Assessing Autonomous Ship Navigation using Bridge Simulators Enhanced by Cycle-Consistent Adversarial Networks

Journal Title
XX(X):1–9

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Andreas Brandsæter[1,2] and Ottar L. Osen[3]

Abstract

The advent of artificial intelligence and deep learning has provided sophisticated functionality for sensor fusion and object detection and classification which have accelerated the development of highly automated and autonomous ships as well as decision support systems for maritime navigation. It is, however, challenging to assess how the implementation of these systems affects the safety of ship operation. We propose to utilize marine training simulators to conduct controlled, repeated experiments allowing us to compare and asses how functionality for autonomous navigation and decision support affects navigation performance and safety. However, although marine training simulators are realistic to human navigators, it cannot be assumed that the simulators are sufficiently realistic for testing the object detection and classification functionality, and hence this functionality cannot be directly implemented in the simulators. We propose to overcome this challenge by utilizing Cycle-Consistent Adversarial Networks (Cycle-GANs) to transform the simulator data before object detection and classification is performed. Once object detection and classification are completed, the result is transferred back to the simulator environment. Based on this result, decision support functionality with realistic accuracy and robustness can be presented and autonomous ships can make decisions and navigate in the simulator environment.

Keywords

Simulator-Based Assessment, Autonomous Ships, MASS, Cycle-GAN, Validation and Verification, Safety Assessment, Marine Bridge Training Simulators

Introduction

The rapid development of new sensors and software, utilizing artificial intelligence and deep learning, facilitates the optimization and automation of a range of tasks in the maritime industry, including navigation. Several research and development projects are proposed and developed targeting concepts at different degrees of automation including ships with automated processes and decision support, remotely controlled ships with or without seafarers on board, and fully autonomous ships where the operating system of the ship is able to make decisions and determine actions by itself.[1]

The development of highly automated and autonomous ships and sophisticated decision support systems is motivated by the promise of benefits such as optimized operation, reduced crew costs and increased safety. It is, however, challenging to assess how the implementation of these systems affects the safety of ship operations. Although a large number of accidents are caused by human errors, removing humans will not automatically increase the safety of ship operations. This depends on the quality of the functionality and how it is used. Misunderstandings, wrong use of the functionality, malfunctioning user-interface, as well as bad or wrong decision proposals can have severe consequences, even if the functionality is used purely as decision support to a human navigator.[2] This is partly due to confirmation bias which can easily arise in complex navigation situations.[3] Over-trust in the technology is also identified as a challenge, causing humans not to monitor the situation carefully enough to be able to safely take control when needed.[4]

Obviously, the safety also depends on the accuracy and robustness of the system, and it is therefore utmost important that the functionality is thoroughly tested and that its capabilities and limitations are well documented. Unfortunately, verifying and testing this functionality, which is based on algorithms and methods from artificial intelligence and inductive learning, is inherently difficult.[5]

To demonstrate and prove the autonomous system's applicability and performance, controlled experiments should be executed in a simulator,[6,7] where experiments with identical initial conditions can be conducted repeatedly, allowing us to compare and analyze the performance of conventional ships and ships with different degrees of automation; such as human navigator supported by enhanced situational awareness or suggested decisions as well as fully autonomous unmanned navigation (Fig. 1). Note, however, that even when the initial conditions are identical,

[1] Department of Science and Mathematics, Volda University College
[2] Group Research & Development, DNV
[3] Department of ICT and Natural Sciences, NTNU – Norwegian University of Science and Technology

Corresponding author:
Andreas Brandsæter,
Volda University College, P.O. Box 500, 6101 Volda, Norway
Email: andreas.brandsaeter@hivolda.no

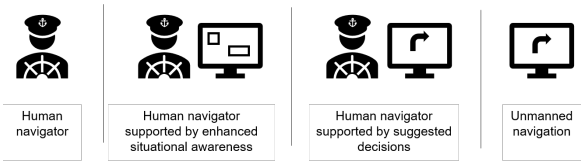


Figure 1. A navigation scenario can be repeated in the simulator to assess and compare how the introduction of autonomous functionality affects safety.

quantifying and assessing the performance and safety is indeed challenging.[8]

The use of marine simulators are well-established in seafarer training,[7] recognized by the International Convention on Standards of Training, Certification and Watchkeeping of Seafarers (STCW) of the International Maritime Organization (IMO) to demonstrate competence (assessment) and continued proficiency.[9] Maritime simulators are offered for various parts of seafarer training including full mission and multi-task bridge navigation simulators which are the focus of this paper. Such simulators are capable of simulating total shipboard bridge operations,[9] and includes instrumentation that "looks, feels and has functions like the real equipment used on board vessels [...] including necessary controls and functions for training ship handling, navigation, and communication",[10] and screens or 3D-views to visualize the surrounding world.[11]

A simulator used for mandatory training shall be approved by the relevant maritime administration, ensuring that the simulator includes an appropriate level of physical and behavioral realism.[9] But although maritime bridge simulators are realistic to human navigators, it cannot be assumed that the simulators are sufficiently realistic for testing the object detection and classification functionality of an autonomous system. Hence, we propose a novel approach utilizing Cycle-Consistent Adversarial Networks (Cycle-GANs) to transfer the simulator data to a real-world-like environment before autonomous functionality such as object detection and classification is performed. The inverse mapping is utilized to transfer the result of the autonomous functionality, such as for example bounding boxes, back to the simulator environment. The information which is transferred back to the simulator environment can now be utilized by human navigators with access to decision support functionality or by fully autonomous vehicles.

In the literature, the use of Cycle-Consistent Adversarial Networks related to autonomous ship navigation is limited to the development and improvement of detection methods.[12] We believe this paper is the first to propose and explain how this powerful image translation technique can be utilized to facilitate fair comparison and assessment of autonomous functionality in maritime bridge simulators.

In the following, we first discuss challenges to assurance with emphasis on machine learning based functionality. We also briefly describe a selection of currently available recommended practices, standards and assurance frameworks. Following this, we discuss challenges related to assessing a system's situational awareness, and based on this, we argue that the assessment should be focused directly on the navigation tasks. A short introduction to Cycle-Consistent Adversarial Networks is provided next, followed by a detailed description of our proposed framework for testing autonomous navigation functionality in bridge simulators. Finally, future work is outlined and concluding remarks are offered.

Challenges to Assurance of Machine Learning based Functionality

Assuring the safety of functionality for autonomous navigation and decision support is by many identified as one of the key barriers to large scale implementation. Koopman et al.[13] identifies a set of topics that must be specifically addressed for highly automated vehicles, including defining operational domain such as various weather conditions, machine learning faults, external operational faults such as other vehicles violating rules, high residual unknowns such as requirements gaps, and lack of human oversight and malfunctioning human machine interface.

Autonomous navigation systems, as currently envisioned, rely on machine learning, including deep learning, for fundamental functionality. Salay and Czarnecki[14] argue that the extensive use of machine learning approaches is motivated by the fact that functionality like perception is difficult to specify. Instead of being programmed from a specification, software components are therefore implemented by training from examples. The training (and test) dataset enumerates a set of input values and correct system outputs, and this functions as a proxy for "something akin to requirements".[15,16] Advances in machine learning have proven extremely successful in many tasks where a clear specification is lacking, however for assurance and validation, the lack of specification is still a fundamental challenge.

Currently Available Guidelines, Standards and Recommendations

Wood et al.[17] argue that the safety standards available within the automotive industry and any other industry have been defined without explicitly considering the specifics of machine learning algorithms and data-driven models. These models represent technology that are "inherently incompatible with legacy safety standards approaches".[13,16] The functional reasoning of such black-box models are challenging or even impossible to understand and predict,[18] and an inductive learning approach is often followed, making the verification inherently difficult.[5]

In recent years, the maritime industry has invested much research effort in developing methodologies and standards for testing, verification and validation of autonomous functionality and its use.[19] Class guidelines are published, offering process and technology guidance to the design and arrangements of systems supporting autonomous and remote operation of vessels, with the objective to ensure safe implementation of novel technologies.[20,21] The IMO's *Interim guideline for MASS trials*[22] provides a set of general principles and main objectives aiming to assist relevant authorities and stakeholders to ensure that trails are conducted safely, securely and with due regard for protection of the environment.

A range of standards, guidelines and methodologies for measuring or demonstrating safety of autonomous vehicles

are proposed also in other industries, including automotive which is in a more advanced state than maritime.[19] These standards, guidelines and methodologies should be adopted and utilized by the maritime industry when applicable, and, when necessary, be modified to fit the maritime domain. The automotive industry might be ahead of the maritime industry, but further research and development are needed also in this domain. Waymo, formerly known as Google Self-Driving Car Project, argue that currently there exists no ”definitive, widely accepted, empirical methodology for answering the question often asked with regard to AVs: *How safe is safe enough?*”[4] Waymo applies industry standards where appropriate, but relies on its own approach aiming to incorporating ”safety at every system level and every development stage, from design to testing and validation.”[4,23] Extensive testing is conducted through driving in simulation, on closed courses, and on public roads. Additionally, national crash-data and driving studies are utilized to provide insights into potential hazards.

ISO Standard 26262:2018[24] *Road vehicles – Functional safety* describes a framework for functional safety to assist the development of safety-related electrical or electronic systems, addressing possible hazards caused by malfunctioning behavior of safety-related systems. However, traditional requirements-based verification, such as the V-model referred to in the Standard, typically assumes that the requirements of a component are completely specified and that ”each refinement can be verified with respect to its specification”.[14] But due to the lack of specification of tasks such as perception, the V-model is not applicable.[15]

ISO/PAS 21448:2019(E)[25] *Road vehicles — Safety of the intended functionality* is designed to complement ISO 26262,[24] and provides a list of recommended measures in

- the design phase (such as requirement on sensor performance),
- the verification phase (such as technical reviews, test cases with high coverage of relevant scenarios, injection of potential triggering events, in the loop testing), and
- the validation phase (such as long term vehicle test, simulations).

Furthermore, Appendix D of the standard proposes practices for the verification and validation of automotive perception systems, listing considerations regarding data collection, variation in drivers and driving habits and testing, including dedicated testing in extreme conditions, production tolerance testing and testing of the interaction between systems and on multiple versions.

ANSI/UL 4600 *Standard for Safety for the Evaluation of Autonomous Products*[26] is a goal-based, technology-agnostic safety standard approach based on an overarching safety case.[13] A safety case can be defined as: “A documented body of evidence that provides a convincing and valid argument that a system is adequately safe for a given application in a given environment.”[27] An implementation of a safety case includes claims, evidence and a set of safety arguments linking the claims to the evidence. The lack of specification is a challenge as each claim needs a stringent specification. A claim regarding detection of kayaks should, for example, include detection distance capabilities, detection rates, the kayaks size, color and shape, external conditions such as weather, waves and lighting, etc.[2] Similarly, the International Regulations for Preventing Collisions at Sea (COLREG) do not provide sufficient details to qualitatively assess autonomous collision avoidance maneuvers.

Figure 2. Illustration of the safe and unsafe regions for a ship navigating with and without functionality for autonomous navigation and decision support.

Assessment

Another important challenge is determining the level of safety required for autonomous navigation functionality, answering the question “How safe is safe enough for AVs?”[4] It is often claimed that the approval of autonomous navigation functionality used in the maritime industry, including decision support systems and functionality to enhance situational awareness, can be based on the equivalence principle: the use of the novel functionality must make the operation safer or at least as safe as conventional operation.[2] However Ringbom[28] argue that this principle, as described in SOLAS req. I/5, ”has mainly been limited to technical arrangements”, and ”neither the COLREGs nor the watchkeeping parts of the STCW [International Convention on Standards of Training, Certification and Watchkeeping for Seafarers] include this option”. Furthermore, Ringbom claims that the ”existing regulatory framework offers no guidance at all” regarding the automated processing of the observations made or data transmitted. We do not dispute this interpretation. It should, however, be noted that the IMO’s Interime Guidelines on MASS trials[22] also refer to equivalent safety in their argumentation: ”Trials should be conducted in a manner that provides at least the same degree of safety, security and protection of the environment as provided by the relevant instruments”. Furthermore, we do believe, regardless of how current rules and regulations are interpreted, that the safety level of future autonomous navigation functions will have to be equivalent to or surpass the safety level of current solutions. Measuring and testing equivalence is challenging, but deciding concrete acceptance criteria without current solutions as a reference is perhaps even more difficult.

When different functionality and solutions are compared, it is important to note that some scenarios, both in real-world and in simulators, are likely to favor the use of autonomous navigation functionality, while in other scenarios the performance will decrease. It is important that the test design takes this into consideration, ensuring a fair and balanced exploration of the safe and unsafe regions. We believe relevant risk trade-offs should be permitted, following the tolerability principle ”globally at least as good”.[25] This means that the implementation of

Figure 3. Illustration of a decision support system with object detection and classification. Photo: Orca AI, all right reserved.

functionality for autonomy can be approved even though the implementation contributes to decreased performance and safety in some scenarios, if this is compensated by an increase in performance and safety in other scenarios (See Figure 2).

Assessing Situational Awareness

Although various tools and techniques are developed to assess the situational awareness of humans,[29–31] lack of an objective ground truth and determining the relevance of different objects are key challenges when assessing situational awareness. As an illustration, we consider a complex scenario with several ships and boats, some at berth and some sailing, small moving objects and aids to navigation. A snapshot of a decision support system with object detection and classification is shown in Figure 3. Investigation of this snapshot indicates that the system has an impressive capability in detecting and classifying both large ships and smaller objects, although some ships and boats are undetected. More importantly, however, the bridge pillars are also undetected. How should this be incorporated in the total assessment? Some would argue that the bridge pillars are easily detected by a human navigator, and hence for a decision support system it is not important that this is detected by the system. Others would argue that the bridge pillars are highly relevant information, and claim that systems which fails to communicate such information contribute to increase the risk of collisions, and should not be implemented. This example illustrates that an assessment of functionality for autonomous navigation and decision support systems has to consider the relevance and importance of the different objects. But as the relevance of an object depends on the navigation task, it is difficult to assess the system's situational awareness in isolation.

Assessing Navigation Performance

To avoid some the challenges associated with assessing situational awareness, we rather concentrate on assessing the navigation tasks directly. Note that adequate situational awareness is a prerequisite for safe navigation. Assessment of the navigation can be performed by subjective domain experts or by automated assessment systems, such as the prototype presented by Øvergård et al.[32] whose evaluations strongly correlates with evaluations performed by subject matter experts. Note however, that safe navigation is also a task which lacks a clear specification, and the STCW Code neither provides sufficient detail for all the necessary competencies for safe navigation nor the methods for assessing them.[33]

Simulator-based Testing

To demonstrate the capabilities and performance of autonomous functionality, we can execute controlled experiments where scenarios are replicated and the navigation performance and safety of a ship which is controlled by or assisted by autonomous navigation functionality is assessed and compared with the safety of a conventionally operated ship. To be able to control the experiment and make sure that the scenarios can be repeated, the experiment should be performed in a simulator.[6] It is fair to assume that these simulators replicate the real-world with adequate detail for a human navigator, but this assumption is not necessarily valid for the machine learning methods. As mentioned before, the machine learning methods used for image detection and classification, usually neural networks, can be considered "black box" models. We can observe the behavior of these models, but we don't know exactly how they work or what kind of features they find in an image. Is it the shape, the color, the derivative in some direction or some other statistical analysis mixed together? Various methods are developed to explain and interpret the results of machine learning methods.[34,35] Nevertheless, for advanced methods such as neural networks, which comprise myriads of interconnected functions, it is challenging or even impossible to predict its behavior even in simple and trivial cases. Since simulators typically simplify the real-world scenarios with straight lines, sharp angles, simple shapes and uniform colors, without the noise typically present in real-world sensor data, we cannot know how this affects the result of the machine learning methods. The results also depend on whether the initial training was done on real or synthetic images.

Implementing Camera-Based Object Detection and Classification in the Simulator

Our goal is to understand how well an autonomous system will function in the real-world, but we aim to test this in a simulator. Due to the above-mentioned difference between the real-world data and the data from the simulator, applying functionality for autonomous navigation and decision support in the simulator is not straight forward. We propose to overcome this challenge by utilizing a Cycle-Consistent Adversarial Network to transform the simulator data from a synthetic dataset into something that is real-world like before the autonomous functionality is applied.

In the following, we briefly define image-to-image translation problems and provide a brief introduction to Cycle-Consistent Adversarial Networks and explain how this approach can be used for translations between real-world and simulator environments. Our focus in this paper is on image-to-image translations. The presented approach is, however, not necessarily limited to image and video data. For future studies we suggest investigating if, and to what extent, the proposed approach can apply to other sensor data, including radar.

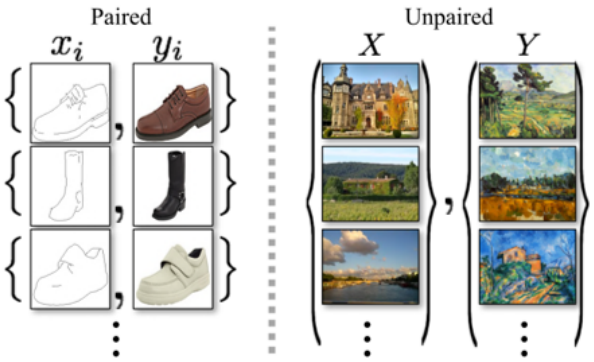


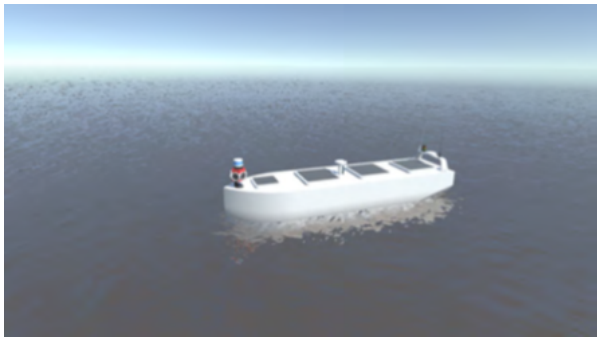
Figure 4. Paired training data (left) consists of training examples where the correspondence between x_i and y_i exists. Cycle-Consistent Adversarial Network instead consider unpaired training data (right). Figure and description from Zhu et al.[36]

Image-to-Image Translation

An image-to-image translation method aims to find a mapping $G : X \rightarrow Y$ between an input space of images X and an output space Y such that the set of images $G(X)$ is indistinguishable from the set of images Y. Many such translation methods require a training set of aligned image pairs (x_i, y_i), that is every output image $y_i \in Y$ in the training data set corresponds to an input image $x_i \in X$ (see Figure 4 (left)). Successful methods for this application includes for example Conditional Adversarial Networks as proposed by Isola et al.[37] However, in many applications (including translation between real-world and simulator environment) paired datasets are often unavailable. This motivates the development of methods which allows translation of images in the absence of paired examples (see Figure 4 (right)). Approaches which utilize unpaired images do not typically require less images than approaches which require paired images, but the effort of collecting samples is reduced, since the samples do not need to correspond. In our case, collecting images from a simulator environment and from a real-world environment is cheap and easy, and requires almost no work. Collecting corresponding images can be very challenging, at best expensive and time consuming. The total number of images required to achieve desired performance will always depend on the set of images available. Images which are very similar, almost duplicates, will for example not contribute much.

Cycle-Consistent Adversarial Network

Cycle-Consistent Adversarial Networks are successfully applied for various image-to-image translation problems. This includes mappings between photos and paintings, aerial photos and maps, and between different objects such as horses and zebras or apple and pears.[36] Cycle-Consistent Adversarial Networks are for example also used for de-noising of networks for multi-phase coronary CT angiography,[38] augmenting data and improve generalizability in CT segmentation tasks,[39] creating realistic snow-covered scenes of multispectral Sentinel-2 imagery,[40] and unsupervised adaptation from synthetic (computer game) to real-world driving domains.[41] De



(a) Image from a simulator

(b) Image generated based on (a)

(c) Real-world image of the ship model

Figure 5. Shows how Cycle-Consistent Adversarial Networks can be applied on an image from a simulator (a) to generate a new image (b), and compares this with a real-world image of the physical ship model (c). Figures from Bekkeheien.[12]

Curtó and Duvall[42] successfully utilize Cycle-Consistent Adversarial Networks in the context of space science and planetary exploration, presenting a framework for neural style transfer based on rendered simulator images of the Moon. Simulator data are used to train and test localization algorithms, and to test software design. Similarily, Bekkeheien[12] utilize Cycle-Consistent Adversarial Networks to improve the quality of data acquired by a marine simulator, and use this data to train a detection algorithm (see Figure 5). It is claimed that this approach improves the detection algorithms in the marine environment.

A Cycle-Consistent Adversarial Networks (Cycle-GAN), as proposed by Zhu et al.,[36] aims to learn an image-to-image translation mapping $G : X \rightarrow Y$ between the input space of images X and the output space Y in the absence of paired examples. The aim for the translation mapping G is that images $G(X)$ are indistinguishable from the set of images Y. To accomplish this, we utilize Generative Adversarial

Networks (GANs) as proposed by Goodfellow,[43] which have shown strong performance in image generation.[43–49] A discriminator D_Y is defined, which is trained to distinguish between real images $y \in Y$ and translated images $G(x)$ for $x \in X$. Simultaneously, the generative model, G, is trained to create realistic images such that the discriminator fails to distinguish between real and translated images. This gives the following objective

$$\mathcal{L}_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{\text{data}}(y)}[\log D_Y(y)] + \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log(1 - D_Y(G(x)))]. \quad (1)$$

The discriminative model, D_Y, is trained to maximize the probability of assigning the correct label to real samples y and translated samples $G(x)$, simultaneously as G is trained to create images which minimize the probability of D_Y assigning the correct label. We think of G and D_Y as players in a two-player minimax game, that is

$$\min_G \max_{D_Y} \mathcal{L}_{GAN}(G, D_Y, X, Y). \quad (2)$$

Because the mapping G is highly under-constrained, G is coupled with an inverse mapping $F : Y \to X$ which is also trained simultaneously as a discriminative model, D_X, which is trained to maximize the probability of assigning the correct label to real samples x and translated samples $F(y)$. Simultaneously, F is trained to create images which minimize the probability of D_X assigning the correct label, that is

$$\min_F \max_{D_X} \mathcal{L}_{GAN}(F, D_X, Y, X). \quad (3)$$

Additionaly, a cycle consistency loss function , defined as

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1], \quad (4)$$

is included in the optimization routine to ensure consistency when an image is translated from one domain to the other and back again. This means that when an image $x \in X$ is translated from X to Y, $(G(x))$, and then translated back to X again $(F(G(x)))$, the result is approximately equal to the original image, that is

$$F(G(x)) \approx x \qquad \text{for } x \in X, \quad (5)$$

and vice versa $(G(F(y)) \approx y)$. For details we refer to Zhu et al.[36]

Mappings Between Real-World and Simulator Environment

We aim to assess the navigation performance in a simulator environment, but since we cannot apply the autonomous functionality directly in the simulator environment, we utilize Cycle-Consistent Adversarial Networks to transform the simulator data to a real-world environment before the autonomous functionality is applied. We let R be real-world sensor data environment, and let S be sensor data in a simulator environment, and let the mappings A and B be translations between real-world sensor data and the simulator environment such that

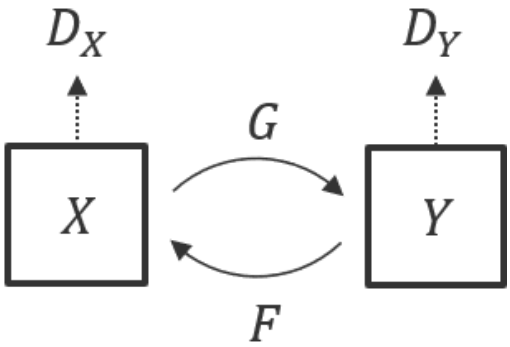


Figure 6. The model contains two mapping functions $G : X \to Y$ and $F : Y \to X$, and associated adversarial discriminative models D_Y and D_X. D_Y encourages G to translate X into outputs indistinguishable from domain Y , and vice versa for D_X and F. Figure and explanation derived from Zhu et al.[36]

$$A : R \to S \qquad B : S \to R \quad (6)$$

We let these translations be based on Cycle-Consistent Adversarial Networks which learn the mapping between the real-world and the simulator environment from a training dataset of unpaired images. We define a discriminator, D_R, which is trained to distinguish whether a sample is a real-world image or if it is translated from the simulator environment. Simultaneously, B is trained to create images which are real-world like, and by this, minimizing the probability of D_R making the correct assignment. Similarly, a discriminator D_S is defined and trained to distinguish simulator images and images translated from real-world images. This gives the adversarial losses

$$\min_B \max_{D_R} \mathcal{L}_{GAN}(B, D_R, S, R) \quad (7)$$

and

$$\min_A \max_{D_S} \mathcal{L}_{GAN}(A, D_S, R, S). \quad (8)$$

Additionally, a cycle consistency loss, $\mathcal{L}_{cyc}(G, F)$ is introduced to push $B(A(r)) \approx r$ for real-world scenarios $r \in R$.

A set of scenarios should be collected and used to test that the performance of the autonomous functionality, when applied to the translated simulator data, is similar to the performance of the autonomous functionality if applied to real-world data. We define a function z which takes an image as input, and outputs autonomous functionality (for example information regarding the ship's situational awareness such as bounding boxes, segmentation, speed and distance estimates, and suggested maneuvers, etc.). It should be demonstrated that when a scenario $r \in R$ is translated to the simulator environment $(A(r))$, and then back again $(B(A(r)))$, applying autonomous functionality to this output gives approximately the same result as if it was applied on the original real-world scenario. As we are only interested in the performance of this autonomous functionality, in our

application, it is sufficient that

$$z(B(A(r))) \approx z(r) \qquad \text{for scenarios } r \in R \tag{9}$$

and

$$z(A(B(s))) \approx z(s) \qquad \text{for scenarios } s \in S, \tag{10}$$

and it is not necessary to require $B(A(r)) \approx r$ and $A(B(s)) \approx s$. We include this explicitly in the training by modifying the penalty term of the cycle consistency loss function, such that the cycle consistency loss becomes

$$\begin{aligned}\mathcal{L}_{cyc}(A,B) = \mathbb{E}_{r\sim p_{\text{data}}(r)}[\|z(B(A(r))) - z(r)\|_1] \\ + \mathbb{E}_{s\sim p_{\text{data}}(s)}[\|z(A(B(s))) - z(s)\|_1].\end{aligned} \tag{11}$$

Furthermore, information from the real-world scenarios $r \in R$, including the ship's location, location of other vessels, shoreline, weather, waves, etc. should be used to reconstruct simulator scenarios $s \in S$, such that it can be demonstrated that if this scenario is translated from the simulator environment to real-world, the autonomous functionality returns similar output as if it was applied on the original real-world scenario, that is

$$z(B(s)) \approx z(r). \tag{12}$$

Quantifying the success of a style transfer is always challenging. If an image of a horse is transferred to an image of a zebra,[36] it is difficult to assess if and to what degree the transfer was successful. In our case, we are only interested in the autonomous functionality z and do not require $B(s) \approx r$. Still, demonstrating that Eq. (12) holds is non-trivial, and a set of corresponding scenarios from the simulator and the real-world environment is needed. Fortunately, if we are able to produce such corresponding scenarios, we do not need to evaluate the full image $B(s)$, but can focus our evaluation on the autonomous functionality z (such as for example bounding boxes) which often makes the evaluation less complex.

Future Work

Previous successful applications of Cycle-Consistent Adversarial Networks, in the maritime domain as well as other domains, motivate the approach proposed in this paper. Nevertheless, extensive testing is needed to validate the applicability of our proposed test approach, which is obviously dependent on the quality of the simulator in addition to the translation mappings.

In this paper our focus is on Cycle-Consistent Adversarial Networks, but other related methods should also be explored. For example, Liu et al.[50] propose an unsupervised image-to-image translation framework based on Coupled Generative Adversarial Networks to translate street scene images for example between winter and summer, night and day and wet and dry. It is also shown how this can be utilized for translations between simulators and real-world images. Other examples include the Multimodal Unsupervised Image-to-image Translation (MUNIT) framework as proposed by Huang et al.[51], and the Diverse Image-to-Image Translation via Disentangled Representations (DRIT) as proposed by Lee et al.[52] The above-mentioned methods are two-sided, meaning that when a mapping G which translates samples from a domain X to a domain Y is learned, the inverse mapping F from Y to X is learned simultaneously. Notable one-sided methods, where the translation is learned without learning its inverse, include SelfDistance and DistanceGAN[53], and GcGAN[54].

Whenever paired data are available, it can be worthwhile to explore models which are capable of utilizing both paired and unpaired training data simultaneously, such as for example the general-purpose image-to-image translation model proposed by Tripathy.[55] The study demonstrates how the proposed method obtains qualitatively and quantitatively improved results compared to two baselines, outperforming the baselines also in the case of purely paired and unpaired training data.

Conclusion

The development of functionality for autonomous navigation and decision support systems in the maritime domain is motivated by the promise of increased safety. However, assessing and quantifying the safety level is challenging.

In this paper we propose a simulator-based test framework designed to assess and compare how functionality for autonomous navigation and decision support contributes to increase navigation performance and safety. Since we cannot directly apply the autonomous functionality to the images/video generated by the simulator, we propose to first translate the data using Cycle-Consistent Adversarial Networks. Such networks are designed to translate an image from a source domain to a target domain in the absence of paired examples, such that the translated image appears more realistic. A critical assumption in the proposed test framework is that the performance of the autonomous functionality, when applied to the translated simulator data, is similar to the performance of the autonomous functionality if applied to real-world data. This assumption is dependent both on the quality of the simulator and the mappings between the real-world and the simulator environment. Hence, demonstrating that this assumption holds should always be a natural and necessary first step towards implementing the proposed test approach.

With the procedure proposed in this paper, automatic object detection and classification can be performed realistically in the simulator, making it possible to conduct controlled, repeated experiments with identical initial conditions. This allows us to compare the navigation of autonomous ships at various degrees of automation with the navigation of conventional navigation. Furthermore, various assessment methods, both manual and automated, can be utilized to quantify and assess the navigation performance of the different ships, and ultimately, quantify and measure how the use of autonomous functionality affects maritime safety.

Acknowledgements

Criticisms and discussions regarding the topics of this paper with Knut Erik Knutsen (DNV), Dor Raviv (Orca AI), Tobias Rye Torben (Norwegian University of Science and Technology), Tom Arne Pedersen (DNV), Kristian Karolius (DNV), Christian Hovden (University of South-Eastern Norway) and Anete Vagale

(Norwegian University of Science and Technology) are greatly appreciated.

Declaration of conflicting interests

The authors declare that there are no conflict of interest.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Research Council of Norway [269465].

References

1. Maritime safety committee (MSC) 99/wp.9 annex 1. Regulatory scoping exercise for the use of maritime autonomous surface ships (MASS). Report of the working group. Technical report, International Maritime Organization (IMO), 2018.
2. Brandsæter A, Smefjell G, van de Merwe K et al. Assuring safe implementation of decision support functionality based on data-driven methods for ship navigation. In *Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference*. Research Publishing, Singapore, 2020.
3. Accident Investigation Board Norway. Part one report on the collision on 8 November 2018 between the frigate HNoMS Helge Ingstad and the oil tanker Sola TS outside the sture terminal in the hjeltefjord in hordaland county. Technical report, Accident Investigation Board Norway, Defence Accident Investigation Board Norway, 2019.
4. Waymo's safety methodologies and safety readiness determinations. Technical report, Waymo LLC, 2020.
5. Koopman P and Wagner M. Autonomous vehicle safety: An interdisciplinary challenge. *IEEE Intelligent Transportation Systems Magazine* 2017; 9(1): 90–96.
6. Mislevy RJ. Evidence-centered design for simulation-based assessment. *Military medicine* 2013; 178(10): 107–114.
7. Zghyer R and Ostnes R. Opportunities and challenges in using ship-bridge simulators in maritime research. In *Proceedings of Ergoship 2019, Haugesund*. Western Norway University of Applied Sciences (HVL), 2019.
8. Schöner HP. "how good is good enough?" in autonomous driving. In *Electronic Components and Systems for Automotive Applications*. Springer, 2019. pp. 119–142.
9. DNV GL. Maritime simulator systems. *Standard DNVGL-ST-0033* 2017; URL https://rules.dnvgl.com/docs/pdf/DNVGL/ST/2017-03/DNVGL-ST-0033.pdf.
10. Kongsberg Digital. K-sim navigation brochure. Technical report, 2020. URL https://www.kongsberg.com/globalassets/digital/maritime-simulation/k-sim-navigation/docs/k-sim-navigation-brochure.pdf.
11. Porathe T et al. Human-centred design in the maritime domain. *DS 85-1: Proceedings of NordDesign 2016, Volume 1, Trondheim, Norway, 10th-12th August 2016* 2016; : 175–184.
12. Bekkeheien LMW. *Synthesizing Photo-Realistic images from a Marine Simulator via Generative Adversarial Networks*. Master's Thesis, Norwegian University of Science and Technology, 2020.
13. Koopman P, Ferrell U, Fratrik F et al. A safety standard approach for fully autonomous vehicles. In *2019 International Conference on Computer Safety, Reliability, and Security*. Springer, pp. 326–332.
14. Salay R, Queiroz R and Czarnecki K. An analysis of iso 26262: Using machine learning safely in automotive software. *arXiv:170902435* 2017; .
15. Koopman P and Wagner M. Challenges in autonomous vehicle testing and validation. *SAE International Journal of Transportation Safety* 2016; 4(1): 15–24.
16. Koopman P and Wagner M. Toward a framework for highly automated vehicle safety validation. Technical report, SAE Technical Paper, 2018.
17. Wood M, Robbel P, Maass M et al. Safety first for automated driving. Technical report, Aptiv Services US, LLC; AUDI AG; Bayrische Motoren Werke AG; Beijing Baidu Netcom Science Technology Co., Ltd; Continental Teves AG & Co oHG; Daimler AG; FCA US, LLC; HERE Global B.V.; Infineon Technologies AG; Intel; Volkswagen AG., 2019.
18. Brandsæter A and Knutsen KE. Towards a framework for assurance of autonomous navigation systems in the maritime industry. In *Safety and Reliability–Safe Societies in a Changing World : Proceedings of ESREL 2018*. CRC Press, 2018. pp. 449–457.
19. Brinkmann M, Böde E, Lamm A et al. Learning from automotive: Testing maritime assistance systems up to autonomous vessels. In *OCEANS 2017-Aberdeen*. IEEE, pp. 1–8.
20. DNV GL. Autonomous and remotely operated ships. *Class Guideline DNVGL-CG-0264* 2018; URL https://rules.dnvgl.com/docs/pdf/DNVGL/CG/2018-09/DNVGL-CG-0264.pdf.
21. Bureau Veritas. Guidelines for autonomous shipping. *Class Guideline NI641 R01* 2019; URL https://marine-offshore.bureauveritas.com/ni641-guidelines-autonomous-shipping.
22. IMO. Interim guidelines on mass trials. *MSC1/Circ1604* 2019; URL https://www.register-iri.com/wp-content/uploads/MSC.1-Circ.1604.pdf.
23. Waymo safety report. Technical report, Waymo LLC, 2020.
24. ISO 26262:2018. Road vehicles – Functional safety. Standard, International Organization for Standardization, Geneva, CH, 2018.
25. ISO/PAS 21448:2019(E). Road vehicles — Safety of the intended functionality. Standard, International Organization for Standardization, Geneva, CH, 2019.
26. ANSI/UL 4600. ANSI/UL 4600 Standard for Safety for the Evaluation of Autonomous Products. Standard, Underwriters Laboratories Inc., 2020.
27. Bishop P and Bloomfield R. A methodology for safety case development. *Safety and Reliability* 2000; 20(1): 34–42. DOI: 10.1080/09617353.2000.11690698.
28. Ringbom H. Regulating autonomous ships—concepts, challenges and precedents. *Ocean Development & International Law* 2019; 50(2-3): 141–169.
29. Endsley MR. Design and evaluation for situation awareness enhancement. *Proceedings of the Human Factors Society Annual Meeting* 1988; 32(2): 97–101. DOI:10.1177/154193128803200221.
30. Endsley MR. Situation awareness global assessment technique (SAGAT). In *Proceedings of the IEEE 1988 national aerospace and electronics conference*. IEEE, pp. 789–795.

31. Endsley MR. Direct measurement of situation awareness: Validity and use of SAGAT. *Situation awareness analysis and measurement* 2000; 10: 147–173.
32. Øvergård KI, Nazir S and Solberg AS. Towards automated performance assessment for maritime navigation. *TransNav, International Journal on Marine Navigation and Safety of Sea Transportation* 2017; 11(2): 229–234.
33. Kobayashi H. Use of simulators in assessment, learning and teaching of mariners. *WMU Journal of maritime affairs* 2005; 4(1): 57–75.
34. Adadi A and Berrada M. Peeking inside the black-box: A survey on explainable artificial intelligence (xai). *IEEE Access* 2018; 6: 52138–52160.
35. Brandsæter A and Glad IK. Explainable artificial intelligence: How subsets of the training data affect a prediction. *arXiv:201203625* 2020; .
36. Zhu JY, Park T, Isola P et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision, 2017*. pp. 2223–2232.
37. Isola P, Zhu JY, Zhou T et al. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition, 2017*. pp. 1125–1134.
38. Kang E, Koo HJ, Yang DH et al. Cycle-consistent adversarial denoising network for multiphase coronary ct angiography. *Medical physics* 2019; 46(2): 550–562.
39. Sandfort V, Yan K, Pickhardt PJ et al. Data augmentation using generative adversarial networks (cyclegan) to improve generalizability in ct segmentation tasks. *Scientific reports* 2019; 9(1): 1–9.
40. Ren CX, Ziemann A, Theiler J et al. Cycle-consistent adversarial networks for realistic pervasive change generation in remote sensing imagery. In *IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI), 2020*. IEEE, pp. 42–45.
41. Hoffman J, Tzeng E, Park T et al. Cycada: Cycle-consistent adversarial domain adaptation. In *International conference on machine learning, 2018*. PMLR, pp. 1989–1998.
42. de Curtó J and Duvall R. Cycle-consistent generative adversarial networks for neural style transfer using data from chang'e-4. *arXiv:201111627* 2020; .
43. Goodfellow I, Pouget-Abadie J, Mirza M et al. Generative adversarial nets. In *Advances in neural information processing systems, 2014*. pp. 2672–2680.
44. Curtó JD, Zarza IC, De La Torre F et al. High-resolution deep convolutional generative adversarial networks. *arXiv:171106491* 2017; .
45. Antoniou A, Storkey A and Edwards H. Data augmentation generative adversarial networks. *arXiv:171104340* 2017; .
46. Zhu JY, Krähenbühl P, Shechtman E et al. Generative visual manipulation on the natural image manifold. In *European conference on computer vision*. Springer, pp. 597–613.
47. Wang X and Gupta A. Generative image modeling using style and structure adversarial networks. In *European conference on computer vision*. Springer, pp. 318–335.
48. Radford A, Metz L and Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv:151106434* 2015; .
49. Odena A, Olah C and Shlens J. Conditional image synthesis with auxiliary classifier gans. In *International conference on machine learning*. PMLR, pp. 2642–2651.
50. Liu MY, Breuel T and Kautz J. Unsupervised image-to-image translation networks. In *Advances in neural information processing systems, 2017*. pp. 700–708.
51. Huang X, Liu MY, Belongie S et al. Multimodal unsupervised image-to-image translation. In *Proceedings of the European Conference on Computer Vision (ECCV), 2018*. pp. 172–189.
52. Lee HY, Tseng HY, Huang JB et al. Diverse image-to-image translation via disentangled representations. In *Proceedings of the European conference on computer vision (ECCV)*. pp. 35–51.
53. Benaim S and Wolf L. One-sided unsupervised domain mapping. *arXiv:170600826* 2017; .
54. Fu H, Gong M, Wang C et al. Geometry-consistent generative adversarial networks for one-sided unsupervised domain mapping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 2427–2436.
55. Tripathy S, Kannala J and Rahtu E. Learning image-to-image translation using paired and unpaired training samples. In *Asian Conference on Computer Vision, 2018*. Springer, pp. 51–66.