Towards Open World Anomaly Detection in Tabular Data: A Method for Adapting to Normality Shift

Shu Li[1], Yi Lu[1], and Jiong Yu[1*]

[1]School of Computer Science and Technology, Xinjiang University, Urumqi, China
{lsmiao, outman}@stu.xju.edu.cn, {yujiong}@xju.edu.cn

***Abstract*—Current tabular anomaly detection methods can effectively identify unknown and diverse anomalies by learning general patterns of normal data and detecting deviations from these patterns. However, their strong performance relies on a closed-world assumption that the distribution of normal data remains stable over time. In open-world scenarios, external factors such as environmental changes often cause shift in the normal data distribution, known as normality shift, which can lead to significant performance degradation. Thus, we propose a tabular Anomaly Detection method for Adapting to Normality Shift (ADANS). During training, ADANS learns perturbation-invariant representations of normal data to improve robustness and adaptability to normality shift. During testing, it employs an adaptive representation revision process to align the representations of normal data between the training and testing sets. This process is guided by a dynamic weight adjustment mechanism, which prioritizes likely normal samples while suppressing the influence of anomalies. Extensive experiments on four real-world benchmarks demonstrate that the proposed ADANS outperforms ten state-of-the-art methods in adapting to normality shift.**

***Index Terms*—tabular anomaly detection, normality shift, perturbation-invariant, representation revision, dynamic weight**

I. Introduction

Tabular data is a form of structured data arranged in rows and columns, where each row represents an individual observation and each column corresponds to a specific attribute of that observation. Owing to its ease of storage and high manageability, tabular data has been widely used for organizing and presenting information across various high-stakes domains. Consequently, reliable anomaly detection in tabular data has become essential for effectively identifying and mitigating potential risks. Given the inherent rarity and diversity of anomalies, as well as the high cost of labeling them, mainstream tabular anomaly detection methods, commonly referred to as zero-positive anomaly detection [1], prefer to learn the general pattern of normal data during training and identify deviations from this pattern as anomalies during testing [2], demonstrating superior capability in distinguishing between normal and abnormal instances across various domains.

Despite the progress, the effectiveness of these methods largely relies on a closed-world assumption, where normal data in both the training and test sets are assumed to be independent and identically distributed (IID), as illustrated in Fig. 1(a). When normal data distribution shifts during testing (known as normality shift [1], as illustrated in Fig. 1(b)), the model will misclassify normal instances as anomalies, thereby significantly undermining the reliability and practical utility of the anomaly detector.

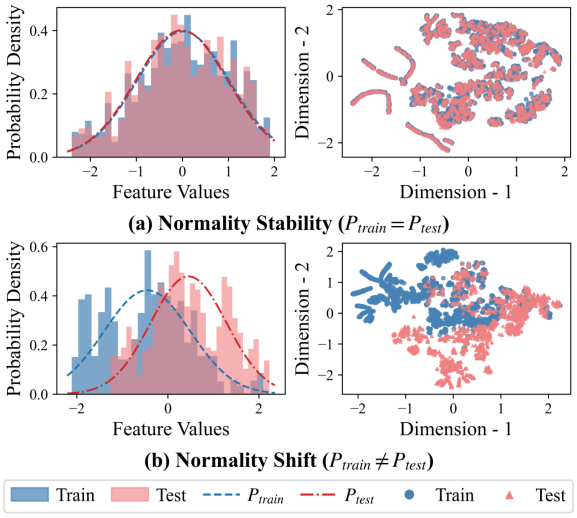

Fig. 1. Illustration of normality stability and shift. P_{train} and P_{test} denote the distributions of normal data in the training and test sets, respectively. (a) Under closed-world assumption, both sets share the same distribution, (b) while in open-world scenarios, the normal data distribution shifts over time.

Given the high cost and impracticality of retraining models in dynamic environments [1], recent studies have explored retraining-free strategies, which can be broadly categorized into two groups. The first focuses on improving training strategies to enhance robustness to normality shift at test time [3]. However, these methods rely heavily on the diversity and quality of training data and often fail to generalize when normal samples are limited or lack variability. Moreover, in open-world scenarios with uncertain normal distributions, adapting without access to test data remains challenging [4]. Thus, the second category leverages test data to adapt models to unknown distribution shifts [1]. Nevertheless, since test data may contain anomalies, such approaches are vulnerable to anomaly interference. In addition, normality shift can cause normal test samples to be misrepresented as anomalies, un-

* Corresponding author: Jiong Yu. This work was supported by the National Natural Science Foundation of China Project (62262064)

For more details, please visit https://github.com/ls-xju/ADANS

dermining adaptation and increasing false positive rates [5].

To address the limitations of existing research, we preserve the strengths of previous approaches while mitigating their shortcomings, and propose a tabular **A**nomaly **D**etection method for **A**dapting to **N**ormality **S**hift (ADANS). ADANS focuses on preventing significant performance degradation under normality shift during training and adapting the model to normality shift during testing.

First, the model is trained to learn perturbation-invariant representations, improving robustness to small variations and providing a foundation for adapting to normality shift. Specifically, a learnable masking generator (LMG) randomly masks feature dimensions and fills them with semantically plausible values, increasing normal data diversity and promoting generalized representations. In addition, FGSM [6] is used to generate boundary-sensitive samples, enhancing data representativeness and stabilizing the decision boundary against variations in normal data.

Second, the well-trained model is fine-tuned at test time using unlabeled data to improve adaptability to normality shift. Specifically, an adaptive representation revision process aligns the latent distributions of training and test data, ensuring that normal test samples are accurately encoded and remain consistent with training representations. Meanwhile, a dynamic weight adjustment mechanism mitigates anomaly interference by weighting samples according to their global and local similarities to the training data, allowing likely normal samples to dominate fine-tuning while suppressing the influence of potential anomalies.

This paper makes the following three contributions:

- We propose **ADANS**, a novel method for tabular anomaly detection that explicitly addresses the *normality shift* problem by enhancing model robustness and adaptability.
- During training, we introduce a *perturbation-invariant learning* strategy that simulates potential normality shifts in open-world scenarios, enhancing robustness to small variations and establishing stable feature representations for effective test-time adaptation.
- During testing, we develop an *adaptive representation revision* process that enables the model to cope with normality shift via unsupervised parameter fine-tuning. Meanwhile, a *dynamic weight adjustment* mechanism is employed to suppress the influence of anomalies.

II. Related Work

To combine deep representation learning with the efficiency of traditional methods, several zero-positive approaches have been proposed for tabular anomaly detection. DIF [7] projects data into latent spaces and applies isolation forest, while LUNAR [8] enhances KNN with graph neural networks. Inspired by self-supervised learning, MCM [9] and SLAD [10] construct multiple data views via feature masking and scale transformations, respectively, while DTPM [11] leverages forward diffusion to generate noisy variants for anomaly classification. However, these methods are built on a closed-world assumption. When the normal data distribution shifts at test time, models relying solely on training distributions tend to suffer from increased false alarms, limiting their applicability in open-world scenarios with dynamic distribution changes [4].

To address normality shift, ACR [3] partitioned training data into chronological subsets and enforced representation consistency by encouraging normal samples to converge to a shared latent center. However, limited data or insufficient diversity may cause overfitting, reducing robustness to unseen distribution shift of normal data. In contrast, OWAD [1] adapted a well-trained model at test time by fine-tuning on a small set of manually labeled test samples. While effective, this approach relies on manual annotation, significantly reducing efficiency and making it difficult to meet the fast-response demands of real-world applications. Moreover, the above approaches optimize either the training or testing phase in isolation, overlooking their interaction.

III. Preliminaries

A. Problem Formulation

Given a training set X that contains only normal data, zero-positive tabular anomaly detection aims to learn the normal pattern from X. During testing, samples that deviate from this learned normal pattern are identified as anomalies. When normality shift occurs, the test set X^t, which contains both normal and anomalous samples, follows a different normal data distribution from that of X.

B. Base Detector

In this work, we adopt a hypergraph-based autoencoder f as the base detector, consisting of an encoder and a decoder. To capture high-order correlations in real-world tabular data, we construct a hypergraph $HG = (V_H, E_H, W_H)$ [12]–[14], where each sample $v_H \in V_H$ forms a hyperedge $e_H \in E_H$ by connecting to its top-K most similar neighbors based on cosine similarity, and the hyperedge weight $w_H \in W_H$ is defined as the average similarity. Then, we leverage the hypergraph neural network (HGNNConv$^+$) [14], which follows a vertex $\rightarrow$ hyperedge $\rightarrow$ vertex message passing mechanism, extracting informative representations from limited training data and model normal patterns via high-order interactions. Morever, to mitigate over-smoothing induced by deep hypergraph convolutions, we further design a hybrid hypergraph convolutional block (HH Block), which fuses shallow and deep representations to balance vertex-specific features and global context. Both the encoder and decoder of f are constructed using the HH Block, which comprises two HGNNConv+ layers and a feature fusion module. The HH Block is defined as:

$$\begin{aligned} \text{HH}(X, HG) &= fusion(H_1, H_2) \\ &= \text{MLP}_1(\text{CONV}(\text{STACK}(\text{MLP}_2(H_1), H_2)), \end{aligned} \tag{1}$$

where MLP denotes a fully connected layer and CONV denotes a convolutional layer. $H_1 = \text{HGNNConv}^+(X, HG)$, $H_2 = \text{HGNNConv}^+(H_1, HG)$. $\text{STACK}(a, b) \in \mathbb{R}^{N \times 2 \times d}$, with $a, b \in \mathbb{R}^{N \times d}$, stacks a and b along the second dimension.

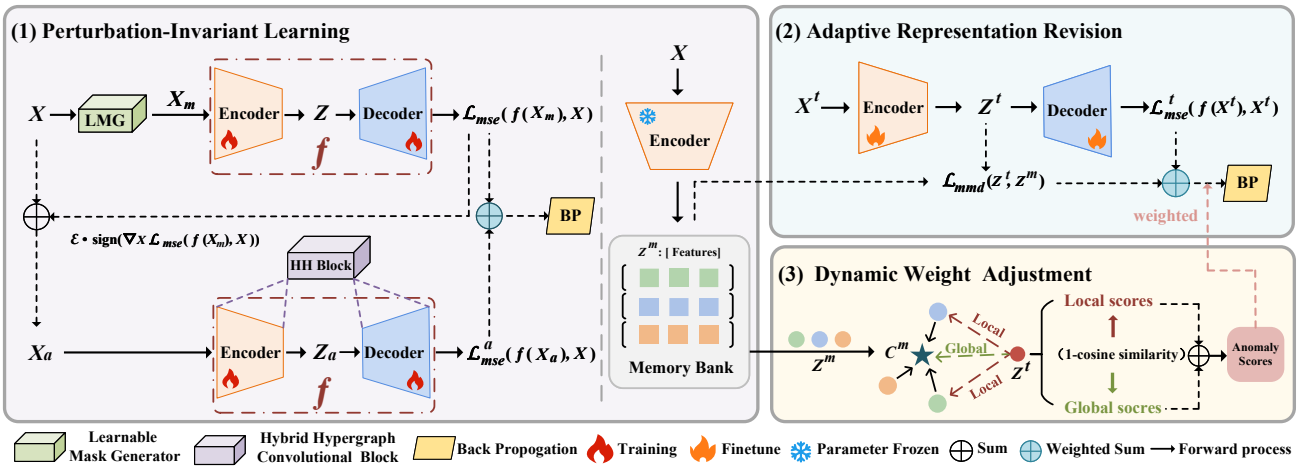

Fig. 2. Overview of ADANS. (1) f is trained to be invariant to transformations of X by reconstructing both the masked input X_m and adversarial examples X_a back to X. Then, (2) f is fine-tuned to mitigate normality shift by minimizing the distribution discrepancy between Z^t and Z^m, along with the reconstruction error of X^t. (3) Local anomaly scores are computed from the similarity between Z^t and its K-nearest neighbors in Z^m, while global anomaly scores are derived from the similarity between Z^t and the center C^m of Z^m. These scores are then used for instance-level weighting to guide the fine-tuning of f.

IV. Methodology

In practical applications, an ideal anomaly detector should accurately identify unknown anomalies without prior knowledge while maintaining reliable detection of normal instances under distributional shift. To this end, we propose ADANS, an anomaly detector tailored to handle normality shift in tabular data. As illustrated in Fig. 2, ADANS employs an autoencoder as its base detector and incorporates three key modules. Specifically, during training, the perturbation-invariant learning module enhances the autoencoder's robustness to potential normality shift, thereby reducing significant performance degradation at test time. During testing, the adaptive representation revision and dynamic weight adjustment modules jointly fine-tune the well-trained model using unlabeled test data, enabling effective adaptation to distributional changes in normal samples. Details of ADANS are elaborated in the following sections.

A. Perturbation-Invariant Learning (PIL)

When normality shift occurs, test-time normal samples may lie outside the decision boundaries learned during training, leading to their misclassification as anomalies. Thus, ADANS introduces two perturbation strategies during training to simulate potential variations in normal data distributions, thereby improving robustness to subtle changes and enhancing adaptability to real-world normality shift.

First, to promote more generalized normal representations, we propose LMG, which randomly masks input features using a binary mask matrix $\text{Mask} \in 0, 1^{N\times D}$. The masked entries are replaced with a learnable feature vector $L \in \mathbb{R}^{1\times D}$, yielding the masked data X_m:

$$X_m = \text{Mask} \odot L + (1 - \text{Mask}) \odot X, \tag{2}$$

where $\odot$ denotes element-wise multiplication with broadcasting. L is optimized to preserve semantic consistency between masked and original data, preventing the introduction of unrealistic patterns that could impair normal pattern learning. Mask is generated as:

$$\text{Mask}_{ij} = \mathbb{I}(r_{ij} > p), \quad r_{ij} \sim U(0, 1), \tag{3}$$

where $I(\cdot)$ represents indicator function, $p \in [0, 1]$ indicates the probability of not masking a feature, and U denotes the uniform distribution.

Second, to simulate boundary cases under distribution shifts, we adopt FGSM [6] to generate adversarial samples X_a by adding gradient-based perturbations to the input X. These perturbations do not alter the semantics of X but are sufficient to mislead the model f, thereby reducing its sensitivity to distribution shift. The adversarial samples are computed as:

$$X_a = X + \epsilon \cdot sign(\nabla_X \mathcal{L}_{mse}(f(X_m), X)), \tag{4}$$

where ϵ controls the perturbation magnitude, $sign(\cdot)$ denotes the sign function, $\mathcal{L}_{mse}$ is the mean squared error loss between $f(X)$ and X, and ∇_X is the gradient of the loss with respect to X.

The overall training objective of ADANS is to reconstruct both X_m and X_a back to X by minimizing their reconstruction losses:

$$J_t(\theta) = \lambda_1 \mathcal{L}_{mse} + \lambda_2 \mathcal{L}^a_{mse}, \tag{5}$$

where $\mathcal{L}_{mse} = \frac{1}{N}\sum_{i=1}^{N} ||X - f(X_m))||_2^2$, $\mathcal{L}^a_{mse} = \frac{1}{N}\sum_{i=1}^{N} ||X - f(X_a))||_2^2$, λ_1 and λ_2 are weighting coefficients used to balance the contributions of the two loss functions.

After training, the latent representations Z^m of X are stored in a memory bank to guide the adaptation of f during testing and serve as a reference for identifying anomalies.

B. Adaptive Representation Revision (ARR)

Although f is well trained, normality shift at test time may produce inaccurate latent representations Z^t, causing a mismatch between training and test normal data in the latent space

and the misclassification of normal samples as anomalies. Thus, ADANS adaptively revise test-time representations via unsupervised fine-tuning with a regularization term, aligning normal samples in X^t with those in X in the latent space.

First, to adapt to the normal data in X^t, we fine-tune f during testing. Since X^t is unlabeled and may contain anomalies, directly fine-tuning on it risks overfitting to abnormal patterns and reduce discriminability. Thus, we introduce sample-level weights w (defined in Eq 10) to dynamically control each sample's contribution, emphasizing likely normal samples while suppressing potential anomalies. Accordingly, the reconstruction loss is reformulated as a weighted objective:

$$\mathcal{L}^t_{mse} = \sum_{i=1}^{M} w_i ||f(x^t_i) - x^t_i||_2^2. \tag{6}$$

Second, to maintain consistent latent representations for normal samples across training and testing, we introduce a regularization term that aligns the distributions of Z^t and Z^m. Specifically, we employ MMD [15], a widely used metric for measuring distributional differences [16], as the regularization objective. To simplify computation, we apply the kernel trick and use the squared MMD as the regularization loss, defined as:

$$\mathcal{L}_{\text{mmd}} = \text{MMD}^2(Z^m, Z^t). \tag{7}$$

The overall testing objective of ADANS is optimized by jointly minimizing the weighted reconstruction loss and the MMD-based regularization loss:

$$J_a(\theta) = \gamma_1 \mathcal{L}^t_{mse} + \gamma_2 \mathcal{L}_{mmd}, \tag{8}$$

where γ_1 and γ_2 are weighting coefficients used to balance the contributions of the two loss functions.

C. Dynamic Weight Adjustment (DWA)

To ensure normal samples dominate the unsupervised fine-tuning process and reduce the influence of potential anomalies in the test data, we introduce a dynamic weighting mechanism that assigns higher weights to likely normal samples and lower weights to likely anomalies during model adaptation.

First, to estimate sample weights, we compute anomaly scores at each test-time iteration. As adaptation proceeds, normal samples in Z^t gradually align with the training representations Z^m, while anomalies remain misaligned. This discrepancy provides a natural signal for anomaly identification. Accordingly, we estimate anomaly scores from both local and global perspectives. The local score measures dissimilarity to the top-K nearest neighbors in Z^m, while the global score measures deviation from the training center C^m. The final anomaly score is defined as

$$AS = \sum \big(1 - \text{TopK}(cs(Z^t, Z^m))\big) + \big(1 - cs(Z^t, C^m)\big), \tag{9}$$

where $cs(\cdot,\cdot)$ denotes cosine similarity, $TopK_{values}$ selects the top-K similarity values, and C^m is the mean of Z^m.

Second, to emphasize normal samples, we apply a softmax function to the negated anomaly scores $AS = \{as_i \mid i = 1, \ldots, M\}$, yielding normalized, contrast-enhancing weights to guide the fine-tuning of f:

$$w_i = \text{softmax}(as.neg()) = \frac{e^{-as_i}}{\sum_{j=1}^{M} e^{-as_j}}. \tag{10}$$

These weights are incorporated into the reconstruction loss (Eq. (6)) to modulate each sample's contribution.

V. Experiments

A. Experimental Setup

1) Datasets.: We evaluate our method on four tabular datasets with temporal distribution shifts. The Net-Instruction dataset[1] is from the Anoshift benchmark [17], while Ecom-Offers[2], Site-Insurance[3], and Home-Credit[4] are sourced from the TabReD benchmark [18]. Following prior protocols [3], [17], the majority class is treated as normal and the minority as abnormal. Each dataset is temporally split into a training set and four test sets: IID, Near, Far, and Avg. The Train and IID sets are collected during the same period and share identical normal distributions. The Near set, collected shortly after the Train set, shows minor normality shift. The Far set, collected much later, exhibits a larger normality shift. The Avg set spans both the Near and Far periods.

2) Baselines and metrics.: We compare our method with ten SOTA approaches, including traditional methods (LOF [19], OCSVM [20]), deep learning–based methods (DIF [7], LUNAR [8], DTPM [11], SLAD [10], MCM [9]), and three methods specifically designed to address normality shift (OWAD [1], ACR_NTL [3], and ACR_DSVDD [3]). Model performance is evaluated using AUC-ROC and AUC-PR.

3) Implementation details.: All baselines are reproduced using official implementations with carefully tuned hyperparameters for fair comparison. For ADANS, both training and fine-tuning are conducted for 100 epochs, with learning rates of $1e^{-2}$ and $5e^{-3}$, respectively. The number of vertices per hyperedge and nearest neighbors in Eq. (9) are both set to 20. p in Eq. (3) is 0.8, ϵ in Eq. (4) is linearly increased from 0.05 to 1.0 during training. The embedding dimension is set to 24, and the hidden layer width of HGNNConv+ is three times the data dimension. Loss weights in Eq. (8) and Eq. (5) are computed using adaptive weighting [21].

B. Performance comparison

Normality stability. We first evaluate ADANS under normality stability using the IID setting, where the normal data distributions in the training and test sets are identical. We compute the average ranking of each method across four datasets using AUC-ROC and AUC-PR, with lower rankings indicating better performance. As shown in Fig. 3, ADANS achieves the lowest average rank under both metrics, demonstrating superior performance when no distribution shift is present.

[1]https://www.takakura.com/Kyoto_data/.
[2]https://kaggle.com/competitions/acquire-valued-shoppers-challenge.
[3]https://kaggle.com/competitions/homesite-quote-conversion.
[4]https://kaggle.com/competitions/home-credit-credit-risk-model-stability.

TABLE I
AUC-ROC AND AUC-PR RESULTS ON FOUR DATASETS UNDER VARYING DEGREES OF NORMALITY SHIFT

	Methods	Net-Intrusion			Ecom-Offers			Site-Insurance			Home-Credit		
		NEAR	FAR	AVG	NEAR	FAR	AVG	NEAR	FAR	AVG	NEAR	FAR	AVG
AUC-ROC	LOF	0.7616	0.4708	0.6440	0.4335	0.4411	0.4121	0.5579	0.5587	0.5599	0.5669	0.5029	0.5205
	OCSVM	0.8098	0.4801	0.6412	0.6525	0.5799	0.6124	0.5097	0.4945	0.5115	0.5409	0.5279	0.5234
	DIF	0.7521	0.6232	0.5479	0.6513	0.5824	0.6095	0.5025	0.5018	0.5172	0.4985	0.4997	0.4870
	LUNAR	0.4583	0.2871	0.3777	0.6082	0.5494	0.5725	0.4840	0.4555	0.4707	0.5308	0.5076	0.5107
	DTPM	0.6267	0.3915	0.5108	0.5975	0.5423	0.5536	0.5002	0.4717	0.4998	0.5078	0.5267	0.5079
	SLAD	0.2342	**0.6374**	0.3647	0.6375	0.5848	0.6121	0.5403	0.5428	0.5323	0.5575	0.4974	0.5348
	MCM	0.8168	0.4667	0.6619	0.6314	0.5710	0.5934	0.5377	0.5230	0.5300	0.5431	0.5100	0.5234
	OWAD	0.4138	0.4643	0.4396	0.6537	0.5803	0.6127	0.4873	0.4765	0.4954	0.5364	0.5081	0.5190
	ACR_DSVDD	0.7883	0.5289	0.6962	0.5298	0.5486	0.5475	0.5486	0.5805	0.5435	0.4198	0.5136	0.4872
	ACR_NTL	0.7967	0.6177	0.6936	0.6107	0.5675	0.6231	0.5374	0.4814	0.5441	0.5385	0.5100	0.5222
	ADANS	**0.9113**	0.5709	**0.7765**	**0.6675**	**0.6072**	**0.6416**	**0.6613**	**0.6198**	**0.6535**	**0.5686**	**0.5422**	**0.5398**
AUC-PR	LOF	0.4448	0.1931	0.2665	0.1701	0.1769	0.1617	0.2606	0.2431	0.2617	0.2143	0.2047	0.2088
	OCSVM	0.4419	0.2278	0.2599	0.2994	0.2645	0.2830	0.2368	0.2125	0.2409	0.2078	0.2096	0.2080
	DIF	0.3561	**0.3054**	0.2265	0.2974	0.2671	0.2827	0.2336	0.2174	0.2440	0.1958	0.2028	0.1987
	LUNAR	0.4625	0.1508	0.1754	0.2825	0.2413	0.2515	0.1888	0.1873	0.1933	0.2013	0.2080	0.2070
	DTPM	0.3361	0.2218	0.2264	0.2720	0.2355	0.2453	0.2238	0.2067	0.2308	0.1918	0.2094	0.2052
	SLAD	0.1553	0.2930	0.1613	0.2975	0.2685	0.2844	0.2425	0.2342	0.2412	0.2167	0.1939	0.2063
	MCM	0.4953	0.2172	0.2723	0.2909	0.2544	0.2715	0.2445	0.2278	0.2498	0.2095	0.2090	0.2107
	OWAD	0.2555	0.2529	0.2149	0.3017	0.2630	0.2823	0.2261	0.2044	0.2337	0.2071	0.2081	0.2098
	ACR_DSVDD	0.4584	0.2324	0.3656	0.2413	0.2495	0.2449	0.2492	0.2698	0.2502	0.1694	0.2152	0.2020
	ACR_NTL	0.3984	0.2679	0.3295	0.2713	0.2552	0.2747	0.2472	0.2044	0.2460	0.2298	0.2112	0.2248
	ADANS	**0.6791**	0.2610	**0.4039**	**0.3499**	**0.3172**	**0.3037**	**0.3109**	**0.2734**	**0.3201**	**0.2302**	**0.2187**	**0.2313**

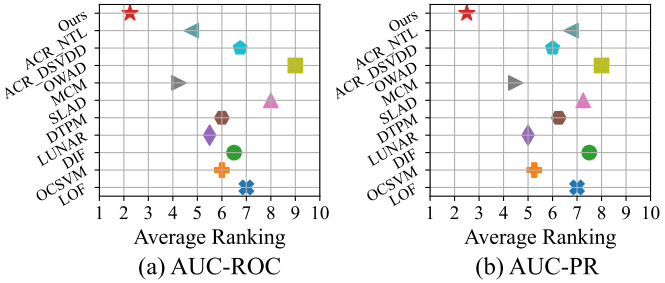

Fig. 3. Average rankings of eleven methods across four datasets under normality stability, where lower values indicate better performance.

Normality shift. Table I reports performance under increasing degrees of normality shift. As the shift intensifies from NEAR to FAR, all methods experience performance degradation, confirming the adverse impact of normality shift on anomaly detection. ADANS consistently delivers competitive or superior performance, particularly under NEAR and AVG settings, indicating strong robustness to mild and moderate shifts. Under the more challenging FAR setting, ADANS outperforms all baselines on three of the four datasets, highlighting its effectiveness under severe distribution shifts. A notable exception is observed on the Net-Intrusion dataset, where performance degrades noticeably. This is likely due to the substantial distribution shift caused by an approximately ten-year temporal gap between the training and test data, which makes it difficult for the ARR module to align the two distributions. While this reveals a limitation of ADANS under extreme long-term drift, the method remains robust across most scenarios. Improving adaptability to long-term distribution evolution is a promising direction for future work.

TABLE II
ABLATION STUDY ON THREE MAJOR MODULES. ✓ AND × DENOTE THE INCLUSION AND EXCLUSION OF A MODULE, RESPECTIVELY.

Modules			Net-Intrusion (AUC-PR)		
PIL	ARR	DWA	NEAR	FAR	AVG
×	×	×	0.4093	0.1694	0.2443
✓	×	×	0.4429	0.2090	0.2514
×	✓	×	0.5291	0.2067	0.3045
×	✓	✓	0.5508	0.2244	0.3232
✓	✓	✓	**0.6791**	**0.2610**	**0.4039**

C. Ablation Study

ADANS addresses normality shift through three components: PIL, ARR, and DWA. An ablation study on the Net-Intrusion dataset (Table II) shows that removing all modules results in poor performance. Adding PIL alone yields moderate improvements, suggesting enhanced representation robustness but insufficient adaptation capability. In contrast, enabling ARR leads to a substantial performance gain, highlighting the importance of test-time representation alignment. Further incorporating DWA consistently enhances performance by suppressing anomaly interference. The full model achieves the best results, demonstrating the complementary roles of PIL, ARR, and DWA in adapting to normality shift.

D. Robustness Analysis

To evaluate robustness under varying test data sizes, we analyze performance across batch sizes from 20 to 4000 on the Ecom-Offers dataset (AVG setting, Fig. 4). ADANS consistently outperforms all baselines and remains stable even with small batches, where limited samples often hinder reliable adaptation. This stability results from the combined effect

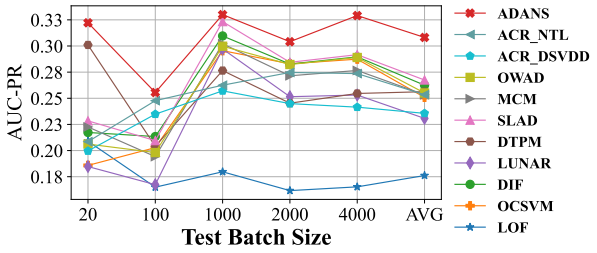

Fig. 4. Performance comparison across varying test set sizes.

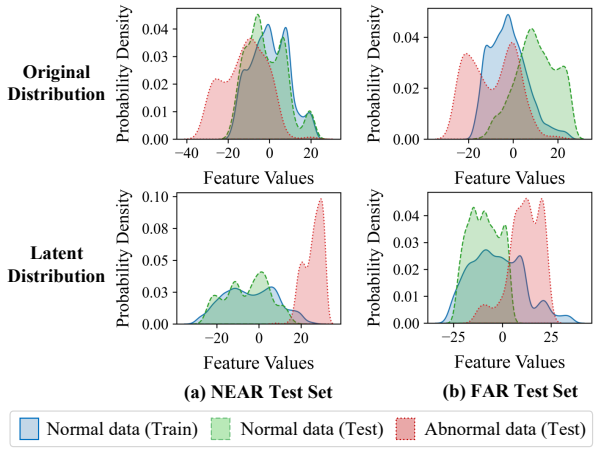

Fig. 5. Distribution of normal and abnormal data in original and latent spaces.

of representation alignment and dynamic weighting, which mitigates overfitting to noisy or abnormal samples. Such robustness is particularly valuable in real-world scenarios where test data size is limited or evolves over time.

E. Visualization Results

We visualize the original and latent data distributions using t-SNE (Fig. 5). In the original space, normal data exhibit clear distribution shifts and anomalies are poorly separated. After applying ADANS, normal representations become compact and well aligned across training and testing, while anomalies are more clearly separated, demonstrating the effectiveness of ADANS in mitigating normality shift.

VI. Conclusion

Existing zero-positive anomaly detection methods often focus on distribution shifts of anomalies, while overlooking the impact of normality shift. Although some recent studies have begun to address this issue, their effectiveness remains limited by various constraints. To overcome this challenge, we propose ADANS. During training, the PIL module enhances robustness to minor distributional variations, while during testing, the ARR module fine-tunes the model on unlabeled data guided by the DWA module, which emphasizes normal samples and suppresses anomalies. Extensive experiments on four real-world tabular datasets with normality shift demonstrate the effectiveness of ADANS.

References

[1] Dongqi Han, Zhiliang Wang, Wenqi Chen, Kai Wang, Rui Yu, Su Wang, et al., "Anomaly detection in the open world: Normality shift detection, explanation, and adaptation.," in *Proc. Netw. Distrib. Syst. Secur. Symp. (NDSS)*, 2023, pp. 1–18.

[2] Zongyuan Huang, Baohua Zhang, Guoqiang Hu, Longyuan Li, Yanyan Xu, and Yaohui Jin, "Enhancing unsupervised anomaly detection with score-guided network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 10, pp. 14754–14769, 2024.

[3] Aodong Li, Chen Qiu, Marius Kloft, Padhraic Smyth, Maja Rudolph, and Stephan Mandt, "Zero-shot anomaly detection via batch normalization," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2023, vol. 36, pp. 40963 – 40993.

[4] Yifan Zhang, Xue Wang, Kexin Jin, Kun Yuan, Zhang Zhang, Liang Wang, et al., "Adanpc: Exploring non-parametric classifier for test-time adaptation," in *Proc. Int. Conf. Mach. Learn.*, 2023, pp. 41647–41676.

[5] A Tuan Nguyen, Thanh Nguyen-Tang, Ser-Nam Lim, and Philip HS Torr, "Tipi: Test time adaptation with transformation invariance," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2023, pp. 24162–24171.

[6] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy, "Explaining and harnessing adversarial examples," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–11.

[7] Hongzuo Xu, Guansong Pang, Yijie Wang, and Yongjun Wang, "Deep isolation forest for anomaly detection," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 12, pp. 12591–12604, 2023.

[8] Adam Goodge, Bryan Hooi, See-Kiong Ng, and Wee Siong Ng, "Lunar: Unifying local outlier detection methods via graph neural networks," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, 2022, vol. 36, pp. 6737–6745.

[9] Jiaxin Yin, Yuanyuan Qiao, Zitang Zhou, Xiangchao Wang, and Jie Yang, "Mcm: Masked cell modeling for anomaly detection in tabular data," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2024, pp. 1–23.

[10] Hongzuo Xu, Yijie Wang, Juhui Wei, Songlei Jian, Yizhou Li, and Ning Liu, "Fascinating supervisory signals and where to find them: Deep anomaly detection with scale learning," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2023, p. 38655–38673.

[11] Victor Livernoche, Vineet Jain, Yashar Hezaveh, and iamak Ravanbakhsh, "On diffusion modeling for anomaly detection," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2024, pp. 1–31.

[12] Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao, "Hypergraph neural networks," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, 2019, vol. 33, pp. 3558–3565.

[13] Yue Gao, Zizhao Zhang, Haojie Lin, Xibin Zhao, Shaoyi Du, and Changqing Zou, "Hypergraph learning: Methods and practices," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 5, pp. 2548–2566, 2022.

[14] Yue Gao, Yifan Feng, Shuyi Ji, and Rongrong Ji, "Hgnn+: General hypergraph neural networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 3181–3199, 2023.

[15] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan, "Learning transferable features with deep adaptation networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2015, pp. 97–105.

[16] Zhiming Cheng, Shuai Wang, Defu Yang, Jie Qi, Mang Xiao, and Chenggang Yan, "Deep joint semantic adaptation network for multi-source unsupervised domain adaptation," *Pattern Recognit.*, vol. 151, pp. 110409, 2024.

[17] Marius Dragoi, Elena Burceanu, Emanuela Haller, Andrei Manolache, and Florin Brad, "Anoshift: A distribution shift benchmark for unsupervised anomaly detection," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2022, vol. 35, pp. 32854–32867.

[18] Ivan Rubachev, Nikolay Kartashev, Yury Gorishniy, and Artem Babenko, "Tabred: Analyzing pitfalls and filling the gaps in tabular deep learning benchmarks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2025, pp. 1–35.

[19] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander, "Lof: identifying density-based local outliers," in *Proc. ACM SIGMOD Int. Conf. Manag. Data (SIGMOD)*, 2000, pp. 93–104.

[20] Bernhard Schölkopf, Robert C Williamson, Alex Smola, John Shawe-Taylor, and John Platt, "Support vector method for novelty detection," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 1999, vol. 12.

[21] Roberto Cipolla, Yarin Gal, and Alex Kendall, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 7482–7491.