Deep Anomaly Detection Of Intelligent Robot Via Acoustic Signal Recognition

Yu Jiang

Shenyang Normal University, ShenYang, China

Corresponding author. E-mail: jiangyu@synu.edu.cn

Received: Jan. 01, 2024; Accepted: Jan. 27, 2024

With the advancement of artificial intelligence, intelligent robots, equipped with capabilities such as autonomous perception, learning adaptability, autonomous decision-making, and human-machine interaction, are widely deployed across various industries. In this paper, an acoustic signal analysis within the deep encoding-decoding architecture is devised for intelligent robot anomaly detection (DAD-IR-ASR), which comprises acoustic signal data preprocessing of intelligent robot, deep encoding-decoding architecture, accumulation-based anomaly detection. Specifically, DAD-IR-ASR designs an acoustic sensor collection device to collect acoustic signal data of robots. Subsequently, it utilizes Fourier transformation and filters to extract meaningful spectral features. Simultaneously, it designs a deep encoding-decoding architecture, employing unsupervised reconstruction training solely with normal data to adaptively learn an error threshold. Furthermore, DAD-IR-ASR conducts an accumulation-based anomaly detection strategy to determine if the intelligent robot is anomalous by comparing the cumulative sum of reconstruction errors within deep encoding-decoding architecture. Finally, the effectiveness of DAD-IR-ASR is demonstrated through comparisons with multiple existing unsupervised detection methods in intelligent robot anomaly detection task.

Keywords: Artificial intelligence; intelligent robots; deep anomaly detection

http://dx.doi.org/10.6180/jase.202412_27(12).0013

1. Introduction

The industries are embracing the artificial intelligence through pattern mining [1], knowledge discovery [2], information acquisition [3], data analysis [4], and intelligent decision making [5, 6], to elevate production capabilities, which is devoted to achieving superior quality, heightened efficiency, and reduced energy consumption. Against this evolving backdrop, intelligent robots have risen to prominence as a crucial technology. Through seamless integration and connection with various devices, sensors, and systems, they have significantly improved production efficiency and reduced manufacturing costs. For example, automobile manufacturers employ intelligent robots to complete the assembly process of cars. These robots are equipped with advanced vision systems that enable them to recognize the positions of components, performing high-precision welding, assembly, and painting operations without the need for human intervention. Intelligent robots have been widely deployed across various aspects of the Industrial Internet of Things, making a substantial contribution to the advancement of modern manufacturing practices [7, 8].

However, it is inevitable that intelligent robots may experience performance degradation, malfunctions, and even shutdowns during operation, which leads to serious productivity issues and even dangerous accidents. Therefore, extensive research has been conducted on the detection of abnormal behavior in intelligent robots. Usually, anomaly detection in intelligent robots can be divided

into traditional-based abnormal detection methods and machine learning-based anomaly detection methods abnormal detection methods. The former typically considers various static and dynamic parameters, delving into prior knowledge and rules of robots, to construct precise models, which mainly includes thresholding, rule engines, and state monitoring [9–11]. Thresholding triggers anomaly alerts by setting critical values, rule engines establish a set of rules defining normal and abnormal behavior, and state monitoring focuses on parameters such as the robot's motion state. The latter emphasizes feature extraction, extracting compressed and meaningful information from raw sensor signals, which primarily employs supervised, unsupervised, or semi-supervised learning strategies [12–14]. Supervised learning utilizes labeled data for training, unsupervised learning relies on unlabeled data, and semi-supervised learning combines both labeled and unlabeled data. In recent years, with advancements in sensor and computing technologies, the effective generation of large amounts of data has significantly improved the accuracy of models built using machine learning-based anomaly detection methods.

In spite of substantial advancements made by machine learning-based anomaly detection methods, there are still five hurdles that need to be surmounted.

1. Data Acquisition Challenges: Collecting machine data under abnormal conditions is challenging, and artificially introducing defects may lead to severe machine failures, making effective data acquisition a significant issue.

2. Uncertainty in Category Definition: Building a predictive model requires predefining fault categories based on machine states. However, it's not always feasible to define each type of anomaly, which may result in diagnostic failures or false alarms.

3. Imbalance in Data Distribution: In real industrial scenarios, anomalies are relatively rare, causing an imbalance in the quantity of normal and abnormal data. This imbalance biases the decision boundary of machine learning models toward normal data, reducing the detection rate for anomalies.

4. Limitations of Supervised Learning Methods: Many machine learning-based anomaly detection methods rely on supervised classification methods such as SVM. However, due to the scarcity of abnormal data for mechanical arm anomalies and the imbalanced ratio with normal data, these methods can only detect known anomaly situations, struggling to adapt to new and unforeseen anomalies.

5. High Cost of Data Acquisition: They rely on approaches such as force, vibration, or optical sensors to gather data on the status of robots. However, these methods incur substantial costs, making them less practical for industrial applications.

To this end, an acoustic signal analysis within the deep encoding-decoding architecture is devised for intelligent robot status anomaly detection (DAD-IR-ASR), which comprises acoustic signal data preprocessing of intelligent robot, deep encoding-decoding architecture, accumulation-based anomaly detection. Specifically, DAD-IR-ASR designs an acoustic sensor collection device that does not require signal conditioning equipment such as amplifiers and analog-to-digital converters, enabling the cost-effective acquisition of acoustic signals by attaching it to points of interest on the robot. Subsequently, it utilizes Fourier transformation and filters to extract meaningful spectral features from the acoustic information. Simultaneously, it designs a deep encoding-decoding architecture, employing unsupervised reconstruction training solely with normal data to adaptively learn an error threshold., which effectively mitigates the issue of imbalance in the number and categories between normal and anomalous data. Furthermore, DAD-IR-ASR conducts an accumulation-based anomaly detection strategy to determine if the intelligent robot is anomalous by comparing the cumulative sum of reconstruction errors from the deep autoencoder architecture with the threshold derived from normal data. Finally, the effectiveness of DAD-IR-ASR is demonstrated through comparisons with multiple existing unsupervised detection methods in intelligent robot anomaly detection task.

The rest of this paper: Section II comprehensively explores the seminal works in abnormal detection for intelligent robots, providing an in-depth understanding of the existing literature and notable contributions in this domain. Moving to Section III, the paper elucidates the intricate details of deep anomaly detection in intelligent robots through acoustic signal recognition. This section delves into the specific methodologies, techniques, and architectures employed in leveraging acoustic signals for effective anomaly detection. In Section IV, the experimental evaluations are meticulously illustrated. This section offers a detailed account of the experiments conducted, the datasets utilized, and the results obtained. It provides a critical analysis of the performance of DAD-IR-ASR. Lastly, Section V serves as the conclusion of the entire paper.

2. Related works

In the industrial sector, the anomaly detection capabilities of intelligent robots are crucial for enhancing production efficiency and maintaining the stable operation of equipment. By continuously monitoring key parameters and equipment status in real-time during the production process, intelligent robots can swiftly identify and report any abnormal conditions. This capability enables production teams to intervene promptly, preventing potential issues from escalating and thereby reducing production line downtime, ultimately improving overall production efficiency. Abnormal detection in intelligent robots is categorized into two types: traditional-based abnormal detection methods and machine learning-based anomaly detection methods abnormal detection methods.

Traditional-based abnormal detection methods

This type of approach involves establishing a dynamic model for intelligent robots through physical analysis to conduct anomaly detection along with incorporating defensive measures to counter the threat of external attacks. For example, a physical analysis of the robotic arm is conducted to establish a model for accurately identifying anomalies based on the current control commands and intelligent robot motion states [9]. After deploying the model, the operation data derived from intelligent robots is estimated using control commands and the model. If the threshold exceeds the deviation, the detector issues an alarm. Handling the tracking control issue of intelligent robots amidst Denial-of-Service attacks, a hybrid dynamic model is established via combining with the explicit characteristics of malicious DoS attacks, which achieves closed-loop tracking control based on a given time-varying trajectory [10]. An anomaly-based attack detection model is devised via conducting an in-depth analysis from the perspectives of secure networks, control systems, physical processes, and the interactions between these aspects, implementing multifaceted defense to effectively detect deceptive attacks [11]. A combined anomaly detection system is proposed, incorporating communication, task, resource, and control data flow models [15]. After detecting anomalies in the system, a Markov classifier is constructed for alarm classification, distinguishing between system anomalies and faults. Jaber et al. acquire vibration signals from the repetitive operation of the PUMA 560 robot. Statistical features are extracted from these signals, and a statistical control chart is established to monitor the distribution of these features and detect any inconsistencies with normal conditions [16]. Sathish et al. utilize Principal Component Analysis to investigate the impact of training data sources and types on anomaly detection in industrial robots [17]. A system is stimulated by adding random noise to the control inputs to examine its corresponding response, and then based on the observed sensor measurements, the system's normal operating state is determined using a chi-square test [18].

Although the strategy of introducing random noise is successful in identifying replay attacks, it concurrently led to a decline in the control performance of the system, exerting a notable impact on industrial control systems. Most traditional-based abnormal detection methods face several challenges in the field of intelligent robotics. Firstly, the requirement for in-depth domain knowledge of robotic arms and production environments makes these methods complex in practical applications. The motion state information of intelligent robots typically exhibits highly nonlinear characteristics, rendering traditional physical modeling quite challenging. This complexity makes it difficult to accurately capture the dynamic features of the system through traditional manual modeling methods. A more intricate issue arises when the modeling process does not consider system features that attackers might exploit. In such cases, the established detection model may exhibit biases. This implies that the model might only detect specific types of attacks that were considered during the modeling process and may not adapt well to new types of attacks or different threat scenarios. Consequently, this bias can lead to vulnerabilities in the system when faced with unknown threats.

Machine learning-based anomaly detection methods

This type of approach is highly favored in both industrial and academic settings. For example, Narayanan et al. proposed a support vector machine-based anomaly detection method to identify these anomalous changes. The method incorporates the concept of tolerance enveloping to further enhance the detection accuracy, especially concerning product tolerance specifications [19]. Zhou et al. divide the operation of intelligent robots into five continuous states and then utilize the particle swarm updating strategy to classify the robotic operating states [20]. During the detection process, the classifier provided real-time predictions of intelligent robot operating state. This involved assessing whether robotic executed tasks in the specified order, enabling intrusion detection. The accuracy of classification and detection are 96.02% and 90%, respectively, demonstrating the effectiveness and reliability of detecting physical logic attacks. A probability-based outlier detector using Deep Neural Networks (DNN) was implemented and evaluated on the SWaT dataset [21]. The performance of this detector was compared with OCSVM [13]. The results indicated that the DNN method outperformed in terms of the F-measure and precision but incurred higher computa-

tional costs. On the other hand, the OCSVM method exhibited faster computation, better recall, but occasional false positives. A sliding window was applied to filter the latest m input samples for each robotic arm dataset, reducing noise [12]. Subsequently, dimensionality reduction was performed by segmenting the data into relevant attribute sets. For each set, Mahalanobis distance was calculated, and if it exceeded a predefined threshold, an abnormal operation was determined. Common unsupervised learning methods for distance-based anomaly detection and density-based classification methods, such as the K-Nearest Neighbor (KNN) algorithm, demonstrated effectiveness in detecting the aforementioned data. However, these approaches encounter constraints about time and computational burden, especially when confronted with high-dimensional data. Unlike traditional SVM methods, OCSVM generates a discriminative boundary based solely on normal operating data. Data points falling outside this boundary are considered anomalies. While OCSVM does not require a large amount of pre-training with anomalous data, its efficiency may decrease due to significant data variations and noise. Some existing detection methods leverage autoencoders and their variants to extract latent representations of normal data, reconstructing it for anomaly detection. Xu et al. utilizes a variational autoencoder to key performance indicator time-series data for anomaly detection [14]. Vibration signals are captured using a three-axis accelerometer to monitor abnormal crystal vibrations [22]. Maintenance alerts were issued before vibration loss of control affected production, reducing the impact on the industrial control system production line. An unsupervised learning approach incorporates a monocular camera to train a feature extractor, enabling collaborative learning of network parameters across various pseudo-classes without the need for additional data. The outcome is the generation of dual-feature representations [23]. Grayscale images of a moving robotic arm were converted into continuous frames, and the signal's similarity to a reference signal was computed [24]. Higher similarity indicated that the motion trajectory at that moment aligned with expected normal data.

Machine learning-based strategies are widely adopted for various anomaly detection tasks. However, practical industrial environments pose challenges as anomalies are infrequent and challenging to capture, resulting in an imbalance between normal and anomalous data quantities. The model's decision boundary often prioritizes the larger volume of normal data, leading to a significant reduction in anomalous data detection rates. Many machine learning-driven anomaly detection methods rely on supervised classification approaches, such as Support Vector Machines (SVM). These methods demand a substantial amount of anomalous data, yet anomalies in mechanical arms are rare, resulting in an imbalance in normal-to-anomalous data ratio. Furthermore, these detection models are limited to recognizing known anomaly patterns, making them ineffective against new and unforeseen anomalies. Additionally, current methods predominantly focus on anomaly detection in individual dimensions, lacking the capability to identify anomalies across multiple dimensions. This limitation hinders the comprehensive detection of complex anomalies that may span multiple aspects of the system. In summary, while machine learning-based strategies are widely used for anomaly detection, challenges arise in industrial settings due to the scarcity of anomalous data, imbalance issues, and limitations in recognizing new and multidimensional anomalies. Addressing these challenges is crucial for enhancing the effectiveness of anomaly detection methods in practical applications.

3. Deep anomaly detection of intelligent robot via acoustic signal recognition

An acoustic signal analysis within the deep encoding-decoding architecture is devised for intelligent robot status anomaly detection (DAD-IR-ASR), which comprises acoustic signal data preprocessing of intelligent robot, deep encoding-decoding architecture, accumulation-based anomaly detection, as shown in Fig. 1. The main mathematical notations used in the paper are listed in Table 1.

In the training stage, DAD-IR-ASR utilizes normal acoustic signals X_i as the input to extract features E_i based on SIFT and the filter function, and then conducts the deep encoding-decoding architecture to employ unsupervised reconstruction training solely with normal features E_i to adaptively learn an error threshold τ. In the testing stage, DAD-IR-ASR utilizes normal and abnormal acoustic signals X_i as the input to extract features E_i based on SIFT and the filter function, and then conducts an accumulation-based anomaly detection strategy to determine if the intelligent robot is anomalous by comparing the cumulative sum s_i of reconstruction errors r_i with the threshold.

3.1. Acoustic signal data preprocessing of intelligent robot

We utilize acoustic sensors to gather data on the working state of intelligent robots.

Subsequently, we employ STFT to convert the collected data into a frequency-domain spectrogram. By applying filtering, we compress the spectrogram into features that are processable by deep autoencoders, achieving preprocessing of the acoustic signal data.

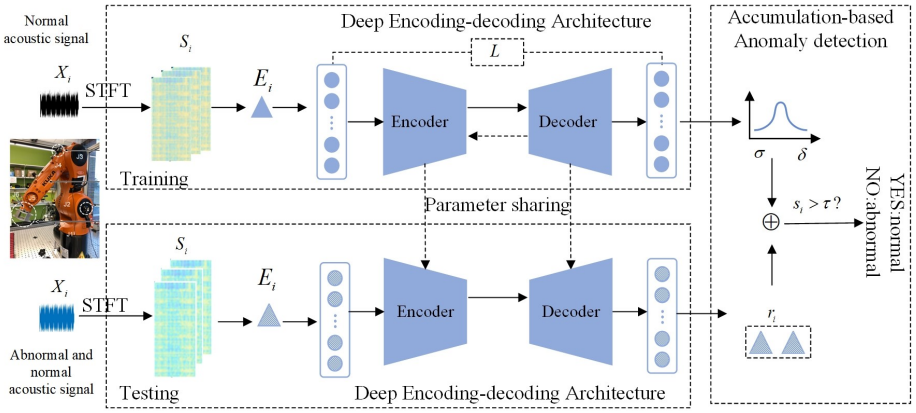

Fig. 1. The illustration of DAD-IR-ASR.

Table 1. Frequently used notations.

Notations	Description
$x(t)$	Acoustic signal data.
$x(\omega, t)$	Spectrogram.
ω	Frequency.
t	Time.
$H(\omega)$	Butterworth bandpass filter function.
$E(t)$	Feature.
z^i	Representation.
f_{encoder}	Encoder function.
f_{decoder}	Decoder function.
$\hat{E}(t)$	Reconstruction feature.
η	Learning rate.
θ	Model parameter.
r_i	Reconstruction error.
τ	Threshold.

Specifically, given an acoustic signal data $x(t)$ collected from intelligent robots, where t represents time, we firstly apply STFT to transform the signal data from the time to the frequency domain:

$$X(\omega, \tau) = \int_{-\infty}^{\infty} x(t)\omega(t-\tau)e^{-j\omega t}dt \tag{1}$$

where $X(\omega, \tau)$ is the spectrogram, ω is frequency, t is time, and $\omega(t)$ is the window function. The spectrogram information is generated via $S(\omega, \tau) = |X(\omega, \tau)|^2$, representing the distribution of frequencies over time and forming the spectrogram. Subsequently, filtering is performed using the operation $S_{\text{filtered}}(\omega, \tau) = H(\omega) \cdot S_{\text{filtered}}(\omega, \tau)$, where $H(\omega)$ is the butterworth bandpass filter function. This step compresses the frequency range in the spectrogram to the region of interest, emphasizing specific frequency bands. Finally, features are extracted from the filtered spectrogram, expressed as:

$$E(\tau) = \frac{1}{N}\sum_{\omega} S_{\text{filered}}(\omega, \tau) \tag{2}$$

where $E(\tau)$ represents the feature (average energy) extracted within the time window τ.N denotes the spectrum number, indicating the quantity of discrete sampling points in the frequency domain.

3.2. Deep encoding-decoding architecture

Deep encoding-decoding architecture is an unsupervised deep learning model, containing an encoder and a decoder. It operates by mapping data from a high-dimensional data space to a low-dimensional representation space via the encoder and reconstructing it back to the high-dimensional data space via the decoder, whose process is based on approximating the identity function to extract low-dimensional abstract features from the data.

Specifically, given a sample set $\left\{E^{(i)}\right\}_{i=1}^{n}$, where n denotes the sample number, $E^{(i)}$ represents the i-th sample, and d is the data dimension. The encoder extracts an abstract representation of the samples, denoted as $z^{(i)}$, which can be described using the following formula:

$$z^i = f_{\text{encoder}}\left(E^{(i)}\right) \tag{3}$$

where $f_{\text{encoder}}(\cdot)$ is the encoder function, mapping data $E^{(i)}$ as abstract representation $z^{(i)}$.

This mapping can be nonlinear and typically involves an activation function, such as sigmoid or tanh. Next, the decoder reconstructs the original high-dimensional data by mapping the abstract representation $z^{(i)}$ back to the original space, resulting in the reconstructed sample:

$$\hat{E}^{(i)} = f_{\text{decoder}}\left(z^i\right) \tag{4}$$

where $\hat{E}^{(i)}$ is the reconstructed sample, and $f_{\text{decoder}}(\cdot)$ is the decoder function.

The overall training objective of the deep encoding-decoding architecture is usually to minimize the recon-

struction error, making $E^{(i)}$ and $\hat{E}^{(i)}$ as similar as possible. This is achieved by defining a loss function L :

$$L(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left\|E^{(i)} - \hat{E}^{(i)}\right\|^2 \tag{5}$$

This training objective helps to learn a compact representation $z^{(i)}$ that captures the main features of the input data. Finally, the Stochastic Gradient Descent algorithm is employed to update θ :

$$\theta = \theta - \eta \cdot \nabla_\theta L(\theta) \tag{6}$$

where η is the learning rate, determining the step size in the parameter space. ∇_θ denotes the gradient with respect to the model parameters θ.

To use deep encoding-decoding architecture to classify abnormal signals, we design a training-testing mechanism:

- Training: the deep encoding-decoding architecture is trained exclusively with normal state signals. By exclusively using normal signals, the model is capable of capturing the features of normal states, resulting in an encoder and decoder that can effectively reconstruct normal data.

- Testing: the deep encoding-decoding architecture faces data from "unknown" anomalous states. As the model hasn't been exposed to abnormal signals during training, it struggles to accurately reconstruct these abnormal signals, leading to a substantial error between the generated signal and the original signal. By observing the increase in reconstruction error, it can be used to identify and differentiate between normal and anomalous signals. The core idea of this training-testing mechanism is rooted in the normal state of the training data, detecting patterns that deviate from the normal mode to identify anomalous signals.

3.3. Accumulation-based anomaly detection

In the testing phase, we employ a method to assess whether the output estimated by the autoencoder deviates from the normal input. Once a deviation is detected, an alarm is immediately issued, and operations of the intelligent robot is halted. The accumulation-based detection method involves following four steps:

(1) Calculate the residuals between the reconstructed output $\hat{E}^{(i)}$ and the input $E^{(i)}$:

$$r_i = E_i - \hat{E}_i \tag{7}$$

(2) Compute the expectation δ and standard deviation σ of reconstruction for normal data in the normal production process:

$$\begin{aligned}\delta &= \frac{1}{n}\sum_{i=1}^{n} r_i \\ \sigma &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(r_i - \delta)^2}\end{aligned} \tag{8}$$

(3) For each residual r_i at every time step, calculate the cumulative Sum s_i, considering significant noise in motion state data, especially acceleration data. Set $\delta' = 3\delta$

$$S_i = \max\left(0, s_{i-1} + r_i - 3\delta\right) \tag{9}$$

(4) Check whether the cumulative sum s_i exceeds the Threshold τ. If $s_i > \tau$, indicate an abnormal motion state, trigger an alarm, halt the robotic operation, and await further evaluation by operators.

4. Results

4.1. Setup

Data Collection: We have designed and constructed a safety testing platform for an intelligent robot palletizing process production line. This platform is based on real industrial scenarios, allowing for the complete configuration of robots to closely replicate industrial settings. Simultaneously, we simulated three typical anomalies: malicious command injection anomaly, manipulation of motion state data anomaly, and replay anomaly. We utilized acoustic sensors to record both 90 abnormal and 300 normal data, as shown in Table 2, during these scenarios, Here are detailed steps for the data collection process as illustrated in the Fig. 2.

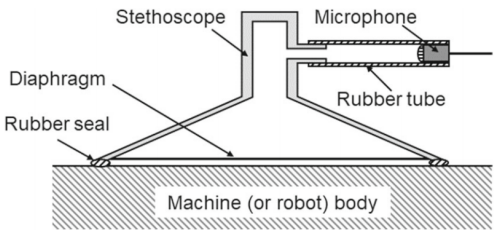

Fig. 2. The detailed device information for the data collection process.

Sensor Connection: The USB microphone is ingeniously connected to the acoustic sensor via a rubber tube, ensuring stable signal transmission.

Table 2. Statistical details in model training and testing, abnormal_1:malicious command injection anomaly, abnormal_2: manipulation of motion state data anomaly, abnormal_3: replay anomaly.

model	normal_1	normal_2	abnormal_1	abnormal_2	abnormal_3
training	200	0	0	0	0
testing	0	100	30	30	30

Diaphragm Design: The chest piece incorporates a diaphragm structure designed to isolate external noise to the maximum extent, ensuring that captured audio signals are closely related to the robot's operational state.

Secure Attachment of Chest Piece: The chest piece is securely attached to the surface of the intelligent robot through a specialized rubber seal. This ensures a snug fit between the chest piece and the robot, preventing the generation of stray noises during movement.

No Signal Conditioning Devices: The sensor design eliminates the need for additional signal conditioning devices, such as amplifiers and analog-to-digital converters. This means that we can directly obtain audio signals from the personal computer (PC) through USB communication, simplifying the data collection process.

Through meticulous design and detailed steps, we not only ensure the reliability of the acoustic sensor in the data collection process but also reduce costs, making it more economically viable compared to other sensors.

Implementation Details: The hardware and training environment configurations for the anomaly detection model based on the deep encoding-decoding architecture are as follows: The CPU is an Intel i7-10870H 2.20GHz, with 32GB of RAM, and the GPU is an NVIDIA GeForce GTX 1650 Ti. The operating system is Windows 10, the programming environment is Python version 3.7.0, and the deep learning framework is TensorFlow-GPU version 2.7.1. The deep encoding-decoding architecture is shown in Tables 3 and 4.

Evaluation Metrics: The anomaly detection performance of DAD-IR-ASR is assessed by the Detection Rate (DR) and False Alarm Rate (FAR). quantifies the percentage of samples identified by the model among the total number of anomaly samples. FAR represents the probability of the robotic arm being detected as anomalous during normal operation. A lower False Alarm Rate signifies better model performance, minimizing frequent false alarms that could inconvenience operators.

$$\mathrm{DR} = \frac{100 \times \mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{10}$$

$$\mathrm{FAR} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{11}$$

where TP (True Positives) denotes the count of samples correctly identified as anomalies, TN (True Negatives) is the count of samples correctly classified as normal, FP (False Positives) is the count of samples incorrectly classified as anomalies when they are normal, and FN (False Negatives) is the count of samples incorrectly classified as normal when they are anomalies.

4.2. Comparison with baselines

Comparison methods: Eight baseline methods are compared including OCSVM [13], DCUSUM [25], CNN-AE [26], LSTM+CUSUM [27], CLF-AIAD [28], ADS-IR [29], RMC-CNN [30], and DOCSVM [31].

Comparison results: Figs. 3 and 4 show experiment outcomes regarding DR and FAR in the model testing, for the four configurations, i.e., the combination of malicious command injection anomaly and normal_2, the combination of manipulation of motion state data anomaly and normal_2, the combination of replay anomaly and normal_2, and the combination of three anomaly and normal_2. From experiment outcomes, it can be seen that DAD-IR-ASR outperforms other methods or is on par with the top-performing method across both settings, which proves effectiveness and superiority of DAD-IR-ASR. The seasons are twofold: (1) DAD-IR-ASR designs a deep encoding-decoding architecture, employing unsupervised reconstruction training solely with normal data to adaptively learn an error threshold, which effectively mitigates the issue of imbalance in the number and categories between normal and anomalous data. (2) DAD-IR-ASR conducts an accumulation-based anomaly detection strategy to determine if the intelligent robot is anomalous by comparing the cumulative sum of reconstruction errors from the deep autoencoder architecture with the threshold derived from normal data. In addition, there are some observations. (1) Compared to our approach, OCSVM exhibits a higher false alarm rate and lower detection rate. This can be attributed to the significant variations in the speed and acceleration data of the intelligent robot, which belong to highly nonlinear datasets. OCSVM struggles to effectively fit a hyperplane in such conditions, leading to suboptimal detection performance. (2) The detection rate of LSTM+CUSUM is comparable to DAD-IR-ASR, show-

Table 3. Statistical details of the encoder (FC: Fully Connected).

Layer Name	Layer Type	Number of Neurons	Number of Parameters
Layer1	FC	128	2432
Layer2	FC	64	8256
Layer3	FC	32	2080
Embedded	FC	15	495

Table 4. Statistical details of the decoder (FC: Fully Connected).

Layer Name	Layer Type	Number of Neurons	Number of Parameters
Layer4	FC	32	512
Layer5	FC	64	2112
Layer6	FC	128	8320
Output	FC	-	-

ing effective detection performance. However, the LSTM+CUSUM method has a higher false alarm rate than DeepAE+CUSUM. Autoencoders detect anomalies by comparing the reconstruction of input data with the input itself, learning the most crucial information about the data. On the other hand, LSTM utilizes past data to predict the future trend, and compared to the uncertainty in predicting future data trends by LSTM, autoencoders generate smaller residuals when reconstructing the input data.

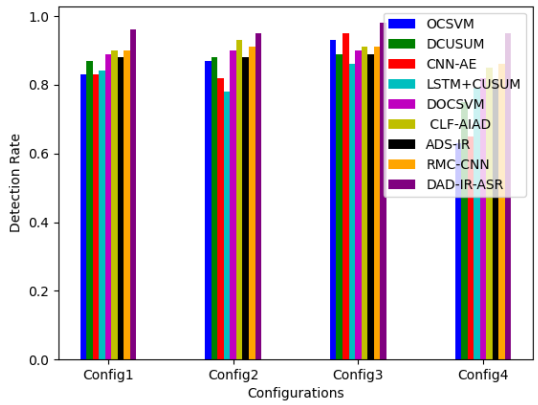

Fig. 3. The Detection Rate of DAD-IR-ASR and compared methods.

4.3. Parameter analysis

Feature dimension. In autoencoders, selecting the dimensionality of latent features is one of the most critical parameters. A too-small dimensionality may lead to severe information loss. In the case of undercomplete autoencoders, the dimensionality of latent features is smaller than the input data dimension. While

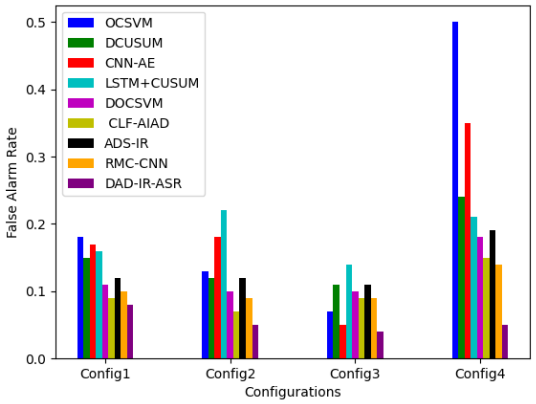

Fig. 4. The False Alarm Rate of DAD-IR-ASR and compared methods.

setting the model dimensionality to be greater than the input data dimension might yield better results, it could also render the autoencoder meaningless in terms of extracting essential features and removing noise. In this study, latent data feature dimensions were set to 3, 6, 9, 12, 15, 18, and 21. The evaluation of the average estimation error in input-output data was carried out to understand the effects of different dimensions on the detection performance of DAD-IR-ASR. Experimental results shown in Fig. 5 indicate that the average estimation error is minimized when the latent dimension is set to 15, suggesting optimal reconstruction performance at this dimension.

Layer number. Compared to traditional autoencoders, deep autoencoders have multiple hidden layers. The deep structure's layer-wise reduction facilitates better feature compression and extraction of more representative hidden features. In deep autoencoders with

multiple hidden layers, including input and output layers, the average estimation error in input-output data is generally lower. In comparison to simple autoencoders with fewer layers, deep autoencoders tend to exhibit lower estimation errors, as shown in Fig. 5. Therefore, deep autoencoders, with their superior capability for extracting data features in comparison to shallow and simple autoencoders, result in better overall detection performance.

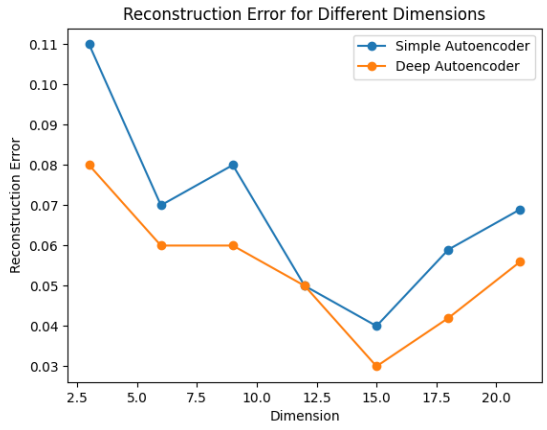

Fig. 5. Reconstruction error for different dimensions.

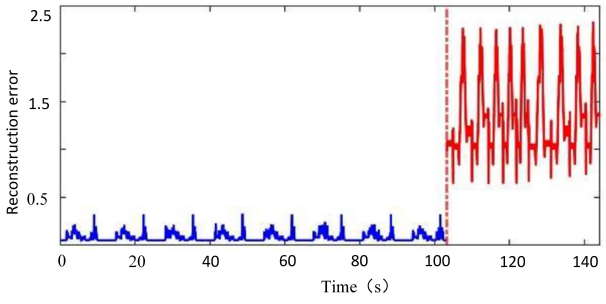

Fig. 6. Reconstruction error for different dimensions.

4.4. The effectiveness of reconstruction error

Fig. 6 illustrates a comparative chart of the model reconstruction error for robot data before and after the occurrence of anomalies. The blue curve represents the data curve during normal operation of the robot, while the red portion represents the data situation after the occurrence of anomalies. It can be observed that once an attack occurs, there is a significant change in the residual of the model reconstruction error. Therefore, the model is capable of promptly detecting anomalous situations.

5. Conclusions

This study proposed an innovative approach named Deep Acoustic Detection for Intelligent Robot Anomaly Status (DAD-IR-ASR). DAD-IR-ASR introduces an acoustic signal analysis within a deep encoding-decoding architecture, offering a comprehensive solution to the outlined issues. The methodology includes acoustic signal data preprocessing for intelligent robots, the implementation of a deep encoding-decoding architecture, and an accumulation-based anomaly detection strategy. Specifically, DAD-IR-ASR addresses the challenge of costly data acquisition by designing an acoustic sensor collection device that eliminates the need for additional signal conditioning equipment, ensuring a cost-effective acquisition of acoustic signals. The Fourier transformation and filtering techniques are employed to extract meaningful spectral features from the acoustic information. The deep encoding-decoding architecture utilizes unsupervised reconstruction training exclusively with normal data, adapting dynamically to varying error thresholds. This adaptation effectively mitigates the issue of imbalance in the quantity and categories between normal and anomalous data. Furthermore, the study introduces an accumulation-based anomaly detection strategy, which compares the cumulative sum of reconstruction errors from the deep autoencoder architecture with a threshold derived from normal data. This method enhances the sensitivity of anomaly detection in the presence of imbalanced datasets. The effectiveness of DAD-IR-ASR is demonstrated through comprehensive comparisons with multiple existing unsupervised detection methods in the field of intelligent robot anomaly detection. However, DAD-IR-ASR can only provide alerts without further information. In the future, it is advisable to consider automatically identifying the source of attacks and implementing proper recovery procedures for the system. Additionally, it would be meaningful to detect and prevent malicious code, especially commands that directly impact the system, from being downloaded and executed on the robotic arm before the attack is initiated.

Acknowledgements

This work was supported by “Horizontal project: Appearance design of intelligent cooperative robot SCR seri”.

References

[1] S. Yin, (2023) *“Object Detection Based on Deep Learning: A Brief Review”* **IJLAI Transactions on Science and Engineering** ***1***(02): 1–6.

[2] J. Gao, M. Liu, P. Li, J. N. Zhang, and Z. K. Chen, (2023) *"Deep Multiview Adaptive Clustering With Semantic Invariance"* **IEEE Trans. Neural Netw. Learn. Syst.** DOI: 10.1109/TNNLS.2023.3265699.

[3] P. Li, A. A. Laghari, M. Rashid, J. Gao, T. R. Gadekallu, A. R. Javed, and S. Yin, (2023) *"A Deep Multimodal Adversarial Cycle-Consistent Network for Smart Enterprise System"* **IEEE Transactions on Industrial Informatics *19***(1): 693–702. DOI: 10.1109/TII.2022.3197201.

[4] I. Kotenko, K. Izrailov, and M. Buinevich, (2022) *"Static analysis of information systems for IoT cyber security: a survey of machine learning approaches"* **Sensors *22***(4): 1335. DOI: 10.3390/s22041335.

[5] P. Li, J. Gao, J. N. Zhang, S. Jin, and Z. K. Chen, (2022) *"Deep Reinforcement Clustering"* **IEEE Trans. Multimedia**: DOI: 10.1109/TMM.2022.3233249.

[6] L. Teng, (2023) *"Brief Review of Medical Image Segmentation Based on Deep Learning"* **IJLAI Transactions on Science and Engineering *1***(02): 01–08.

[7] M. Mukhasheva, K. Ybyraimzhanov, K. Naubaeva, A. Mamekova, and B. Almukhambetova, (2023) *"The Impact of Educational Robotics on Cognitive Outcomes in Primary Students: A Meta-Analysis of Recent Studies."* **European Journal of Educational Research *12***(4): DOI: 10.12973/eu-jer.12.4.1683.

[8] A. Modliński, P. Fortuna, and B. Rożnowski, (2023) *"Human–machine trans roles conflict in the organization: How sensitive are customers to intelligent robots replacing the human workforce?"* **International Journal of Consumer Studies *47***(1): 100–117. DOI: 10.1111/ijcs.12811.

[9] H. Alemzadeh, D. Chen, X. Li, T. Kesavadas, Z. T. Kalbarczyk, and R. K. Iyer. "Targeted attacks on teleoperated surgical robots: Dynamic model-based detection and mitigation". In: *2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. IEEE. 2016, 395–406. DOI: 10.1109/DSN.2016.43.

[10] Y. Tang, D. Zhang, D. W. Ho, W. Yang, and B. Wang, (2018) *"Event-based tracking control of mobile robot with denial-of-service attacks"* **IEEE Transactions on Systems, Man, and Cybernetics: Systems *50***(9): 3300–3310. DOI: 10.1109/TSMC.2018.2875793.

[11] C. Zhou, B. Hu, Y. Shi, Y.-C. Tian, X. Li, and Y. Zhao, (2020) *"A unified architectural approach for cyberattack-resilient industrial control systems"* **Proceedings of the IEEE *109***(4): 517–541. DOI: 10.1109/JPROC.2020.3034595.

[12] E. Khalastchi, M. Kalech, G. A. Kaminka, and R. Lin, (2015) *"Online data-driven anomaly detection in autonomous robots"* **Knowledge and Information Systems *43***: 657–688. DOI: 10.1007/s10115-014-0754-y.

[13] F. Angiulli and C. Pizzuti. "Fast outlier detection in high dimensional spaces". In: *European conference on principles of data mining and knowledge discovery*. Springer. 2002, 15–27. DOI: 10.1007/3-540-45681-3_2.

[14] H. Xu, W. Chen, N. Zhao, Z. Li, J. Bu, Z. Li, Y. Liu, Y. Zhao, D. Pei, Y. Feng, et al. "Unsupervised anomaly detection via variational auto-encoder for seasonal kpis in web applications". In: *Proceedings of the 2018 world wide web conference*. 2018, 187–196. DOI: 10.1145/3178876.3185996.

[15] C. Zhou, S. Huang, N. Xiong, S.-H. Yang, H. Li, Y. Qin, and X. Li, (2015) *"Design and analysis of multimodel-based anomaly intrusion detection systems in industrial process automation"* **IEEE Transactions on Systems, Man, and Cybernetics: Systems *45***(10): 1345–1360. DOI: 10.1109/TSMC.2015.2415763.

[16] A. A. Jaber and R. Bicker, (2016) *"Industrial robot fault detection based on statistical control chart"* **Am. J. Eng. Applied Sci *9***: 251–263.

[17] V. Sathish, S. Ramaswamy, and S. Butail, (2016) *"Training data selection criteria for detecting failures in industrial robots"* **IFAC-PapersOnLine *49***(1): 385–390. DOI: 10.1016/j.ifacol.2016.03.084.

[18] Y. Zhao and C. Smidts, (2020) *"A control-theoretic approach to detecting and distinguishing replay attacks from other anomalies in nuclear power plants"* **Progress in Nuclear Energy *123***: 103315. DOI: 10.1016/j.pnucene.2020.103315.

[19] W. Zhu, W. Li, H. Liao, and J. Luo, (2021) *"Temperature network for few-shot learning with distribution-aware large-margin metric"* **Pattern Recognition *112***: 107797. DOI: 10.1016/j.patcog.2020.107797.

[20] Y. Zhou, L. Xie, and H. Pan, (2022) *"Research on a PSO-H-SVM-Based Intrusion Detection Method for Industrial Robotic Arms"* **Applied Sciences *12***(6): 2765. DOI: 10.3390/app12062765.

[21] J. Inoue, Y. Yamagata, Y. Chen, C. M. Poskitt, and J. Sun. "Anomaly detection for a water treatment system using unsupervised machine learning". In: *2017 IEEE international conference on data mining workshops (ICDMW)*. IEEE. 2017, 1058–1065. DOI: 10.1109/ICDMW.2017.149.

[22] Ö. Gültekin, E. Cinar, K. Özkan, and A. Yazıcı, (2022) *"Real-time fault detection and condition monitoring for industrial autonomous transfer vehicles utilizing edge artificial intelligence"* **Sensors 22**(9): 3208. DOI: 10.3390/s22093208.

[23] A. Munawar, P. Vinayavekhin, and G. De Magistris. "Spatio-temporal anomaly detection for industrial robots through prediction in unsupervised feature space". In: *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE. 2017, 1017–1025. DOI: 10.1109/WACV.2017.118.

[24] M. Said Elsayed, N.-A. Le-Khac, S. Dev, and A. D. Jurcut. "Network anomaly detection using LSTM based autoencoder". In: *Proceedings of the 16th ACM Symposium on QoS and Security for Wireless and Mobile Networks*. 2020, 37–45. DOI: 10.1145/3416013.3426457.

[25] L. Mhamdi, D. McLernon, F. El-Moussa, S. A. R. Zaidi, M. Ghogho, and T. Tang. "A deep learning approach combining autoencoder with one-class SVM for DDoS attack detection in SDNs". In: *2020 IEEE Eighth International Conference on Communications and Networking (ComNet)*. IEEE. 2020, 1–6. DOI: 10.1109/ComNet47917.2020.9306073.

[26] A. Binbusayyis and T. Vaiyapuri, (2021) *"Unsupervised deep learning approach for network intrusion detection combining convolutional autoencoder and one-class SVM"* **Applied Intelligence *51***(10): 7094–7108. DOI: 10.1007/s10489-021-02205-9.

[27] E. Mushtaq, A. Zameer, M. Umer, and A. A. Abbasi, (2022) *"A two-stage intrusion detection system with autoencoder and LSTMs"* **Applied Soft Computing *121***: 108768. DOI: 10.1016/j.asoc.2022.108768.

[28] K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Soderstrom. "Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding". In: *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 2018, 387–395. DOI: 10.1145/3219819.3219845.

[29] C. Yukawa, K. Toyoshima, Y. Nagai, M. Niihara, Y. Yamashita, T. Oda, and L. Barolli. "An Anomaly Detection System for Intelligent Robot Vision Using LSTM". In: *International Conference on Advanced Information Networking and Applications*. Springer. 2023, 192–198. DOI: 10.1007/978-3-031-28451-9_17.

[30] Z. Liu, Y. Hou, H. Tang, Á. López-Chilet, S. Michiels, D. Botteldooren, J. A. Gómez, and D. Hughes. "CLF-AIAD: A Contrastive Learning Framework for Acoustic Industrial Anomaly Detection". In: *International Conference on Neural Information Processing*. 2023, 125–137. DOI: 10.1007/978-981-99-8126-7_10.

[31] K. S. Sankaran and B.-H. Kim, (2023) *"Deep learning based energy efficient optimal RMC-CNN model for secured data transmission and anomaly detection in industrial IOT"* **Sustainable Energy Technologies and Assessments *56***: 102983. DOI: 10.1016/j.seta.2022.102983.