{fp}} + n_{\text{tn}}}, \tag{8}$$

where n_{tp} is the number of correctly identified failure signals, n_{fp} is the number of incorrectly identified failure signals, n_{tn} is the number of correctly monitored normal signals, and n_{fn} is the number of incorrectly monitored normal signals.

D. Main Experimental Results

In this subsection, we extensively assess the SSAA algorithm in both MultiDG and SingleDG scenarios.

1) MultiDG: The cross-domain classification results on N-PWD and AWD are shown in Table I. The yellow background marks the classic ERM model. The red marks the DA methods, which perform MultiDG through domain-invariant representation learning. The purple marks the DG methods. It can be found that the proposed SSAA is superior to other comparison methods by large margins. Specifically, our SSAA achieves the best classification performance in four out of five tasks. Furthermore, the SSAA method improves average accuracy by 6.32% and 4.17% compared to the best algorithm CORAL of DA and the best algorithm DIFEX of DG, respectively. Based on the above results, we can draw three important conclusions:

- As a baseline without any generalization operations, ERM can achieve results comparable to some of the most advanced DA or DG methods in specific tasks, which means that achieving stable performance in different DG situations is difficult. Despite this, the SSAA still achieves inspiring performance on most MultiDG tasks.
- Simple distribution alignment methods, including DANN and DeepCORAL, may be effective in DA with available source and target domains, but they fail to achieve the expected results in DG tasks due to the difficulty in reducing distribution differences between multiple source domains.
- Compared with methods such as GroupDRO, MLDG, and DIFEX, which achieve DG by learning the distribution differences among multiple domains, the proposed SSAA algorithm depicts the shape of the decision boundary by correctly identifying the support samples, and therefore effectively improves the generalization ability.

2) SingleDG: In this part, we further perform SingleDG to verify the superiority of our method. The experimental results on NPWD and AWD are shown in Table II and Table III. Table II presents the transfer tasks between three different domains within NPWD. Table III records the transfer tasks between different domains in NPWD and AWD.

In Table II, the proposed SSAA achieves the highest accuracy in five tasks: HP→MP, HP→LP, MP→HP, LP→HP, and LP→MP. For the task MP→LP, the SSAA performs slightly worse than MLDG. A possible reason for this phenomenon is that the limited number of samples in a single domain increases the difficulty of training a stable model that can generalize well to the target domain. Despite this, the proposed

TABLE I: MultiDG results on NPWD and AWD. The best results are highlighted in **bold**.(%)

Source	Target	ERM	DANN	CORAL	GroupDRO	MLDG	DIFEX	SSAA
HP, MP, AW	LP	74.38	76.98	77.32	75.44	73.70	75.96	**83.28**
HP, MP, LP	AW	73.75	73.75	75.08	75.50	73.32	76.18	**79.04**
HP, LP, AW	MP	70.80	74.42	72.68	**80.82**	72.06	77.88	78.54
MP, LP, AW	HP	73.90	73.38	75.54	75.08	70.30	75.58	**78.78**
Ave.		73.21	74.63	75.16	76.71	72.35	76.40	**79.91**

SSAA still achieves the highest average accuracy in five of the six tasks.

In Table III, it can be found that the SSAA method achieves superior or comparable performance to other state-of-the-art methods with an average accuracy of 53.57%. Compared to the second-best algorithm GroupDRO, the SSAA achieves an additional gain of 18.25% in average accuracy due to its full learning from pseudo-target domain samples that closely resemble real-world samples. The results show that our SSAA is beneficial in improving generalization capability compared to some advanced algorithms, achieving the highest performance in five out of six tasks.

TABLE II: SingleDG results on NPWD.(%)

Source	Target	GroupDRO	MLDG	DIFEX	SSAA
HP	MP	75.10	74.86	75.70	**80.74**
HP	LP	71.42	67.48	75.90	**80.94**
MP	HP	66.98	71.17	72.30	**79.02**
MP	LP	72.12	**72.66**	72.24	70.18
LP	HP	74.98	68.34	75.04	**78.06**
LP	MP	72.28	66.48	70.90	**81.70**
Ave.		72.15	70.17	73.68	**78.44**

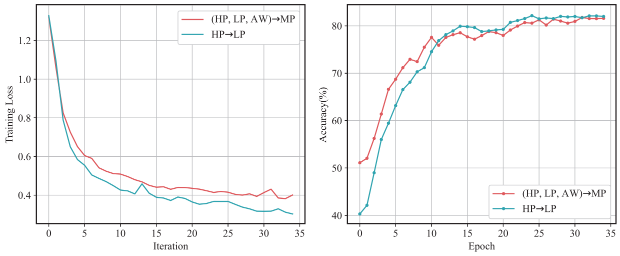
TABLE III: SingleDG results on NPWD and AWD.(%)

Source	Target	GroupDRO	MLDG	DIFEX	SSAA
HP	AW	44.64	43.90	39.18	**54.34**
MP	AW	**54.78**	45.12	49.95	53.48
LP	AW	45.84	43.64	53.52	**55.34**
AW	HP	44.08	45.06	24.48	**51.84**
AW	MP	45.20	39.06	45.36	**57.32**
AW	LP	37.26	37.24	36.70	**49.12**
Ave.		45.30	42.34	41.53	**53.57**

E. Performance Analysis

1) Parameter Sensitivity: The sensitivities of all hyperparameters in the proposed method, including λ_a, λ_d, and λ_s, are investigated in this section. As shown in Fig. 5(a), a large λ_a will result in a negative transfer, as overpowering attacks generally lead to the gradient exploding issue, and therefore an appropriate λ_a is important to ensure the effectiveness of the SSAA. According to the Fig. 5(b), we can find that a relatively small λ_d tends to result in a good performance, while $\lambda_d > 0.01$ causes a significant drop in accuracy. A possible reason for this phenomenon is that the inherent differences between the source domain and target domain may arise from the discriminability characteristics in each domain, and an overemphasis on minimizing distribution discrepancies may lead to negative transfer. Furthermore, Fig. 5(c) shows that the proposed SSAA is robust to the variations of λ_s when $\lambda_s < 0.01$, which indicates the dominant role of the weighting mechanism for learning discriminative features from pseudo-target samples. Generally, our SSAA method is insensitive to hyper-parameters in the appropriate range.

2) Training Stability: To verify the stability of our method, we report the loss and validation accuracy during the training process on the MultiDG task (HP, LP, AW)$\rightarrow$MP and the SingleDG task HP$\rightarrow$LP, as shown in Fig. 6. Fig. 6(a) shows that the training loss smoothly and rapidly reduces to a small value in both tasks. Additionally, the validation accuracies of MultiDG and SingleDG can quickly converge to a satisfactory value in Fig. 6(b). Thus, we can find that our SSAA is able to achieve stable performance in generalizing valuable knowledge from the source domain to the unseen target domain.

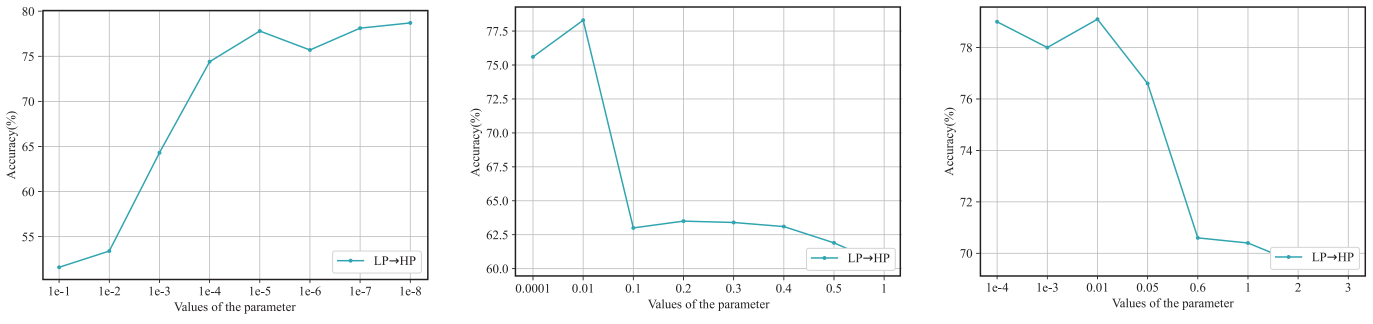


(a) Training Loss (b) Validation Accuracy

Fig. 6: The training loss and validation accuracy of the SSAA on NPWD and AWD. We take (HP, LP, AW)$\rightarrow$MP and HP$\rightarrow$LP as examples.

3) Significance Test: In this paper, a statistical significance test is performed on SingleDG results to quantitatively verify the significant improvement brought by the proposed SSAA in contrast to the strong baselines. Following the common practice, the significance threshold is set to 0.05, which can lead to two results including p-value$\leq$0.05 (reject the null hypothesis) and p-value>0.05 (accept the null hypothesis). The former implies that the improvement of the SSAA algorithm is significant, while the latter means that the performance of the SSAA algorithm is similar to that of the comparison methods.

In this experiment, we refer to the transfer tasks in Table II as Case 1 and the transfer tasks in Table III as Case 2. The statistical results in Table IV show that the SSAA has a significant effect on improving performance compared to state-of-the-art techniques.



(a) λ_a

(b) λ_d

(c) λ_s

Fig. 5: Sensitivity analysis of various hyperparameters, including λ_a for adversarial attacks, λ_d for reducing distributional discrepancies, and λ_s for defending against attacks.

TABLE IV: Results of paired t-test.

Method	Case1			Case2		
	t-value	p-value	Result	t-value	p-value	Result
GroupDRO	3.00	0.03	+	4.08	9.50E-03	+
MLDG	3.21	0.02	+	6.94	9.56E-04	+
DIFEX	2.75	0.04	+	2.96	3.15E-02	+

[1] "+" represents the SSAA algorithm significantly outperforms the compared algorithm.

V. Conclusion

In this paper, a novel and effective SSAA approach has been proposed to tackle the challenging yet common DG problem where only a single source domain is available. Initially, pseudo-target domains have been created to disrupt the target classifier by introducing diverse and smooth perturbations to the source distribution, a process known as the non-targeted attack stage. Subsequently, an auxiliary classifier and a target classifier have been employed to select adversarial and support samples from the generated pseudo-target samples, which aids in minimizing unnecessary interference and avoiding the wastage of computational resources. Ultimately, the generalization capability of the SSAA has been enhanced by extending the initial decision boundary to approach support samples, which can form a new decision boundary that encompasses all unseen samples. Leveraging the attack-defense strategy, the target classifier is expected to learn discriminative and transferable feature representations from these pseudo-target samples, thereby achieving desired DG. Extensive experiments conducted on two pipeline datasets have demonstrated that the proposed SSAA provides comparable or superior cross-domain classification performance as compared to existing advanced techniques in both MultiDG and SingleDG scenarios. In the future, advanced attack-defense strategies and corresponding theoretical proofs will be introduced to achieve more precise model generalization.

References

[1] Y. Dong, F. Liao, T. Pang, H. Su, J. Zhu, X. Hu, and J. Li, Boosting adversarial attacks with momentum, In: *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 9185–9193.

[2] Z. Fan, Q. Xu, C. Jiang, and S. X. Ding, Deep mixed domain generalization network for intelligent fault diagnosis under unseen conditions, *IEEE Transactions on Industrial Electronics*, vol. 71, no. 1, pp. 965–974, Jan. 2024.

[3] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, Domain-adversarial training of neural networks, *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2096–2030, 2016.

[4] 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.

[5] I. Goodfellow, J. Shlens, and C. Szegedy, Explaining and harnessing adversarial examples, *arXiv preprint arXiv:1412.6572*, 2014.

[6] D. Ji, C. Wang, J. Li, and H. Dong, A review: Data driven-based fault diagnosis and RUL prediction of petroleum machinery and equipment, *Systems Science & Control Engineering*, vol. 9, no. 1, pp. 724–747, 2021.

[7] A. Kurakin, I. Goodfellow, and S. Bengio, Adversarial examples in the physical world, In: *Proceedings of 5th International Conference on Learning Representations (ICLR)*, Toulon, France, 2017.

[8] D. Li, Y. Yang, Y.-Z. Song, and T. Hospedales, Learning to generalize: Meta-learning for domain generalization, In: *Proceedings of the Thirty-Second Conference on Artificial Intelligence (AAAI)*, New Orleans, Louisiana, USA, vol. 32, no. 1, 2018, pp. 3490–3497.

[9] J. Li, Z. Du, L. Zhu, Z. Ding, K. Lu, and H. T. Shen, Divergence-agnostic unsupervised domain adaptation by adversarial attacks, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 8196–8211, Nov. 2022.

[10] X. Li, M. Li, P. Yan, G. Li, Y. Jiang, H. Luo, and S. Yin, Deep learning attention mechanism in medical image analysis: Basics and beyonds, *International Journal of Network Dynamics and Intelligence*, vol. 2, no. 1, pp. 93–116, Mar. 2023.

[11] S. Lu, Z. Gao, Q. Xu, C. Jiang, A. Zhang, and X. Wang, Class-imbalance privacy-preserving federated learning for decentralized fault diagnosis with biometric authentication, *IEEE Transactions on Industrial Informatics*, vol. 18, no. 12, pp. 9101–9111, Dec. 2022.

[12] W. Lu, J. Wang, H. Li, Y. Chen, and X. Xie, Domain-invariant feature exploration for domain generalization, *arXiv preprint arXiv: 2207.12020*, 2022.

[13] G. Ma, S. Xu, B. Jiang, C. Cheng, X. Yang, Y. Shen, T. Yang, Y. Huang, H. Ding, and Y. Yuan, Real-time personalized health status prediction of lithium-ion batteries using deep transfer learning, *Energy & Environmental Science*, vol. 15, no. 10, pp. 4083–4094, 2022.

[14] G. Ma, S. Xu, T. Yang, Z. Du, L. Zhu, H. Ding, and Y. Yuan, A transfer learning-based method for personalized state of health estimation of lithium-ion batteries, *IEEE Transactions on Neural Networks and Learning Systems*, in press, DOI: 10.1109/TNNLS.2022.3176925.

[15] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, Towards deep learning models resistant to adversarial attacks, In: *Proceedings of 6th International Conference on Learning Representations (ICLR)*, Vancouver, BC, Canada, 2018.

[16] Z. H. Pang, L. Z. Fan, H. Guo, Y. Shi, R. Chai, J. Sun, and G. Liu, Security of networked control systems subject to deception attacks: A survey, *International Journal of Systems Science*, vol. 53, no. 16, pp. 3577–3598, 2022.

[17] M. Ragab, Z. Chen, W. Zhang, E. Eldele, M. Wu, C-K. Kwoh, and X. Li, Conditional contrastive domain generalization for fault diagnosis, *IEEE Transactions on Instrumentation and Measurement*, vol. 71, art. no. 3506912, 2022.

[18] R. Rahimilarki, Z. Gao, N. Jin, and A. Zhang, Convolutional neural network fault classification based on time-series analysis for benchmark wind turbine machine, *Renewable Energy*, vol. 185, pp. 916–931, Feb. 2022.

[19] S. Sagawa, P. W. Koh, T. B. Hashimoto, and P. Liang, Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization, In: *Proceedings of the International Conference on Learning Representations (ICLR)*, Millennium Hall, Addis Ababa, Ethiopia, 2020.

[20] B. Sun and K. Saenko, Deep coral: Correlation alignment for deep domain adaptation, In: *Proceedings of European Conference on Computer Vision*, Berlin, Germany: Springer, 2016, pp. 443–450.

[21] C. Wang, Z. Wang, and H. Dong, A novel prototype-assisted contrastive adversarial network for weak-shot learning with applications: Handling weakly labeled data, *IEEE/ASME Transactions on Mechatronics*, in press, DOI: 10.1109/TMECH.2023.3287070.

[22] C. Wang, Z. Wang, Q.-L. Han, F. Han, and H. Dong, Novel leader-follower-based particle swarm optimizer inspired by multiagent systems: Algorithm, experiments, and applications, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 3, pp. 1322–1334, Mar. 2023.

[23] J. Wang, C. Lan, C. Liu, Y. Ouyang, T. Qin, W. Lu, Y. Chen, W. Zeng, and P. S. Yu, Generalizing to unseen domains: A survey on domain generalization, *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 8, pp. 8052–8072, Aug. 2023.

[24] J. Wang, Y. Zhuang, and Y. Liu, FSS-Net: A fast search structure for 3D point clouds in deep learning, *International Journal of Network Dynamics and Intelligence*, vol. 2, no. 2, art. no. 100005, Jun. 2023.

[25] D. Yang, J. Lu, H. Dong, and Z. Hu, Pipeline signal feature extraction method based on multi-feature entropy fusion and local linear embedding, *Systems Science & Control Engineering*, vol. 10, no. 1, pp. 407–416, 2022.

[26] X. Yi, H. Yu, Z. Fang, and L. Ma, Probability-guaranteed state estimation for nonlinear delayed systems under mixed attacks, *International Journal of Systems Science*, vol. 54, no. 9, pp. 2059–2071, 2023.

[27] W. Zellinger, T. Grubinger, E. Lughofer, T. Natschläger, and S. S. Platz, Central moment discrepancy (cmd) for domain-invariant representation learning, *arXiv preprint arXiv:1702.08811*, 2019.

[28] N. Zeng, P. Wu, Z. Wang, H. Li, W. Liu, and X. Liu, A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection, *IEEE Transactions on Instrumentation and Measurement*, vol. 71, art. no. 3507014, 2022.

[29] J. Zhang and X. He, A partial-label U-net learning method for compound-fault diagnosis with fault-sample class imbalance, *IEEE Transactions on Industrial Informatics*, in press, DOI: 10.1109/TII.2023.3281660.

[30] C. Zhao and W. Shen, A domain generalization network combing invariance and specificity towards real-time intelligent fault diagnosis, *Mechanical Systems and Signal Processing*, vol. 173, art. no. 108990, 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. From August 2022 to August 2023, he was a Visiting Ph.D. Student with the Department of Computer Science, Brunel University London, Uxbridge, U.K. 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.

Qinyuan Liu received the B.Eng. degree in measurement and control technology and instrumentation from Huazhong University of Science and Technology, Wuhan, China, in 2012, and the Ph.D. degree in control science and engineering from Tsinghua University, Beijing, China, in 2017. He is currently a Professor in the Department of Computer Science and Technology, Tongji University, Shanghai, China. From Jul. 2015 to Sep. 2016, he was a Researcher Assistant in the Department of Electronic & Computer Engineering, Hong Kong University of Science and Technology, Hong Kong. From Jan. 2016 to Jan. 2017, he was an international researcher in the Department of Computer Science, Brunel University London, UK. His research interests include networked control systems, multi-agent systems, and distributed filtering. He is an active reviewer for many international journals.

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.