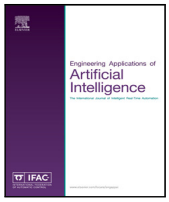
Engineering Applications of Artificial Intelligence 104 (2021) 104379

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Unsupervised anomaly detection for underwater gliders using generative adversarial networks

Peng Wu [a,*], Catherine A. Harris [b], Georgios Salavasidis [b], Alvaro Lorenzo-Lopez [b], Izzat Kamarudzaman [b], Alexander B. Phillips [b], Giles Thomas [a], Enrico Anderlini [a,*]

[a] Department of Mechanical Engineering, University College London, London, WC1E 7JE, UK
[b] Marine Autonomous and Robotic Systems, National Oceanography Centre, Southampton, SO14 3ZH, UK

ARTICLE INFO

Keywords:
Anomaly detection
Underwater gliders
Marine autonomous systems
Generative adversarial networks

ABSTRACT

An effective anomaly detection system is critical for marine autonomous systems operating in complex and dynamic marine environments to reduce operational costs and achieve concurrent large-scale fleet deployments. However, developing an automated fault detection system remains challenging for several reasons including limited data transmission via satellite services. Currently, most anomaly detection for marine autonomous systems, such as underwater gliders, rely on intensive analysis by pilots. This study proposes an unsupervised anomaly detection system using bidirectional generative adversarial networks guided by assistive hints for marine autonomous systems with time series data collected by multiple sensors. In this study, the anomaly detection system for a fleet of underwater gliders is trained on two healthy deployment datasets and tested on other nine deployment datasets collected by a selection of vehicles operating in a range of locations and environmental conditions. The system is successfully applied to detect anomalies in the nine test deployments, which include several different types of anomalies as well as healthy behaviour. Also, a sensitivity study of the data decimation settings suggests the proposed system is robust for Near Real-Time anomaly detection for underwater gliders.

1. Introduction

Whilst autonomous systems are predicted to become pervasive in the maritime industry (Department for Transport, 2019), this growth is currently heavily constrained by the challenges of fully independent remote operation in hazardous and dynamic marine environments. Marine Autonomous Systems (MAS), such as Underwater Gliders (UGs), can be at sea for up to months at a time, during which they periodically surface and communicate via satellite with remote expert operators known as pilots. The transmission of data to and from the MAS is severely constrained by low-bandwidth satellite, making it challenging for pilots to monitor MAS time series data and behaviour during operation manually. If the underlying cause of observed adverse behaviour cannot be correctly diagnosed and the situation remedied, e.g. via the remote adjustment of piloting parameters or mission scope, the MAS and its data cargo can be lost or present a hazard to shipping (Thieme and Utne, 2017). As a result, to reduce operational costs, increase reliability and scale-up the use of MAS within the maritime industry, strategies must be developed for automated anomaly detection and fault diagnosis.

The code of practice for maritime autonomous surface systems developed by Maritime UK (Maritime UK, 2019) recognises the need for MAS to support on-board signal processing with remote condition monitoring to interpret the impact of faults and adverse conditions on the vehicle's safety and performance. On-board systems are limited by power and computational constraints, whilst current manual detection and diagnosis approaches are limited by the experience of the individual pilot and are subject to human error, especially when MAS require pilot attention around the clock. In the absence of general on-board anomaly detection and diagnosis systems, the ability to transmit sensor data in a timely manner to an off-board system and to receive appropriate commands in response becomes of critical importance for MAS safety and performance.

The operation of MAS platforms beyond the visual line of sight requires a suitable command and control system. For example, the UK's National Oceanography Centre (NOC) Oceanids C2 system is a platform to support the over-the-horizon operation of MAS within the National Marine Equipment Pool for efficient fleet management (Farley et al., 2019; Harris et al., 2020). Another example is the LSTS Neptus and Dune over-the-horizon command-and-control environment (Dias et al., 2005; Madureira et al., 2013; Pinto et al., 2013). This work aims to develop a holistic automated anomaly detection system[1], well-suited to the limited availability of multivariate time series data during MAS operations.

* Corresponding authors.
E-mail addresses: peng.wu.14@ucl.ac.uk (P. Wu), e.anderlini@ucl.ac.uk (E. Anderlini).

[1] https://github.com/pwu01/ALADDIN-BiGAN-anomaly-detection

https://doi.org/10.1016/j.engappai.2021.104379
Received 31 March 2021; Received in revised form 17 June 2021; Accepted 30 June 2021
Available online 12 July 2021

The contributions of this study are twofold. First, this work proposes an improved Bidirectional Generative Adversarial Networks (BiGAN) anomaly detection system guided by periodic assistive hints to achieve effective and stable training of generative adversarial models. Second, this work introduces a novel holistic anomaly detection system for MAS to be integrated within remote control centres to monitor operations over the horizon. The system is based on BiGAN to detect faults by tracking the anomaly score. As compared with state-of-the-art anomaly detection systems for MAS that exploit steady-state conditions, deep neural networks are used to capture dynamic effects from the time series data. Two healthy deployment datasets in time series are used to train the system via unsupervised learning. The developed system is tested using actual datasets from nine deployments collected by a selection of vehicles operating in a range of locations and environmental conditions with varying mission length. A sensitivity analysis on the data decimation settings for satellite communication suggests the proposed approach is insensitive to these settings, making it suitable for Near Real-Time (NRT) anomaly detection for UGs that are known to be under-observed systems. Such an unsupervised approach requires minimum training data preparation efforts and successfully detects anomalies for the nine test MAS deployments.

2. Related work

Methods for condition monitoring can be subdivided into model-based and data-driven diagnostics (Fink, 2020; Fink et al., 2020). The former rely on dynamic models of the physical systems, whereas the latter on the analysis of actual sensor data. Whilst model-based solutions are better for condition monitoring of new systems where available data is limited, data-driven methods show significant improvements in accuracy in the cases where significant prior data exists (Michau and Fink, 2021).

A review of nonlinear model-based methods for condition monitoring can be found in Hong et al. (2008). These approaches are useful for systems that present strong nonlinearities or coupling. A summary of fault detection methods for aircraft based on signal-processing and dynamic models can be found in Zolghadri et al. (2015). These techniques are robust, simple and computationally relatively inexpensive. Data-driven methods can be generalised to different fault detection and diagnosis problems and scaled to a large number of sensors. These approaches would require the data to be appropriately collected and processed.

Crestani et al. (2015) integrated fault tolerance into the design of a robot real-time control architecture showing the benefits of including condition monitoring considerations from the initial stages of design of a new prototype. Specifically for Autonomous Underwater Vehicles (AUVs), many fault detection studies involve thrusters, inclusive of model-based solutions (Freddi et al., 2013), radial basis function networks (Wang and Zhang, 2006), Gaussian particle filter (Sun et al., 2016) and artificial immune system (Yao et al., 2018). Clustering solutions are also investigated by Ray Harris (2015) to determine faults in an unsupervised way. Raanan et al. (2016, 2018) have developed an automatic fault detection system for long-range AUVs based on Bayesian nonparametric topic modelling techniques. Although the dataset focuses on the identification of bottoming events, the behaviour of the analysed long-range AUV is similar to that of UGs. The nearest neighbour classifier presents particularly high accuracy over two different test sets. A system to develop safety indicators for the operation of MAS is described in Thieme and Utne (2017), with a case study on an AUV. Hamilton et al. (2007) propose an integrated fault detection and diagnosis architecture for AUVs, although the focus is on on-board systems. Anderlini et al. (2020a) have designed rule- and model-based methods for the detection of the loss of wings on UGs and the onset of high levels of marine growth on UGs (Anderlini et al., 2020b). Further work by Anderlini et al. (2021a) has developed and tested an anomaly detection system that blends model- and data-based solutions to detect both simulated and naturally accumulated biofouling.

Pang et al. (2021) provide a comprehensive review of deep learning for anomaly detection, and propose a taxonomy by classifying the state-of-the-art deep anomaly detection techniques into three categories, i.e. feature extraction, learning feature representations of normality and end-to-end anomaly score learning. In the category of learning feature representations of normality, models based on Auto-Encoder (AE) are proposed to detect anomalies by learning low-dimensional feature representations to reconstruct given data instances (Sakurada and Yairi, 2014; Zhou and Paffenroth, 2017; Borghesi et al., 2019). For the anomaly detection of UGs, Anderlini et al. (2021b) have developed different data-driven solutions, including feedforward neural networks and AEs that can detect anomalies such as wing loss and marine growth. However, the features learned by AE-based models can be biased by infrequent normalities in the training dataset. With data instances encoded by a prior distribution over the latent space, Variational Auto-Encoder (VAE) enables a better reconstruction of input data instances; hence improved anomaly detection performance can be achieved. For anomaly detections of multivariate sequence data, variants of VAE have been developed (Su et al., 2019; Li et al., 2020). The Generative Adversarial Networks (GAN) proposed by Goodfellow et al. (2014) can capture the data distribution via generative and adversarial processes. The improved capability of capturing data distribution is particularly useful for anomaly detection applications (Di Mattia et al., 2019). The superior feature representation learning capability makes GAN particularly promising for remotely operated MAS (e.g. UGs) that can be highly under-observed due to low data transmission bandwidth and limited sensing ability to reduce the on-board space and power requirements. However, GAN are constrained by issues such as training instability (Pang et al., 2021).

Based on work of Goodfellow et al. (2014), Donahue et al. (2016) and Dumoulin et al. (2016) have developed variants of the original GAN, i.e. Adversarial Learned Inference GAN (ALI-GAN) and Bidirectional-GAN (BiGAN), respectively, to additionally learn a latent representation of the data, which have become the basis of several GAN-based anomaly detection systems. Zenati et al. (2018) developed a BiGAN-based anomaly detection system for high-dimensional real-world data such as images. Schlegl et al. (2017, 2019) have developed a series of BiGAN based anomaly detection models for medical image anomaly detection. An earlier study by Li et al. (2018) applied GAN to detect cyber attacks, using multivariate time series with the need of the inference process to map the test data back to latent space. Although these GAN-based anomaly detection systems appear successful in the applied domains, the GAN-based anomaly detection system is still relatively difficult to train for reasons including its unsupervised nature and the generative and adversarial process between multiple deep neural networks (Mutlu and Alpaydin, 2020). Despite GAN-based anomaly detection systems' success in other domains, they have not been applied to MAS, which are subjected to limited accessibility to system data and require a high level of generality to detect unpredicted anomalies in highly dynamic ocean environments.

3. Underwater gliders and data description

3.1. Slocum underwater gliders

All data used in this study are from deployments of Slocum G2 UG (Teledyne Webb Research, 2012), manufactured by Teledyne Webb Research (Webb et al., 2001; Schofield et al., 2007). As shown in Fig. 1, a Slocum UG is actuated by a Variable Buoyancy Device (VBD), which enables the vehicle's displacement and thus its buoyancy to be varied. Pitch is controlled by shifting the position of a movable battery pack, and the yaw angle is controlled using a rudder. Using fixed wings to provide lift, gliders can perform a sawtooth-like profile through the water-column. A Slocum UG starts a "yo", or cycle, by reducing its buoyancy and shifting the battery forward to initiate the descent, then extends the VBD and shifts the battery afterwards to climb to a

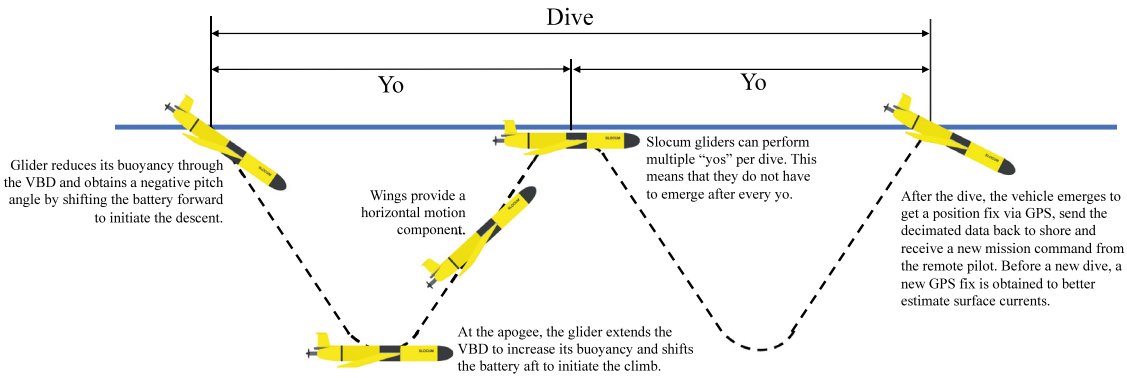


Fig. 1. Diagram showing the concept of operation of a Slocum UG. The drawing is not to scale: the analysed vehicles reach their apogee at a maximum depth of either 200 m or 1000 m and have glide path angles with a magnitude in the range of 15° to 30°.

Table 1
The data applied in this study is measured by a number of Slocum G2 gliders over eleven developments.

No.	ID	Date	Project	Organisation	Location	Duration [days]	Status	Purpose
1	unit 345	2014	AtlantOS, CaNDyFloSS	NOC	The Celtic Sea	123.9	Healthy	Training
2	unit 397	2015	AtlantOS, CaNDyFloSS	NOC	The Celtic Sea	45.9	Healthy	Training
3	unit 419	2015	AtlantOS, CaNDyFloSS	NOC	The Celtic Sea	11.0	Healthy	Testing
4	unit 399	2015	AtlantOS, CaNDyFloSS	NOC	The Celtic Sea	84.6	Possible biofouling	Testing
5	unit 423	2015	AtlantOS, CaNDyFloSS	NOC	The Celtic Sea	6.8	OMG	Testing
6	unit 424	2015	AtlantOS, CaNDyFloSS	NOC	The Celtic Sea	20.8	OMG	Testing
7	unit 194	2017	ALTERECO	NOC	The North Sea	83.9	Angle of list	Testing
8	unit 304	2019	ALTERECO	NOC	The North Sea	76.9	Loss of right wing	Testing
9	unit 345	2019	ALTERECO	NOC	The North Sea	76.8	Strong disturbances	Testing
10	unit 436	2019	ALTERECO	NOC	The North Sea	89.8	Loss of left wing	Testing
11	unit 492	2020	IDUG	PLOCAN	Gran Canaria	9.5	Simulated biofouling	Testing

designated depth at the apogee, completing the glider's "yo". A single dive can comprise multiple yos as shown in Fig. 1.

Once deployed, UGs operate fully autonomously, with pilots relying on a limited snapshot of multivariate time series data sent via satellite whilst the vehicle is on the surface, which includes system health variables, current instructions, last GPS position and decimated data from past dives. During normal operation, the remote pilots will first manually check this dataset for subsystem errors, warning and oddities reported by the glider itself (e.g. glider stalls, behaviour errors, and communication interruptions Schofield et al., 2007) along with the sawtooth dive profile to ensure each yo is symmetrical and the glider is reaching the target depth. Progress towards the target waypoint is also considered, along with a check of the battery health and consumption. This check is usually performed once per day, with the pilot making smaller observations more regularly after each dive. Therefore, pilots are only likely to look into the flight parameters in detail if the glider: is reporting errors in these subsystems, is failing to dive correctly or is clearly not making expected progress. Hence, issues that gradually emerge or that affect parameters outside those routinely monitored, such as roll, can go unnoticed, resulting in significant impacts on vehicle endurance and safety. Consequently, a smart anomaly detection system is crucial.

3.2. Datasets

Table 1 lists the multivariate time series datasets used in this study. The datasets are measured by Slocum G2 gliders over ten deployments operated by the NOC (1–10) (BODC, 2019) and one deployment (11) operated by the Oceanic Platform of the Canaries (PLOCAN) giving a total of 11 deployments. The labels correspond to healthy or standard baseline conditions, natural and simulated biofouling, angle of list, loss of one wing, strong environmental disturbances, e.g. due to ocean currents, and bulky sensory packs, e.g. the turbulence probes for the Ocean Microstructure Gliders (OMG). Biofouling caused by marine growth in shallow, warm and tropical waters can lead to an increase of a UG's weight, a significant drop in speed, even possible premature retrieval at sea (Anderlini et al., 2020b). OMG would lead to a higher drag coefficient and a higher negative buoyancy offset. The first two deployments (healthy) are used to train the anomaly detection system and the remaining nine datasets, which include several anomalies as well as healthy behaviour, are used for testing purposes to assess the generality of the approach.

Table 2
Sensor list.

No.	Sensor	No.	Sensor	No.	Sensor
1	Battery position	6	Pressure	11	Heading
2	Battery voltage	7	Roll	12	Temperature
3	State of charge	8	Rudder angle	13	Vacuum
4	Leak detection voltage	9	Pitch	14	VBD
5	Vertical velocity	10	Conductivity	15	sci_pressure[a]

[a]sci_pressure is the pressure as measured by the scientific payload pressure sensor.

The variables are detailed in Table 2. Mission specific geographical positions, control signals (heading control, pitch control, rudder angle control and VBD control) are not included in the datasets to ensure generality. The vertical velocity is calculated from the depth signal, which is in turn obtained from the pressure signal. The scientific pressure sensor measures the sci_pressure signal.

4. Anomaly detection using BiGAN

The presented anomaly detection method is based upon BiGAN (Donahue et al., 2016), with additional training hints guiding more effective generator (G) and discriminator (D) training. In each training iteration, the discriminator, generator and encoder (E) are trained concurrently. The assistive hint loss function is applied periodically to guide the encoder and generator using the errors terms of data patch reconstruction and discriminator feature. This approach is inspired by Schlegl et al. (2019) and Mutlu and Alpaydin (2020) but has been

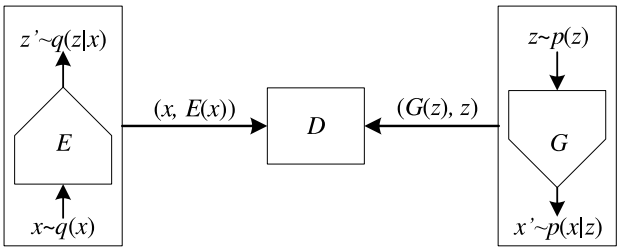


Fig. 2. Structure of the BiGAN (Donahue et al., 2016).

improved. In Schlegl et al. (2019), the encoder E is trained only after D and G have been trained, whereas in this study the discriminator, generator and encoder are trained concurrently. In Mutlu and Alpaydin (2020), the hint loss is added directly to the BiGAN loss function; our approach instead applies a periodic update step to the parameters of the generator and encoder.

4.1. Problem statement

Goodfellow et al. (2014) proposed Generative Adversarial Networks for estimating generative models via an adversarial process training a generative model G to capture the data distribution, and a discriminative model D that estimates the probability that a data sample comes from the training data or is generated by G. This framework trains D and G concurrently such that D maximises the probability of assigning the correct label to both training samples from data x and generated samples from G. D and G play a minimax game with the value function $V(G, D)$:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}\left[\log D(x)\right] + \mathbb{E}_{z\sim p_z(z)}\left[\log\left(1 - D(G(z))\right)\right] \quad (1)$$

where $x \sim p_{data}(x)$ is the data distribution, $z \sim p_z(z)$ is a prior on input noise variables. Although the ability of the original GAN framework to learn generative models mapping from simple latent distributions to arbitrarily complex data distributions has been demonstrated, it cannot project data back into the latent space. The BiGAN (Donahue et al., 2016) and ALI-GAN (Dumoulin et al., 2016) adopt a similar approach using an encoder with a generator to learn this inverse mapping.

Fig. 2 shows the structure of the BiGAN, which includes an additional encoder E that maps data x to its latent representations z. A trained BiGAN encoder can serve as a useful feature representation for related semantic tasks, i.e. the latent representation z can be regarded as a representation of data x. Unlike the standard GAN (Goodfellow et al., 2014), the discriminator D of the BiGAN discriminates $(x, E(x))$ and $(G(z), z)$. The training objective of the BiGAN is:

$$\min_{G,E} \max_D V(D,E,G) = \mathbb{E}_{x\sim p_{data}(x)}\left[\log D\left(x, E(x)\right)\right] + \mathbb{E}_{z\sim p_z(z)}\left[\log\left(1 - D(G(z), z)\right)\right] \quad (2)$$

Algorithm 1 details the training process of the BiGAN-based anomaly detection system. Donahue et al. (2016) have proven that the encoder E and generator G must learn to invert one another in order to fool the discriminator D. The encoder and generator of the BiGAN structure behave similarly to the encoder and decoder of an independent auto-encoder which learns a representation for a set of input data and reconstructs the data samples as closely as possible to their original inputs. Inspired by this feature of auto-encoders, we use the reconstruction difference between the input sample and reconstructed data to assist the BiGAN training, i.e. using the L_2 norm between the input data x and its reconstruction $G(E(x))$ via E and G:

$$\mathcal{L}_{re} = \frac{1}{n_x}\left\|x - G(E(x))\right\|_2 \quad (3)$$

where n_x is the number of input data elements.

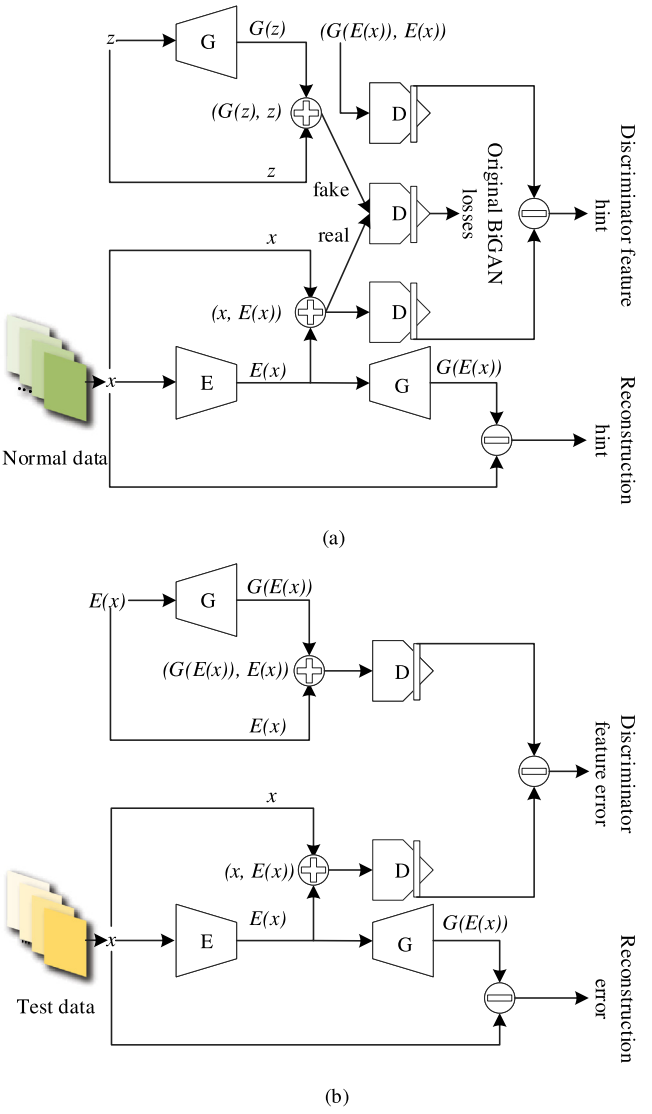


Fig. 3. Anomaly detection using BiGAN for underwater gliders: (a) training using normal data and (b) testing using unseen deployment data.

It is worth noting that Mutlu and Alpaydin (2020) have proposed training generative adversarial models using several assistive hints. However, such hints are added to the BiGAN loss function directly in their approach. This work proposes applying such hints periodically to achieve higher training efficiency. In addition, in the discriminator network, the neural network layer right before the final output layer is defined as a feature layer, which outputs a feature f. With this feature f provided by the discriminator, an additional hint loss is defined as:

$$\mathcal{L}_{fe} = \frac{1}{n_f}\left\|f(x, E(x)) - f(G(E(x)), E(x))\right\|_2 \quad (4)$$

where n_f is the feature layer's number of neurons.

Combining $\mathcal{L}_{re}$ and $\mathcal{L}_{fe}$, the assistive hint loss function is thus:

$$\mathcal{L}_{hint} = \frac{\kappa}{n_x}\left\|x - G(E(x))\right\|_2 + \frac{1}{n_f}\left\|f(x, E(x)) - f(G(E(x)), E(x))\right\|_2 \quad (5)$$

where κ is a hyperparameter which can be adjusted. Note that in the validation and test phases, the residual of $\mathcal{L}_{hint}$ is defined as the anomaly score that represents the degree of anomalies. Ideally, the residual should be near zero if the query data patch is normal. A high anomaly score indicates the input data patch deviates severely from healthy deployment data pattern.

Algorithm 1 The training procedure for underwater glider anomaly detection using GAN.

Prepare the training dataset
Initialise D, G and E parametrised by θ_D, θ_G and θ_E, respectively
for $i = 1$ to t **do** ▷ t is the number of total training iterations
  procedure TRAIN D, G AND E
    for $j = 1$ to k **do** ▷ Assistive hint is applied every k training steps
      Sample a mini-batch with M data samples x from the training dataset
      Sample z with a size of M from a prior Gaussian distribution $z \sim p_z(z)$
      $\mathcal{L}_{BiGAN} \leftarrow \frac{1}{M}\sum_{s=1}^{M}\log\left(D\left(x^{(s)}, E(x^{(s)})\right)\right) + \frac{1}{M}\sum_{s=1}^{M}\log\left(1 - D(G(z^{(s)}), z^{(s)})\right)$
      $\theta_D \leftarrow \theta_D + \nabla_{\theta_D}\mathcal{L}_{BiGAN}$, $\theta_G \leftarrow \theta_G + \nabla_{\theta_G}\mathcal{L}_{BiGAN}$, $\theta_E \leftarrow \theta_E + \nabla_{\theta_E}\mathcal{L}_{BiGAN}$
    end for
  end procedure
  procedure TRAIN THE G AND E WITH HINT
    Sample M data patches x from the training dataset
    $\mathcal{L}_{hint} \leftarrow \frac{\kappa}{n_x}\|x - G(E(x))\|_2 + \frac{1}{n_f}\|f(x, E(x)) - f(G(E(x)), E(x))\|_2$ ▷ Hint loss
    $\theta_E \leftarrow \theta_E + \nabla_{\theta_E}\mathcal{L}_{hint}$, $\theta_G \leftarrow \theta_G + \nabla_{\theta_G}\mathcal{L}_{hint}$
  end procedure
  if i mod $n = 0$ **then** ▷ Test anomaly detection performance using synthetic sensor faults every n training iterations
    Anomaly score $\leftarrow \mathcal{L}_{hint}$
  end if
end for

4.2. GAN for underwater glider anomaly detection

Fig. 3 shows the proposed anomaly detection framework using BiGAN. In the training phase, the pre-processed healthy deployment datasets are applied to train the generator G, encoder E and discriminator D concurrently. Assistive hints are applied to guide the generator G and encoder E training periodically. In the test phase, the reconstruction error and discriminator feature hint error jointly represent the degree of an anomaly.

Fig. 4 shows the workflow of the proposed anomaly detection system underwater gliders. The model is trained using normal deployment data, i.e. no anomalies are included in the training dataset. It should be noted that this framework is unsupervised, as no labelling is required on the normal deployment data (Radford et al., 2015; Schlegl et al., 2017). During training, the model is tested periodically with synthetic sensor faults by manually setting one or more sensor readings to their lower bounds to check whether the model can detect synthetic faults. Once the model has been checked, it will be applied to detect anomalies for vehicles of the same type within a fleet. If the model has learned the distribution $x \sim p_{data}(x)$ of the training data, it should be able to output a high anomaly score that represents the degree of an anomaly. The anomaly score should be close to zero if the input data is normal. When anomalies happen, the system is expected to output a high anomaly scores that represent the degree of anomalies.

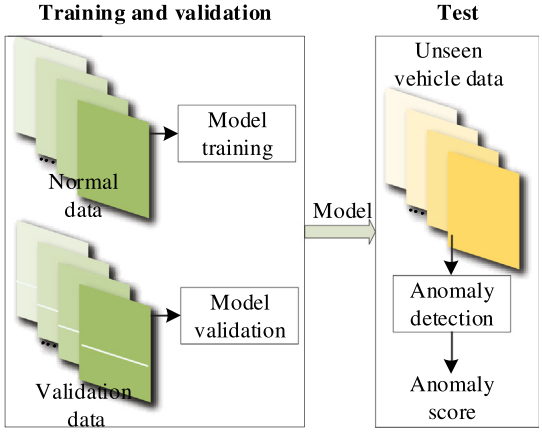


Fig. 4. Workflow of unsupervised anomaly detection using GAN for underwater gliders.

5. Training and validation

5.1. Data processing

Fig. 5 illustrates the data processing process preparing the training and validation datasets, using the two healthy deployments of glider units 345 and 397 in 2014 and 2015. The dive cycles of the multivariate time series datasets are filtered to remove cycles with insufficient data points (less than ten data points for each sensor) to maintain cycles carrying sufficient features. Unified timelines with a time step of 5 s are subsequently applied to the filtered cycles by linear interpolation of all the remaining sensor measurements. The interpolated data are then normalised to the range of $[0, 1]$. Random data patches with 64 time steps are sampled evenly from the valid dive cycles in each dataset. The data patches are augmented as $a \times b$ matrices, where a is the number of sensors, b is the number of time steps (64) for each data patch, so that the training dataset is ready for the training of the anomaly detection system. To monitor and check the performance of the anomaly detection system performance during training, synthetic sensor anomalies are injected into the data patches by setting a number of sensor measurements to their minimum values. Note that the sensors with anomalies are randomly chosen for each validation data patch. For the test datasets, a similar data processing flow has been followed. However, it should be noted that ten random data patches are sampled from each dive cycle in the test datasets.

5.2. Training

The proposed anomaly detection system is implemented in Python 3.8 and TensorFlow v2.4.1. The encoder network consists of: an input layer that receives the flattened data patches; four sequentially connected hidden dense layers that are followed by their own batch normalisation, leaky ReLU activation and dropout (0.1 dropout rate) layers; and an output layer with the size of 256 activated by sigmoid. The encoder and generator network structures are inversely similar to each other. The discriminator processes (x, z) to output a feature f from the feature layer and its prediction (a scalar) on whether (x, z) is from the training dataset. The hidden layers of the discriminator

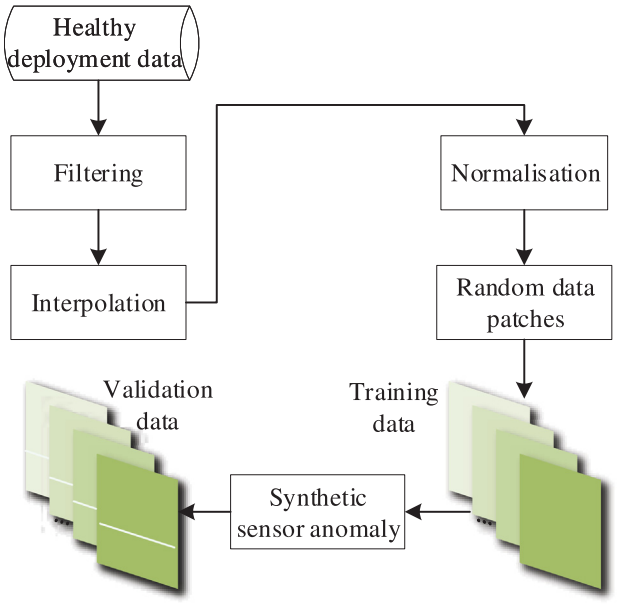


Fig. 5. Data processing procedure applied to prepare the training and validation datasets using the two deployments of units 345 and 397.

are configured the same as that of the encoder. The Adam optimiser is applied to update $\mathcal{L}_{BiGAN}$ and $\mathcal{L}_{hint}$ is with a learning rate of 1.0×10^{-5}. Note that $\mathcal{L}_{hint}$ is updated every $k = 10$ training iterations of $\mathcal{L}_{BiGAN}$. The coefficient κ that adjusts the weights of the reconstruction and feature losses is set as $\kappa = 2$. The mini-batch size is $M = 256$. The training dataset includes 1.0×10^5 data patches extracted from two healthy deployments of UGs. The test dataset includes 4.3×10^4 data patches extracted from the nine test deployments (see Table 1). The training is terminated after $t = 30{,}000$ training iterations of $\mathcal{L}_{BiGAN}$ (51 min on a Nvidia V100 GPU).

5.3. Validation using synthetic anomalies

Fig. 6 shows the validation process using synthetic sensor faults (every $n = 100$ training iterations of D, G and E). For the validation without faults, the anomaly score stabilises to a value slightly less than 0.002. For the validation data samples with 1, 2 and 3 abnormal sensors, the anomaly scores stabilise to the values around 0.016, 0.022 and 0.028, respectively. The stabilised anomaly scores suggest that the training of the anomaly detection system has converged. It should be noted that the converged anomaly score (0.002) for normal data is not exactly zero, which suggests the reconstruction and discriminator feature residuals still exist in low magnitudes. Nevertheless, this value is an order of magnitude lower compared to the ones with synthetic sensor anomalies.

5.4. Ablation study

An ablation study has been conducted to confirm the effectiveness of the assistive hints added to guide the BiGAN-based anomaly detection system training. Removing the assistive hint $\mathcal{L}_{hint}$ (Eq. (5)) leads to the BiGAN unable to reconstruct normal enquiry data accurately. As shown in Fig. 7, when $\mathcal{L}_{hint}$ is removed, the model has converged to a state of not being able to differentiate the four synthetic anomalies, suggesting the added hint has effectively guided the training of the BiGAN model (also see Fig. 6).

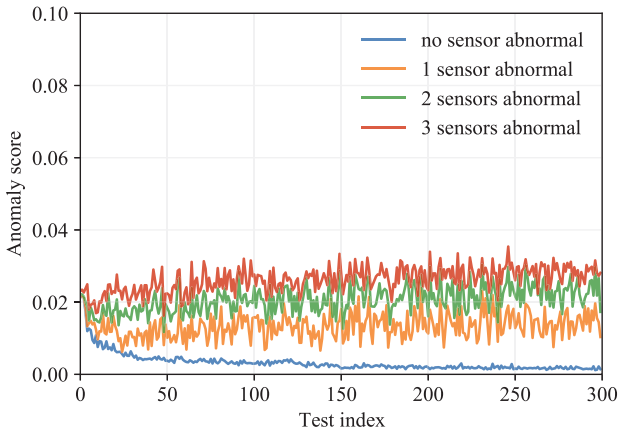


Fig. 6. Periodic algorithm validation in training using synthetic sensor faults with 0, 1, 2 and 3 abnormal sensors randomly chosen. The model is tested every 100 training iterations of $\mathcal{L}_{BiGAN}$. The records of the randomly chosen sensors are manually set to their minimum values.

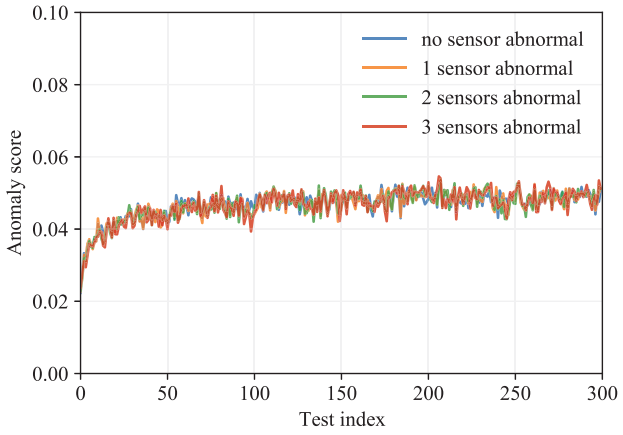


Fig. 7. Periodic algorithm validation in training using synthetic sensor faults when no hints are added to guide the training.

Table 3
Anomaly scores of the test deployments.

UG	Mean	Standard deviation
unit 419	0.00188	0.00026
unit 399	0.00235	0.00053
unit 492	0.00470	0.00113
unit 423	0.00353	0.00081
unit 424	0.00434	0.00024
unit 194	0.00303	0.00026
unit 304	0.00296	0.00067
unit 436	0.00319	0.00080
unit 345	0.00530	0.00097

6. Field test results and discussion

The anomaly detection system is tested using the deployment datasets detailed in Table 1. Ten data patches are randomly sampled from each dive cycle, which are then applied to get an averaged anomaly score for a dive cycle. Hence, the anomaly detection system applies to any dive cycles with a sufficient number of data points and can output an anomaly score for NRT monitoring when the dive cycle data is received. Table 3 details the mean and standard deviation values of the test deployments (calculated using dive cycle values in the deployments). The test deployments include one healthy deployment of unit 419. The average anomaly score of unit 419 will be applied as the baseline to assess the relative levels of the anomalies.

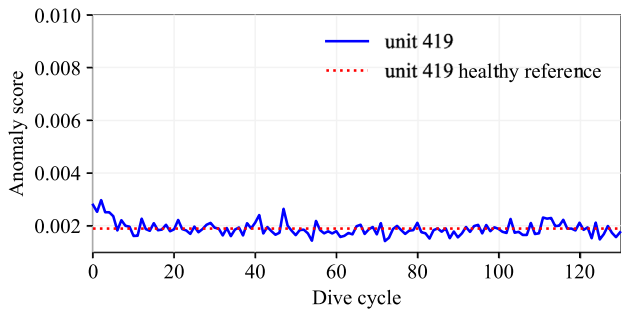


Fig. 8. Test using a dataset of a healthy glider deployment dataset collected by unit 419 in 2015. The healthy reference is the average anomaly score of unit 419 over this deployment and will be applied as the baseline to assess the anomaly levels of other deployments.

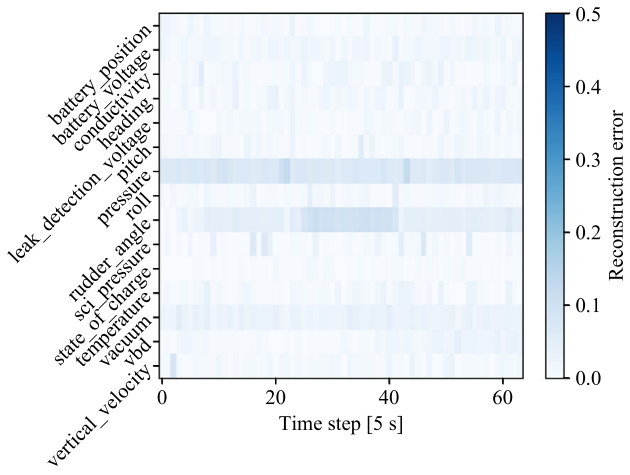


Fig. 9. Reconstruction errors of a typical data patch of unit 419 in a healthy deployment. Note that the maximum reconstruction error could be up to 1.0; the upper limit of the colour bar is set as 0.5 to better visualise the errors.

The reconstructed sensor data is compared against the original input query data. The sensors with high anomaly score contributions are annotated by the system to alert the pilot. For the ith sensor's reading at time step j, its reconstruction error $\delta_{i,j}$ is defined as the absolute value of the difference between the enquiry data $x_{i,j}$ and the reconstructed data via BiGAN:

$$\delta_{i,j} = |x_{i,j} - G(E(x))_{i,j}| \tag{6}$$

which will be highlighted by its magnitude to visualise anomalies on sensor readings.

6.1. Healthy glider deployment

As shown in Fig. 8, for a healthy deployment of unit 419 in 2015, the anomaly scores distribute evenly around their average value in general (0.00188 ± 0.00026). The initial anomaly score starts from the value of 0.0030 and decreases to the average anomaly score after ten dive cycles. The initial high anomaly score is likely caused by the shallow trial dives with significant dynamic effects at the start of a glider deployment. Note that the average anomaly score of 0.00188 very close to the converged anomaly score in the validation test without anomaly. The slight variance in the anomaly score throughout the cycles for this healthy glider deployment suggests that the proposed anomaly detection system can accurately reconstruct the data samples similar to those it encountered in the training phase without giving false indications of anomalies.

Fig. 9 shows the reconstruction errors of a typical data patch from unit 419 in a healthy deployment. Errors with small magnitudes distribute among most sensors, suggesting that the anomaly detection system has reconstructed the data patch with high accuracy. Ideally, the anomaly detection system should learn the healthy deployment datasets' data distributions applied for training. The low magnitudes of reconstruction errors for unit 419 suggest that the anomaly detection system has learned the patterns of the training datasets.

6.2. Deployments with biofouling

Fig. 10 details the anomaly scores of two deployments with biofouling, i.e. deployments of unit 399 (0.000235 ± 0.00053) with naturally accumulated biofouling in 2015 (Fig. 10a) and unit 492 (0.00470 ± 0.00113) with simulated biofouling in 2020 (Fig. 10b). For unit 399, shallow trial dives with dynamic effects at the beginning of the deployment lead to high initial anomaly scores. The anomaly score increases gradually from dive cycle 200, which is likely associated with marine growth (Haldeman et al., 2016). The growing anomaly score is in line with the increasing drag coefficient deduced through model-based and other data-driven approaches in Anderlini et al. (2020b). The similar growth of the anomaly score and the drag coefficient suggest that the proposed BiGAN-based anomaly detection system can capture slowly growing anomalies, even though it is trained with deployment datasets collected by other gliders in different missions. As shown in Fig. 10b, for the deployment of unit 492 with drag simulators added to the UG to simulate extreme levels of marine growth (Anderlini et al., 2020b), the anomaly scores of the dive cycles are distributed around their average value of 0.00470 which is high above the baseline deduced from the healthy deployment of unit 419. The average anomaly score of unit 492 is close to the anomaly values of the final dive cycles of unit 399.

As shown in Fig. 11a, the system has reconstructed a data patch from the final stage of this deployment with high reconstruction errors due to possible biofouling. The relatively high reconstruction errors can be observed from the sensors including VBD, state_of_charge, depth and pressure. Even stronger highlights of reconstruction errors can be observed in Fig. 11b from a data patch of unit 492 with simulated biofouling.

6.3. Deployments with OMG

Fig. 12 shows the anomaly scores of two deployments (unit 423 and unit 424) with OMG. The two deployments' anomaly scores are similar to each other, i.e. 0.00353 ± 0.00081 for unit 423 and 0.00434 ± 0.00024 for unit 424, respectively. The additional turbulence probes have affected the gliders' hydrodynamic characteristics, leading to increased drag coefficients and high negative buoyancy offsets and, consequently, to higher anomaly scores. It is worth noting that no dramatic change in the anomaly score magnitudes can be observed from the two deployments with OMG, suggesting that it is unlikely other anomalies occurred during the two deployments. The proposed anomaly detection system appears to be effective in detecting anomalies caused by increased drag coefficient consistently throughout the deployments of the glider units.

Fig. 13 shows the reconstruction errors of a typical data patch of unit 423 and unit 424. Obviously, the rudder_angle, battery_position, battery state_of_charge and VBD sensors that are the most highlighted ones, as the vehicle becomes less manoeuvrable due to the higher drag. Although errors in lower magnitudes can also be observed for other sensors, the most striking difference is the vacuum sensor which indicates a different pressure inside the pressure hull was used for the OMG as compared with the standard gliders.

6.4. Deployment with angle of list

Fig. 14 details the anomaly scores of the unit 194 deployment (0.00303 ± 0.00026) in 2017 which experienced an angle of approximately 9° list due to a pre-deployment error. Fluctuations of the anomaly scores can be observed, which suggests that the operational

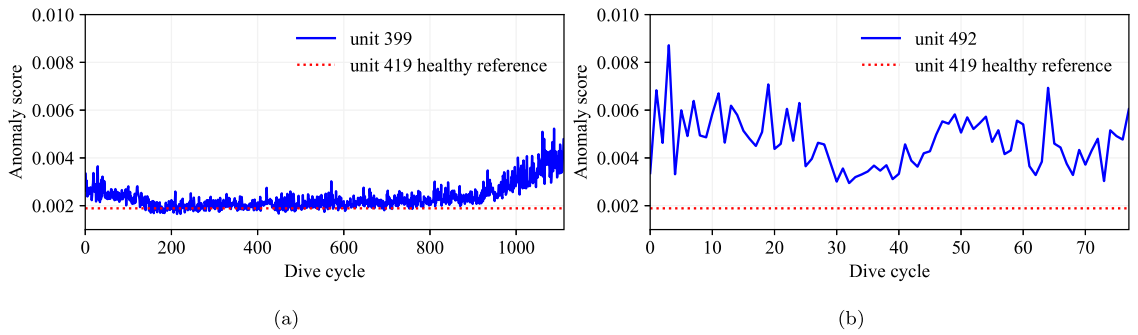


Fig. 10. Test using datasets of two deployments with biofouling: (a) unit 399 in its final stage of deployment with a high anomaly score caused by naturally accumulated biofouling and (b) unit 492 with simulated high levels of biofouling.

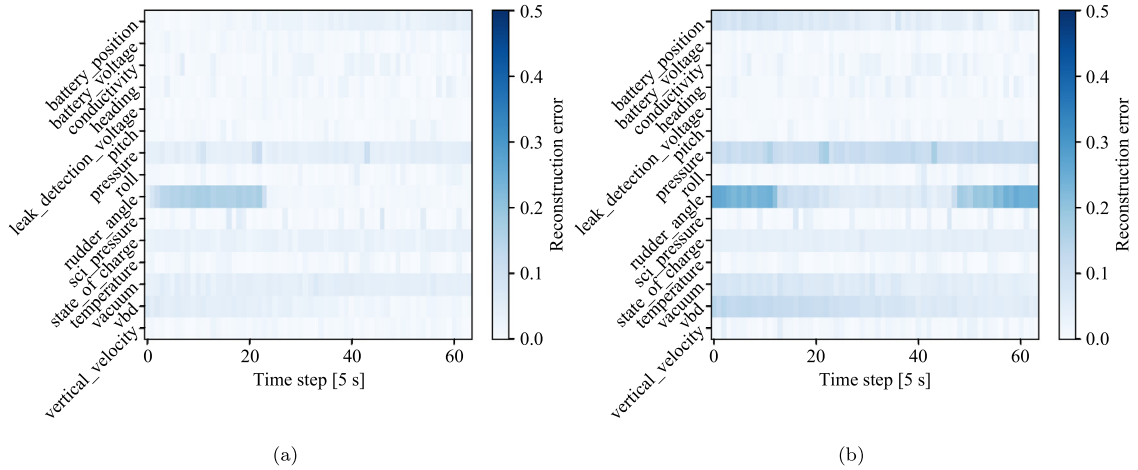


Fig. 11. Reconstruction errors of typical data patches with biofouling: (a) unit 399 in its final stage of deployment with a high anomaly score caused by naturally accumulated biofouling and (b) unit 492 with simulated high levels of biofouling.

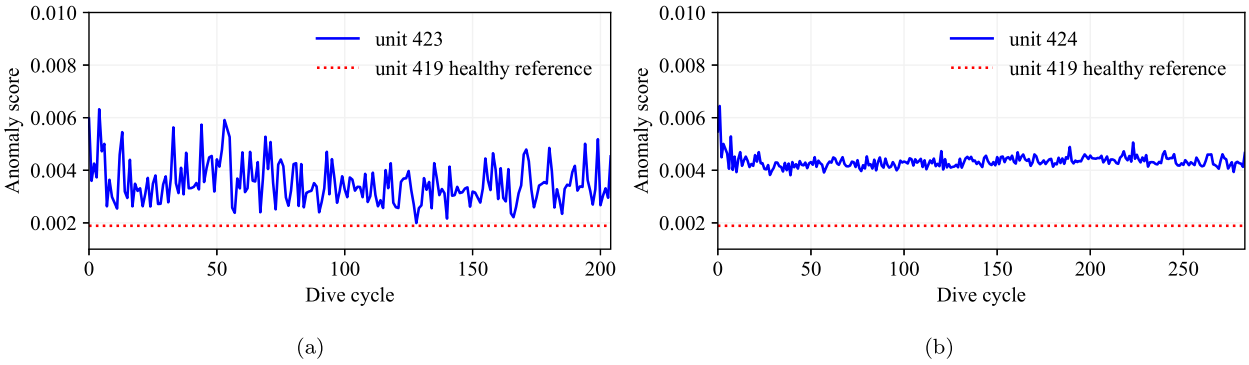


Fig. 12. Test using datasets of two deployments with OMG: (a) deployment of unit 423 in 2015 and (b) deployment of unit 424 in 2015.

status of this glider were relatively unstable compared to the deployments discussed in Section 6.1. In this deployment, the angle of list caused asymmetric drag and lift forces. Consequently, the average anomaly scores presented in Fig. 14 are 61.2% higher than that of the unit 419 baseline deployment. In addition, an increasing trend of the anomaly score can be observed in the final stage of the deployment (after 680 dive cycles), which is possibly due to the pilot's control decisions. It appears that the anomaly detection system can detect the transient effects in the abnormal glider status.

Fig. 15 presents the reconstruction errors of a data patch extracted from the deployment of unit 194, which encountered a significant angle of list of approximately 9°. The underwater glider attempted to compensate the list angle by applying control actions to the rudder. However, no apparent errors can be observed from the roll and pitch signals, which could be due to the errors are low in magnitudes compared to the maximum readings of these signals.

6.5. Deployments with loss of wing

Fig. 16 shows the anomaly scores of two deployments with the loss of one wing, i.e. unit 304 (0.00296 ± 0.00067) with the loss of the right wing (Fig. 16a) and unit 436 (0.00319 ± 0.00080) with the loss of the left wing (Fig. 16b) in 2019. As shown in Fig. 16a, fluctuating high anomaly scores are present for the early dive cycles (before 500 dive cycles) of unit 304, which is mainly due to the very shallow dive depth, where dynamic effects of the vehicle and the variation in the oceanic sensor data are more significant (Anderlini et al., 2020a). The anomaly scores of unit 304 start to increase abruptly from dive cycles

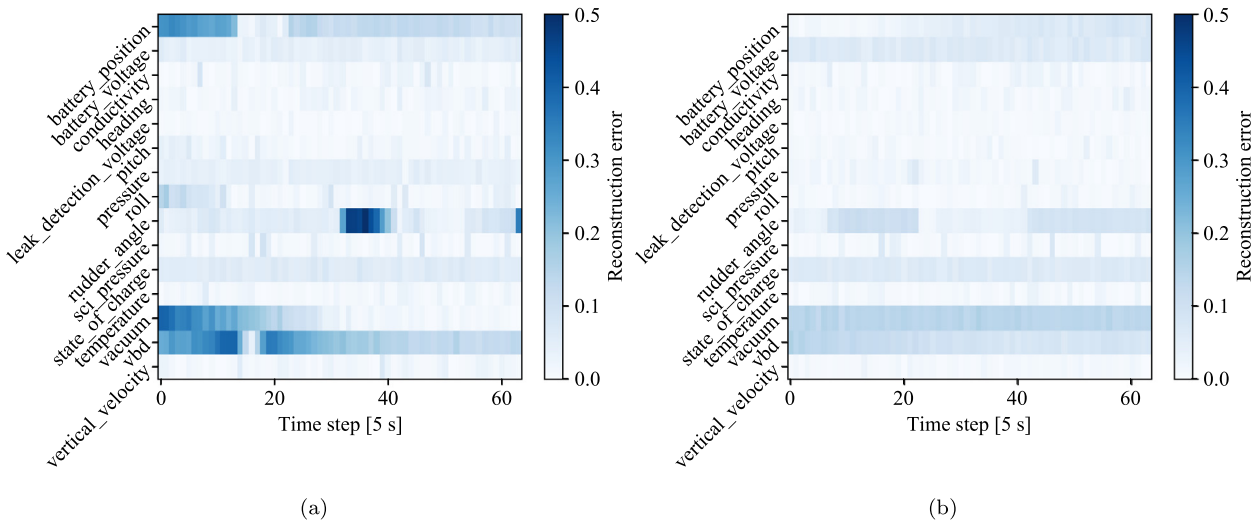


Fig. 13. Reconstruction errors of typical data patches with OMG: (a) unit 423 and (b) unit 424.

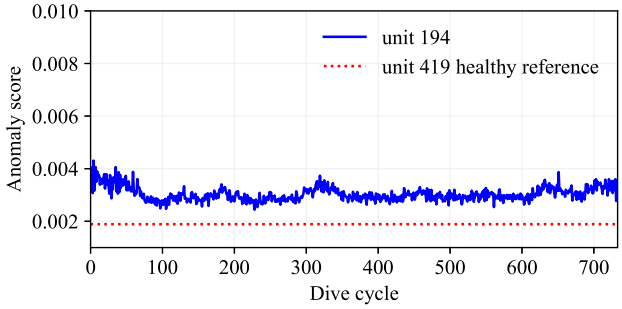


Fig. 14. Test using dataset of a deployment dataset collected by unit 194 with angle of list in 2017.

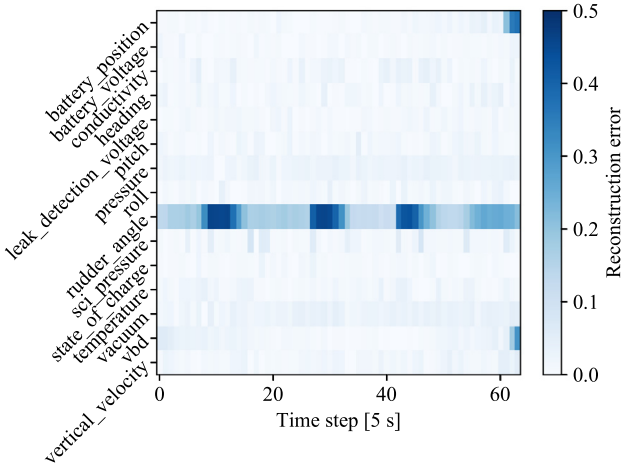


Fig. 15. Reconstruction errors of typical data patch of unit 194 with angle of list.

510 to 560, suggesting that the anomaly detection system has detected an unusual pattern. This anomaly has also been detected in Anderlini et al. (2020a), which corresponds to unit 304's loss of right wing in this deployment. As shown in Fig. 16b, the anomaly score of unit 436 jumps to 0.0060 within one dive cycle after 230 dives, suggesting something very unusual happened within that dive cycle. In addition, the anomaly scores of dive cycles from 120 to 220 suggest that unit 436 may have encountered an unusual event, which subsequently caused a delayed but instant loss of its left wing.

Fig. 17 highlights the reconstruction errors of data patches from units 304 and 436, which lost the right wing and the left wing in their deployments, respectively. The rudder_angle signal is the most highlighted for both units 304 and 436, suggesting the gliders had frequently attempted to use their rudders to compensate for the imbalance caused by the wing loss. In addition, the battery state_of_charge unit 304 signal also shows a high level of reconstruction errors due to the compensating control actions that consumed excessive energy from the battery. Note that the reconstruction error shown in Fig. 17a is a late-stage data patch of the deployment; the accumulated excessive energy consumption from the battery becomes apparent in the case of unit 304.

6.6. Deployment with strong environmental disturbances

Fig. 18 details the anomaly scores of the dive cycles within the deployment of unit 345 in 2019. This deployment was subject to strong disturbances (e.g. transverse ocean currents). Consequently, the average of anomaly scores (0.00530 ± 0.00097) is 181.9% higher than the baseline of unit 419. It should be noted that another dataset collected by unit 345 in 2014 has been included in the training dataset. The system has identified the anomalies of unit 345 in 2019.

Fig. 19 shows the reconstruction errors of a data patch of the deployment of unit 345 that experienced strong ocean disturbances such as transverse ocean currents. Apparent errors can be observed from the sensors, including battery_voltage, rudder_angle, VBD and state_of_charge. Unlike other scenarios that have been discussed, battery voltage signal deviates from the pattern the anomaly detection system has learned, which is probably due to the glider had to make frequent adjustments to its rudder to overcome the strong disturbances causing excessive power consumption from the battery, which led to a lower battery voltage.

6.7. Summary

The proposed anomaly detection system successfully identifies anomalies for a fleet of gliders operating in different deployments over different times. In addition, the highest observed anomaly score among all tests is around 0.010, which is significantly lower compared to the anomaly score obtained for simulated sensor faults. This indicates that it is significantly more challenging for humans to detect actual anomalies with abnormal patterns distributed over multiple sensor readings in low magnitudes.

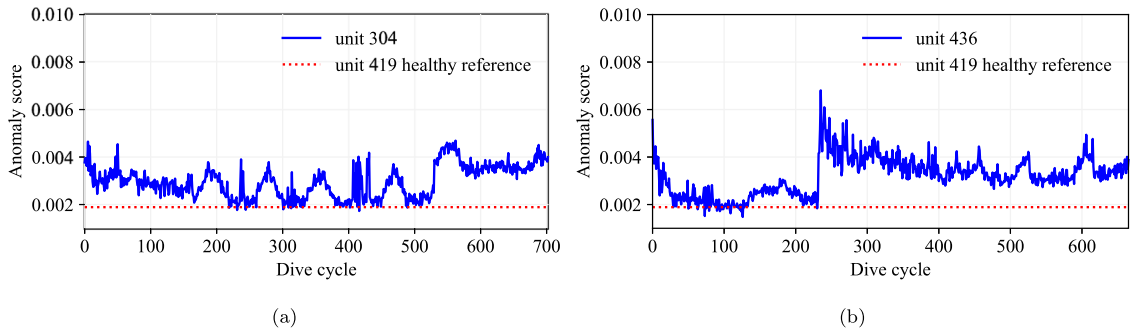


Fig. 16. Test using datasets of two deployments with wing loss: (a) deployment of unit 304 with the loss of the right wing in 2019 and (b) deployment of unit 436 with the loss of the left wing in 2019.

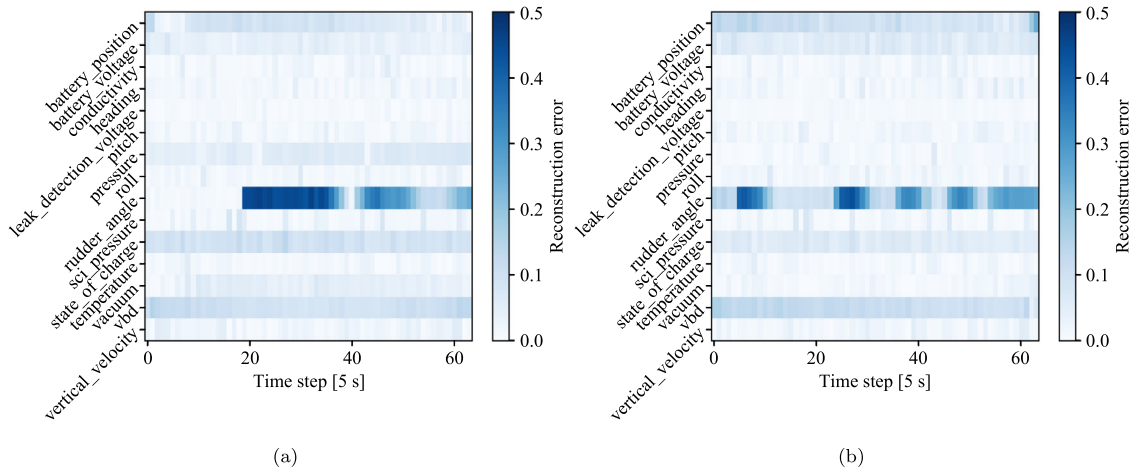


Fig. 17. Reconstruction errors of typical data patches of losses of wings: (a) unit 304—loss of the right wing and (b) unit 436—loss of the left wing.

7. Sensitivity study of NRT data decimation settings

In this section, the influence of NRT data decimation settings over anomaly detection accuracy is investigated. The sensitivity study is implemented by varying the data sampling intervals to simulate the gliders' decimation process. The sampling intervals of dt include 5 s, 10 s, 30 s, 60 s, 120 s and 240 s. The gaps between two adjacent timestamps are filled by linear interpolation to match the enquiry data structure with a fixed dt of 5 s. It should be noted that the results acquired with the decimation interval of 5 s are deemed as ground truth as they are completely based on the most detailed recovery mode data. For each decimation setting, the anomaly detection result of a dive cycle (either positive or negative with anomaly) is compared with its corresponding ground truth. A result is deemed as accurate if it matches its ground truth, otherwise it is inaccurate. The anomaly detection accuracy is calculated by dividing the number of accurate detections with total detection times.

7.1. Individual sensors

In Group A, for each of the 15 sensors under investigation, at one time, one of the sensor readings are re-sampled with different decimation settings whilst other sensor data remain as originally recorded. As a result, the influences of individual sensors can be explored. The anomaly detection accuracy is calculated using all the nine test deployments, i.e. it is an overall accuracy over all the test deployments. As shown in Table 4, only 3 of the sensors, i.e. battery_position, rudder_angle and VBD have minor impacts on the anomaly detection accuracy, whist the decimation of all the other sensors do not present observable impacts on the anomaly detection accuracy. It should be noted that battery_position, rudder_angle and VBD signals correspond to the actuators directly controlling the UG. Hence, this suggests that the anomalies included in the test datasets are highly likely to have caused the actuators to operate differently than in normal scenarios.

Table 4
Sensitivity study results on individual sensor decimations.

Sensor	dt [s]					
	5	10	30	60	120	240
battery_position	1.000	0.998	0.992	0.986	0.981	0.972
battery_voltage	1.000	1.000	1.000	1.000	1.000	1.000
conductivity	1.000	1.000	1.000	1.000	1.000	1.000
heading	1.000	1.000	1.000	1.000	1.000	1.000
leak_detection_voltage	1.000	1.000	1.000	1.000	1.000	0.998
pitch	1.000	1.000	0.999	0.999	0.999	0.998
pressure	1.000	1.000	1.000	1.000	0.999	0.998
roll	1.000	1.000	1.000	1.000	1.000	1.000
rudder_angle	1.000	0.996	0.987	0.975	0.968	0.968
sci_pressure	1.000	1.000	1.000	1.000	1.000	1.000
state_of_charge	1.000	1.000	1.000	1.000	1.000	1.000
temperature	1.000	1.000	1.000	1.000	1.000	1.000
vacuum	1.000	1.000	1.000	0.999	0.996	0.991
VBD	1.000	0.999	0.992	0.984	0.976	0.972
vertical_velocity	1.000	0.998	0.997	0.995	0.993	0.993

7.2. All sensors

Table 5 presents the sensitivity study results on decimations of all sensors. Data from all the 15 sensors are decimated simultaneously for the sampling intervals of 5 s, 10 s, 30 s, 60 s, 120 s and 240 s, respectively. Note that $dt = 5$ s is still deemed as ground truth. The minimum anomaly detection accuracy achieved is 90.2% for the

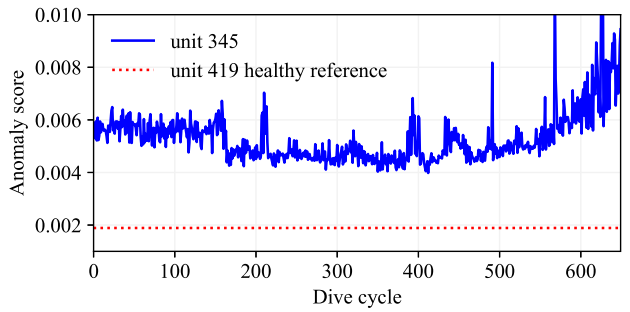


Fig. 18. Test using the dataset of a deployment collected by unit 345 with strong environmental disturbances in 2019.

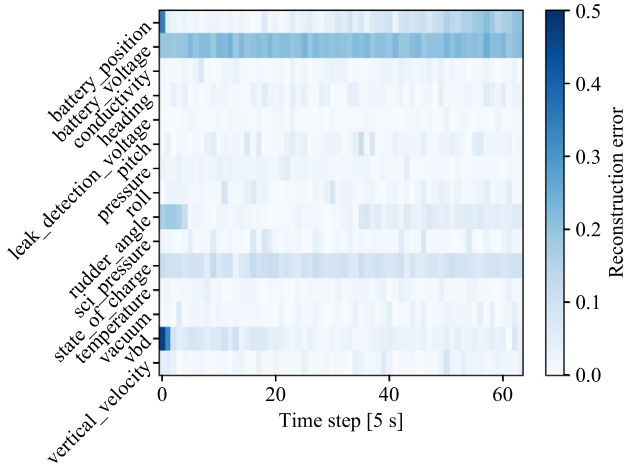


Fig. 19. Reconstruction errors of a typical data patch of unit 345 encountered strong environmental disturbances.

Table 5
Sensitivity study results on decimations of all sensors.

dt [s]	5	10	30	60	120	240
Accuracy	1.000	0.995	0.976	0.951	0.916	0.902

sampling interval of 240 s, which suggests that the proposed anomaly detection system is insensitive to the data decimation settings.

8. Conclusions

This work presents an unsupervised anomaly detection system with an improved training procedure of existing BiGAN. Data reconstruction and discriminator feature losses are adopted as assistive hints to periodically guide effective training of the BiGAN-based anomaly detection system. A novel data augmentation strategy of multi-sensor time series data is proposed to capture the transient features within data. The anomaly detection system can provide a thorough evaluation of a dive profile using multiple data patches extracted from the profile and is flexible with dive lengths. Compared to the authors' previous work, this study has endeavoured to apply more signals with improved signal reconstruction capability provided by the enhanced BiGAN structure. Although the method is proposed for the anomaly detection of UGs, the workflow developed does not require application-specific features. Therefore, it can be adapted for other application scenarios with multivariate time series data, e.g., high-performance computing system anomaly detection and aircraft turbulence detection.

The proposed anomaly detection system is trained using two healthy Slocum G2 glider deployments. Synthetic sensor faults are injected to the training dataset to check the anomaly detection system performance. Real-world collected datasets are applied to test the proposed anomaly detection system. The test results suggest that the BiGAN-based anomaly detection has successfully detected anomalies caused by biofouling, bulky sensors, angle of list, losses of wings and strong disturbances, whilst without giving false detections for healthy deployment. The unsupervised anomaly detection system has achieved satisfactory anomaly detection performance over a fleet of underwater gliders with minimal training data preparation. A sensitivity analysis of the decimation settings has shown that the anomaly detection system is insensitive to the data decimation settings. The outcome of the work will support the over-the-horizon operation of marine autonomous systems within the National Marine Equipment Pool at the National Oceanography Centre.

Although the system can highlight the anomalies on sensor readings, it can still be challenging for humans to accurately determine the types of anomalies, including known and unknown ones. An intelligent anomaly classification system will be developed to classify anomalies automatically using deep learning in further work. The unsupervised learning anomaly detection method proposed in this study requires only healthy deployment datasets, making it generic to different types of anomalies, even unknown ones. Such a unique feature also makes it ideal for annotating dive cycles with only deployment-level anomaly information. The annotated dive cycles can be further applied to train supervised or semi-supervised fault diagnostics models.

CRediT authorship contribution statement

Peng Wu: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing - original draft. **Catherine A. Harris:** Resources, Writing - review & editing. **Georgios Salavasidis:** Resources, Writing - review & editing. **Alvaro Lorenzo-Lopez:** Resources, Writing - review & editing. **Izzat Kamarudzaman:** Resources, Writing - review & editing. **Alexander B. Phillips:** Resources, Writing - review & editing. **Giles Thomas:** Writing - review & editing. **Enrico Anderlini:** Conceptualization, Methodology, Formal analysis, Data curation, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to acknowledge the support received by Stephen A. Woodward, an expert glider pilot at the National Oceanography Centre, in identifying suitable deployments of underwater gliders to be used in this study.

The research has been funded by the Lloyd's Register Foundation International Programme at the University of York as part of demonstrator project "Assuring Long-term Autonomy through Detection and Diagnosis of Irregularities in Normal operation (ALADDIN)". C. A. Harris, G. Salavasidis, A. Lorenzo Lopez, I. Kamarudzaman and A. B. Phillips' contributions are also funded under the NERC/ISCF Oceanids programme, UK. The results for simulated biofouling (unit 492) were obtained thanks to funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731103.

References

Anderlini, E., Harris, C.A., Salavasidis, G., Lorenzo, A., Phillips, A.B., Thomas, G., 2020a. Autonomous detection of the loss of a wing for underwater gliders. In: IEEE/OES Autonomous Underwater Vehicle Symposium (AUV). IEEE/OES, St John's, NF, Canada, pp. 1–6. http://dx.doi.org/10.1109/AUV50043.2020.9267895.

Anderlini, E., Real-arce, D.A., Morales, T., Barrera, C., Thomas, G., 2021a. An innovative marine growth detection system for underwater gliders. J. Ocean. Eng. (June), 1–15. http://dx.doi.org/10.1109/JOE.2021.3066373.

Anderlini, E., Salavasidis, G., Harris, C.A., Wu, P., Lorenzo, A., Phillips, A.B., Thomas, G., 2021b. A remote anomaly detection system for slocum underwater gliders. Ocean Eng. unpublished.

Anderlini, E., Thomas, G., Woodward, S.C.A., Real-Arce, D.A., Morales, T., Barrera, C., Hernández-Brito, J.J., 2020b. Identification of the dynamics of biofouled underwater gliders. In: IEEE/OES Autonomous Underwater Vehicle Symposium (AUV). IEEE/OES, St John's, NF, Canada, pp. 1–6. http://dx.doi.org/10.1109/AUV50043.2020.9267919.

BODC, 2019. Glider inventory. URL: https://www.bodc.ac.uk/data/bodc_database/gliders/.

Borghesi, A., Bartolini, A., Lombardi, M., Milano, M., Benini, L., 2019. A semisupervised autoencoder-based approach for anomaly detection in high performance computing systems. Eng. Appl. Artif. Intell. 85, 634–644. http://dx.doi.org/10.1016/j.engappai.2019.07.008.

Crestani, D., Godary-Dejean, K., Lapierre, L., 2015. Enhancing fault tolerance of autonomous mobile robots. Robot. Auton. Syst. 68, 140–155. http://dx.doi.org/10.1016/j.robot.2014.12.015.

Department for Transport, 2019. Technology and Innovation in UK Maritime: The Case of Autonomy. Technical Report, London, UK.

Di Mattia, F., Galeone, P., De Simoni, M., Ghelfi, E., 2019. A survey on gans for anomaly detection. arXiv preprint arXiv:1906.11632.

Dias, P.S., Fraga, S.L., Gomes, R.M., Goncalves, G.M., Pereira, F.L., Pinto, J., Sousa, J.B., 2005. Neptus-a framework to support multiple vehicle operation. In: Europe Oceans 2005, Vol. 2. IEEE, pp. 963–968. http://dx.doi.org/10.1109/OCEANSE.2005.1513187.

Donahue, J., Krähenbühl, P., Darrell, T., 2016. Adversarial feature learning. arXiv preprint arXiv:1605.09782.

Dumoulin, V., Belghazi, I., Poole, B., Mastropietro, O., Lamb, A., Arjovsky, M., Courville, A., 2016. Adversarially learned inference. arXiv preprint arXiv:1606.00704.

Farley, J., Morris, A.W., Jones, O.D., Harris, C.A., Lorenzo, A., 2019. Marine science from an armchair: A unified piloting framework for autonomous marine vehicles. In: IEEE Oceans. Marseille, France, pp. 1–10.

Fink, O., 2020. Data-driven intelligent predictive maintenance of industrial assets. In: Women in Industrial and Systems Engineering. Springer, pp. 589–605. http://dx.doi.org/10.1007/978-3-030-11866-2_25.

Fink, O., Wang, Q., Svensén, M., Dersin, P., Lee, W.-J., Ducoffe, M., 2020. Potential, challenges and future directions for deep learning in prognostics and health management applications. Eng. Appl. Artif. Intell. 92, 103678. http://dx.doi.org/10.1016/j.engappai.2020.103678.

Freddi, A., Longhi, S., Monteriù, A., 2013. Actuator fault detection system for a remotely operated vehicle. In: IFAC Proceedings Volumes (IFAC-PapersOnline), Vol. 46. (33 PART 1), IFAC Secretariat, pp. 356–361. http://dx.doi.org/10.3182/20130918-4-JP-3022.00050.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. In: Advances in Neural Information Processing Systems. pp. 2672–2680.

Haldeman, C.D.I., Aragon, D.K., Miles, T., Glenn, S.M., Ramos, A.G., 2016. Lessening biofouling on long-duration AUV flights: Behavior modifications and lessons learned. In: MTS/IEEE Oceans. MTS/IEEE, Monterey, California, USA, http://dx.doi.org/10.1109/OCEANS.2016.7761236.

Hamilton, K., Lane, D.M., Brown, K.E., Evans, J., Taylor, N.K., 2007. An integrated diagnostic architecture for autonomous underwater vehicles. J. Field Robotics 24 (6), 497–526. http://dx.doi.org/10.1002/rob.20202.

Harris, C., Lorenzo, A., Jones, O., Buck, J., Kokkinaki, A., Loch, S., Gardner, T., Phillips., A.B., 2020. Oceanids C2: An integrated command,control and data infrastructure for the over-the-horizon operation of marine autonomous systems. Front. Mar. Sci. (submitted in 2019).

Hong, X., Mitchell, R.J., Chen, S., Harris, C.J., Li, K., Irwin, G.W., 2008. Model selection approaches for non-linear system identification: A review. Internat. J. Systems Sci. 39 (10), 925–946. http://dx.doi.org/10.1080/00207720802083018.

Li, D., Chen, D., Goh, J., Ng, S.-k., 2018. Anomaly detection with generative adversarial networks for multivariate time series. arXiv preprint arXiv:1809.04758.

Li, L., Yan, J., Wang, H., Jin, Y., 2020. Anomaly detection of time series with smoothness-inducing sequential variational auto-encoder. IEEE Trans. Neural Netw. Learn. Syst. http://dx.doi.org/10.1109/TNNLS.2020.2980749.

Madureira, L., Sousa, A., Braga, J., Calado, P., Dias, P., Martins, R., Pinto, J., Sousa, J., 2013. The light autonomous underwater vehicle: Evolutions and networking. In: 2013 MTS/IEEE OCEANS-Bergen. IEEE, pp. 1–6. http://dx.doi.org/10.1109/OCEANS-Bergen.2013.6608189.

Maritime UK, 2019. Being a Responsible Industry: Maritime Autonomous Surface Ships (MASS) UK Industry Conduct Principles and Code of Practice. Technical Report.

Michau, G., Fink, O., 2021. Unsupervised transfer learning for anomaly detection: Application to complementary operating condition transfer. Knowl.-Based Syst. 216, 106816. http://dx.doi.org/10.1016/j.knosys.2021.106816.

Mutlu, U., Alpaydin, E., 2020. Training bidirectional generative adversarial networks with hints. Pattern Recognit. 107320. http://dx.doi.org/10.1016/j.patcog.2020.107320.

Pang, G., Shen, C., Cao, L., Hengel, A.V.D., 2021. Deep learning for anomaly detection: A review. ACM Comput. Surv. 54 (2), 1–38. http://dx.doi.org/10.1145/3439950.

Pinto, J., Dias, P.S., Martins, R., Fortuna, J., Marques, E., Sousa, J., 2013. The LSTS toolchain for networked vehicle systems. In: 2013 MTS/IEEE OCEANS-Bergen. IEEE, pp. 1–9. http://dx.doi.org/10.1109/OCEANS-Bergen.2013.6608148.

Raanan, B.-Y., Bellingham, J.G., Zhang, Y., Kemp, M., Kieft, B., Singh, H., Girdhar, Y., 2016. Automatic fault diagnosis for autonomous underwater vehicles using online topic models ben-yair. In: MTS/IEEE OCEANS. Monterey, CA, USA, http://dx.doi.org/10.1109/OCEANS.2016.7761139.

Raanan, B.Y., Bellingham, J., Zhang, Y., Kemp, M., Kieft, B., Singh, H., Girdhar, Y., 2018. Detection of unanticipated faults for autonomous underwater vehicles using online topic models. J. Field Robotics 35 (5), 705–716. http://dx.doi.org/10.1002/rob.21771.

Radford, A., Metz, L., Chintala, S., 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.

Ray Harris, W., 2015. Anomaly Detection Methods for Unmanned Underwater Vehicle Signature redacted Signature redacted Signature redacted (M.Sc. thesis). Massachusetts Institute of Technology.

Sakurada, M., Yairi, T., 2014. Anomaly detection using autoencoders with nonlinear dimensionality reduction. In: Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis, pp. 4–11.

Schlegl, T., Seeböck, P., Waldstein, S.M., Langs, G., Schmidt-Erfurth, U., 2019. F-anogan: Fast unsupervised anomaly detection with generative adversarial networks. Med. Image Anal. 54, 30–44. http://dx.doi.org/10.1016/j.media.2019.01.010.

Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G., 2017. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: International Conference on Information Processing in Medical Imaging. Springer, pp. 146–157. http://dx.doi.org/10.1007/978-3-319-59050-9_12.

Schofield, O., Kohut, J., Aragon, D., Creed, L., Graver, J., Haldeman, C., Kerfoot, J., Roarty, H., Jones, C., Webb, D., Glenn, S., 2007. Slocum gliders: Robust and ready. J. Field Robotics 24 (6), 474–485. http://dx.doi.org/10.1002/rob.20200, arXiv:10.1.1.91.5767.

Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W., Pei, D., 2019. Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 2828–2837. http://dx.doi.org/10.1145/3292500.3330672.

Sun, Y.S., Ran, X.R., Li, Y.M., Zhang, G.C., Zhang, Y.H., 2016. Thruster fault diagnosis method based on Gaussian particle filter for autonomous underwater vehicles. Int. J. Nav. Archit. Ocean Eng. 8 (3), 243–251. http://dx.doi.org/10.1016/j.ijnaoe.2016.03.003.

Teledyne Webb Research, 2012. Slocum G2 Glider Operators Manual. Technical Report, Teledyne Webb Research, www.webbresearch.com.

Thieme, C.A., Utne, I.B., 2017. Safety performance monitoring of autonomous marine systems. Reliab. Eng. Syst. Saf. 159, 264–275. http://dx.doi.org/10.1016/j.ress.2016.11.024.

Wang, Y., Zhang, M., 2006. Research on test-platform and condition monitoring method for AUV. In: IEEE International Conference on Mechatronics and Automation. Luoyang, China, http://dx.doi.org/10.1109/ICMA.2006.257448.

Webb, D.C., Simonetti, P.J., Jones, C.P., 2001. SLOCUM: An underwater glider propelled by environmental energy. IEEE J. Ocean. Eng. http://dx.doi.org/10.1109/48.972077.

Yao, F., Wang, F., Zhang, M., 2018. Weak thruster fault detection for autonomous underwater vehicle based on artificial immune and signal pre-processing. Adv. Mech. Eng. 10 (2), http://dx.doi.org/10.1177/1687814018758739.

Zenati, H., Romain, M., Foo, C.-S., Lecouat, B., Chandrasekhar, V., 2018. Adversarially learned anomaly detection. In: 2018 IEEE International Conference on Data Mining. ICDM, IEEE, pp. 727–736. http://dx.doi.org/10.1109/ICDM.2018.00088.

Zhou, C., Paffenroth, R.C., 2017. Anomaly detection with robust deep autoencoders. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 665–674.

Zolghadri, A., Cieslak, J., Efimov, D., Henry, D., Goupil, P., Dayre, R., Gheorghe, A., Leberre, H., 2015. Signal and model-based fault detection for aircraft systems. In: IFAC-PapersOnLine, Vol. 28. (21), pp. 1096–1101. http://dx.doi.org/10.1016/j.ifacol.2015.09.673.