

2.2.4.1. Verification of sensing requirements

Practical guidance - cross-domain and maritime

Authors: ALADDIN demonstrator project

Background

Sensing the RAS's state and its operating environment is crucial for its safe operation and is fundamental for the other elements in the RAS architecture. However, anomalies can develop during the RAS's operations, either due to the RAS's internal faults or external factors in its operating environment. Therefore, it is critical to embed autonomous systems that can correctly sense the anomalies so that mitigation actions can be in place with an updated maintenance schedule to minimise the safety hazards to the RAS and its operating environment. Regulations for condition monitoring systems can be found in the ISO standards 13372:2012 [2] and 26262-1:2018 [3] for machines and road vehicles, respectively. The data collection and management specification for automated vehicle trials can be found from [4].

This Body of Knowledge entry details the verification of the sensing requirements defined in Section ref¹. The verification is implemented through a field test (the FRONTIERS project) using an underwater glider deployed for 19 days in Mallorca, Spain, July 2021, to test the data-driven anomaly detection method developed in [1]. The developed and tested anomaly detection method is demonstrated through the application to underwater gliders (i.e. a Marine Autonomous System (MAS)). However, the method can be adapted and transferred to any other RAS for its generality.

Stage input and output artefacts

Required input/existing knowledge

- List of all sensors on the RAS the list should clearly label any redundant sensors on over-observed systems
- Signal output from all sensors, including both readings and the associated time stamps, synced for the RAS
- Knowledge of the dynamics of the RAS (desirable, especially for under-observed systems)
- Failure Mode Effect Analysis (FMEA) for the RAS (desirable), which can be completed according to the IEC 60812:2018 standards [5]
- Hazard and Operability study (HAZOP) for the RAS (desirable), which can be completed according to the IEC 61882:2016 standards [6]
- Formal description of the baseline RAS behaviour
- Metadata (e.g. calibration, system configuration, deployment reports or operator logs) to enable understanding of wider context

¹ https://www.york.ac.uk/assuring-autonomy/guidance/body-of-knowledge/implementation/2-2/2-2-1/2-2-1-1/cross-domain-maritime/

- Training and validation datasets for baseline RAS behaviour data from multiple RAS covering a wide range of normal missions would be ideal
- Testing datasets (including datasets collected previously and the datasets to be collected by the scheduled field tests) for anomalous RAS behaviour data of a wide range of anomalies would be ideal

Assumptions

- Sensing requirements, as defined in ref², are met (i.e. the installed sensors are appropriate for the RAS)
- The training, validation and testing datasets are collected by similar systems
- The incoming volume of data is manageable in real-time with the installed processing power

Expected outputs

• The system able to autonomously detect anomalies and inform the operator accordingly.

Stage input and output artefacts



Figure 1: Summary procedure to verify sensing requirements.

- 1. **Data cleaning**: the data applied for the training, validation and test datasets, as well as the RAS real-time implementation data need to be cleaned through the following steps
 - a. Signal processing and data treatment according to the data collection and management specification for automated vehicle trials [4]
 - b. Feature engineering:
 - i. Design additional virtual signals derived from dynamic models to better present the RAS's operating status and its environmentii. Data fusion to combine the virtual and actual signals
 - c. Depending on the type of RAS and characteristics of the data, some application may need filtering of transient effects and measurement noise to reduce the computational cost and improve real-time performance
 - i. In the case of Unmanned Aerial Vehicles, reserving full dynamic effects can be important
 - ii. In the case for underwater gliders, the RAS can spend significant time in steady-state operations. Reducing/removing transient effects can beneficial

² https://www.york.ac.uk/assuring-autonomy/guidance/body-of-knowledge/implementation/2-2/2-2-1/2-2-1-1/cross-domain-maritime/

- 2. **Dataset creation**: creation of suitable datasets for the development of the anomaly detection system. Data augmentation (see Section 2.3.1) is required in many applications to achieve desired testing/deployment performance of the model
 - a. Training dataset
 - b. Validation dataset for hyperparameter selection and training
 - c. Test dataset:
 - i. Previous deployment data has not been seen by the model during training
- 3. Development: the anomaly detection method is developed as described in [1]
- 4. **Training**: the anomaly detection system is trained using the previously prepared training dataset. Different signals can be selected as input during this stage to assess improvements in performance (prediction accuracy and computational cost)
- 5. **Validation**: the hyperparameters of the network, e.g. size of the Deep Neural Networks (DNNs), are selected with a sensitivity analysis to improve prediction accuracy and computational cost and prevent overfitting on the validation dataset
- Test: the ability of the anomaly detection system in correctly identifying and labelling baseline and anomalous behaviour is assessed for the test datasets. In particular, sensing deviations are quantified and visualised (see an example in Figure 2)
- 7. **Sensitivity analysis**: handling large quantities of data in real-time is challenging or there can be constraints associated with data transmission (e.g. if undertaken by satellite). Therefore, the data needs to be decimated, i.e. collected at larger time steps, during actual deployment. The sensitivity analysis on the data decimation settings indicates whether the proposed anomaly detection system is insensitive to the selected sample time in the data decimation
- 8. **Deployment**: the anomaly detection system's architecture is updated to enable its real-time deployment for the RAS operations



Figure 2: Example of anomaly detection results of underwater gliders on the test datasets: (a) unit 194 which experienced angle of list (internal anomaly) in 2017, and (b) unit 345 which had encountered strong environmental disturbances (external anomaly) in 2019

Method

Figure 3 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, the discriminator D of the BiGAN discriminates (x, E(x)) and (G(z), z), where G is the generator. Additionally, we add assistive hints to guide the BiGAN training. Two features (i.e. the reconstruction error and the discriminator D feature) jointly represent the anomaly score that indicates the degree of an anomaly [1]. A higher anomaly score suggests significant deviations from normal operating characteristics.



Figure 3: Structure of the BiGAN (Donahue et al., 2016)

Figure 4 illustrates the data processing process preparing the training and validation datasets, using the healthy deployment data. 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.



Figure 4: Data processing procedure applied to prepare the training and validation datasets using healthy deployment data.

Results

As shown in Figure 5A, the test started at t_0 (Figure 5B-a) and ended at t_8 (Figure 5B-b). The anomaly detection system based upon Bidirectional Generative Adversarial Networks (BiGAN) has successfully output anomaly scores over the test. The pitch angles for t_0 - t_1 , t_1 - t_2 , and t_2 - t_3 were set as 30°, 18°, and 26°, respectively. The glider's starboard wing was removed at t_3 (see Figure 5B-c). At t_4 , the starboard wing was restored while the port wing was removed (see Figure 5B-d). At t_5 , the port wing was restored while the balancing weight setting in the wing rails was adjusted from left-2 & right-5 to left-5 & right-2 (each pill is 15.5 g) (see the vehicle status in Figure 5B-e). At t_6 , the wrong battery position was applied. At t7, the battery position servo mode was set, and the balancing weight setting was changed to left-0 & right-3 (2 extra pills removed along the length of the vehicle in each wing rail, see Figure 5B-f). The glider was recovered at t_8 .





Figure 5: (A) anomaly scores over the FRONTIERS test in Mallorca, Spain, July 2021. (B) a: the glider at the beginning of the test, b: the glider before recovery at the end of the test, c: the glider with its starboard wing removed, d: the glider with its port wing removed, e: incorrectly ballasted glider, f: the balancing weight setting for the simulated trimming fault.

A data-driven anomaly detection system based on a BiGAN architecture with added hints was trained with data from deployments from the British Oceanographic Data Centre and the SOCIB portal. The system uses the decimated semi-real-time data signals from each dive of the glider sent ashore to calculate an anomaly score that can be used to determine whether anomalies are present on board the vehicle. Once trained, the system was validated using the data stream from the JERICO deployment. As can be seen in Figure 5A, as the 30° and 18° pitch settings were not included in the training dataset, high anomaly scores have been incorrectly returned at the start of the deployment for normal behaviour. However, the system was able to clearly detect the loss of wing, as removing the starboard and port wings resulted in high anomaly scores of similar magnitudes. Additionally, relatively high anomaly scores can be observed from t_5 to t_8 for the incorrect ballasting and trimming. In conclusion, the simulated faults were correctly detected, validating the proposed anomaly detection solution. The false positive at the start of the deployment suggest that the training data needs broadly capture the normal operating patterns of the RAS.

B

Advantages of the approach

- The applied approach is generic and is designed with an expandable architecture and can be extended to any RAS.
- Anything differs from the normal pattern can be detected.

Limitations of the approach

- The training dataset needs to broadly capture the normal pattern of the RAS operating state and the operating environment, states that vary from the pattern existing in the training dataset will be sensed as anomaly. False positive anomaly detection results may be provided if the training dataset is not sufficient.
- Training and validation require significant computational resources and time.
- A pre-trained anomaly detection model requires the same sensor/feature list in deployment

References

[1] P. Wu, C. A. Harris, G. Salavasidis, A. Lorenzo-Lopez, I. Kamarudzaman, A. B. Phillips, G. Thomas and E. Anderlini, "Unsupervised anomaly detection for underwater gliders using generative adversarial networks," Engineering Applications of Artificial Intelligence, vol. 104, p. 104379, 2021.

[2] ISO, 13372:2012(en): Condition monitoring and diagnostics of machines, 2012.

[3] ISO, 26262-1:2018(en): Road vehicles — Functional safety, 2018.

[4] British Standards Institution, PAS 1881:2020 Assuring the safety of automated vehicle trials and testing – Specification, 2021.

[5] IEC, 60812:2018: Failure modes and effects analysis (FMEA and FMECA), 2018.

[6] IEC, 61882:2016: Hazard and operability studies (HAZOP studies) - Application guide,2016.