UKAEA-RACE-PR(22)03

Luigi Pangione, Guy Burroughes, Rob Skilton

Use of VAE to identify anomalous data in robots

Use of VAE to identify anomalous data in robots

Luigi Pangione, Guy Burroughes, Rob Skilton

This is a preprint of a paper submitted for publication in Robotics

Article

Use of VAE to identify anomalous data in robots

Luigi Pangione [1,†], Guy Burroughes [1] and Rob Skilton [1,]

[1] RACE - UKAEA
* Correspondence: luigi.pangione@ukaea.uk
† Current address: RACE - United Kingdom Atomic Energy Authority, Culham Science Centre, Abingdon, Oxon, OX14 3DB, UK.

Abstract: For robotic systems involved in challenging environments it is crucial to be able to identify faults as early as possible. In challenging environments it is not always possible to explore all of the fault space, thus anomalous data can act as a broader surrogate, where a anomaly may represent a fault or a predecessor to a fault. This paper proposes a method for identifying anomalous data from a robot, whilst using minimal nominal data for training. A Monte-Carlo ensemble sampled Variational Autoencoder is utilised to determine nominal and anomalous data through reconstructing live data. This has been tested on simulated anomalies on real data, demonstrating the technique being capable of reliable identifying anomaly, with no pre-knowledge of the system. With the proposed system getting an F1-score of 0.85 in testing.

Keywords: Condition monitoring, robot, VAE, anomaly detection

1. Introduction

In robotic systems involved in nuclear operations, it is crucial to be able to identify anomalies as early as possible. In radiation environments, where human access is not possible, being able to identify a problem in early stages can allow the operator to stop operations and relocate the robot to a place where it is possible to perform necessary maintenance. Moreover, in such environments, robots suffer early ageing due to the radiation dose they are exposed to. Effects of radiations can develop in a gradual degradation of robot performance as well as a sudden failures. Radiation can have diverse effects on a range of components of the robot including those that would be considered robust in normal operations. It is clear then, in a radiation environment the appearance of a fault can be a dramatic event. A robot unable to move can have a dramatic impact on safety. Moreover, such conditions can have serious impact on operational costs as it may be highly difficult to recover the robot for repair. It is worth noting, in fact, that often robotic systems for nuclear operations require bespoke solutions difficult to be replaced.

A good example of a challenging environment is given by nuclear gloveboxes. They provide to the operator a very limited workspace, prone to clutter, with a vision from the outside not always optimal. Moreover, an operator is equipped with personal protective equipment such as coveralls and masks which reduce the ability to move and see. Also, the processed object can contain hazardous material difficult to assess in such conditions. In typical glovebox operations, objects that need processing are inserted inside the glovebox through a sealed door, once the objects are secured inside the glovebox, the operator executes all the required tasks; at the end of them, the processed objects are posted out and the glovebox is prepared for the subsequent task. In this work sequence is extremely important that the operator can complete all the tasks assigned without interruptions. It is clear then, in a robotic glovebox, information on the status of the robot and its ability to complete the tasks without occurring in faults is of paramount importance.

Citation: Pangione, L.; Burroughes, G.; Skilton, R. Use of VAE to identify anomalous data in robots. *Journal Not Specified* **2021**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

The purpose of a Condition Monitoring System (CMS) is to monitor measurements taken from the robotic system, and infer the health of the device, including possibility of unusual behaviour or degrading performance and report these findings to the robot's operator. Traditional CMS uses dedicated sensors to identify a fault in components, for example a vibration sensor to identify faults on a motor. In our work, we make use of already existing data provided by the robot hardware to the operator and to the control system. As it will be more clear later, an anomaly in this data is not necessarily related to a fault in a robot component, but represents an unexpected event in a wider sense. Differences in measurements like position and velocity can be noted by an expert operator without any additional system. Other measurements, like motor current, torque and temperature are usually hidden to the operator to avoid distractions. Moreover, the wealth of information available to operators during a complex robot operation may be overwhelming. Variations in such measurements are therefore impossible to be noted by an operator, even the most experienced one. From the operator point of view, it is important to remain focused on performing the task and be able to be informed only with the most relevant information in case a fault is developing.

In our work we use Variational AutoEncoder to identify anomalies in our robotic glovebox setup. This choice is motivated by the highly complex and structured nature of the relationship between the measured signals and the robot's health. We use real data to train the Variational AutoEncoder and then test it using simulated faults. We score samples by using loss function scoring and we make use of F1 score and ROC score to sensitivity in to discriminate anomalies.

This paper is organized as follow. In next section we give a background of anomaly detection and Variational AutoEncoder. In Section 3 we introduce a technique to use them anomaly detection. In subsequent section (section 4) we introduce our experimental setup. In sections 5 and 6 we respectively report and discuss our results. In final section we report our conclusions and outline future works.

2. Background

2.1. Anomaly Detection

Traditional fault detection techniques require a detailed a priori knowledge of all the possible faults that a robot may encounter. However, this is not always possible in challenging environments, as access to extensive characterisation is rarely feasible. This leads to many faults occurring in a nuclear environment (for example) being novel. However, the existence of a fault can be inferred by a discrepancy with respect to the usual behaviour in the robot's data. Such discrepancy, or anomaly, in data can represent different type of data anomaly. In [1] the authors classify anomalies in the following three categories:

- Point anomalies – where a single instance of the data is anomalous with respect to all the rest of the data.
- Contextual anomalies – where an instance of the data is anomalous with respect to the specific context of the data; i.e. data that would be nominal in context a robot linear motion would be anomalous when the robot is doing an accelerating motion.
- Collective anomalies – where a collection of the data is anomalous with respect to the data set; e.g. data from the current sensor and thermometer are individually nominal, but not both at the same time.

2.2. Variational AutoEncoder

In the last few years, deep learning based generative models have gained more and more interest due to (and implying) some amazing improvements [2] in the field. One such technique is the Variational AutoEncoder (VAE). In probability model terms, the VAE refers to approximate inference in a latent Gaussian model where the approximate posterior and model likelihood are parametrized by neural networks (the inference and generative networks). In neural network language, a VAE consists of an encoder, a decoder, and a loss function.

The purpose of the encoder is to map the information included in the sample into a reduced dimension space, called latent space. This space is meant to contain the main characteristics of the samples. The decoder, on the other hand, maps a sample from its latent space representation back to the original form. The peculiarity of the VAE is that each dimension of the latent space consists of a Gaussian distribution, each of them characterised by a mean and a logarithmic variance value. This implies that, once a sample is mapped into the latent space, it is possible to draw multiple times to obtain multiple reconstructions of the original sample. In a VAE the loss function is the sum of two parts: reconstruction loss and latent loss. The reconstruction loss is a metric of the VAE ability to reproduce the desired output; for example, such loss can be the mean square error (MSE) or the mean absolute percentage error (MAPE). The latent loss encourages the latent space to have a form of Gaussian distribution; an example of latent loss is the KL divergence loss.

In recent years VAE have been used in anomaly or fault detection in a wide range of applications, from images to bank transactions. In [3] the authors combine VAE and Long Short-Term Memory (LSTM) to detect anomalies in time series. In [4], the authors use VAE model in detecting anomalies in videos. It is interesting to note that in the paper the latent space is modelled as Gaussian Mixture Model (GMM) rather than a single Gaussian distribution. In [5], the authors take advantage of multiple draws from the latent space to map the reconstruction error, i.e. the difference between input sample and its reconstruction, into Gaussian distribution. We do not think it is possible to apply the same techniques to our data and therefore, as it will be clearer later in the paper, we adopt a different method to identify anomalies. In [6] the authors discriminate anomalies by clustering the latent space. Also in this case, we do not believe it is possible to apply this technique to our data.

3. VAE for Anomaly recognition

3.1. Reconstruction

We train the VAE to reproduce in output the sample presented in input. The main idea is that the VAE will be able to reproduce a sample that already appeared during training, while it will fail if a sample contains any kind of anomaly. A VAE sample is made by measurements collected from all the joints. In case of an anomaly in a joint, only some of the measurements will be affected.

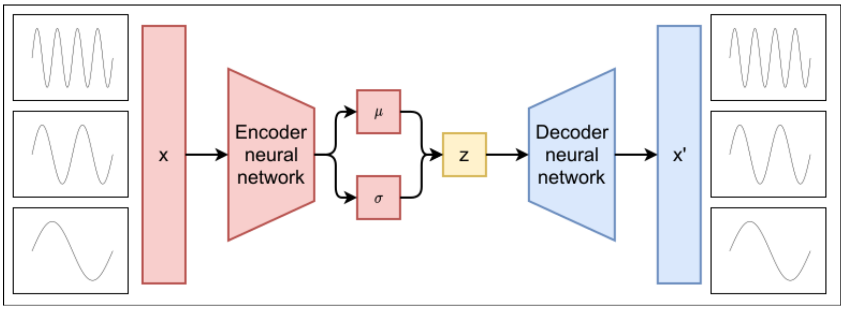
Figures 1a, 1b illustrates a simplification of the reconstruction concept. In particular, in normal conditions represented by Figure 1a, measurements collected from the robot are collectively known, therefore the VAE is able to reproduce all of them correctly. In case of measurements not collectively already presented during training, Figure 1b, the VAE will not be able to correctly reproduce them.

It is important to note that in 1b, the anomalous measurement not necessary must be novel or containing values never seen before. The VAE will not be able to reproduce all of them as long as they are not collectively the same.

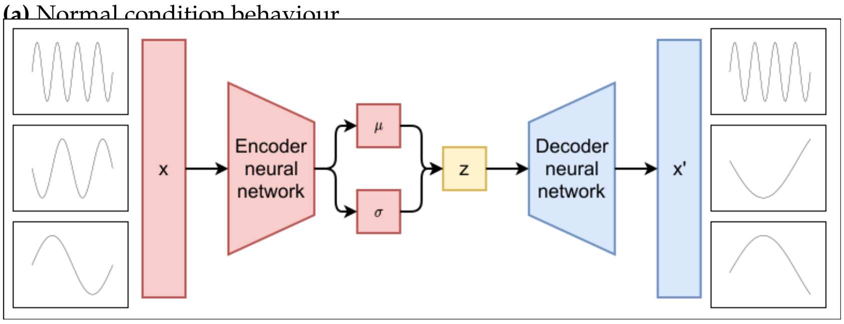
One way of seeing this is that the state of the machine is then encoded in the latent space. If the encoder, encodes a region of the latent space that hasn't been trained, the decoder will not be able to decode and thus reproduce the values coherently/correctly. Following this analogy, using a VAE allows the system to account for sensor noise, the latent space can encode a covariance to the probability of values based on the region.

3.2. Monte-carlo Reconstruction

As already stated, having a stochastic process as part of the latent space permits to generate multiple reconstructions of the predicted signal starting from a single point in the latent space. By separating the encoder and the decoder components of a trained VAE, it is possible to use the encoder to obtain a latent space representation of a sample.



(a) Normal condition behaviour



(b) Anomalous condition behaviour

Figure 1. Simplified schema of VAE reproducing data in normal condition and anomalies.

From there it is then possible to sample the decoder multiple times, in a Monte-carlo fashion, to collect a statistic of the expected reconstruction behaviour. The reconstruction of any sample which is not compliant with this statistic can be interpreted as an anomaly. This would enable the system to be tolerant to sensor noise. It is inevitable that they will be a base level of noise on any sensor reading, as this entropy-like, it will not be possible for the decoder to reproduce this signal component exactly. However, the level of noise compared to the signal could be encoded into the covariance of the VAE.

It is assumed that the noise can be approximated as a Gaussian in the latent space. As the latent space would approximate to underlying parameters of the system. A Monte-Carlo ensemble decoded from the latent space can then approximate a nominal stochastic distribution.

For example, it is possible to generate a zone around each signal showing nominal behaviour. This zone can be calculated as the convex hull of all the expected reconstructed measurements. Each signal reconstructed within this zone can be considered as nominal behaviour. A Gaussian mixture model was investigated but deemed unnecessary.

In particular, it is possible to use samples of the training set to obtain the worst possible nominal reconstruction cases over multiple draws. This effectively creates a band around each measurement in which we expect the VAE to reconstruct it. Reconstructed samples from the testing set outside this band can be considered anomalies.

4. Experimental setup

4.1. Glovebox use-case

The setup consists of two Kinova Gen 3, seven degrees of freedom [7] robotic arms, installed in the glovebox demonstrator [8]. The two robotic arms are identical, but the end effectors attached to them differ considerably in dimensions and weights. The experiment will perform a series of CMS dedicated moves that are executed at the beginning and at the end of operations. There are many advantages in using dedicated moves, but the main one is that they provide a solid and reliable base of data not affected by any external factor, such as human error. Moreover, in a robotic glovebox context and from the operational point of view, it is convenient to have "warm-up" and “cool-down”

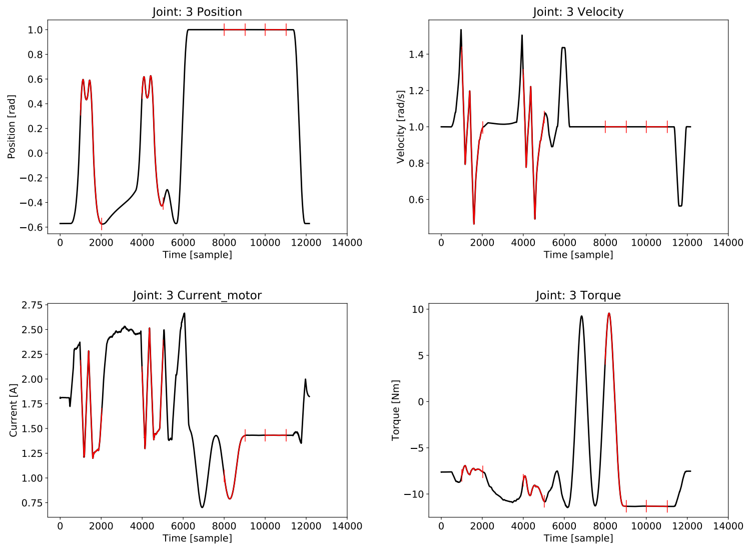


Figure 2. An example of time-windowed data sampled from a single joint for one glovebox robot.

phases at the beginning and the end of operations, in which make sure the system is respectively ready to operate or has not been damaged during operations.

The glovebox system is controlled using the ROS framework [9,10]. The Kinova ROS package provides a large amount of real-time data for each joint. In particular, in our work we use joint position, joint velocity, motor current and motor torque of each joint.

It is very important to note that the measurements in use are very different from each other and their range of extension is very diverse. This makes very difficult to scale them to a common range, i.e. between 0.0 and 1.0. Scaling each individual measurement with a common scaling factor would have made features disappear; on the other hand, scaling them with a measurement's specific factor would have modified the relationship between measurements; unless all data was scaled equally, which would potentially suppress some values.

Not having all the input normalised between 0.0 and 1.0, creates a constraint in the definition of the reconstruction loss. In fact, the same mean square error in two different measurements can have different effects. For this reason, this work evaluates the reconstruction error using mean absolute percentage error (MAPE). Using MAPE reconstruction loss the same error in two different measurements is considered differently according to the measurement magnitude.

As previously explained, the data consists of a set of time series measurements coming from each joint. To be able to capture the dynamic behaviour of the system, it has been chosen to consider, as a single sample in input or in output of the VAE, data collected during a configurable time-window. A sample is therefore a matrix where each row represents a measurement and columns represent acquisition times. In Figure 2 is presented an example of data coming from joint 3 sectioned in time-window.

It is important to note that this choice does not affect in any way the capability of the system to identify anomalies while working online as the current sample at time t_{now} should be the set of data collected during time-window $[t_{now} - t_{window}, t_{now})$. It is also important to note that the longer the time-window, the highest is the amount of information each sample contains. On the other hand, using a long time-window, the system becomes less sensitive to short time perturbations. As will be explained later, the length of the time-window changes the behaviour of the system in identifying different types of faults.

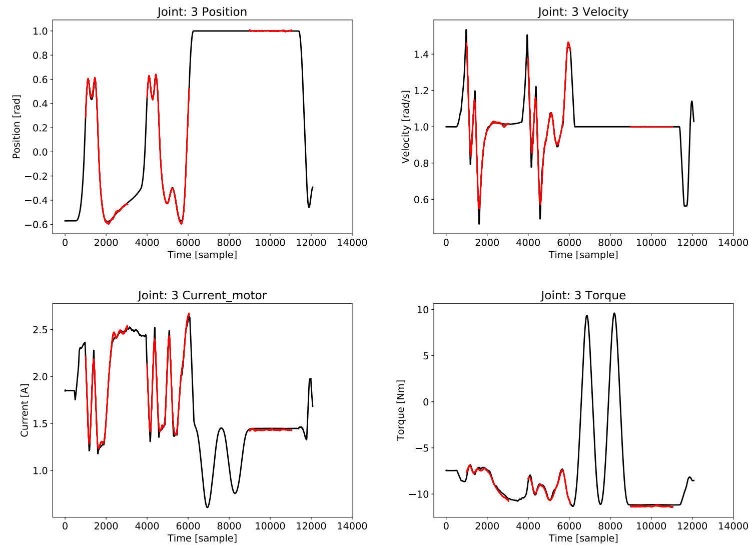


Figure 3. Example of data reconstructed by VAE. In black are reported original measurements not used during training. In red are reported different samples of output reproduced by the trained VAE.

4.2. Implementation

Our VAE layout consists of a fully connected multiple layers neural network. Through testing, an optimal number of layers and their size have been determined. Good results in reproducing the input samples have been obtained with encoder's layers dimensioned respectively [512, 256, 128, 64, 32] and with a latent space with dimension of 6. As sample's values are not bonded between 0.0 and 1.0, a Leaky ReLu activation function has been used. The decoder has been implemented in a symmetric way. For operational reasons, the VAE has been trained using data coming from only one of the two robots installed in the glovebox, from now on the training robot. Data collected from the other robot, from now on the testing robot, have been used for comparison purposes.

In Figure 3 it is reported data reconstructed by the trained VAE. In particular, in the figure are reported reconstructed measurements of Joint 3 using data collected from the training robot, but not used during VAE training. Black lines represent original measurements, while red lines represent reconstructed measurement. To increase the readability of the figure, it has been reported reconstructed data only during some time windows.

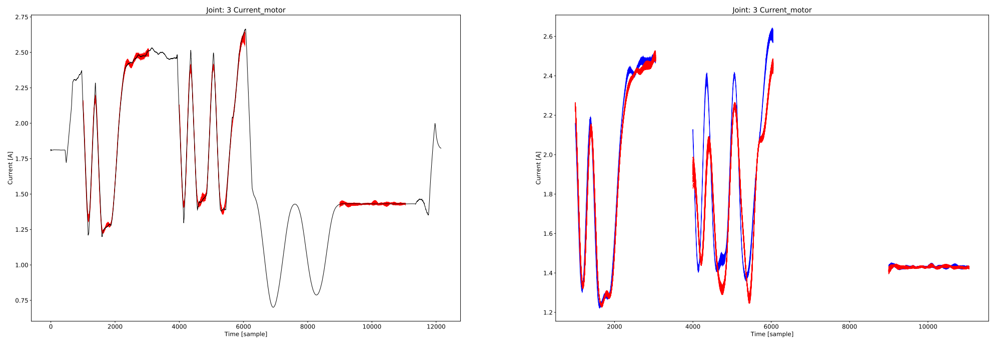
5. Results

We have used the Monte-carlo technique explain in section 3.2 to generate a zone in which expect signal with nominal behaviour.

In figure 4a is shown in red the calculated zone in which the motor current of joint 3 is expected to be reconstructed, while the black line represents the actual measure. As it is possible to note, the zone of expected nominal behaviour is very narrow.

In figure 4b multiple draws of samples collected from the testing robot are reported in blue and are compared to the expected behaviour zone. Data from the testing robot are clearly different from nominal and definitively outside the zone of expected behaviour. This is because the two robots are equipped with two different end effectors.

To improve the identification of anomalies we have opted to use the VAE loss function to score a sample. We focused on the training robot only as data coming from the testing robot have been already proved as anomalous. As we did not experience any anomaly in the training robot over time, we modified our data to generate simulated anomalies. In particular, we modified our data in two different ways:



(a) Training **(b)** Testing

Figure 4. Example of calculated nominal behaviour zone and its comparison against multiple reconstruction of same same in case of testing robot data.

a) variation perturbation - a variation of 20% in some of a joint's measurements.
b) swap perturbation - a swap between two time windows of some of the joint's measurements.

The first type of data perturbation is intended to reproduce a "point anomaly", where the perturbed data is anomalous with respect to all the rest of the data. In this sense, we can imagine this anomaly as data that assumes values never seen before during the training.

The second type of data perturbation is intended to reproduce a "contextual anomaly", where the perturbed data is anomalous in its context. In this sense, we can imagine this anomaly as data that is not new, but anomalous because in the wrong context.

We consider an anomaly as a sample for which the loss score is higher than a predefined threshold. We calculate the threshold as the value that provides the maximum F1 score. It is important to note that this is possible because we are generating simulated faults and therefore we have ground truth information on data.

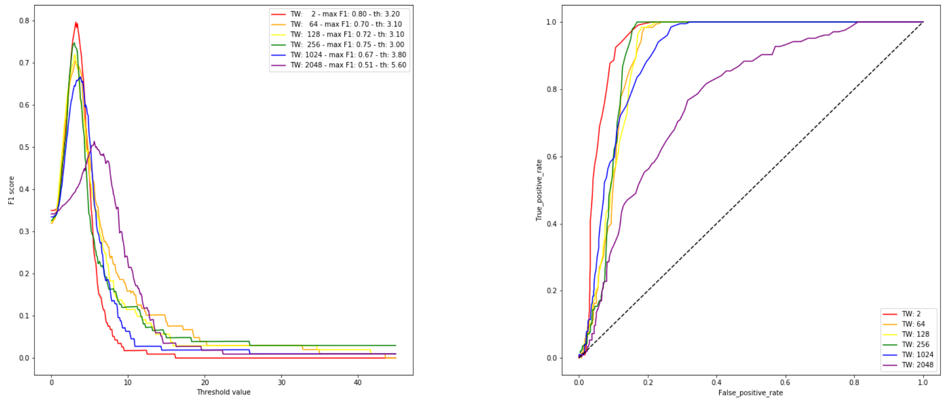
In Figure 5a and Figure 5b it is reported different values of F1 score and ROC score for different values of the time-windows in case data are modified using a "variation perturbation".

Our results show that for this type of anomaly the best F1 score is 0.80 and it is obtained with a time-window of 2 samples. Small time-windows (2, 64, 128, 256) do have similar good performances, while longer time-windows (1024 and 2048) have worst performances. This results are confirmed also by the ROC score, in which small time-windows curves are closer to the top left corner. Intuitively this is equivalent to say that as point anomalies are data never seen before, they are easier to recognise using short time-windows.

Similarly, in Figure 6a and Figure 6b, it is reported different values of F1 score and ROC score for different values of the time windows in case data are modified using a "swap perturbation".

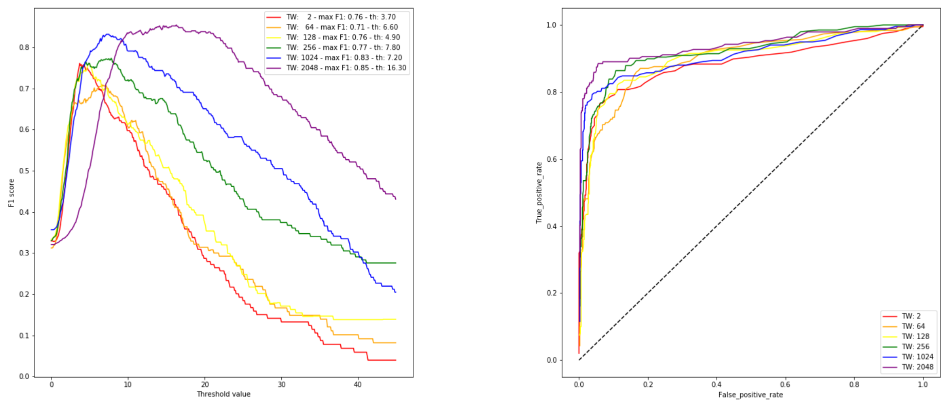
Our results show that for this type of anomaly the best F1 score is 0.85 and it is obtained with a time-window of 2048, which corresponds to about 2 seconds. Also in this case it is possible to observe how long time-windows and short time-windows perform in opposite way. Intuitively long time-windows perform better with "swap perturbation" anomaly because data are not novel in value, but in context, therefore the system needs more information to identify the anomalies.

Overall these results show that the VAE provides very good accuracy in both types of simulated anomalies.



(a) F1 scores **(b)** ROC scores

Figure 5. Trends of F1 score (a) and ROC score (b) curves for different values of time-windows in case of variation perturbation simulated anomalies.



(a) F1 scores **(b)** ROC scores

Figure 6. Trends of F1 score (a) and ROC score (b) curves for different values of time-windows in case of swap perturbation simulated anomalies.

6. Discussion

The stochastic process in the latent space allowed us to encode a sample into a latent space and then draws multiple times from it, to obtain a statistic of the VAE ability to reconstruct that sample. Contrary to [5] we decided to do not assume the reconstruction error was Gaussian. Instead, we decided to use these statistics differently. Unfortunately, the zone we obtained was too narrow to be practically useful. However, this may not be true for less stable, more fault-prone systems.

As alternative we opted to use the VAE score to assess whether the sample was an anomaly or not. To calculate the threshold to discriminate normal from anomalous behaviour we made use of F1 and ROC scores. Interestingly the length of the time window influenced the ability to identify different types of simulated anomalies. In particular, in simulated context anomalies, obtained by swapping values of different instant in time for some measurements, long time windows performed better. This is expected as in this type of anomaly values have been presented during training, but not collectively at the same time. This required more information to be available in the sample. On the other hand, in simulated point anomalies, obtained by increasing by 20% values of some measurements, VAE did not need much information to identify values never been presented before. In this case short time windows performed better.

7. Conclusions

In this paper we have investigated the use of VAE in identifying anomalous data collected from our robotic glovebox setup. We defined a Monte-carlo based technique to produce a statistic of expected nominal behaviour results. We have applied this technique to data collected from two identical robots equipped with different end effector. To improve our results, we used loss function score against simulated anomalies in data. We proved both techniques can be used in detection of anomalies in data.

One of the weaknesses of the study is the lack of real anomaly data vs actual anomaly data. During our time with the Kinovas we have only witnessed one anomaly in 100s of hours of operations, that was identified by the proposed system. For future studies it proposed to use a more anomaly prone system, rather than an industrial robot, which are known for their robustness.

In future works we will investigate more the possibility of using statistics from multiple reconstructions of the same sample to identify anomalies.

Funding: This project has been supported by the RAIN Hub, which is funded by the Industrial Strategy Challenge Fund, part of the government's modern Industrial Strategy. The fund is delivered by UK Research and Innovation and managed by EPSRC [EP/R026084/1].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly detection: A survey. *ACM Comput. Surv.* **2009**, *41*, 15:1–15:58. doi:10.1145/1541880.1541882.
2. Kingma, D.P.; Welling, M. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* **2013**.
3. Lin, S.; Clark, R.; Birke, R.; Schönborn, S.; Trigoni, N.; Roberts, S. Anomaly Detection for Time Series Using VAE-LSTM Hybrid Model. ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020, pp. 4322–4326.
4. Fan, Y.; Wen, G.; Li, D.; Qiu, S.; Levine, M.D.; Xiao, F. Video anomaly detection and localization via gaussian mixture fully convolutional variational autoencoder. *Computer Vision and Image Understanding* **2020**, *195*, 102920.
5. An, J.; Cho, S. Variational Autoencoder based Anomaly Detection using Reconstruction Probability. 2015.
6. Xu, H.; Feng, Y.; Chen, J.; Wang, Z.; Qiao, H.; Chen, W.; Zhao, N.; Li, Z.; Bu, J.; Li, Z.; et al.. Unsupervised Anomaly Detection via Variational Auto-Encoder for Seasonal KPIs in Web Applications. *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18* **2018**. doi:10.1145/3178876.3185996.
7. Kinova Gen 3 product description. https://www.kinovarobotics.com/en/products/gen3-robot. Accessed: 2020-05-21.
8. Tokatli, O.; Das, P.; Nath, R.; Pangione, L.; Altobelli, A.; Burroughes, G.; Skilton, R. Robot Assisted Glovebox Teleoperation for Nuclear Industry **2021**.

9. Quigley, M.; Conley, K.; Gerkey, B.; Faust, J.; Foote, T.; Leibs, J.; Wheeler, R.; Ng, A.Y.; others. ROS: an open-source Robot Operating System. ICRA workshop on open source software. Kobe, Japan, 2009, Vol. 3, p. 5.
10. ROS: Robot Operating System. https://www.ros.org/. Accessed: 2020-05-21.