An Energy Consumption Auditing Anomaly Detection System of Robotic Manipulators based on a Generative Adversarial Network

Ge Song[1], Seong Hyeon Hong[2], Tristan Kyzer[1], and Yi Wang[1]

[1]*University of South Carolina, Columbia, SC, 29208, USA*

gsong@email.sc.edu
Tristan.Kyzer@integer-tech.com
yiwang@cec.sc.edu

[2]*Florida Institute of Technology, Melbourne, FL, 32901, USA*

shhong@fit.edu

ABSTRACT

Unexpected anomalies pose significant risks to the health and security of intelligent manufacturing systems. This paper proposes a generative adversarial network (GAN)-based anomaly detection framework specifically for monitoring robotic manipulator operation using a side-channel energy auditing mechanism. To tackle the limitation arising from the lack of labeled data, the GAN model is trained by a semi-supervised learning approach that identifies anomalies during online operations as outliers. The overfitting is purposely utilized during the model training to enlarge the difference between normal energy consumption patterns used for training and anomalous profiles in real-time testing. In addition, the GAN model is modified to use multiple discriminators to analyze the individual energy profile associated with each joint or motor. The anomaly is detected by evaluating the mean and standard deviation values of anomaly scores' distribution, and both values are continuously updated by Welford's algorithm in real time to take into account the effect of environmental variations during operations. The detection performance on our custom dataset demonstrates the feasibility of the proposed pipeline. Specifically, for physical attacks, the framework can achieve an accuracy of approximately 0.93 for instant-wise detection and 0.84 for event-wise detection.

1. INTRODUCTION

Robotic manipulators have become a popular tool in modern manufacturing and production processes. They are used extensively in various industries, such as automotive, aerospace, electronics, among others, to perform repetitive and complex tasks with high precision and speed. However, one of the concomitant challenges is the around-the-clock and reliable monitoring of their security and health due to their vulnerability to attacks. The built-in sensors of robotic manipulators can also be the target of attacks and hence generate faked data, reducing the trustworthiness of the monitoring system. One feasible solution is implementing a side-channel energy consumption auditing mechanism that directly interfaces with the robotic manipulator to avoid network communication. Thus, it is only reflective of the energy consumption signal under surveillance. Therefore, it is necessary to develop a system that can accurately monitor and detect anomalies through energy consumption patterns.

Recently, deep learning algorithms have shown great promise in detecting anomalies in various applications, such as abnormal pedestrian behaviors detection (Jiang, Song, Qian, & Wang, 2022) and rocket engines (Yan, Liu, Chen, Feng, & Wang, 2023). However, due to the significant variations and uncertainty in attack patterns, it is challenging to create comprehensive labeled datasets containing all potential anomalies. As one solution, the semi-supervised learning approach is frequently applied in anomaly detection tasks (Memarzadeh, Matthews, Templin, Sharif Rohani, & Weckler, 2023), where only patterns for the normal class are employed for model training. Anomalies can, hence, be identified as outliers. The model by Nguyen et al. (Nguyen, Hum, Do, & Tran, 2023) was developed in such a manner to detect anomalies that occurred in laser powder bed fusion products. The model can classify surface appearances in the reference monitoring data and correlate them to post-process characteristics to assess the quality of printed samples.

On the other hand, the semi-supervised method requires models to have a salient pattern-learning ability to digest normal patterns fully. The generative adversarial network (GAN), which was developed by Goodfellow et al. (Goodfellow, et al., 2020), achieves a good balance between performance and inference speed. It exploits a second neural network, called the discriminator, to perform adversarial learning along with the main neural network, named the generator. Starting from the work of Sabokrou et al. (Sabokrou, Khalooei, Fathy, & Adeli, 2018), GAN becomes a prevalently used technique of anomaly detection. One example is the research from Contreras-Cruz et al. (Contreras-Cruz, Correa-Tome, Lopez-Padilla, & Ramirez-Paredes, 2023), where a GAN model was applied to perform abnormal region detection tasks on satellite or aerial photographs. The MTAD-GAN (Multivariate Time Series Data Anomaly Detection with GAN) proposed by Lian et al. (Lian, Geng, & Tian, 2023) is applied to oil and gas stations. The station-operating logic of behavior is described by a stochastic Petri net. The MTAD-GAN is utilized to reconstruct multivariate time series by combining knowledge graph attention and temporal attention.

In this paper, we propose an energy consumption auditing and anomaly detection system for robotic manipulators based on a generative adversarial network. The GAN is used to learn the normal energy consumption patterns and then detect anomalies as outliers. The criterion for detection utilizes the mean and standard deviation values calculated from the anomaly scores' distribution of the normal energy consumption, and the values are updated by Welford's algorithm in real-time to accommodate the effects of environmental changes. In addition, considering the unique physical aspects of the robotic manipulator, a novel GAN with multiple discriminators is also developed. The system is evaluated by a custom dataset, and the results demonstrate salient anomaly detection performance.

The remainder of this paper is organized as follows. The GAN-based anomaly detection system is described in detail in Section 2. Section 3 presents a specific manufacturing task using a robot manipulator to simulate threats likely occurring during industrial operations. Experimental results and discussions are given in Section 4, and the paper is concluded in Section 5.

2. METHODOLOGY

As presented in Figure 1, the proposed energy consumption monitoring system is composed of two main stages:

1. Offline model development. In this stage, the robotic manipulator is only allowed to perform the predefined task normally. The corresponding energy consumption is the sole input to the GAN model. To improve model generalization, simulated noise following the normal distribution is added to the raw input. The key component, the GAN model, learns to reconstruct normal input patterns with low errors and assign low anomaly scores.
2. Online anomaly detection. The three components, data preprocessing, GAN model, and anomaly detector, run sequentially to process the input at each time instant. The well-trained GAN model is supposed to reconstruct inputs with low errors and produce low anomaly scores when inputs represent normal operations while yielding large reconstruction errors and high anomaly scores for anomalous inputs. The detector adopts dynamic thresholding on anomaly scores to detect outliers or anomalies. The input is treated as an anomaly if its associated anomaly score does not follow the normal distribution, i.e., it deviates from the mean by three standard deviations whose values are updated in real-time using Welford's algorithm to mitigate the influence of the environmental changes.

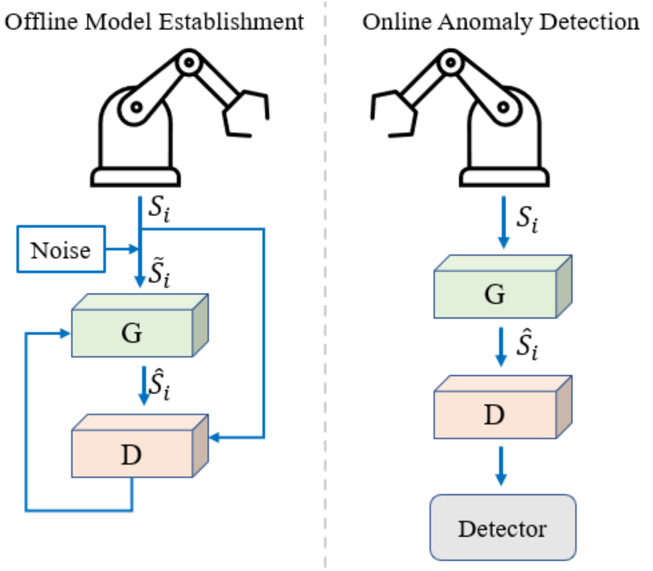


Figure 1. The pipeline of the anomaly detection system

2.1. Data Preprocessing

For a robotic manipulator consisting of N_j joints or motors, at each time instant, the energy consumption has the form:

$$e_t = \left\{e_t^1, \cdots, e_t^{N_j}\right\} \tag{1}$$

Given that one operation starts at the first time instant and ends at the N^{th} instant, the associated energy consumption can be described by a time-dependent data sequence $\{e_1, \cdots, e_N\}$. To enrich representations and features for learning, a sliding window technique with a fixed window size T is applied to split the entire sequence into several overlapped segments. Therefore, the input segment at time i, S_i, is $\{e_{i-T+1}, \cdots, e_i\}$. Using such an input segment rather than a single data instance can include more statistical feature representations and, hence, improve the detection performance.

2.2. Generative Adversarial Network

The GAN consists of two subnetworks, a generator and a discriminator. The two subnetworks perform adversarial learning to boost the performance of both networks.

The autoencoder model of the generator is shown in Figure 2, which is developed based on the Long-Short Term Memory (LSTM) structure. It consists of an encoder (colored in blue) and a decoder (colored in yellow). The former module maps input segments to the features in the latent space, while the latter module reconstructs inputs using the corresponding features from the latent space. Each node represents an LSTM cell. As aforementioned, the input to the generator is the energy consumption data in each joint of the robotic manipulator. Given the sliding window size T and joints number N_j, the number of input data points at each time instant is $N_j \times T$. The well-trained generator is supposed to produce low errors for reconstructing normal samples while distinctly large errors for abnormal patterns.

As presented in Figure 3, the discriminator is a multilayer perceptron (MLP) structure. It contains two sequential fully connected layers with a different number of neurons followed by a sigmoid activation function. During the offline model construction, this subnetwork takes raw input S_i and the associated reconstructed samples $\hat{S}_i$, respectively, in two separate runs. The discriminator intends to distinguish them by enlarging the difference between their outputs, called the anomaly score. That is, for raw inputs, the discriminator aims to generate scalar values close to zero. On the other hand, scalar values close to one will be produced for reconstructed samples. Conversely, the generator receives outputs from the discriminator and endeavors to minimize the difference. Such adversarial learning can improve both models' performance. Thus, the anomaly scores in the online anomaly detection step will be close to zero if inputs represent normal operations following training dataset distribution, while it yields large anomaly scores for inputs containing anomalous patterns. Therefore, detecting anomalies can be achieved solely by interrogating the anomaly score.

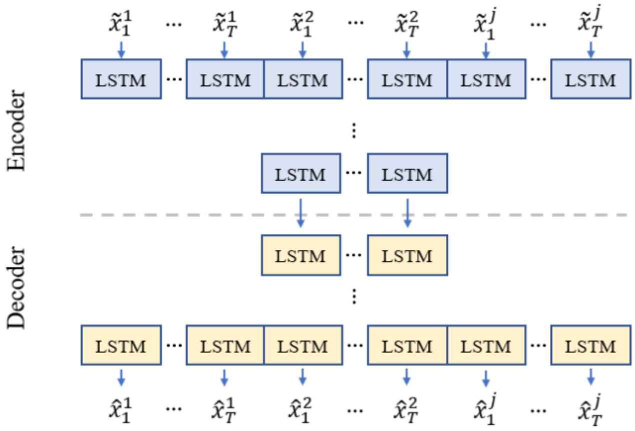

$\tilde{x}_1^1$ ··· $\tilde{x}_T^1$ $\tilde{x}_1^2$ ··· $\tilde{x}_T^2$ $\tilde{x}_1^j$ ··· $\tilde{x}_T^j$
LSTM
Encoder
Decoder
$\hat{x}_1^1$ ··· $\hat{x}_T^1$ $\hat{x}_1^2$ ··· $\hat{x}_T^2$ $\hat{x}_1^j$ ··· $\hat{x}_T^j$

Figure 2. The model architecture of the generator

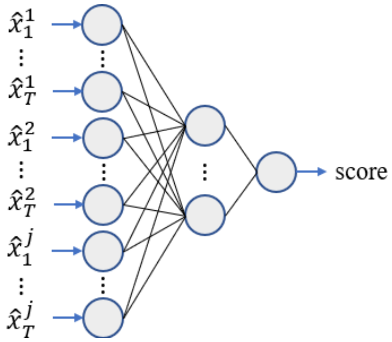


Figure 3. The model architecture of the discriminator

To train the discriminator, the following binary cross-entropy loss is used, which needs to be maximized:

$$l_D = \log\left(1 - D(\hat{S}_i)\right) + \log(D(S_i)) \tag{2}$$

where $G(\cdot)$ and $D(\cdot)$ denote the operation by the generator and the discriminator, respectively. Again, the original input segment is S_i, and $\hat{S}_i$ denotes the reconstructed samples. For the generator, the following 2-norm of the error between the raw inputs and reconstructed samples is employed as part of the loss function to evaluate its reconstruction performance:

$$l_{rec} = \|S_i - \hat{S}_i\|_2 \tag{3}$$

For adversarial learning, the final loss function to be minimized by the generator is the combination of l_{rec} and l_D

$$l_G = l_{rec} + \max\ l_D \tag{4}$$

The model training process purposely exploits overfitting to further extend the gap between normal and abnormal patterns. However, the overfitted model also has the possibility of rejecting some normal inputs and, hence, compromises the accuracy. To tackle this limitation and increase the model's generalization, simulated noise following the normal distribution is added to the raw input segment, which is given as follows

$$\tilde{S}_i = S_i + \left(\eta \sim N(0, \sigma^2)\right) \tag{5}$$

As a result, the generator receives the segments of noisy energy consumption ($\tilde{S}_i$) and then reconstructs it ($\hat{S}_i$),

$$\hat{S}_i = G(\tilde{S}_i) \tag{6}$$

2.3. Generative Adversarial Network with Multiple Discriminators

To take into account the physical aspects of the robotic manipulator for accuracy improvement, a variant of the GAN model above is developed by incorporating multiple discriminators. This model modification is based on the fact that some attacks could lead to more dramatic impacts on certain joints. One example is the physical attack, which imposes more influences to the end effector motor of the

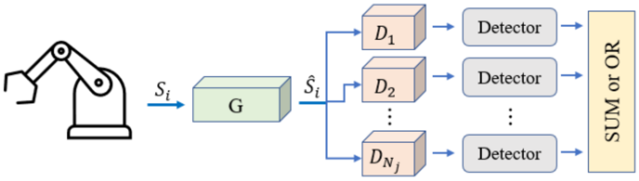


Figure 4. The pipeline of the GAN with multiple discriminators

robotic manipulator and its associated energy consumption. In addition, if anomaly detection relies on the combined energy consumption of all motors, the differences in energy between motors can be reduced, resulting in inaccurate or incomplete detection of outlier samples.

Figure 4 demonstrates the pipeline of the GAN with multiple discriminators. Each manipulator motor is monitored by an individual discriminator. Therefore, in this case, the number of discriminators is identical to the number of joints of the robotic manipulator. All discriminators share the same structure and size as presented in Figure **3**. The reconstructed samples are split into several subsets, each for one corresponding discriminator. Different operations are applied to process outputs from discriminators. During the model training, the SUM function is used to drive all discriminators to converge to the desired state. While in online testing, discriminators' outputs are processed by the OR function, so that even an anomaly occurring in one motor can trigger a positive anomaly detection.

2.4. Detection Criterion Based on Welford's Algorithm

During robotic manipulator operations, minor variations in the manipulator's states and environment can cause fluctuations in energy consumption. These changes can be caused by factors such as motor heat generation, operational time, and room temperature, resulting in shifts in the distribution of normal operation patterns. To address this issue, a dynamic thresholding approach is used in this work instead of a static threshold on discriminator outputs. Anomalies are detected if the associated anomaly scores y deviate from the mean of the distribution by three standard deviations. The mean and standard deviation values are initialized from the training dataset and updated during online testing. Only data samples classified as normal are utilized and anomalies are excluded from the update process. To be more specific, in this study, Welford's algorithm (Welford, 1962) is utilized to update distribution parameters by employing the following equations:

$$\mu_n = \mu_{n-1} + \frac{y_n - \mu_{n-1}}{n} \tag{7}$$

$$M_n = M_{n-1} + (y_n - \mu_{n-1})(y_n - \mu_n) \tag{8}$$

$$\sigma_n^2 = \frac{M_n}{n} \tag{9}$$

where the mean and standard deviation are denoted by μ and σ, respectively. The anomaly score y_n corresponds to the input at the n^{th} time instant. To eliminate the impact of the noise during testing, a moving average operation is utilized to filter out high-frequency oscillations and noise in the original output.

3. EXPERIMENT

This section contains information on the experiments performed to evaluate the proposed framework, which covers the robotic manipulator, operational tasks, and simulated attacks implemented in the system to replicate potential anomalies in an industrial setting.

3.1. Robotic Manipulator

For this study, the Lynxmotion robot arm with four degrees of freedom, as depicted in Figure 5, was chosen as the main component of the manufacturing system. The robot arm is composed of five Lynxmotion smart servo motors, also considered as five joints. Each motor has an adapter board attached performing as the communication module. The normal operating speed range is between one to four revolutions per minute (RPM) considering safe operations. The side-channel sensors, which are independent of the task control signal, are applied to each motor to monitor and record the associated states, such as position, speed, current, and voltage. However, only the current and voltage values are used to calculate energy consumption. A separate computer is wirelessly connected to the robotic manipulator to control operations by sending commanded control signals. In addition, a Raspberry Pi 3 Model B+ is used to monitor and record the manipulator's state and input signal.

3.2. Tasks

In general, the robotic manipulator in industrial scenes needs to perform repetitive actions, such as moving to the target configuration, extending the end effector, and gripping items. Therefore, in order to simulate such scenarios, the robot arm is commanded to execute the same task at different locations, where attacks can be injected. As shown in Figure 6, there are six small areas in a circle numbered from 1 to 6. The robotic manipulator initially starts at Position 1 followed by moving

Figure 5. Lynxmotion robot arm

in a counterclockwise direction until arriving at Position 6, and then starts a new circle. At each position, the manipulator is commanded to perform the activities discussed above, that is, extending the end effector to the center of each small area, performing gripping action, and returning to the original posture. In addition, to augment the dataset and enhance the model's generalization, random movements, which produce multiple variations in trajectories for accomplishing the same task, are inserted between each position. As a result, despite identical task execution, the associated energy consumption still has slight variations.

3.3. Simulated Attacks

This study considers two types of attacks that have the highest possibility of occurrence in real-world industrial scenes and pose security concerns. The cyber-attack (abbreviated as CA) involves a scenario, where configurations of the manufacturing system are altered remotely. The undesired forces affecting the system are classified as physical attacks (abbreviated as PA). The above two anomalies are intended to be unnoticeable by the networked sensors during task execution or in the manufactured product. For easy implementation, the speed range of each motor is changed to either [1,12] or [1,24] RPM. Consequently, increasing the speed range of the motor may cause it to reach the desired states earlier than the scheduled time. To impose more challenges on cyber-attack detection, the duration of each task under cyber-attack is forced to remain the same as the normal operation by commanding the motor to stay at the desired state following the schedule, even if it completes the tasks earlier. The robotic manipulator is supposed to carry no objects during normal operations. Therefore, to mimic the unexpected physical forces, a PVC pipe segment is placed at a random position and gripped by the robotic manipulator. The weight of the pipe is either 33 grams or 250 grams to simulate the physical attack at different intensity levels. For simplicity, cyber attacks and physical attacks are added to two separate experiments, respectively. The first experiment, which lasts for approximately 30 minutes, has cyber-attacks that are temporally randomly injected into the system with a duration of 10 or 20 seconds. The second experiment has the same duration as the first one where physical attacks are also randomly added and last for 10 seconds.

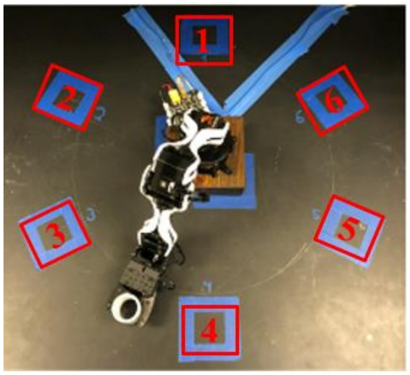


Figure 6. Tasks performed by the robotic manipulator

4. RESULTS AND DISCUSSION

As discussed above, the proposed framework is trained using the data representing normal operations only. Our custom training dataset contains over 6,000 data instances collected at 3 HZ for a total of 35 minutes. The size of the sliding window equals 3, which means that at each time, the input energy consumption segment contains the current patterns plus two consecutive previous values for each motor. As a result, the number of input data instances to the anomaly detection is 15 for each model run (recall that the manipulator has 5 motors/joints), which essentially means that each data entry to our model captures the operation for one second. Both the encoder and decoder of the generator have one hidden layer which was found to achieve a great balance between detection accuracy and inference speed. Each layer has the rectified linear unit (ReLU) as the activation function given non-negative energy consumption. The discriminator that follows the MLP structure has 15 and 8 neurons for the two consecutive fully connected layers, respectively. The noise following a normal distribution has a mean of 0 and a standard deviation of 0.05, which is found to yield the best performance. Given that MinMaxScalar is applied to preprocess the energy consumption patterns, the distribution can produce noise around 10% of the input maximum amplitude. The training dataset is split into two subsets for training and validation with a ratio of 80% and 20%, respectively. The proposed framework is coded in PyTorch and is trained using the Adam optimizer with 1000 epochs. The model training and evaluation are completed by a GPU workstation with Intel® Core™ i9-9820X CPU @ 3.3GHZ and NVIDIA GeForce RTX 2080Ti GPU.

4.1. Evaluation Metric

For instance-wise detection tasks, there are three frequently used evaluation metrics to describe the model's performance, accuracy, precision, and recall. The numerical equations are given as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \tag{10}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{11}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{12}$$

where values of true positive, true negative, false positive, and false negative are denoted by the symbol TP, TN, FP, and FN, respectively. Accuracy measures the proportion of correct detections made on all samples, including both normal and abnormal ones. Accurately detected abnormal samples out of all the samples detected as positive is descri-

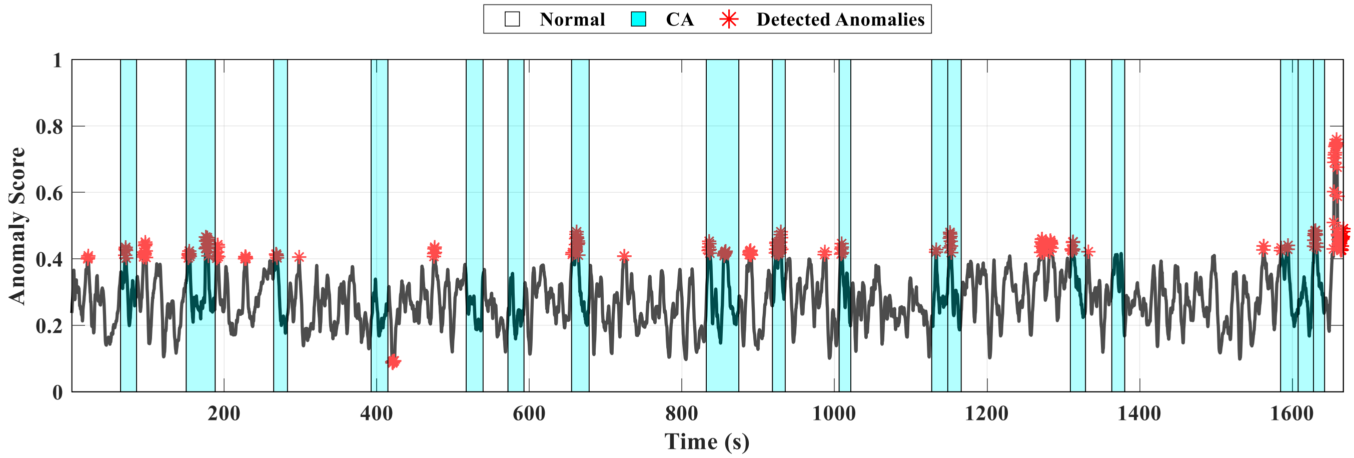


Figure 7. Cyber-attack detection result using GAN with 1 discriminator

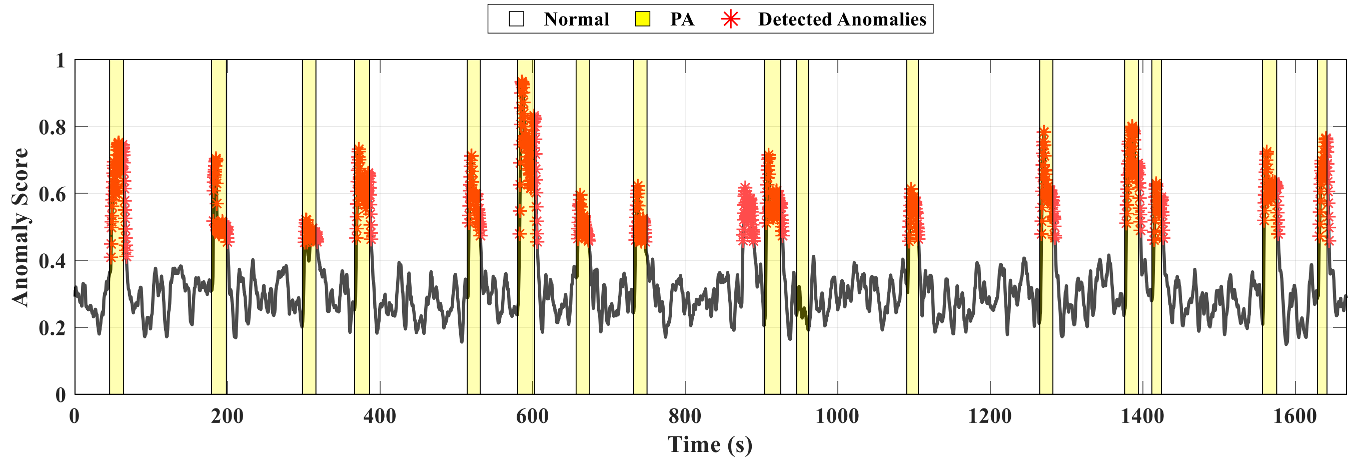


Figure 8. Physical-attack detection result using GAN with 1 discriminator

bed by the precision. Recall, on the other hand, is the ratio of detected anomalies to all the abnormal samples present in the ground truth. In general, higher values in these three evaluation metrics mean better detection performance.

4.2. Experimental Results

Figure 7 illustrates the detection result of cyber attack using the proposed GAN-based framework with a single discriminator. Values in the *x*-axis and *y*-axis stand for the time and anomaly scores produced by the discriminator. The white background represents the operations with normal settings. On the other hand, periods with the background colored in cyan are operations under cyber-attacks. Detected anomalous instances, via Welford's algorithm-based dynamic thresholding, are marked as red asterisks. It is observed that most cyber-attack events are detected by identifying some abnormal samples within each attack period. The attack occurred between 1350^{th} and 1400^{th} instants and the two attacks between 500^{th} and 600^{th} are detected as false negative samples. A few false positives can also be observed during normal operations, which contain only one incorrectly identified data instance in each normal operation period. However, given that a cyber-attack event contains several anomalous time instants, it is almost impossible to capture each of them considering variations in energy consumption and fluctuations in anomaly scores and, hence, cause high values in false negative. The associated confusion matrix is given as follows:

Table 1. The confusion matrix of cyber-attack detection using GAN with 1 discriminator

		Detected Label	
		Positive	Negative
Ground Truth	Positive	197	925
	Negative	66	3813

Similarly, the physical attack detection result is shown in Figure 8. Again, the *x*-axis demonstrates the time of the

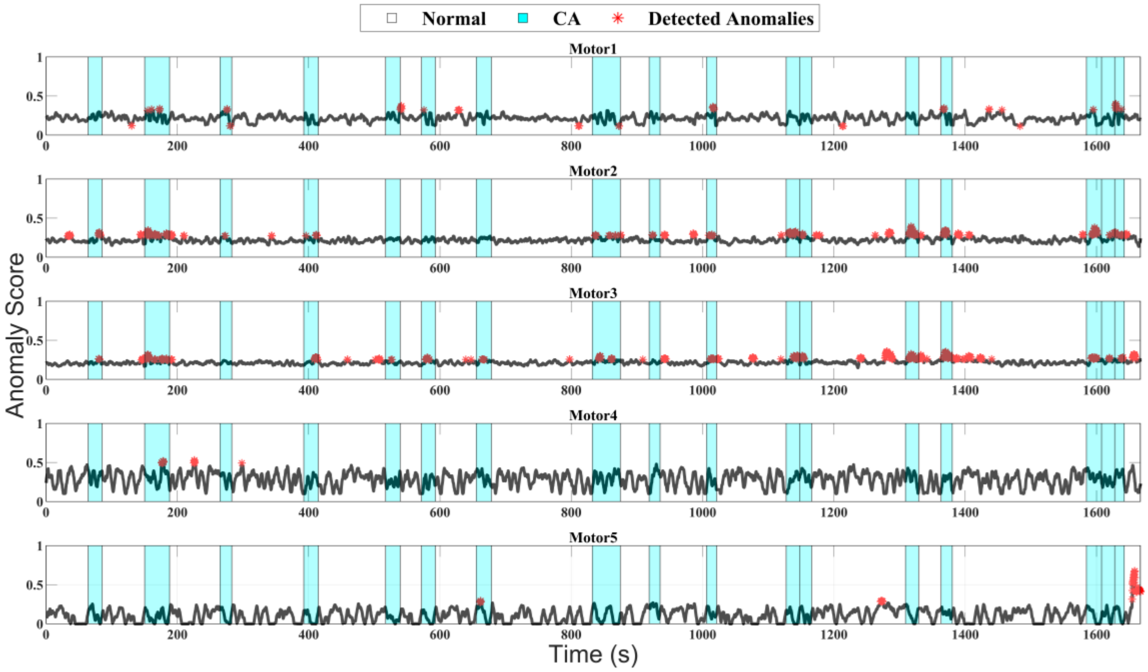


Figure 9. Cyber-attack detection result using GAN with 5 discriminators

experiment and the y-axis is for the anomaly score. The background colors (white and yellow) represent the event status (performing as the ground truth) of the robotic manipulator when performing tasks normally or under physical attacks, respectively. Same as in Figure 7, the abnormal samples found during instance-level detection are denoted as red asterisks. The associated confusion matrix is presented in Table 2. The most obvious misdetection lies in the false positive samples around the 900^{th} second yielding more false positives for physical-attack detection compared to cyber-attack. On the other hand, most data points within the period under physical attacks remain at a high level, so that most of them can be captured leading to a high true positive score.

Table 2. The confusion matrix of physical-attack detection using GAN with 1 discriminator

		Detected Label	
		Positive	Negative
Ground Truth	Positive	665	180
	Negative	136	4019

Figure 9 and Figure 10 present detection results using GAN with 5 discriminators for cyber-attack and physical attack, respectively. The background colors in the figures still share the same representation with Figure 7 and Figure 8, where periods of normal operations without attacks are denoted by the white color and periods of operations under cyber and physical attacks are colored in cyan and yellow, respectively. The index of motors is assigned in ascending order from the bottom to the top of the robotic manipulator. In Figure 9, all 17 attack events have at least one time instance that is detected as abnormal, which exceeds the GAN with only one discriminator. However, the issue of the false positive in Figure 7 is still present in most normal operation periods, and there exist some instances that are incorrectly identified as anomalies. As for the physical attack in Figure 10, a salient performance in distinguishing abnormal patterns from normal ones can be observed simply by the 5^{th} motor where noticeable peaks in anomaly scores indicate the presence of anomalies. On the other hand, the GAN framework of multi-discriminators still fails to detect the attack occurring around the 950^{th} second and produces false positive detections around the 900^{th} second. The associated confusion matrices are given in Table 3 and Table 4. Compared to detection results using GAN with a single discriminator, using more discriminators results in higher true positives. Nevertheless, the instances of false positives in both attacks also increase proportionally.

Table 3. The confusion matrix of cyber-attack detection using GAN with 5 discriminators

		Detected Label	
		Positive	Negative
Ground Truth	Positive	266	856
	Negative	153	3725

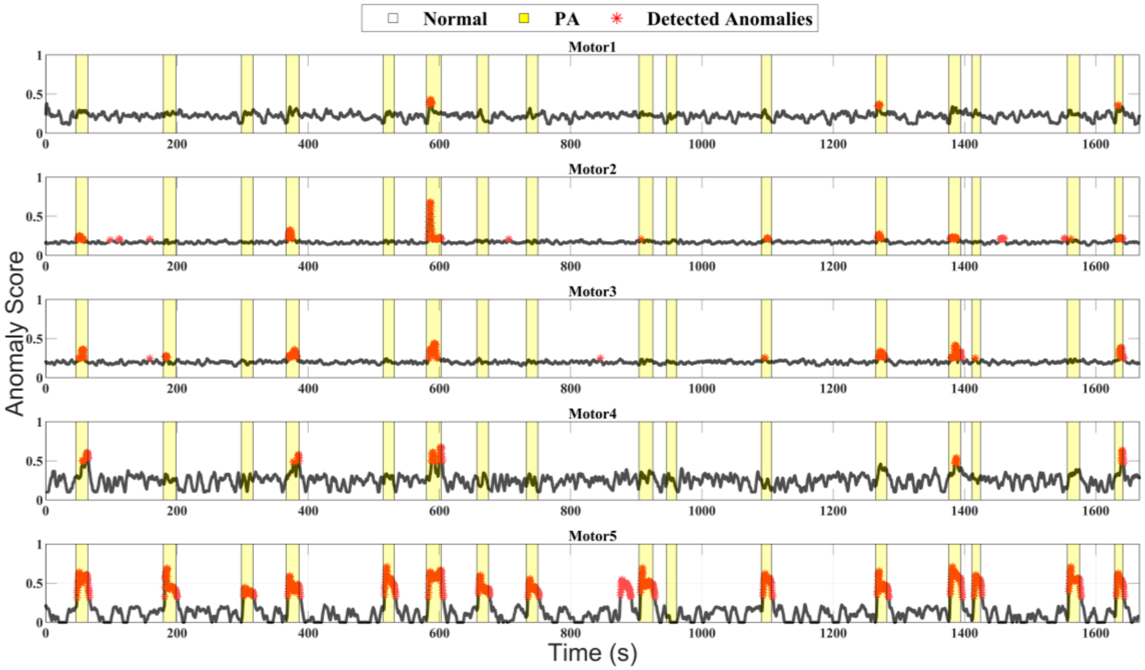


Figure 10. Physical-attack detection result using GAN with 5 discriminators

Table 4. The confusion matrix of physical-attack detection using GAN with 5 discriminators

		Detected Label	
		Positive	Negative
Ground Truth	Positive	687	158
	Negative	167	3988

Table 5. Anomaly detection results

Attack	Model	Evaluation Metric		
		Accuracy	Precision	Recall
Cyber	GAN_1D	0.802	0.749	0.176
	GAN_5D	0.798	0.635	0.237
Physical	GAN_1D	0.937	0.83	0.787
	GAN_5D	0.935	0.804	0.813

The values of the three evaluation metrics discussed above, accuracy, precision, and recall of both frameworks are illustrated in Table 5. The value of the model with better performance in each metric is highlighted in red color. For cyber-attack detection, both models can achieve an accuracy of approximately 0.8. The GAN with 5 discriminators achieves a higher value in the recall (0.2371) compared to the model with only one discriminator (0.176). However, more false positive samples also cause a lower precision value of the GAN with multiple discriminators. On the other hand, excellent performance in physical attack detection can be achieved by both models, where the accuracies are over 0.93. The corresponding precision values also maintain at a high level (more than 0.8 for both models). Similarly, high scores (approximately 0.8) are also in the recall evaluation metric.

4.3. Analysis

From Table **5**, the limitation of the proposed framework lies in the low recall score of cyber-attack detection. A statistical analysis of the energy consumption dataset is conducted to reveal the underlying reason causing this limitation. The distributions of combined energy consumption of all motors under normal operations vs. under cyber-attack and physical attack are shown in **Error! Reference source not found.** and **Error! Reference source not found.**, respectively. The value in *x*-axis in both figures represents bins of energy consumption, and *y*-axis displays the percentage of data samples that lie in the specific bin of energy consumption relative to the total number of samples. The histogram in black color represents the normal operation without any attacks, and the one in red color stands for cyber-attack in **Error! Reference source not found.** and physical attack in **Error! Reference source not found.**. An evident separation between the distribution of normal operations and physical attacks can be observed, leading to the prominent performance in physical attack detection. However, such a difference does not occur in the case of cyber-attacks, which explains the low score in the recall metric.

In the conducted experiment, the built-in algorithm and module of the robotic manipulator for motion planning is utilized to compute its trajectories to reach the desired task configurations. As a result, it may only slightly adjust the trajectory of each joint in response to the changes in the speed range under cyber attacks, leading to the hardly noticeable variation in energy consumption. Therefore, to improve cyber-attack detection performance, additional features from the side-channel mechanism need to be incorporated, such as measurement of the torque of each joint, which will be pursued in future work.

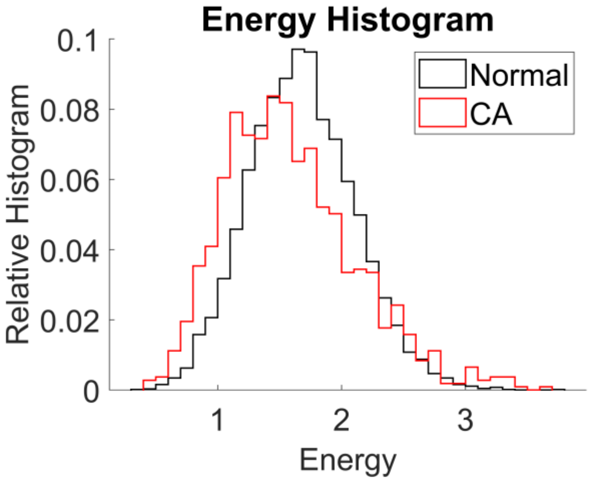


Figure 11. Histogram of energy consumption under normal operations vs. cyber-attack

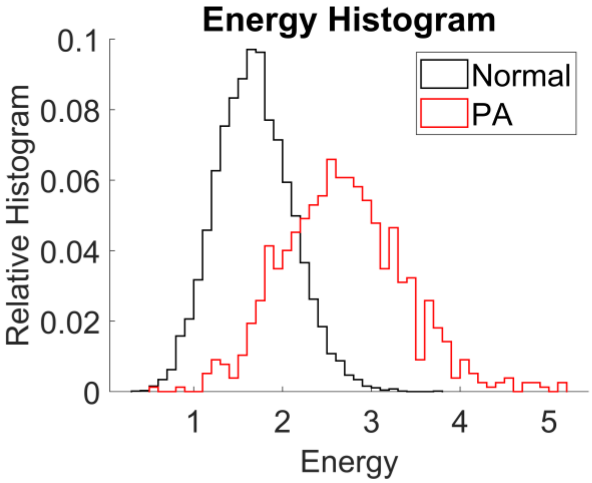


Figure 12. Histogram of energy consumption under normal operations vs. physical attack

5. CONCLUSION

In this paper, a side-channel anomaly detection system based on energy consumption auditing and a generative adversarial network (GAN) is proposed for robotic manipulators. The system uses the GAN to capture essential patterns of normal operations without attacks and, hence, detects anomalies as outliers of the model. The proposed framework is evaluated by a custom dataset, which involves temporal energy consumption profiles under normal operations and simulated cyber and physical attacks. The anomaly detection system is able to achieve high detection accuracy, precision, and recall in the case of physical attacks. However, in cyber-attack cases, the model can only detect a few anomalous instances because of the built-in trajectory planning algorithm that could potentially mitigate the changes in energy consumption due to attacks.

Future work will focus on enhancing both experimentation and detection methods to overcome the limitations in cyber-attack detection discussed above. For example, incorporating additional information from other components in the side-channel monitoring mechanism, e.g., the torque of each joint. It is also worth a trial to apply supervised learning-based approaches to enable simultaneous detection of both cyber and physical attacks.

References

Contreras-Cruz, M. A., Correa-Tome, F. E., Lopez-Padilla, R., & Ramirez-Paredes, J.-P. (2023). Generative Adversarial Networks for anomaly detection in aerial images. *Computers and Electrical Engineering, 106*, 108470.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., . . . Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM, 63*(11), 139-144.

Jiang, Z., Song, G., Qian, Y., & Wang, Y. (2022). A deep learning framework for detecting and localizing abnormal pedestrian behaviors at grade crossings. *Neural Computing and Applications*, 1-15.

Lian, Y., Geng, Y., & Tian, T. (2023). Anomaly Detection Method for Multivariate Time Series Data of Oil and Gas Stations Based on Digital Twin and MTAD-GAN. *Applied Sciences, 13*(3), 1891.

Memarzadeh, M., Matthews, B., Templin, T., Sharif Rohani, A., & Weckler, D. (2023). Semi-Supervised Active Learning for Anomaly Detection in Aviation. *Journal of Aerospace Information Systems, 20*(4), 181-194.

Nguyen, N. V., Hum, A. J., Do, T., & Tran, T. (2023). Semi-supervised machine learning of optical in-situ monitoring data for anomaly detection in laser powder bed fusion. *Virtual and Physical Prototyping, 18*(1), e2129396.

Sabokrou, M., Khalooei, M., Fathy, M., & Adeli, E. (2018). Adversarially learned one-class classifier for novelty detection. *IEEE conference on computer vision and pattern recognition*, (pp. 3379-3388).

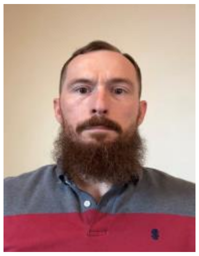
Welford, B. (1962). Note on a method for calculating corrected sums of squares and products. *Technometrics*, *4*(3), 419-420.

Yan, H., Liu, Z., Chen, J., Feng, Y., & Wang, J. (2023). Memory-augmented skip-connected autoencoder for unsupervised anomaly detection of rocket engines with multi-source fusion. *ISA transactions*, *133*, 53-65.

BIOGRAPHIES

Ge Song received his B.S. degree in Mechanical Engineering from the Nanjing University of Science and Technology, Nanjing, China in 2019, and M.S. degree from the Boston University in 2021. He is currently working towards the Ph. D. degree in the department of Mechanical Engineering with the University of South Carolina, Columbia, SC, USA. His research interests include computer vision and robotics and autonomous system.

Seong Hyeon Hong obtained his Ph.D. in Mechanical Engineering from the University of Florida (UF). He is currently serving as an assistant professor at the Florida Institute of Technology (FIT) in the department of Mechanical and Civil Engineering. He has extensive research experience in dynamics, control, estimation, machine learning, and numerical analysis. His recent research focuses on physics-informed learning, intelligent control and perception, human-robot interaction, robot-based inspection, and defect/anomaly detection.



Tristan Kyzer obtained his B.S. and M.S. degrees in Mechanical Engineering from the University of South Carolina. Currently he is a mechanical engineer with Integer Technologies. His research focuses on robotics and computer vision applications for anomaly detection and mitigation in the defense industry. Prior to transitioning into a career in engineering, Tristan served nearly a decade in the US Marine Corps as an infantryman stationed at Camp Lejeune and Camp Pendleton.

Yi Wang obtained the B.S. and M.S. degrees in Power Machinery and Energy Engineering from Shanghai Jiao Tong University, China, in 1998 and 2000, respectively, and Ph.D. degree in Mechanical Engineering from Carnegie Mellon University, USA, in 2005. From 2005 and 2017, he had been in the R&D position with CFD Research Corporation, USA. He is currently an Associate Professor in Mechanical Engineering Department with the University of South Carolina. His research focuses on computational and data-enabled science and engineering, including data-driven multi-fidelity surrogate modeling for multidisciplinary optimization, massive and real-time data analytics and computer vision, and autonomous systems.

.